Dr. Murtadha Alshareifi e-Library Pathophysiology

EIGHTH EDITION

Concepts of Altered Health States

the Point,

Carol Mattson Porth Glenn Matfin

Welpers Kluwer LippinCott Williams & Wilking

Contributors

Diane Book, MD

Assistant Professor, Neurology Medical College of Wisconsin Milwaukee, Wisconsin (Chapter 37)

Edward W. Carroll, MS, PhD

Clinical Assistant Professor Department of Biomedical Sciences, College of Health Sciences Marquette University Milwaukee, Wisconsin (Chapters 1, 3, 36, 40)

Kathryn Ann Caudell, RN, PhD

Clinical Research and Education Manager Amgen-Oncology Professional Services Edgewood, New Mexico (Chapters 5, 11)

Robin Curtis, PhD

Professor, Retired Department of Cellular Biology, Neurobiology, and Anatomy The Medical College of Wisconsin Milwaukee, Wisconsin (Chapter 38)

Elizabeth C. Devine, RN, PhD, FAAN

Professor, School of Nursing University of Wisconsin- Milwaukee Milwaukee, Wisconsin (Chapter 39)

Susan A. Fontana, RN, PhD, CS-FNPNET

Certified Family Nurse Practitioner Associate Professor, School of Nursing University of Wisconsin-Milwaukee Milwaukee, Wisconsin (Chapter 40)

Kathryn J. Gaspard, PhD

Clinical Associate Professor, School of Nursing University of Wisconsin-Milwaukee Milwaukee, Wisconsin (Chapters 11, 12, 13)

Kathleen E. Gunta, RN, MSN, ONC

Clinical Nurse Specialist, Orthopaedics St. Luke's Medical Center Milwaukee, Wisconsin (Chapters 42, 43)

Safak Guven, MD, FACP, FACE

Assistant Professor, Obesity/Metabolic Syndrome Clinic Director Medical College of Wisconsin Milwaukee, Wisconsin (Chapters 30, 31, 32)

Candace Hennessy, RN, PhD

Vice President, Patient Care Services St. Michaels Hospital Milwaukee, Wisconsin (Chapter 18)

Julie A. Kuenzi, RN, MSN, CDE, BC-ADM

Diabetes Care Center Manager Diabetes Nurse Specialist Froedtert Hospital and Medical College of Wisconsin Milwaukee, Wisconsin (Chapters 30, 31, 32)

Mary Pat Kunert, RN, PhD

Associate Professor School of Nursing Marquette University Milwaukee, Wisconsin (Chapter 7)

Glenn Matfin, BSc (Hons), MB ChB, DGM, MFPM,

MRCP(UK), FACE, FACP Medical Director and Vice President Medical Research and Development Inverness Medical Innovations USA and Unipath, UK (Chapters 30, 31, 32)

Patricia McCowen Mehring, RNC, MSN, OGNP

Department of Obstetrics and Gynecology Division of Reproductive Medicine Medical College of Wisconsin Milwaukee, Wisconsin (Chapters 34, 35)

viii Contributors

Joan Pleuss, MS, RD, CDE

Bionutrition Research Manager General Clinical Research Center Medical College of Wisconsin Milwaukee, Wisconsin (Chapter 29)

Debra Ann Bancroft Rizzo, RN, MSN, FNP-C

Rheumatology Nurse Practitioner Rheumatic Disease Center Milwaukee, Wisconsin (Chapter 43)

Gladys Simandl, RN, PhD

Professor Columbia College of Nursing Milwaukee, Wisconsin (Chapter 44)

Cynthia V. Sommer, PhD, MT (ASCP)

Associate Professor Emerita, Department of Biological Sciences University of Wisconsin-Milwaukee Milwaukee, Wisconsin (Chapters 8, 9)

Kathleen A. Sweeney, RN, BSN, ACRN

Women's Statewide Outreach Coordinator Wisconsin HIV Primary Care Support Network Medical College of Wisconsin Milwaukee, Wisconsin (Chapter 10)

Reviewers

Kevin Branch, BHSc.PHC, ACP

Coordinator, Paramedic Programs and Pre-Service Firefighter Education and Training Program Cambrian College Sudbury, Ontario, Canada

Deborah A. Hallingstad, MSN, RN, OCN

Nursing Faculty Chippewa Valley Technical College Eau Claire, Wisconsin

Marguerite J. Murphy, RN, MS

Assistant Professor School of Nursing Medical College of Georgia Barnesville, Georgia

Gila Strauch, BSc, MSc, MEd, PhD Director, Department of Education

Director, Department of Education ICT Schools Toronto, Ontario, Canada

Preface

The widespread acceptance of our textbook, *Pathophysiology*: Concepts of Altered Health States, has led to requests for an essentials version that can be used by students who do not need the extensive breadth or detail of content provided in the larger book. In preparing the essentials version, we have reorganized content, attempted to eliminate nonessential information, and developed or modified illustrations in an effort to produce a comprehensive text that is accurate, up-to-date, and appealing to the reader. Great care also was taken to integrate text and design in a manner that will aid the reader's exploration. Although the chapters are authored by a variety of experts, there is a unified voice and a consistent level of readability.

As a nurse-physiologist, my major emphasis in the preparation of the essentials version, as with current and previous editions of Pathophysiology: Concepts of Altered Health States, has been to relate normal body functioning to the physiologic changes that participate in disease production and occur as a result of disease, as well as the body's remarkable ability to compensate for these changes. The beauty of a physiologic approach is that it integrates all of the aspects of the individual cells and organs of the human body into a total functional whole that can be used to explain both the physical and psychological aspects of altered health. Indeed, it has been my goal to share the beauty of the human body and to emphasize that in disease as in health, there is more "going right" in the body than is "going wrong." This book is an extension of my career and, as such, of my philosophy. It is my hope that readers will learn to appreciate the marvelous potential of the body, incorporating this belief into their own philosophy and ultimately sharing it with their clients.

In the preparation of Essentials of Pathophysiology: Concepts of Altered Health States, every attempt has been made to develop a text that is current, accurate, and well organized. The content has been arranged so that concepts build on one another. Words are defined as content is presented. Concepts from physiology, biochemistry, physics, and other sciences are reviewed as deemed appropriate. Common health problems, including the special needs of children and older adults, are addressed. Although intended as a course textbook, Essentials of Pathophysiology: Concepts of Altered Health States is also intended to serve as a reference book that students can take with them and use in their practice once the course is finished.

The use of full color in the design and illustrations offers visual appeal and enhances conceptual learning, linking text content to illustration content. Icons identify content related to children 📲 , pregnant women 🚔 , and older adults 🐔 Some of the most popular features found in the sixth edition of Pathophysiology: Concepts of Altered Health States have been retained in this Essentials version:

- List of suffixes and prefixes
- Tables of Normal Laboratory Values
- Glossary
- Key Concept Boxes and Summary Boxes, which help the reader retain and use information by showing how individual facts come together to form a larger conceptual unit

Additional features that are new to the Essentials version include:

- Review Questions at the end of each chapter
- An appendix of Website Resources, to encourage readers to investigate areas of individual interest and update knowledge in a world of ever-changing information.

STUDENT AND INSTRUCTOR RESOURCES

A variety of ancillary materials have been developed to support students and instructors alike.

Resources for Students

- A back-of-the-book CD-ROM. This free CD-ROM contains multiple-choice questions for review and animations of selected pathophysiologic processes.
- Study Guide to Accompany Porth's Essentials of Pathophysiology: Concepts of Altered Health States, by Kathleen Schmidt Prezbindowski. This study guide, available for purchase, reinforces the information in the text using a variety of question styles, including multiple choice, fill-in-theblank, matching, short answer, and figure completion exercises.

Resources for Instructors

- Instructor's Resource CD-ROM to Accompany Essentials of Pathophysiology: Concepts of Altered Health States, by Patricia S. Bowne. This CD-ROM contains the following:

 - An Instructor's Manual with lecture outlines and teaching suggestions, plus printable worksheets for student assignments

Preface

- A Test Bank, containing approximately 1100 multiplechoice questions
- An Image Bank, containing approximately 300 images from the text in formats suitable for printing, projecting, and incorporating into web sites
- PowerPoint presentations with incorporated images.

Resources for Students and Instructors

Visit the Porth: *Essentials of Pathophysiology: Concepts of Altered Health States* **Connection site** at http://connection.LWW.com/ go/porth.

connection-

The writing of this book has been a meaningful endeavor for the authors. It was accomplished through an extensive review of the literature and through the use of critiques provided by students, faculty, and content specialists. As this vast amount of information was processed, inaccuracies or omissions may have occurred. Readers are encouraged to contact us about such errors. Such feedback is essential to the continual development of the book.

СМР

To the Reader

This book was written with the intent of making the subject of pathophysiology an exciting exploration that relates normal body functioning to the physiologic changes that participate in disease production and occur as a result of disease, as well as the body's remarkable ability to compensate for these changes. Indeed, it is these changes that represent the signs and symptoms of disease.

Using a book such as this can be simplified by taking the time to find what is in the book and how to locate information when it is needed. The *table of contents* at the beginning of the book provides an overall view of the organization and content of the book. It also provides clues as to the relationships among areas of content. For example, the location of the chapter on neoplasia within the unit on cell function and growth indicates that neoplasms are products of altered cell growth. The *index*, which appears at the end of the book, can be viewed as a road map for locating content. It can be used to quickly locate related content in different chapters of the book or to answer questions that come up in other courses.

ORGANIZATION

The book is organized into units and chapters. The *units* identify broad areas of content, such as alterations in the circulatory system. The *chapters* focus on specific areas of content, such as heart failure and circulatory shock. The *chapter outline* that appears at the beginning of each chapter provides an outline of the content in the chapter. *Icons* identify specific content related

to infants and children [🚆 , pregnant women, 🚉 , and older adults 🦨 .

Many of the units have a chapter that contains essential information about the structures being discussed in the unit. These chapters provide the foundation for understanding the pathophysiology content presented in the subsequent chapters.

READING AND LEARNING AIDS

In an ever-expanding world of information you will not be able to read, let alone remember, everything that is in this book, or in any book, for that matter. With this in mind, we have developed a number of special features that will help you focus on and master the essential content for your current as well as future needs.

It is essential for any professional to use and understand the vocabulary of his or her profession. Throughout the text, you will encounter terms in italics. This is a signal that a word and the ideas associated with it are important to learn. In addition, two aids are provided to help you expand your vocabulary and improve your comprehension of what you are reading: the glossary and the list of prefixes and suffixes.

The *glossary* contains concise definitions of frequently encountered terms. If you are unsure of the meaning of a term you encounter in your reading, check the glossary in the back of the book before proceeding.

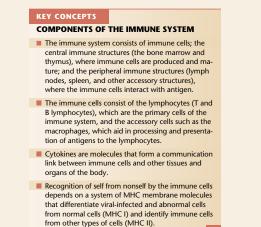
The *list of prefixes and suffixes* is a tool to help you derive the meaning of words you may be unfamiliar with and increase your vocabulary. Many disciplines establish a vocabulary by affixing one or more sounds or letters to the beginning or end of a word or base to form a derivative word. Prefixes are added to the beginning of a word or base, and suffixes are added to the end. If you know the meanings of common prefixes and suffixes, you can usually derive the meaning of a word, even if you have never encountered it before. A list of prefixes and suffixes common to pathophysiology can be found on the inside covers.

BOXES

Boxes are used throughout the text to summarize and highlight key information. You will encounter two types of boxes: Key Concept Boxes and Summary Boxes.

One of the ways to approach learning is to focus on the major ideas or concepts rather than trying to memorize a list of related and unrelated bits of information. As you have probably already discovered, it is impossible to memorize everything that is in a particular section or chapter of the book. First, your brain has a difficult time trying to figure out where to store all the different bits of information. Secondly, your brain doesn't know how to retrieve the information when you need it. Thirdly, memorized lists of content can seldom be applied directly to an actual clinical situation. The *Key Concept Boxes* guide you in identifying the major ideas or concepts that form the foundation for truly understanding the major areas of content. When you understand

the concepts in the Key Concept boxes, you will have a framework for remembering and using all of the facts given in the text.



The *Summary Boxes* at the end of each section provide a review of the main content that has been covered. Use the summaries to assure that you have covered and understand what you have read.

In summary, heart failure occurs when the heart fails to pump sufficient blood to meet the metabolic needs of body tissues. The physiology of heart failure reflects an interplay between a decrease in cardiac output that accompanies impaired function of the failing heart and the compensatory mechanisms designed to preserve the cardiac reserve. Compensatory mechanisms that contribute to maintenance of the cardiac reserve include the Frank-Starling mechanism, sympathetic nervous system responses, the renin-angiotensinaldosterone mechanism, and myocardial hypertrophy. In the failing heart, early decreases in cardiac function may go unnoticed because these compensatory mechanisms maintain the cardiac output. This is called *compensated heart failure*. Unfortunately, the mechanisms were not intended for longterm use, and in severe and prolonged heart failure, the compensatory mechanisms no longer are effective and further impair cardiac function.

Heart failure may be described as high-output or low-output failure, systolic or diastolic failure, and right-sided or leftsided failure. With high-output failure, the function of the heart may be supernormal but inadequate because of excessive metabolic needs, and low-output failure is caused by disorders that impair the pumping ability of the heart. With systolic dysfunction, there is impaired ejection of blood from the heart during systole; with diastolic dysfunction, there is impaired filling of the heart during diastole. Right-sided failure is characterized by congestion in the peripheral circulation, and left-sided failure by congestion in the pulmonary circulation.

The manifestations of heart failure include edema, nocturia, fatigue and impaired exercise tolerance, cyanosis, signs of increased sympathetic nervous system activity, and impaired gastrointestinal function and malnutrition. In rightsided failure, there is dependent edema of the lower parts of the body, engorgement of the liver, and ascites. In left-sided failure, shortness of breath and chronic, nonproductive cough are common.

Acute pulmonary edema is a life-threatening condition in which the accumulation of fluid in the interstitium of the lung and alveoil interferes with lung expansion and gas exchange. It is characterized by extreme breathlessness, crackles, frothy sputum, cyanosis, and signs of hypoxemia. In cardiogenic shock, there is failure to eject blood from the heart, hypotension, inadequate cardiac output, and impaired perfusion of peripheral tissues. Mechanical support devices, including the intra-aortic balloon purmp (for acute failure) and the VAD, sustain life in persons with severe heart failure. Heart transplantation remains the treatment of choice for many persons with end-stage heart failure.

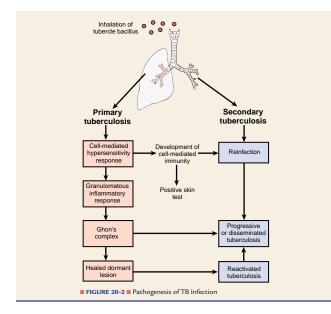
TABLES AND CHARTS

Tables and charts are designed to present complex information in a format that makes it more meaningful and easier to remember. Tables have two or more columns, and are often used for the purpose of comparing or contrasting information. Charts have one column and are used to summarize information.

Hypokalemia Hyperkalemia		
Laboratory Values Laboratory Values Serum potassium <3.5 mEq/L		
Increased thirst Inability to concentrate urine with polyuria and urine with low specific gravity		
Effects of Changes in Membrane Potentials in Neural and Muscle Function Effects of Changes in Membrane Potential on Neural and Muscle Function Gastrointestinal Gastrointestinal Anorexia, nausea, vomiting Nausea, vomiting Abdominal distention Intestinal cramps Paralytic ileus (severe hypokalemia) Diarrhea Nuscle exakness, flabbiness, fatigue Weakness, dizziness Muscle evakness, flabbiness, fatigue Muscle cramps Paratysis (severe hypokalemia) Paresthesias Paratysis (severe hypokalemia) Paresthesias Paratysis (severe hypokalemia) Paresthesias Confusion, depression Cardiovascular Condivoscular Risk of cardiac arrest with severe hyperkalem Predisposition to digitalis toxicity Electrocardiogram changes Cardidox dyshrythmias Cardiac dyshrythmias Acid-Base Balance Metabolic alkalosis		
CHART 5-2 TNM Classification System		
Tx Tumor cannot be adequately assessed T0 No evidence of primary tumor Tis Carcinoma in situ T1-4 Progressive increase in tumor size or involvement		
N (nodes)		
Nx Regional lymph nodes cannot be assessed N0 No evidence of regional node metastasis N 1-3 Increasing involvement of regional lymph nodes		
M (metastasis)		
Mx Not assessed M0 No distant metastasis M1 Distant metastasis present, specify sites		

ILLUSTRATIONS

The full-color illustrations will help you to build your own mental image of the content that is being presented. Each drawing has been developed to fully support and build upon the ideas in the text. Some illustrations are used to help you picture the complex interactions of the multiple phenomena that are involved in the development of a particular disease; others can help you to visualize normal function or understand the mechanisms whereby the disease processes exert their effects. In addition, photographs of pathologic processes and lesions provide a realistic view of selected pathologic processes and lesions.



MATERIALS FOR REVIEW

Several features have been built into the text to help you verify your understanding of the material that you have just read. After you have finished reading and studying the chapter, write down your answers to the *review questions* at the end of the chapter. If you are unable to answer a question, reread the relevant section in the chapter. In addition, you will find *multiple-choice questions* on the back-of-the-book CD-ROM and on the Connection site (http://connection.LWW.com/go/porth). Use these quizzes to test yourself as well.

APPENDICES

Your book contains two appendices. Appendix A, "Laboratory Values," provides rapid access to normal values for many laboratory tests, as well as a description of the prefixes, symbols, and factors (*e.g.*, micro, μ , 10⁻⁶) used for describing these values. Knowledge of normal values can help you to put abnormal values in context. Appendix B, "Weblinks," is provided to help you in obtaining additional information on topics that are of interest to you. The websites in Appendix B will also help you to keep abreast of new information that is constantly evolving. For your convenience, these weblinks are also found on the book's Connection site.

We hope that this guide has given you a clear picture of how to use this book. Good luck and enjoy the journey! 1

24

36

50

64

84

Contents

UNIT ONE Mechanisms of Disease

1	Cell Structure and Function
	Edward W. Carroll Functional Components of the Cell 2 Cell Metabolism and Energy Sources 7 Cell Membrane Transport, Signal Transduction, and Generation of Membrane Potentials 9 Body Tissues 14
2	Tissue Adaptation and Injury
	Cellular Adaptation 24 Cell Injury and Death 27
3	Genetic Control of Cell Function
	Edward W. CarrollGenetic Control of Cell Function36Chromosomes41Patterns of Inheritance44Gene Mapping and Technology46
4	Genetic and Congenital Disorders
	Genetic and Chromosomal Disorders 50 Disorders Due to Environmental Influences 58
5	Alterations in Cell Growth
	and Replication: Neoplasia
	Kathryn Ann Caudell Concepts of Cell Growth 64
	Characteristics of Benign
	and Malignant Neoplasms 67 Carcinogenesis and Causes of Cancer 72
	Clinical Features 77
	Childhood Cancers 81
6	Alterations in Fluids, Electrolytes,
	and Acid-Base Balance
	Composition and Compartmental Distribution of Body Fluids 84
	Sodium and Water Balance 90
	Potassium Balance 99 Calcium and Magnesium Balance 103
	Acid-Base Balance 108

UNIT TWO Alterations in Body Defenses

7	Stress and AdaptationMary Pat KunertHomeostasis122Stress123Disorders of the Stress Response128	121
8	The Immune ResponseCynthia V. SommerThe Immune System134Developmental Aspects of the Immune System148	134
9	Inflammation, Tissue Repair, and Fever Cynthia V. Sommer, Carol M. Porth The Inflammatory Response 150 Tissue Repair and Wound Healing 157 Temperature Regulation and Fever 161	150
10	Alterations in the Immune Response Carol M. Porth, Kathleen Sweeney Allergic and Hypersensitivity Disorders 168 Transplantation Immunopathology 174 Autoimmune Disorders 175 Immunodeficiency Disorders 178	168
Alte	UNIT THREE erations in the Hematologic Syster	n
11	Alterations in White Blood Cells Kathryn Ann Caudell, Kathryn J. Gaspard Hematopoietic and Lymphoid Tissue 191 Non-neoplastic Disorders of White Blood Cells 194 Neoplastic Disorders of Hematopoietic and Lymphoid Origin 197	191
12	Alterations in Hemostasis and Blood Coagulation Kathryn J. Gaspard Mechanisms of Hemostasis 205 Hypercoagulability States 209 Bleeding Disorders 210	205

13 Alterations in Red Blood Cells Kathyn J. Gaspard The Red Blood Cell 216 Anemia 220 Polycythemia 227 Age-Related Changes in Red Blood Cells 227 UNIT FOUR Iterations in the Cardiovascular Sys	216 tem	21 Alterations in Respiratory Function: Disorders of Gas Exchange Disorders of Lung Inflation 376 Obstructive Airway Disorders 380 Interstitial Lung Diseases 390 Pulmonary Vascular Disorders 392 Respiratory Failure 395 UNIT SIX Alterations in the Urinary System
14 Structure and Function of the Cardiovascular System Organization of the Circulatory System 232 Principles of Blood Flow 233 The Heart as a Pump 236 Blood Vessels and the Peripheral Circulation 246	231	 22 Control of Kidney Function Kidney Structure and Function 402 Tests of Renal Function 412 23 Alterations in Renal Function
Neural Control of Circulatory Function 251 Alterations in Blood Flow Disorders of the Arterial Circulation 254 Disorders of the Venous Circulation 266 Disorders of Blood Flow Caused by Extravascular Forces 269	254	Congenital and Hereditary Disorders of the Kidney Obstructive Disorders 417 Urinary Tract Infections 421 Disorders of Glomerular Function 425 Tubulointerstitial Disorders 429 Neoplasms 430
6 Alterations in Blood Pressure Control of Blood Pressure 274 Hypertension 279 Orthostatic Hypotension 287	274	24 Renal Failure Acute Renal Failure 433 Chronic Renal Failure 436 Renal Failure in Children and Elderly Persons 443
 Alterations in Cardiac Function Disorders of the Pericardium 291 Coronary Heart Disease 294 Myocardial and Endocardial Disease 302 Valvular Heart Disease 308 Heart Disease in Infants and Children 312 	291	 25 Alterations in Urine Elimination Control of Urine Elimination 446 Alterations in Bladder Function 449 Cancer of the Bladder 456 UNIT SEVEN Alterations in the
B Heart Failure and Circulatory Shock Candace Hennessy, Carol M. Porth Heart Failure 320 Circulatory Failure (Shock) 330 Heart Failure in Children and the Elderly 336	320	Gastrointestinal System 26 Structure and Function of the Gastrointestinal System Structure and Organization of the Gastrointestinal Tract 460
Exchange of Gases Between the Atmosphere	339 340	Innervation and Motility 463 Hormonal and Secretory Function 466 Digestion and Absorption 468 Anorexia, Nausea, and Vomiting 471 27 Alterations in Gastrointestinal Function Disorders of the Esophagus 473
and the Lungs 344 Exchange and Transport of Gases 351 Control of Breathing 356 O Alterations in Respiratory Function:		Disorders of the Stomach 476 Disorders of the Small and Large Intestines 480

Infectious Disorders and Neoplasia

Respiratory Disorders in Infants and Children

Respiratory Tract Infections

Cancer of the Lung

494	35 Sexually Transmitted Diseases <i>Patricia McCowen Mehring</i> Infections of the External Genitalia 627 Vaginal Infections 630 Vaginal-Urogenital-Systemic Infections 632	27
517	UNIT TEN Alterations in the Nervous System	
517	 36 Organization and Control of the Nervous System Edward W. Carroll, Robin L. Curtis Nervous Tissue Cells 637 Nerve Cell Communication 641 Development and Organization of the Nervous System 644 The Spinal Cord and Brain 651 The Autonomic Nervous System 660 	37
529 538	37 Alterations in Brain Function <i>Diane Book</i> Brain Injury 668 Cerebrovascular Disease 679 Infections and Neoplasms 685 Seizure Disorders 688 Dementias 691	67
560	38 Alterations in Neuromuscular FunctionFunction69Carol M. Porth, Robin L. Curtis696The Organization and Control of Motor Function696Skeletal Muscle and Peripheral Nerve Disorders702Basal Ganglia and Cerebellum Disorders709Upper and Lower Motoneuron Disorders713Spinal Cord Injury716	96
	Elizabeth C. Devine Organization and Control of Somatosensory Function 725 Pain 731 Alterations in Pain Sensitivity and Special Types	25
581 581 602	of Pain 739 Headache 741 Pain in Children and Older Adults 744 40 Alterations in Special Sensory Function 74 Edward W. Carroll, Susan A. Fontana The Eye and Disorders of Vision 747 The Ear and Disorders of Auditory Function 763 The Vestibular System and Maintenance of Equilibrium 772	47
17		

28 Alterations in Hepatobiliary Function

The Liver and Hepatobiliary System 494 Disorders of the Liver 500 Disorders of the Gallbladder and Extrahepatic Bile Ducts 510 Disorders of the Exocrine Pancreas 513

29 Alterations in Body Nutrition

Joan Pleuss Regulation of Food Intake and Energy Metabolism 517 Overnutrition and Obesity 521 Undernutrition 524

UNIT EIGHT

Alterations in the Endocrine System

30 Organization and Control of the Endocrine System

Glenn Matfin, Safak Guven, Julie A. Kuenzi The Endocrine System 529

31 Alterations in Pituitary, Thyroid, Parathyroid, and Adrenal Function

Glenn Mattin, Safak Guven, Julie A. KuenziGeneral Aspects of Altered Endocrine Function538Pituitary and Growth Hormone Disorders539Thyroid Disorders545Parathyroid Hormone Disorders551Disorders of Adrenal Cortical Function552

32 Diabetes Mellitus Safak Guven, Julie A. Kuenzi, Glenn Matfin Energy Metabolism 560 Diabetes Mellitus 565

UNIT NINE

Alterations in the Male and Female Reproductive Systems

33	Alterations in the Male
	Reproductive System 581
	Physiologic Basis of Male Reproductive Function 581
	Disorders of the Penis, the Scrotum and Testes, and the Prostate 587
	Disorders in Childhood and Aging Changes 597
21	Alterations in the Female

34 Alterations in the Female Reproductive System

Patricia McCowen MehringStructure and Function of the Female
Reproductive System602Disorders of the Female Reproductive Organs607Menstrual Disorders619Disorders of the Breast622

814

UNIT ELEVEN Alterations in the Skeletal and Integumentary Systems

41 Structure and Function of the Skeletal System Characteristics of Skeletal Tissue 779 Skeletal Structures 785	779
42 Alterations in the Skeletal System:	
Trauma, Infection, and Developmental Disorders	789
Kathleen E. Gunta Injury and Trauma of Musculoskeletal Structures Bone Infections 799 Neoplasms 802 Disorders of Skeletal Growth and Development in Children 805	789
43 Alterations in the Skeletal System:	

Alterations in the Skeletal Sy Metabolic and Rheumatoid Disorders Debra Bancroft Rizzo, Kathleen E. Gunta

Metabolic Bone Disease 814 Rheumatic Disorders 820 Rheumatic Diseases in Children and the Elderly 832

44	Alterations in Skin Function	
	and Integrity	835
	Gladys Simandl	
	Structure of the Skin 835	
	Manifestations of Skin Disorders 839	
	Skin Damage Caused by Ultraviolet Radiation 841 Primary Disorders of the Skin 842	
	Nevi and Skin Cancers 853	
	Age-Related Skin Manifestations 857	
	5	
	Glossary	863
	·	
	APPENDIX A: Lab Values	873
	APPENDIX B: Weblinks	875
		0/3
	Index	879

One Mechanisms of Disease

CHAPTER



Cell Structure and Function

Functional Components of the Cell The Nucleus The Cytoplasm and Its Organelles **Ribosomes Endoplasmic Reticulum Golgi Complex** Lysosomes and Peroxisomes Mitochondria The Cytoskeleton **Microtubules Microfilaments The Cell Membrane Cell Metabolism and Energy Sources** Anaerobic Metabolism Aerobic Metabolism Cell Membrane Transport, Signal Transduction, and Generation of Membrane Potentials Movement Across the Cell Membrane **Passive Movement** Active Transport and Cotransport Endocytosis and Exocytosis Ion Channels

Signal Transduction and Cell Communication **Cell Membrane Receptors** Messenger-Mediated Control of Nuclear Function **Membrane** Potentials **Electrical Potentials Action Potentials Body Tissues Cell Differentiation Embryonic Origin of Tissue Types Epithelial Tissue Simple Epithelium** Stratified and Pseudostratified Epithelium **Glandular Epithelium Connective or Supportive Tissues** Loose Connective Tissue **Dense Connective Tissue Muscle Tissue Skeletal Muscle** Smooth Muscle **Cell Junctions and Cell-to-Cell Adhesion**

he cell is the smallest functional unit that an organism can be divided into and retain the characteristics necessary for life. Cells with similar embryonic origin or function are often organized into larger functional units called *tissues*. These tissues in turn combine to form the various body structures and organs. Although the cells of different tissues and organs vary in structure and function, certain characteristics are common to all cells. Cells are remarkably similar in their ability to exchange materials with their immediate environment, obtaining energy from organic nutrients, synthesizing complex molecules, and replicating themselves. Because most disease processes are initiated at the cellular level, an understanding of cell function is crucial to understanding the disease process. Some diseases affect the cells of a single organ, others affect the cells of a particular tissue type, and still others affect the cells of the entire organism.

FUNCTIONAL COMPONENTS OF THE CELL

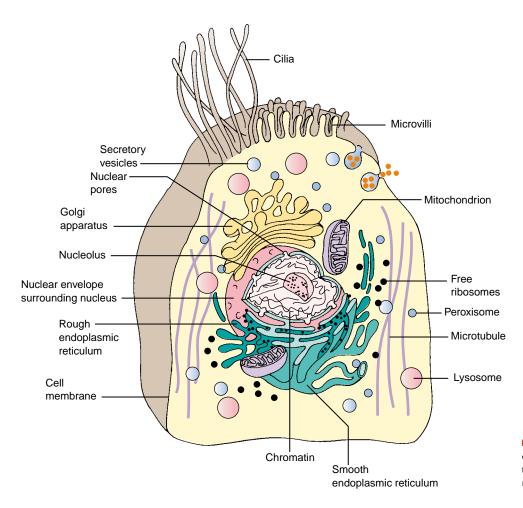
Although diverse in their organization, all eukaryotic cells (cells with a true nucleus) have in common structures that perform unique functions. When seen under a light microscope, three major components of the eukaryotic cell become evident: the nucleus, the cytoplasm, and the cell membrane (Fig. 1-1).

The internal matrix of the cell is called *protoplasm*. Protoplasm is composed of water, proteins, lipids, carbohydrates, and electrolytes. Water makes up 70% to 85% of the cell's protoplasm. The second most abundant constituents (10% to 20%) of protoplasm are the cell proteins, which form cell structures and the enzymes necessary for cellular reactions. Proteins can also be found complexed to other compounds as nucleo-

proteins, glycoproteins, and lipoproteins. Lipids comprise 2% to 3% of most cells. The most important lipids are the phospholipids and cholesterol, which are mainly insoluble in water; they combine with proteins to form the cell membrane and the membranous barriers that separate different cell compartments. Some cells also contain large quantities of triglycerides. In the fat cells, triglycerides can constitute up to 95% of the total cell mass. The fat stored in these cells represents stored energy, which can be mobilized and used wherever it is needed in the body. Few carbohydrates are found in the cell, and these are used primarily for fuel. Potassium, magnesium, phosphate, sulfate, and bicarbonate ions are the major intracellular electrolytes. Small quantities of sodium, chloride, and calcium ions are also present in the cell. These electrolytes facilitate the generation and transmission of electrochemical impulses in nerve and muscle cells. Intracellular electrolytes participate in reactions that are necessary for cellular metabolism.

The Nucleus

The nucleus of the cell appears as a rounded or elongated structure situated near the center of the cell (see Fig. 1-1). It is enclosed in a nuclear membrane and contains chromatin and a distinct region called the *nucleolus*. All eukaryotic cells have at least one nucleus (prokaryotic cells, such as bacteria, lack a nucleus and nuclear membrane). The nucleus is the control center for the cell. It contains deoxyribonucleic acid (DNA) that is



■ FIGURE 1-1 ■ Composite cell designed to show in one cell all of the various components of the nucleus and cytoplasm.

KEY CONCEPTS

THE FUNCTIONAL ORGANIZATION OF THE CELL

- Cells are the smallest functional unit of the body. They contain structures that are strikingly similar to those needed to maintain total body function.
- The nucleus is the control center for the cell. It also contains most of the hereditary material.
- The organelles, which are analogous to the organs of the body, are contained in the cytoplasm. They include the mitochondria, which supply the energy needs of the cell; the ribosomes, which synthesize proteins and other materials needed for cell function; and the lysosomes, which function as the cell's digestive system.

The cell membrane encloses the cell and provides for intracellular and intercellular communication, transport of materials into and out of the cell, and maintenance of the electrical activities that power cell function.

essential to the cell because its genes contain the information necessary for the synthesis of proteins that the cell must produce to stay alive. These proteins include structural proteins and enzymes used to synthesize other substances, including carbohydrates and lipids. The genes also represent the individual units of inheritance that transmit information from one generation to another. The nucleus is also the site of ribonucleic acid (RNA) synthesis. There are three types of RNA: messenger RNA (mRNA), which copies and carries the DNA instructions for protein synthesis to the cytoplasm; ribosomal RNA (rRNA), which moves to the cytoplasm, and becomes the site of protein synthesis; and transfer RNA (tRNA), which also moves into the cytoplasm, where it transports amino acids to the elongating protein as it is being synthesized (see Chapter 3).

The complex structure of DNA and DNA-associated proteins dispersed in the nuclear matrix is called *chromatin*. Each DNA molecule is made up of two extremely long, double-stranded helical chains containing variable sequences of four nitrogenous bases. These bases form the genetic code. In cells that are about to divide, the DNA must be replicated before *mitosis*, or cell division, occurs. During replication, complementary pairs of DNA are generated such that each daughter cell receives an identical set of genes.

The nucleus also contains the darkly stained round body called the *nucleolus*. The rRNA is transcribed exclusively in the nucleolus. Nucleoli are structures composed of regions from five different chromosomes, each with a part of the genetic code needed for the synthesis of rRNA. Cells that are actively synthesizing proteins can be recognized because their nucleoli are large and prominent and the nucleus as a whole is euchromatic.

Surrounding the nucleus is a doubled-layered membrane called the nuclear envelope or nuclear membrane. The nuclear

membrane contains many structurally complex circular pores where the two membranes fuse to form a gap. Many classes of molecules, including fluids, electrolytes, RNA, some proteins, and perhaps some hormones, can move in both directions through the nuclear pores.

The Cytoplasm and Its Organelles

The cytoplasm surrounds the nucleus, and it is in the cytoplasm that the work of the cell takes place. Cytoplasm is essentially a colloidal solution that contains water, electrolytes, suspended proteins, neutral fats, and glycogen molecules. Although they do not contribute to the cell's function, pigments may also accumulate in the cytoplasm. Some pigments, such as melanin, which gives skin its color, are normal constituents of the cell.

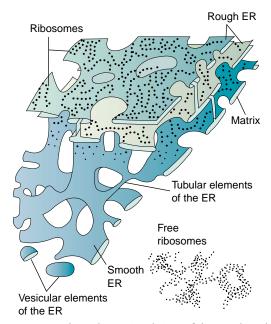
Embedded in the cytoplasm are various *organelles*, which function as the organs of the cell. These organelles include the ribosomes, endoplasmic reticulum, Golgi complex, lysosomes and peroxisomes, and mitochondria.

Ribosomes

The ribosomes serve as sites of protein synthesis in the cell. They are small particles of nucleoproteins (rRNA and proteins) that can be found attached to the wall of the endoplasmic reticulum or as free ribosomes (Fig. 1-2). Free ribosomes are scattered singly in the cytoplasm or joined by strands of mRNA to form functional units called *polyribosomes*. Free ribosomes are involved in the synthesis of proteins, mainly as intracellular enzymes.

Endoplasmic Reticulum

The endoplasmic reticulum (ER) is an extensive system of paired membranes and flat vesicles that connects various parts of the inner cell (see Fig. 1-2). The fluid-filled space, called the



■ FIGURE 1-2 ■ Three-dimensional view of the rough endoplasmic reticulum (ER) with its attached ribosomal RNA and the smooth endoplasmic reticulum.

Unit One: Mechanisms of Disease

matrix, between the paired ER membrane layers is connected with the space between the two membranes of the doublelayered nuclear membrane, the cell membrane, and various cytoplasmic organelles. It functions as a tubular communication system through which substances can be transported from one part of the cell to another. A large surface area and multiple enzyme systems attached to the ER membranes also provide the machinery for a major share of the metabolic functions of the cell.

Two forms of ER exist in cells: rough and smooth. Rough ER is studded with ribosomes attached to specific binding sites on the membrane. The ribosomes, with the accompanying strand of mRNA, synthesize proteins. Proteins produced by the rough ER are usually destined for incorporation into cell membranes and lysosomal enzymes or for exportation from the cell. The rough ER segregates these proteins from other components of the cytoplasm and modifies their structure for a specific function. For example, the production of plasma protein by liver cells take place in the rough ER. All cells require a rough ER for the synthesis of lysosomal enzymes.

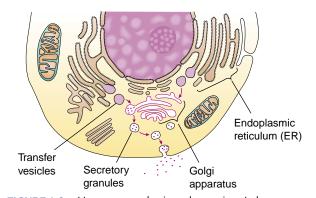
The smooth ER is free of ribosomes and is continuous with the rough ER. It does not participate in protein synthesis; instead, its enzymes are involved in the synthesis of lipid molecules, regulation of intracellular calcium, and metabolism and detoxification of certain hormones and drugs. It is the site of lipid, lipoprotein, and steroid hormone synthesis. The sarcoplasmic reticulum of skeletal and cardiac muscle cells is a form of smooth ER. Calcium ions needed for muscle contraction are stored and released from cisternae of the sarcoplasmic reticulum. Smooth ER of the liver is involved in glycogen storage and metabolism of lipid-soluble drugs.

Golgi Complex

The Golgi apparatus, sometimes called the Golgi complex, consists of stacks of thin, flattened vesicles or sacs. These Golgi bodies are found near the nucleus and function in association with the ER. Substances produced in the ER are carried to the Golgi complex in small, membrane-covered transfer vesicles. Many cells synthesize proteins that are larger than the active product. Insulin, for example, is synthesized as a large, inactive proinsulin molecule that is cut apart to produce a smaller, active insulin molecule within the Golgi complex of the beta cells of the pancreas. The Golgi complex modifies these substances and packages them into secretory granules or vesicles. Enzymes destined for export from the cell are packaged in secretory vesicles. After appropriate signals, the secretory vesicles move out of the Golgi complex into the cytoplasm and fuse to the inner side of the plasma membrane, where they release their contents into the extracellular fluid. Figure 1-3 is a diagram of the synthesis and movement of a hormone through the rough ER and Golgi complex. In addition to its function in producing secretory granules, the Golgi complex is thought to produce large carbohydrate molecules that combine with proteins produced by the rough ER to form glycoproteins.

Lysosomes and Peroxisomes

The lysosomes can be viewed as the digestive system of the cell. They consist of small, membrane-enclosed sacs containing hydrolytic enzymes capable of breaking down worn-out cell parts so they can be recycled. They also break down foreign substances such as bacteria taken into the cell. All of the lysosomal



■ FIGURE 1-3 ■ Hormone synthesis and secretion. In hormone secretion, the hormone is synthesized by the ribosomes attached to the rough endoplasmic reticulum. It moves from the rough ER to the Golgi complex, where it is stored in the form of secretory granules. These leave the Golgi complex and are stored within the cytoplasm until released from the cell in response to an appropriate signal.

enzymes are acid hydrolases, which means that they require an acid environment. The lysosomes provide this environment by maintaining a pH of approximately 5 in their interior. The pH of the cytoplasm is approximately 7.2, which protects other cellular structures from this acidity.

Lysosomal enzymes are synthesized in the rough ER and then transported to the Golgi apparatus, where they are biochemically modified and packaged as lysosomes. Unlike those of other organelles, the sizes and functions of lysosomes vary considerably from one cell to another. The type of enzyme packaged in the lysosome by the Golgi complex determines this diversity. Although enzymes in the secondary lysosomes can break down most proteins, carbohydrates, and lipids to their basic constituents, some materials remain undigested. These undigested materials may remain in the cytoplasm as residual bodies or be extruded from the cell. In some long-lived cells, such as neurons and heart muscle cells, large quantities of residual bodies accumulate as lipofuscin granules or age pigment. Other indigestible pigments, such as inhaled carbon particles and tattoo pigments, also accumulate and may persist in residual bodies for decades.

Lysosomes play an important role in the normal metabolism of certain substances in the body. In some inherited diseases known as *lysosomal storage diseases*, a specific lysosomal enzyme is absent or inactive, in which case the digestion of certain cellular substances does not occur. As a result, these substances accumulate in the cell. In Tay-Sachs disease, an autosomal recessive disorder, the lysosomal enzyme needed for degrading the GM₂ ganglioside found in nerve cell membranes is deficient (see Chapter 4). Although GM₂ ganglioside accumulates in many tissues, such as the heart, liver, and spleen, its accumulation in the nervous system and retina of the eye causes the most damage.

Smaller than lysosomes, spherical membrane-bound organelles called *peroxisomes* contain a special enzyme that degrades peroxides (*e.g.*, hydrogen peroxide). Peroxisomes function in the control of free radicals (see Chapter 2). Unless degraded, these highly unstable chemical compounds would otherwise damage other cytoplasmic molecules. For example, catalase degrades toxic hydrogen peroxide molecules to water. Peroxisomes also contain the enzymes needed for breaking down very-long-chain fatty acids, which are ineffectively degraded by mitochondrial enzymes. In liver cells, peroxisomal enzymes are involved in the formation of the bile acids.

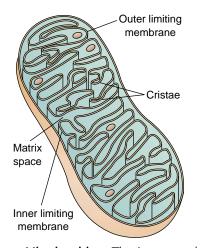
Mitochondria

The mitochondria are literally the "power plants" of the cell because they transform organic compounds into energy that is easily accessible to the cell. Energy is not made here but is extracted from organic compounds. Mitochondria contain the enzymes needed for capturing most of the energy in foodstuffs and converting it into cellular energy. This multistep process requires oxygen and is often referred to as *aerobic metabolism*. Much of this energy is stored in the high-energy phosphate bonds of compounds such as adenosine triphosphate (ATP), which powers the various cellular activities.

Mitochondria are found close to the site of energy consumption in the cell (*e.g.*, near the myofibrils in muscle cells). The number of mitochondria in a given cell type is largely determined by the type of activity the cell performs and how much energy is needed to undertake this activity. For example, large increases in mitochondria have been observed in skeletal muscle that has been repeatedly stimulated to contract.

The mitochondria are composed of two membranes: an outer membrane that encloses the periphery of the mitochondrion and an inner membrane that forms shelflike projections, called *cristae* (Fig. 1-4). The outer and inner membranes form two spaces: an outer intramembranous space and an inner matrix that is filled with a gel-like material. The outer membrane is involved in lipid synthesis and fatty acid metabolism. The inner membrane contains the respiratory chain enzymes and transport proteins needed for the synthesis of ATP.

Mitochondria contain their own DNA and ribosomes and are self-replicating. The DNA is found in the mitochondrial matrix and is distinct from the chromosomal DNA found in the nucleus. Mitochondrial DNA, known as the "other human genome," is a double-stranded, circular molecule that encodes the rRNA and tRNA required for intramitochrondial synthesis of proteins needed for the energy-generating function of the



■ FIGURE 1-4 ■ Mitochondrion. The inner membrane forms transverse folds called cristae, where the enzymes needed for the final step in adenosine triphosphate (ATP) production (*i.e.*, oxidative phosphorylation) are located.

mitochondria. Although mitochondrial DNA directs the synthesis of 13 of the proteins required for mitochondrial function, the DNA of the nucleus encodes the structural proteins of the mitochondria and other proteins needed to carry out cellular respiration.

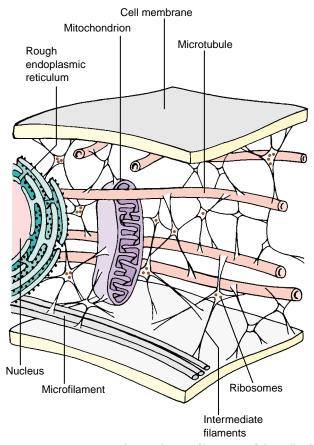
Mitochondrial DNA is inherited matrilineally (*i.e.*, from the mother) and provides a basis for familial lineage studies. Mutations have been found in each of the mitochondrial genes, and an understanding of the role of mitochondrial DNA in certain diseases is beginning to emerge. Most tissues in the body depend to some extent on oxidative metabolism and can therefore be affected by mitochondrial DNA mutations.

The Cytoskeleton

In addition to its organelles, the cytoplasm contains a network of microtubules, microfilaments, intermediate filaments, and thick filaments (Fig. 1-5). Because they control cell shape and movement, these structures are a major component of the structural elements called the *cytoskeleton*.

Microtubules

The microtubules are slender tubular structures composed of globular proteins called *tubulin*. Microtubules function in many ways, including the development and maintenance of



■ FIGURE 1-5 ■ Microtubes and microfilaments of the cell. The microfilaments associate with the inner surface of the cell and aid in cell motility. The microtubules form the cytoskeleton and maintain the position of the organelles.

cell form; participation in intracellular transport mechanisms, including axoplasmic transport in neurons; and formation of the basic structure for several complex cytoplasmic organelles, including the cilia, flagella, centrioles, and basal bodies. Abnormalities of the cytoskeleton may contribute to alterations in cell mobility and function. For example, proper functioning of the microtubules is essential for various stages of leukocyte migration.

Cilia and Flagella. Cilia and flagella are hairlike processes extending from the cell membrane that are capable of sweeping and flailing movements, which can move surrounding fluids or move the cell through fluid media. Cilia are found on the apical or luminal surface of epithelial linings of various body cavities or passages, such as the upper respiratory system. Removal of mucus from the respiratory passages is highly dependent on the proper functioning of the cilia. Flagella form the tail-like structures that provide motility for sperm.

Centrioles and Basal Bodies. Centrioles and basal bodies are structurally identical organelles composed of an array of highly organized microtubules. The centrioles are small, barrel-shaped bodies oriented at right angles to each other. In dividing cells, the two cylindrical centrioles form the mitotic spindle that aids in the separation and movement of the chromosomes. Basal bodies are more numerous than centrioles and are found near the cell membrane in association with cilia and flagella.

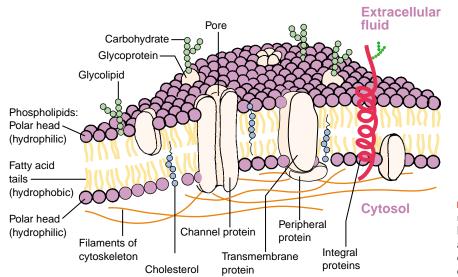
Microfilaments

Microfilaments are thin, threadlike cytoplasmic structures. Three classes of microfilaments exist: (1) thin microfilaments, which are equivalent to the thin actin filaments in muscle; (2) the thick myosin filaments, which are present in muscle cells but may also exist temporarily in other cells; and (3) the intermediate filaments, which are a heterogeneous group of filaments with diameter sizes between the thick and thin filaments. Muscle contraction depends on the interaction between the thin actin filaments and thick myosin filaments. Microfilaments are present in the superficial zone of the cytoplasm in most cells. Contractile activities involving the microfilaments and associated thick myosin filaments contribute to associated movement of the cytoplasm and cell membrane during endocytosis and exocytosis. Microfilaments are also present in the microvilli of the intestine. The intermediate filaments function in supporting and maintaining the asymmetric shape of cells. Examples of intermediate filaments are the keratin filaments that are found anchored to the cell membrane of epidermal keratinocytes of the skin and the glial filaments that are found in astrocytes and other glial cells of the nervous system. The *neurofibrillary tangle* found in the brain in Alzheimer's disease contains microtubule-associated proteins and neurofilaments, evidence of a disrupted neuronal cytoskeleton.

The Cell Membrane

The cell is enclosed in a thin membrane that separates the intracellular contents from the extracellular environment. To differentiate it from the other cell membranes, such as the mitochondrial or nuclear membranes, the cell membrane is often called the *plasma membrane*. In many respects, the plasma membrane is one of the most important parts of the cell. It acts as a semipermeable structure that separates the intracellular and extracellular environments. It provides receptors for hormones and other biologically active substances, participates in the electrical events that occur in nerve and muscle cells, and aids in the regulation of cell growth and proliferation.

The cell membrane consists of an organized arrangement of lipids (phospholipids, glycolipids, and cholesterol), carbohydrates, and proteins (Fig. 1-6). The lipids form a bilayer structure that is essentially impermeable to all but lipid-soluble substances. About 75% of the lipids are phospholipids, each with a hydrophilic (water-soluble) head and hydrophobic (water-insoluble) tails. The phospholipid molecules along with the glycolipids are aligned such that their hydrophobic heads face outward on each side of the membrane and their hydrophobic tails project toward the center of the membrane. The hydrophilic heads retain water and help cells adhere to each other. At



■ FIGURE 1-6 ■ The structure of the cell membrane showing the hydrophilic (polar) heads and the hydrophobic (fatty acid) tails and the position of the integral and peripheral proteins in relation to the interior and exterior of the cell.

Although the basic structure of the cell membrane is provided by the lipid bilayer, most of the specific functions are carried out by proteins. Some proteins, called transmembrane proteins, pass directly through the membrane and communicate with the intracellular and extracellular environments. Many of the transmembrane proteins are tightly bound to lipids in the bilayer and are essentially part of the membrane. These transmembrane proteins are called *integral proteins*. The *peripheral* proteins, a second type of protein, are bound to one or the other side of the membrane and do not pass into the lipid bilayer. Thus, the peripheral proteins are associated with functions involving the inner and outer side of the membrane where they are located. In contrast, the transmembrane proteins can function on both sides of the membrane or transport molecules across it. Many of the integral transmembrane proteins form the ion channels found on the cell surface. These channel proteins have complex structures and are selective with respect to the ions that pass through their channels.

The membrane carbohydrates are incorporated in a fuzzylooking layer, called the *cell coat* or *glycocalyx*, which surrounds the cell surface. The glycocalyx, which is part of the cell membrane, consists of long, complex carbohydrate chains that are attached to proteins and lipids in the form of glycoproteins and glycolipids. The cell coat participates in cell-to-cell recognition and adhesion. It contains tissue transplant antigens that label cells as self or nonself. ABO blood group antigens are contained in the cell coat of red blood cells.

In summary, the cell is a remarkably autonomous structure that functions in a manner strikingly similar to that of the total organism. The nucleus controls cell function and is the mastermind of the cell. It contains DNA, which provides the information necessary for the synthesis of the various proteins that the cell must produce to stay alive and to transmit information from one generation to another.

The cytoplasm contains the cell's organelles. Ribosomes serve as sites for protein synthesis in the cell. The ER functions as a tubular communication system through which substances can be transported from one part of the cell to another and as the site of protein (rough ER), carbohydrate, and lipid (smooth ER) synthesis. Golgi bodies modify materials synthesized in the ER and package them into secretory granules for transport within the cell or for export from the cell. Lysosomes, which can be viewed as the digestive system of the cell, contain hydrolytic enzymes that digest worn-out cell parts and foreign materials. The mitochondria serve as power plants for the cell because they transform food energy into ATP, which is used to power cell activities. Mitochondria contain their own extrachromosomal DNA, which is used in the synthesis of mitochondrial RNAs and proteins used in oxidative metabolism. Microtubules are slender, stiff tubular structures that influence cell shape, provide a means of moving organelles through the cytoplasm, and effect movement of the cilia and of chromosomes during cell division. Several types of threadlike filaments, including actin and myosin filaments, participate in muscle contraction.

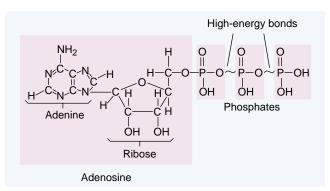
The plasma membrane is a lipid bilayer that surrounds the cell and separates it from its surrounding external environment. It contains receptors for hormones and other biologically active substances, participates in the electrical events that occur in nerve and muscle cells, and aids in the regulation of cell growth and proliferation. The cell surface is surrounded by a fuzzy-looking layer called the cell coat or glycocalyx. The cell coat participates in cell-to-cell recognition and adhesion, and it contains tissue transplant antigens.

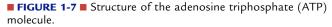
CELL METABOLISM AND ENERGY SOURCES

Energy metabolism refers to the processes by which fats, proteins, and carbohydrates from the foods we eat are converted into energy or complex energy sources in the cell. Catabolism and anabolism are the two phases of metabolism. *Catabolism* consists of breaking down stored nutrients and body tissues to produce energy. *Anabolism* is a constructive process in which more complex molecules are formed from simpler ones.

The special carrier for cellular energy is ATP. ATP molecules consist of adenosine, a nitrogenous base; ribose, a fivecarbon sugar; and three phosphate groups (see Fig. 1-7). The last two phosphate groups are attached to the remainder of the molecule by two high-energy bonds, which are indicated by the symbol ~. Each bond releases a large amount of energy when hydrolyzed. ATP is hydrolyzed to form adenosine diphosphate (ADP) with the loss of one high-energy bond and to adenosine monophosphate (AMP) with the loss of two such bonds. The energy liberated from the hydrolysis of ATP is used to drive reactions that require free energy, such as muscle contraction and active transport mechanisms. Energy from foodstuffs is used to convert ADP back to ATP. ATP is often called the *energy currency* of the cell; energy can be "saved" or "spent" using ATP as an exchange currency.

Two types of energy production are present in the cell: the anaerobic (*i.e.*, without oxygen) glycolytic pathway, occurring in the cytoplasm, and the aerobic (*i.e.*, with oxygen) pathways occurring in the mitochondria. The glycolytic pathway serves as the prelude to the aerobic pathways.

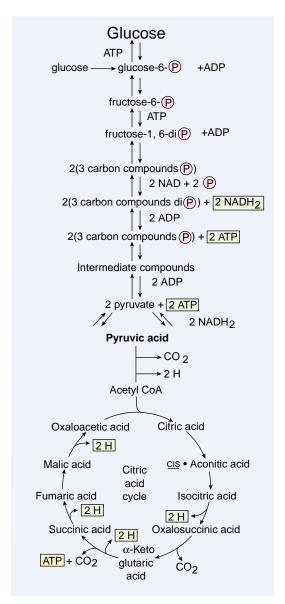




Anaerobic Metabolism

Glycolysis is anaerobic process by which energy is liberated from glucose (Fig. 1-8). It is an important source of energy for cells that lack mitochondria. This process provides energy in situations when delivery of oxygen to the cell is delayed or impaired. Glycolysis involves a sequence of reactions that converts glucose to pyruvate, with the concomitant production of ATP from ADP. The net gain of energy from the glycolysis of one molecule of glucose is two ATP molecules. Although relatively inefficient as to energy yield, the glycolytic pathway is important during periods of decreased oxygen delivery, such as occurs in skeletal muscle during the first few minutes of exercise.

Glycolysis requires the presence of nicotinamide-adenine dinucleotide (NAD⁺), a hydrogen carrier. The end-products of glycolysis are pyruvate and NADH. When oxygen is present, pyruvate moves into the aerobic mitochondrial pathway, and



■ FIGURE 1-8 ■ Glycolytic pathway and citric acid cycle.

NADH subsequently enters into oxidative chemical reactions that remove the hydrogen atoms. The transfer of hydrogen from NADH during the oxidative reactions allows the glycolytic process to continue by facilitating the regeneration of NAD⁺. Under anaerobic conditions, such as cardiac arrest or circulatory shock, pyruvate is converted to lactic acid, which diffuses out of the cells into the extracellular fluid. Conversion of pyruvate to lactic acid is reversible, and after the oxygen supply has been restored, lactic acid is reconverted back to pyruvate and used directly for energy or to synthesize glucose.

Aerobic Metabolism

Aerobic metabolism, which supplies 90% of the body's energy needs, occurs in the cell's mitochondria and requires oxygen. It is here that hydrogen and carbon molecules from dietary fats, proteins, and carbohydrates are broken down and combined with molecular oxygen to form carbon dioxide, and water as energy is released. Unlike lactic acid, which is an end-product of anaerobic metabolism, carbon dioxide and water are relatively harmless and easily eliminated from the body. In a 24-hour period, oxidative metabolism produces 300 to 500 mL of water.

Aerobic metabolism uses the *citric acid cycle*, sometimes called the *tricarboxylic acid* or *Krebs cycle*, as the final common pathway for the metabolism of nutrients (see Fig. 1-8). In the citric acid cycle, each of the two pyruvate molecules formed in the cytoplasm from the glycolysis of one molecule of glucose yields another molecule of ATP along with two molecules of carbon dioxide and eight hydrogen ions. In addition to pyruvate from the glycolysis of glucose, products from amino acid and fatty acid breakdown enter the citric acid cycle.

In the initial stage of the citric acid cycle, acetyl coenzyme A (acetyl-CoA) combines with oxaloacetic acid to form citric acid. The coenzyme A portion of acetyl-CoA can be used again and again to generate more acetyl-CoA from pyruvate, while the acetyl portion becomes part of the citric acid cycle and moves through a series of enzyme-mediated steps that produce carbon dioxide and hydrogen atoms. The carbon dioxide is carried to the lungs and exhaled. The hydrogen atoms are transferred to the electron transport system on the inner mitochondrial membrane for oxidation. Oxidation of hydrogen is accomplished through a series of enzyme-mediated steps that change the hydrogen atoms to hydrogen ions and electrons. The electrons are used to reduce elemental oxygen, which combines with the hydrogen ions to form water. During this sequence of oxidative reactions, large amounts of energy are released and used to convert ADP to ATP. Because the formation of ATP involves the addition of a high-energy phosphate bond to ADP, the process is called oxidative phosphorylation. Cyanide poisoning kills by binding to the enzymes needed for a final step in the oxidative phosphorylation sequence.

In summary, metabolism is the process whereby the carbohydrates, fats, and proteins we eat are broken down and subsequently converted into the energy needed for cell function. Energy is stored in the high-energy phosphate bonds of ATP, which serves as the energy currency for the cell. Two sites of energy conversion are present in cells: the glycolytic or anaerobic pathway in the cell's cytoplasmic matrix and the aerobic or citric acid cycle in the mitochondria. The most efficient of these pathways is the citric acid pathway. This pathway, which requires oxygen, produces carbon dioxide and water as end-products and results in the release of large amounts of energy that is used to convert ADP to ATP. The glycolytic pathway, which is located in the cytoplasm, involves the breakdown of glucose to form ATP. This pathway can function without oxygen by producing lactic acid.

CELL MEMBRANE TRANSPORT, SIGNAL TRANSDUCTION, AND GENERATION OF MEMBRANE POTENTIALS

Movement Across the Cell Membrane

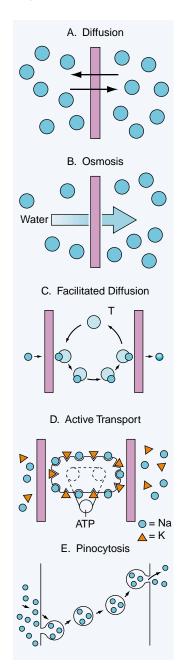
The unique properties of the cell's membrane are responsible for differences in the composition of the intracellular and extracellular fluids. However, a constant movement of molecules and ions across the cell membrane is required to maintain the functions of the cell. Movement through the cell membrane occurs in essentially two ways: passively, without an expenditure of energy, or actively, using energy-consuming processes. The cell membrane can also engulf substances, forming a membrane-coated vesicle; this membrane-coated vesicle is moved into the cell by *endocytosis* or out of the cell by *exocytosis*.

Passive Movement

The passive movement of particles or ions across the cell membrane is directly influenced by chemical or electrical gradients and does not require an expenditure of energy. A difference in the number of particles on either side of the membrane creates a chemical gradient, and a difference in charged particle or ions creates an electrical gradient. Chemical and electrical gradients are often linked and are called *electrochemical gradients*.

Diffusion. Diffusion refers to the process by which molecules and other particles in a solution become widely dispersed and reach a uniform concentration because of energy created by their spontaneous kinetic movements (Fig. 1-9). In the process of reaching a uniform concentration, these molecules and particles move from an area of higher to an area of lower concentration. With ions, diffusion is affected by energy supplied by their electrical charge. Lipid-soluble molecules, such as oxygen, carbon dioxide, alcohol, and fatty acids, become dissolved in the lipid matrix of the cell membrane and diffuse through the membrane in the same manner that diffusion occurs in water. Other substances diffuse through minute pores of the cell membrane. The rate of movement depends on how many particles are available for diffusion and the velocity of the kinetic movement of the particles. Temperature changes the motion of the particles; the greater the temperature, the greater is the thermal motion of the molecules.

Osmosis. Most cell membranes are semipermeable in that they are permeable to water but not all solute particles. Water moves through a semipermeable membrane along a concentration gradient, moving from an area of higher to one of lower con-



■ FIGURE 1-9 ■ Mechanisms of membrane transport. (A) Diffusion, in which particles move to become equally distributed across the membrane. (B) The osmotically active particles regulate the flow of water. (C) Facilitated diffusion uses a carrier system. (D) In active transport, selected molecules are transported across the membrane using the energy-driven (ATP) pump. (E) The membrane forms a vesicle that engulfs the particle and transports it across the membrane, where it is released. This is called pinocytosis.

centration (see Fig. 1-9). This process is called *osmosis*, and the pressure that water generates as it moves through the membrane is called *osmotic pressure*.

Osmosis is regulated by the concentration of nondiffusible particles on either side of a semipermeable membrane. When there is a difference in the concentration of particles, water moves from the side with the lower concentration of particles and higher concentration of water to the side with the higher concentration of particles and lower concentration of water. The movement of water continues until the concentration of particles on both sides of the membrane is equally diluted or until the hydrostatic (osmotic) pressure created by the movement of water opposes its flow.

Facilitated Diffusion. Facilitated diffusion occurs through a transport protein that is not linked to metabolic energy (see Fig. 1-9). Some substances, such as glucose, cannot pass unassisted through the cell membrane because they are not lipid soluble or are too large to pass through the membrane's pores. These substances combine with special transport proteins at the membrane's outer surface, are carried across the membrane attached to the transporter, and then released. In facilitated diffusion, a substance can move only from an area of higher concentration to one of lower concentration. The rate at which a substance moves across the membrane because of facilitated diffusion depends on the difference in concentration between the two sides of the membrane. Also important are the availability of transport proteins and the rapidity with which they can bind and release the substance being transported. It is thought that insulin, which facilitates the movement of glucose into cells, acts by increasing the availability of glucose transporters in the cell membrane.

Active Transport and Cotransport

The process of diffusion describes particle movement from an area of higher concentration to one of lower concentration, resulting in an equal distribution across the cell membrane. However, sometimes different concentrations of a substance are needed in the intracellular and extracellular fluids. For example, the intracellular functioning of the cell requires a much higher concentration of potassium than is present in the extracellular fluid while maintaining a much lower concentration of sodium than in the extracellular fluid. In these situations, energy is required to pump the ions "uphill" or against their concentration gradient. When cells use energy to move ions against an electrical or chemical gradient, the process is called *active transport*.

The active transport system studied in the greatest detail is the sodium–potassium pump, or Na^+/K^+ ATPase pump (see Fig. 1-9). The Na^+/K^+ ATPase pump moves sodium from inside the cell to the extracellular region, where its concentration is approximately 14 times greater than inside; the pump also returns potassium to the inside, where its concentration is approximately 35 times greater than it is outside the cell. If it were not for the activity of the sodium–potassium pump, the osmotically active sodium particles would accumulate in the cell, causing cellular swelling because of an accompanying influx of water (see Chapter 2).

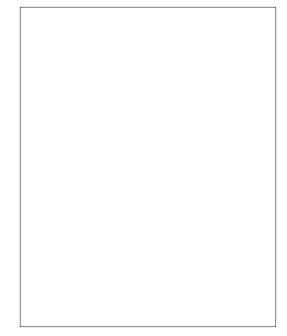
There are two types of active transport: primary active transport and secondary active transport. In *primary active transport*, the source of energy (*e.g.*, ATP) is used directly in the transport of a substance. *Secondary active transport* mechanisms harness the energy derived from the primary active transport of one substance, usually sodium ions, for the cotransport of a second substance. For example, when sodium ions are actively transported out of a cell by primary active transport, a large concentration gradient develops (*i.e.*, high concentration on the out-

side and low on the inside). This concentration gradient represents a large storehouse of energy because sodium ions are always attempting to diffuse into the cell. Similar to facilitated diffusion, secondary transport mechanisms use membrane transport proteins. These proteins have two binding sites: one for sodium ions and the other for the substance undergoing secondary transport. Secondary transport systems are classified into two groups: *cotransport*, in which the sodium ion and solute are transported in the same direction, and *countertransport*, in which sodium ions and the solute are transported in the opposite direction (Fig. 1-10). An example of cotransport occurs in the intestine, where the absorption of glucose and amino acids is coupled with sodium transport.

Endocytosis and Exocytosis

Endocytosis is the process by which cells engulf materials from their surroundings. It includes pinocytosis and phagocytosis. *Pinocytosis* involves the ingestion of small solid or fluid particles. The particles are engulfed into small, membranesurrounded vesicles for movement into the cytoplasm. The process of pinocytosis is important in the transport of proteins and strong solutions of electrolytes (see Fig. 1-9).

Phagocytosis literally means *cell eating* and can be compared with pinocytosis, which means *cell drinking*. Phagocytosis involves the engulfment and subsequent killing or degradation of microorganisms and other particulate matter. During phagocytosis, a particle contacts the cell surface and is surrounded on all sides by the cell membrane, forming a phagocytic vesicle or phagosome. Once formed, the phagosome breaks away from the cell membrane and moves into the cytoplasm, where it eventually fuses with a lysosome, allowing the ingested material to be degraded by lysosomal enzymes. Certain cells, such



■ FIGURE 1-10 ■ Secondary active transport systems. (A) carries the transported solute (S) in the same direction as the Na⁺ ion. (B) Counter-transport carries the solute and Na⁺ in the opposite direction.

as macrophages and polymorphonuclear leukocytes (neutrophils), are adept at engulfing and disposing of invading organisms, damaged cells, and unneeded extracellular constituents (see Chapter 9).

Exocytosis is the mechanism for the secretion of intracellular substances into the extracellular spaces. It is the reverse of endocytosis in that a secretory granule fuses to the inner side of the cell membrane, and an opening occurs in the cell membrane. This opening allows the contents of the granule to be released into the extracellular fluid. Exocytosis is important in removing cellular debris and releasing substances, such as hormones, synthesized in the cell.

During endocytosis, portions of the cell membrane become an endocytotic vesicle. During exocytosis, the vesicular membrane is incorporated into the plasma membrane. In this way, cell membranes can be conserved and reused.

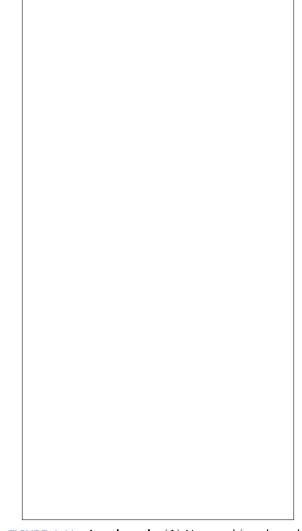
Ion Channels

The electrical charge on small ions such as Na⁺ and K⁺ makes it difficult for these ions to move across the lipid layer of the cell membrane. However, rapid movement of these ions is required for many types of cell functions, such as nerve activity. This is accomplished by facilitated diffusion through selective ion channels. Ion channels are made up of integral proteins that span the width of the cell membrane and are normally composed of several polypeptides or protein subunits that form a gating system. Specific stimuli cause the protein subunits to undergo conformational changes to form an open channel or gate through which the ions can move. In this way, ions do not need to cross the lipid-soluble portion of the membrane but can remain in the aqueous solution that fills the ion channel. Ion channels are highly selective; some channels allow only for passage of sodium ions, and others are selective for potassium, calcium, or chloride ions.

The plasma membrane contains two basic groups of ion channels: nongated and gated channels (Fig. 1-11). Nongated or leakage channels are open even in the unstimulated state, whereas gated channels open and close in response to specific stimuli. There are two types of gated channels: voltage-gated and ligand-gated channels. Voltage-gated channels have electrically operated gates that open when the membrane potential changes beyond a certain point. Ligand-gated channels have chemically operated gates that respond to specific receptorbound ligands, such as the neurotransmitter acetylcholine.

Signal Transduction and Cell Communication

Cells in multicellular organisms need to communicate with one another to coordinate their function and control their growth. Cells communicate with each other by means of chemical messenger systems. In some tissues, messengers move from cell to cell through gap junctions without entering the extracellular fluid. In other tissues, cells communicate by chemical messengers secreted into the extracellular fluid. Many types of chemical messengers that cannot cross the cell membrane bind to receptors on or near the cell surface. These chemical messengers are sometimes called *first messengers* because, by one means or another, their external signal is converted into internal signals carried by a second chemical called a *second messenger*. It is the second messenger that triggers the intracellular changes that produce the desired physiologic effect. Some



■ FIGURE 1-11 ■ Ion channels. (A) Nongated ion channel remains open, permitting free movement of ions across the membrane. (B) Ligand-gated channel is controlled by ligand binding to the receptor. (C) Voltage-gated channel is controlled by a change in membrane potential. (Rhoades R.A., Tanner G.A. [1996]. *Medical physiology.* Boston: Little, Brown)

lipid-soluble chemical messengers move through the membrane and bind to cytoplasmic or nuclear receptors to exert their physiologic effects.

Cell Membrane Receptors

Neurotransmitters, protein and peptide hormones, and other chemical messengers do not exert their effects by entering cells. Instead, they attach to receptors on the cell surface, and their messages are conveyed across the membrane and converted by cell membrane proteins into signals within the cell, a process often called *signal transduction*. Many molecules involved in signal transduction are proteins. A unique property of proteins that allows them to function in this way is their ability to change their shape or conformation, thereby changing their function and consequently the functions of the cell. These conformational changes are often accomplished through

KEY CONCEPTS

CELL COMMUNICATION

- Cells communicate with each other and with the internal and external environments by a number of mechanisms, including electrical and chemical signaling systems that control electrical potentials, the overall function of a cell, and gene activity needed for cell division and cell replication.
- Chemical messengers exert their effects by binding to cell membrane proteins or receptors that convert the chemical signal into signals within the cell, in a process called *signal transduction*.
- Cells regulate their responses to chemical messengers by increasing or decreasing the number of active receptors on their surface.

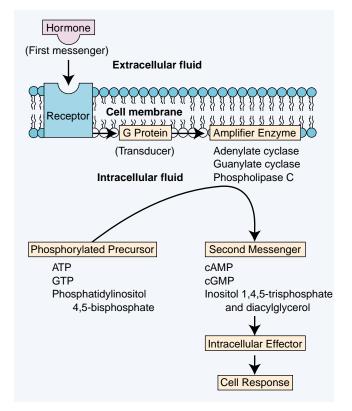
enzymes called *protein kinases* that catalyze the phosphorylation of amino acids in the protein structure.

Each cell type in the body contains a distinctive set of receptor proteins that enable it to respond to a complementary set of signaling molecules in a specific, preprogrammed way. These receptors, which span the cell membrane, relay information to a series of intracellular intermediates that eventually pass the signal to its final destination. Many receptors for chemical messengers have been isolated and characterized. These proteins are not static components of the cell membrane; they increase or decrease in number, according to the needs of the cell. When excess chemical messengers are present, the number of active receptors decreases in a process called *down-regulation*; when there is a deficiency of the messenger, the number of active receptors increases through *up-regulation*. There are three known classes of cell surface receptor proteins: ion channel linked, G protein linked, and enzyme linked.

lon-Channel–Linked Receptors. Ion-channel–linked receptors are involved in the rapid synaptic signaling between electrically excitable cells. This type of signaling is mediated by a small number of neurotransmitters that transiently open or close ion channels formed by integral proteins in the cell membrane. This type of signaling is involved in the transmission of impulses in nerve and muscle cells.

G-Protein–Linked Receptors and Signal Transduction. G proteins constitute the on–off switch for signal transduction. Although there are numerous intercellular messengers, many of them rely on a class of molecules called *G proteins* to convert external signals (first messengers) into internal signals (second messengers). These internal signals induce biochemical changes in the cell that lead to the desired physiologic effects. G proteins are so named because they bind to guanine nucleotides, such as guanine diphosphate (GDP) and guanine triphosphate (GTP).

G-protein-mediated signal transduction relies on a series of orchestrated biochemical events (Fig. 1-12). All signal trans-



■ FIGURE 1-12 ■ Signal transduction pattern common to several second messenger systems. A protein or peptide hormone is the first messenger to a membrane receptor, stimulating or inhibiting a membrane-bound enzyme by means of a G protein. The amplifier enzyme catalyzes the production of a second messenger from a phosphorylated precursor. The second messenger then activates an internal effector, which leads to the cell response. (Redrawn from Rhoades R.A., Tanner G.A. [1996]. *Medical physiology*. Boston: Little, Brown)

duction systems have a receptor component that functions as a signal discriminator by recognizing a specific first messenger. After a first messenger binds to a receptor, conformational changes occur in the receptor, which activates the G protein. The activated G protein, in turn, acts on other membranebound intermediates called *effectors*. Often, the effector is an enzyme that converts an inactive precursor molecule into a second messenger, which diffuses into the cytoplasm and carries the signal beyond the cell membrane.

Although there are differences between the G proteins, all share a number of features. All are found on the cytoplasmic side of the cell membrane, and all incorporate the GTPase cycle, which functions as the on–off switch for G-protein activity. Certain bacterial toxins can bind to the G proteins, causing inhibition or stimulation of its signal function. One such toxin, the toxin of *Vibrio cholerae*, binds and activates the stimulatory G protein linked to the cAMP system that controls the secretion of fluid into the intestine. In response to the cholera toxin, these cells overproduce fluid, leading to severe diarrhea and life-threatening depletion of extracellular fluid volume. There is also interest in the role that G-protein signaling may play in the pathogenesis of cancer. *Enzyme-Linked Receptors.* The receptors for certain protein hormones, such as insulin, and peptide growth factors activate an intracellular domain with enzyme (protein-tyrosine kinase) activity. The enzyme catalyzes the phosphorylation of tyrosine residues of intracellular proteins, thereby transferring an external message to the cell interior. Enzyme-linked receptors mediate cellular responses such as calcium influx, increased sodium/potassium exchange, and stimulation of the uptake of sugars and amino acids.

Growth factors are signal molecules that are similar to hormones in function but act closer to their sites of synthesis. As their name implies, many of the growth factors are important messengers in signaling cell replacement and cell growth. Most of the growth factors belong to one of three groups: factors that foster the multiplication and development of various cell types (*e.g.*, growth factor and epidermal growth factor); lymphokines and cytokines, which are important in the regulation of the immune system; and colony-stimulating factors, which regulate the proliferation and maturation of white and red blood cells.

Messenger-Mediated Control of Nuclear Function

Some messengers, such as thyroid hormone and steroid hormones, do not bind to membrane receptors but move directly across the lipid layer of the cell membrane and are carried to the cell nucleus, where they influence DNA activity. Many of these hormones bind to a cytoplasmic receptor, and together they are carried to the nucleus. In the nucleus, the receptor-hormone complex binds to DNA, thereby increasing transcription of mRNA. The mRNAs are translated in the ribosomes, with the production of increased amounts of proteins that alter cell function.

Membrane Potentials

The human body runs on a system of self-generated electricity. Electrical potentials exist across the membranes of most cells in the body. Because these potentials occur at the level of the cell membrane, they are called *membrane potentials*. In excitable tissues, such as nerve or muscle cells, changes in the membrane potential are necessary for generation and conduction of nerve impulses and muscle contraction. In other types of cells, such as glandular cells, changes in the membrane potential contribute to hormone secretion and other functions.

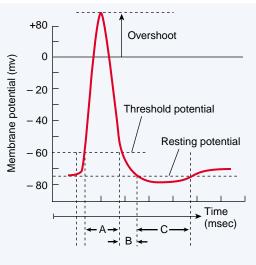
Electrical Potential

Electrical potential, measured in volts (V), describes the ability of separated electrical charges of opposite polarity (+ and –) to do work. The potential difference is the difference between the separated charges. The terms *potential difference* and *voltage* are synonymous. Voltage is always measured with respect to two points in a system. For example, the voltage in a car battery (6 or 12 V) is the potential difference between the two battery terminals. Because the total amount of charge that can be separated by a biologic membrane is small, the potential differences are small and are measured in *millivolts* (1/1000 of a volt). Potential differences across the cell membrane can be measured by inserting a very fine electrode into the cell and another into the extracellular fluid surrounding the cell and connecting the two electrodes to a voltmeter. The movement of charge between two points is called *current*. It occurs when a potential difference has been established and a connection is made such that the charged particles can move between the two points.

Extracellular and intracellular fluids are electrolyte solutions containing approximately 150 to 160 mmol/L of positively charged ions and an equal concentration of negatively charged ions. These are the current-carrying ions responsible for generating and conducting membrane potentials. Usually, a small excess of positively charged ions exists at the outer surface of the cell membrane. This is represented as positive charges on the outside of the membrane and is balanced by an equal number of negative charges on the inside of the membrane. Because of the extreme thinness of the cell membrane, the accumulation of these ions at the surfaces of the membrane contributes to the establishment of a membrane potential.

Action Potentials

Action potentials are abrupt, pulselike changes in the membrane potential that last a few ten thousandths to a few thousandths of a second (Fig 1-13). In a nerve fiber, an action potential can be elicited by any factor that suddenly increases the membrane potential, usually by opening a voltage-gated sodium channel. The *threshold potential* represents the membrane potential at which the ion channels open and neurons and other excitable tissues are stimulated to "fire." In large nerve



A = Absolute refractory period (active potential and partial recovery)

B = Relative refractory period

C=Positive relative refractory period

■ FIGURE 1-13 ■ Time course of the action potential recorded at one point of an axon with one electrode inside and one on the outside of the plasma membrane. The rising part of the action potential is called the spike. The rising phase plus approximately the first half of the repolarization phase is equal to the absolute refractory period (**A**). The portion of the repolarization phase that extends from the threshold to the resting membrane potential represents the relative refractory period (**B**). The remaining portion of the repolarization phase to the resting membrane potential is equal to the negative after potential (**C**). Hyperpolarization is equal to the positive relative refractory period. fibers the sodium channels open at approximately -60 mV. Under normal circumstances, the threshold potential is sufficient to open large numbers of ion channels, triggering massive depolarization of the membrane.

Action potentials can be divided into three phases: (1) resting, (2) depolarization, and (3) repolarization phases. The resting phase is the undisturbed period of the action potential, during which neurons and other excitable tissues are not transmitting impulses. During this phase, the membrane is highly permeable to potassium and there is approximately 70 to 90 mV less charge (-70 mV to -90 mV) on the inside than on the outside of the membrane. This difference in charge is necessary for establishment of current flow once the membrane becomes permeable to the flow of charged ions. During this period, the membrane is said to be polarized because charges of opposite polarity (+ and -) are aligned across the membrane. Depolarization is characterized by the flow of positively charged sodium ions to the interior of the membrane. During the depolarization phase of the action potential, the interior side of the membrane becomes positive (approximately +30 mV to +45 mV). Repolarization is the phase in which the polarity of the resting membrane is re-established. This is accomplished by closure of the sodium channels and opening of the potassium channels. The outflow of positively charged potassium ions returns the membrane potential to negativity. The activity of the Na+/K+ ATPase pump helps to reestablish the resting membrane potential. During repolarization, the membrane remains refractory (i.e., does not fire) until the repolarization is approximately one-third complete. This period, which lasts approximately one half of a millisecond, is called the absolute refractory period (Fig. 1-13). During the relative refractory period, which follows the absolute refractory period, the membrane can be excited, although only by a strongerthan-normal stimuli.

Two main factors alter membrane excitability: (1) the difference in the concentration of ions on the inside and outside of the membrane and (2) changes in membrane permeability. The resting membrane potential is strongly influenced by serum potassium levels and the resulting difference in concentration on the inside and outside of the membrane. When serum levels of potassium are decreased, the resting membrane potential becomes more negative, and nerve and muscle cells become less excitable, sometimes to the extent that they cannot be re-excited (see Chapter 6). An increase in serum potassium has the opposite effect, causing the resting membrane to become more positive, moving closer to threshold. When this happens, the amplitude of the action potential is decreased because the membrane has not been fully repolarized. Should the resting membrane potential reach the level of the threshold potential during the absolute refractory period, the nerve or muscle cell will remain depolarized and unexcitable.

Neural excitability is markedly altered by changes in membrane permeability to current-carrying ions such as sodium. Calcium ions decrease membrane permeability to sodium ions and increase the threshold for initiation of action potentials. If insufficient calcium ions are available, the permeability of the membrane to sodium increases, and as a result, membrane excitability increases, sometimes causing spontaneous muscle movements (tetany) to occur. Local anesthetic agents (*e.g.*, procaine, cocaine) act directly on neural membranes to decrease their permeability to sodium.

In summary, the movement of materials across the cell's membrane is essential for survival of the cell. Diffusion is a process by which substances such as ions move from areas of greater concentration to areas of lesser concentration in an attempt to reach a uniform distribution. Osmosis refers to the diffusion of water molecules through a semipermeable membrane along a concentration gradient. Facilitated diffusion is a passive process, in which molecules that cannot normally pass through the cell's membranes, do so with the assistance of a carrier molecule. Another type of transport, called active transport, requires the cell to expend energy in moving ions against a concentration gradient. The Na⁺/K⁺ ATPase pump is the best-known type of active transport. Endocytosis is a process by which cells engulf materials from the surrounding medium. Small particles are ingested by a process called pinocytosis; larger particles are engulfed by a process called phagocytosis. Exocytosis involves the removal of large particles from the cell and is essentially the reverse of endocytosis. Ion channels are integral transmembrane proteins that span the width of the cell membrane to form a gating system that controls the movement of ions across the cell membrane.

Cells communicate with each other by means of chemical messenger systems. In some tissues, chemical messengers move from cell to cell through gap junctions without entering the extracellular fluid. Other types of chemical messengers bind to receptors on or near the cell surface. There are three known classes of cell surface receptor proteins: ion channel linked, G protein linked, and enzyme linked. Ion-channellinked signaling is mediated by neurotransmitters that transiently open or close ion channels formed by integral proteins in the cell membrane. G-protein-linked receptors rely on a class of molecules called G proteins that function as an on-off switch to convert external signals (first messengers) into internal signals (second messengers). Enzyme-linked receptors interact with certain peptide hormones (e.g., insulin and growth factors) to directly initiate the activity of an intracellular enzyme, which in turn, triggers multiple cellular responses, such as stimulation of glucose and amino acid uptake or transcription of certain genes that control cell proliferation.

Electrical potentials (negative on the inside and positive on the outside) exist across the membranes of most cells in the body. Membrane excitability depends on a separation of charge across the membrane and the permeability of the membrane to the current-carrying ion. Action potentials are abrupt pulselike changes in the membrane potential that last a few ten thousandths to a few thousandths of a second. Action potentials can be divided into three phases: the resting phase, during which neurons and excitable tissues are not generating or transmitting impulses; the depolarization phase, which is characterized by flow of current across the membrane; and the repolarization phase, during which the resting membrane potential is restored.

BODY TISSUES

In the preceding sections, we discussed the individual cell, its metabolic processes, and mechanisms of communication. Although cells are similar, their structure and function vary ac-

cording to the special needs of the body. For example, muscle cells perform functions different from those of skin cells or nerve cells. Groups of cells that are closely associated in structure and have common or similar functions are called *tissues*. Four categories of tissue exist: epithelium, connective (supportive) tissue, muscle, and nerve. These tissues do not exist in isolated units, but in association with each other and in variable proportions, forming different structures and organs. This section provides a brief overview of the cells in epithelial, connective, and muscle tissue. Nervous tissue is described in Chapter 36.

Cell Differentiation

After conception, the fertilized ovum undergoes a series of divisions, ultimately forming approximately 200 different cell types. The formation of different types of cells and the disposition of these cells into tissue types is called cell differentiation, a process controlled by a system that switches genes on and off. Embryonic cells must become different to develop into all of the various organ systems, and they must remain different after the signal that initiated cell diversification has disappeared. The process of cell differentiation is controlled by cell memory, which is maintained through regulatory proteins contained in the individual members of a particular cell type. Cell differentiation also involves the sequential activation of multiple genes and their protein products. This means that after differentiation has occurred, the tissue type does not revert to an earlier stage of differentiation. The process of cell differentiation normally moves forward, producing cells that are more specialized than their predecessors. Usually, highly differentiated cell types,

KEY CONCEPTS

ORGANIZATION OF CELLS INTO TISSUES

- Cells with a similar embryonic origin or function are often organized into larger functional units called *tissues*, and these tissues in turn associate with other, dissimilar tissues to form the various organs of the body.
- Epithelial tissue forms sheets that cover the body's outer surface, lines internal surfaces, and forms glandular tissue. It is supported by a basement membrane, is avascular, and must receive nourishment from capillaries in supporting connective tissues.
- Connective tissue is the most abundant tissue of the body. It is found in a variety of forms, ranging from solid bone to blood cells that circulate in the vascular system.
- Muscle tissue contains actin and myosin filaments that allow it to contract and provide locomotion and movement of skeletal structures (skeletal muscle), pumping of blood through the heart (cardiac muscle), and contraction of blood vessels and visceral organs (smooth muscle).

such as skeletal muscle and nervous tissue, lose their ability to undergo cell division. Cancer is a disorder of cell differentiation in which cells of a single cell line fail to differentiate properly (see Chapter 5).

Embryonic Origin of Tissue Types

All of the approximately 200 different types of body cells can be classified into four basic or primary tissue types: epithelial, connective, muscle, and nervous (Table 1-1). These basic tissue types are often described by their embryonic origin. The embryo is essentially a three-layered tubular structure. The outer layer of the tube is called the *ectoderm*; the middle layer, the *mesoderm*; and the inner layer, the *endoderm*. All of the adult body tissues originate from these three cellular layers. Epithelium has its origin in all three embryonic layers; connective tissue and muscle develop mainly from the mesoderm; and nervous tissue develops from the ectoderm.

Epithelial Tissue

Epithelial tissue forms sheets that cover the body's outer surface, line the internal surfaces, and form the glandular tissue. Underneath all types of epithelial tissue is an extracellular matrix, called the *basement membrane*, which serves to attach the epithelial cells to adjacent connective tissue and provides them with flexible support.

Epithelial cells have strong intracellular protein filaments (*i.e.*, cytoskeleton) that are important in transmitting mechanical stresses from one cell to another. The cells of epithelial tissue are tightly bound together by specialized junctions. These specialized junctions enable these cells to form barriers to the movement of water, solutes, and cells from one body compartment to the next. Epithelial tissue is avascular (*i.e.*, without blood vessels) and must therefore receive oxygen and nutrients from the capillaries of the connective tissue on which the epithelial tissue rests (Fig. 1-14). To survive, the epithelial cells must be kept moist. Even the seemingly dry skin epithelium is kept moist by a nonvitalized, waterproof layer of superficial skin cells called *keratin*, which prevents evaporation of moisture from the deeper living cells. Epithelium is able to regenerate quickly when injured

Epithelial tissues are classified according to the shape of the cells and the number of layers that are present: simple, stratified, and pseudostratified. Glandular epithelial tissue is formed by cells specialized to produce a fluid secretion. The terms *squamous* (thin and flat), *cuboidal* (cube shaped), and *columnar* (resembling a column) refer to the cells' shape (Fig. 1-15).

Simple Epithelium

Simple epithelium contains a single layer of cells, all of which rest on the basement membrane. Simple squamous epithelium is adapted for filtration; it is found lining the blood vessels, lymph nodes, and alveoli of the lungs. The single layer of squamous epithelium lining the heart and blood vessels is known as the *endothelium*. A similar type of layer, called the *mesothelium*, forms the serous membranes that line the pleural, pericardial, and peritoneal cavities and cover the organs of these cavities. A *simple cuboidal epithelium* is found on the surface of the ovary and in the thyroid. *Simple columnar epithelium* lines the intestine. One form of a simple columnar epithelium

TABLE 1-1 Classification of Tissue T	ypes
Tissue Type	Location
Epithelial Tissue	
Covering and lining of body surfaces	
Simple epithelium	
Squamous	Lining of blood vessels, body cavities, alveoli of lungs
Cuboidal	Collecting tubules of kidney; covering of ovaries
Columnar	Lining of intestine and gallbladder
Stratified epithelium	
Squamous keratinized	Skin
Squamous nonkeratinized	Mucous membranes of mouth, esophagus, and vagina
Cuboidal	Ducts of sweat glands
Columnar	Large ducts of salivary and mammary glands; also found in conjunctiva
Transitional	Bladder, ureters, renal pelvis
Pseudostratified	Tracheal and respiratory passages
Glandular	
Endocrine	Pituitary gland, thyroid gland, adrenal, and other glands
Exocrine	Sweat glands and glands in gastrointestinal tract
Neuroepithelium	Olfactory mucosa, retina, tongue
Reproductive epithelium	Seminiferous tubules of testis; cortical portion of ovary
Connective Tissue	
Embryonic connective tissue	
Mesenchymal	Embryonic mesoderm
Mucous	Umbilical cord (Wharton's jelly)
Adult connective tissue	
Loose or areolar	Subcutaneous areas
Dense regular	Tendons and ligaments
Dense irregular	Dermis of skin
Adipose	Fat pads, subcutaneous layers
Reticular	Framework of lymphoid organs, bone marrow, liver
Specialized connective tissue	
Bone	Long bones, flat bones
Cartilage	Tracheal rings, external ear, articular surfaces
Hematopoietic	Blood cells, myeloid tissue (bone marrow)
Muscle Tissue	
Skeletal	Skeletal muscles
Cardiac	Heart muscles
Smooth	Gastrointestinal tract, blood vessels, bronchi, bladder, and others
Nervous Tissue	
Neurons	Central and peripheral neurons and nerve fibers
Supporting cells	Glial and ependymal cells in central nervous system; Schwann and satellite cells in peripheral nervous system
	· · · · · · · · · · · · · · · · · · ·

has hairlike projections called *cilia*, often with specialized mucus-secreting cells called goblet cells. This form of simple columnar epithelium lines the airways of the respiratory tract.

Stratified and Pseudostratified Epithelium

Stratified epithelium contains more than one layer of cells, with only the deepest layer resting on the basement membrane. It is designed to protect the body surface. Stratified squamous keratinized epithelium makes up the epidermis of the skin. Keratin is a tough, fibrous protein existing as filaments in the outer cells of skin. A stratified squamous keratinized epithelium is made up of many layers. The layers closest to the underlying tissues are cuboidal or columnar. The cells become more irregular and thinner as they move closer to the surface. Surface cells become totally filled with keratin and die, are sloughed off, and then replaced by the deeper cells. A stratified squamous nonkeratinized epithelium is found on moist surfaces, such as the mouth and tongue. Stratified cuboidal and columnar epithelia are found in the ducts of salivary glands and the larger ducts of the mammary glands. In smokers, the normal columnar ciliated epithelial cells of the trachea and bronchi are often replaced with stratified squamous epithelium cells that are better able to withstand the irritating effects of cigarette smoke.

Pseudostratified epithelium is a type of epithelium in which all of the cells are in contact with the underlying intercellular matrix, but some do not extend to the surface. A pseudostratified ciliated columnar epithelium with goblet cells forms the lining of most of the upper respiratory tract. All of the tall cells reaching the surface of this type of epithelium are either ciliated cells or mucus-producing goblet cells. The basal cells that do not

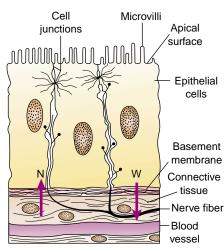


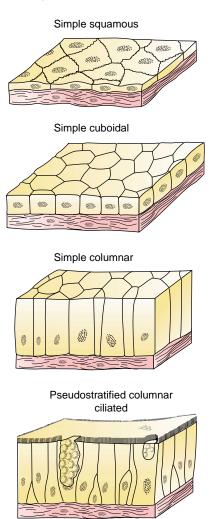
FIGURE 1-14 Typical arrangement of epithelial cells in relation to underlying tissues and blood supply. Epithelial tissue has no blood supply of its own but relies on the blood vessels in the underlying connective tissue for nutrition (N) and elimination of wastes (W).

reach the surface serve as stem cells for ciliated and goblet cells. *Transitional epithelium* is a stratified epithelium characterized by cells that can change shape and become thinner when the tissue is stretched. Such tissue can be stretched without pulling the superficial cells apart. Transitional epithelium is well adapted for the lining of organs that are constantly changing their volume, such as the urinary bladder.

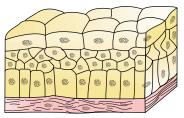
Glandular Epithelium

Glandular epithelial tissue is formed by cells specialized to produce a fluid secretion. This process is usually accompanied by the intracellular synthesis of macromolecules. The chemical nature of these macromolecules is variable. The macromolecules typically are stored in the cells in small, membranebound vesicles called *secretory granules*. For example, glandular epithelia can synthesize, store, and secrete proteins (*e.g.*, insulin), lipids (*e.g.*, adrenocortical hormones, secretions of the sebaceous glands), and complexes of carbohydrates and proteins (*e.g.*, saliva). Less common are secretions such as those produced by the sweat glands, which require minimal synthetic activity.

All glandular cells arise from surface epithelia by means of cell proliferation and invasion of the underlying connective tissue. Epithelial glands can be divided into two groups: exocrine and endocrine glands. Exocrine glands, such as the sweat glands and lactating mammary glands, retain their connection with the surface epithelium from which they originated. This connection takes the form of epithelium-lined tubular ducts through which the secretions pass to reach the surface. Exocrine glands are often classified according to the way secretory products are released by their cells. In holocrine type cells (e.g., sebaceous glands), the glandular cell ruptures, releasing its entire contents into the duct system. New generations of cells are replaced by mitosis of basal cells. Merocrine or eccrine type glands (e.g., salivary glands, exocrine glands of the pancreas) release their glandular products by exocytosis. In apocrine secretions (e.g., mammary glands, certain sweat glands), the apical por-



Transitional



Stratified squamous

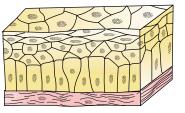


FIGURE 1-15 Representation of the various epithelial tissue types.

tion of the cell, along with small portions of the cytoplasm, is pinched off the glandular cells. *Endocrine glands* are epithelial structures that have had their connection with the surface obliterated during development. These glands are ductless and produce secretions (*i.e.*, hormones) that move directly into the bloodstream.

Connective or Supportive Tissue

Connective tissue (or supportive tissue) is the most abundant tissue in the body. As its name suggests, it connects and binds or supports the various tissues. The capsules that surround organs of the body are composed of connective tissue. Bone, adipose tissue, and cartilage are specialized types of connective tissue that function to support the soft tissues of the body and store fat. Connective tissue is unique in that its cells produce the extracellular matrix that supports and holds tissues together. Connective tissue has a role in tissue nutrition. The close proximity of the extracellular matrix to blood vessels allows it to function as an exchange medium through which nutrients and metabolic wastes pass.

Adult connective tissue proper can be divided into two main types: loose or areolar and dense connective tissue.

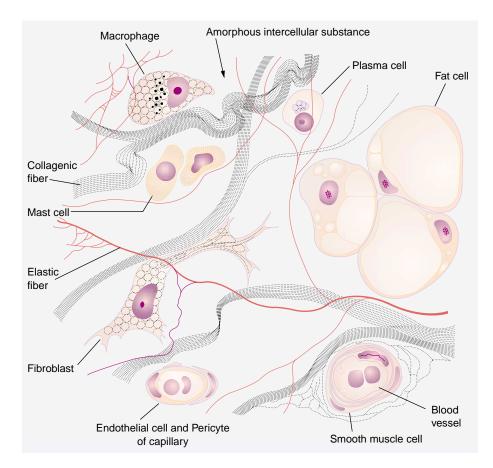
Loose Connective Tissue

Loose connective tissue, also known as areolar tissue, is soft and pliable. Although it is more cellular than dense connective tissue, it contains large amounts of intercellular substance. It fills

spaces between muscle sheaths and forms a layer that encases blood and lymphatic vessels. Areolar connective tissue supports the epithelial tissues and provides the means by which these tissues are nourished. In an organ containing functioning epithelial tissue and supporting connective tissue, the term *parenchymal tissue* is used to describe the functioning epithelium as opposed to the connective tissue framework or stroma.

Cells of loose connective tissue include fibroblasts, mast cells, adipose or fat cells, macrophages, plasma cells, and leukocytes (Fig. 1-16). Loose connective tissue cells secrete substances that form the extracellular matrix that supports and connects body cells. Fibroblasts are the most abundant of these cells. They are responsible for the synthesis of the fibrous and gel-like substance that fills the intercellular spaces of the body and for the production of collagen, elastic, and reticular fibers.

Adipose tissue is a special form of connective tissue in which adipocytes predominate. Adipocytes do not generate an extracellular matrix but maintain a large intracellular space. These cells store large quantities of triglycerides and are the largest repository of energy in the body. Adipose tissue helps fill spaces between tissues and helps to keep organs in place. Subcutaneous layers of fat help to shape the body. Because fat is a poor conductor of heat, adipose tissue serves as thermal insulation for the body. Adipose tissue exists in two forms. Unilocular (white) adipose tissue is composed of cells in which the fat is contained in a single, large droplet in the cytoplasm. Multilocular (brown) adipose tissue is composed of cells that contain multiple droplets of fat and numerous mitochondria.



■ FIGURE 1-16 ■ Diagrammatic representation of cells that may be seen in loose connective tissue. The cells lie in an intercellular matrix that is bathed in tissue fluid that originates in capillaries. (From Cormack D.H. [1987]. *Ham's histology* [9th ed.]. Philadelphia: J.B. Lippincott) *Reticular tissue* is characterized by a network of reticular fibers associated with reticular cells. Reticular fibers provide the framework for capillaries, nerves, and muscle cells. They also constitute the main supporting elements for the blood forming tissues and the liver.

Dense Connective Tissue

Dense connective tissue exists in two forms: dense irregular and dense regular. Dense irregular connective tissue consists of the same components found in loose connective tissue, but there is a predominance of collagen fibers and fewer cells. This type of tissue can be found in the dermis of the skin (*i.e.*, reticular layer), the fibrous capsules of many organs, and the fibrous sheaths of cartilage (*i.e.*, perichondrium) and bone (*i.e.*, periosteum). It also forms the fascia that covers muscles and organs. Dense regular connective tissues are rich in collagen fibers and form the tendons and aponeuroses that join muscles to bone or other muscles and the ligaments that join bone to bone. Tendons and ligaments are white fibers because of an abundance of collagen. Ligaments such as the ligamenta flava of the vertebral column and the true vocal folds are called *yellow fibers* because of the abundance of elastic fibers.

Muscle Tissue

Three types of muscle tissues exist: *skeletal, cardiac,* and *smooth.* Skeletal and cardiac muscles are striated muscles. The actin and myosin filaments are arranged in large parallel arrays in bundles, giving the muscle fibers a striped or striated appearance when they are viewed through a microscope.

Skeletal muscle is the most abundant tissue in the body, accounting for 40% to 45% of the total body weight. Most skeletal muscles are attached to bones, and their contractions are responsible for movements of the skeleton. Cardiac muscle, which is found in the heart, is designed to pump blood continuously. Smooth muscle is found in the iris of the eye, the walls of blood vessels, hollow organs such as the stomach and urinary bladder, and hollow tubes, such as the ureters, that connect internal organs.

Neither skeletal nor cardiac muscle can undergo the mitotic activity needed to replace injured cells. However, smooth muscle may proliferate and undergo mitotic activity. Some increases in smooth muscle are physiologic, as occurs in the uterus during pregnancy. Other increases, such as the increase in smooth muscle that occurs in the arteries of persons with chronic hypertension, are pathologic.

Although the three types of muscle tissue differ significantly in structure, contractile properties, and control mechanisms, they have many similarities. In the following section, the structural properties of skeletal muscle are presented as the prototype of striated muscle tissue. Smooth muscle and the ways in which it differs from skeletal muscle are also discussed. Cardiac muscle is described in Chapter 14.

Skeletal Muscle

Skeletal muscle tissue is packaged into skeletal muscles that attach to and cover the body skeleton. Each skeletal muscle is a discrete organ made up of hundreds and thousands of muscle fibers. Even though muscle fibers predominate, substantial amounts of connective tissue, blood vessels, and nerve fibers are present. In an intact muscle, several different layers of connective tissue hold the individual muscle fibers together. A dense connective tissue covering called the *epimysium* forms the outermost layer surrounding the whole muscle (Fig. 1-17). Each muscle is subdivided into smaller bundles called *fascicles*, which are surrounded by a connective tissue covering called the *perimysium*. The number of fascicles and their size vary among muscles. Fascicles consist of many elongated structures called *muscle fibers*, each of which is surrounded by connective tissue called the *endomysium*.

Skeletal muscles are syncytial or multinucleated structures, meaning there are no true cell boundaries within a skeletal muscle fiber. The cytoplasm or sarcoplasm of the muscle fiber is contained within the sarcolemma, which represents the cell membrane. Embedded throughout the sarcoplasm are the contractile elements actin and myosin, which are arranged in parallel bundles (*i.e., myofibrils*). The thin, lighter-staining myofilaments are composed of actin, and the thicker, darker-staining myofilaments are composed of myosin. Each myofibril consists of regularly repeating units along the length of the myofibril; each of these units is called a sarcomere (see Fig. 1-17). Sarcomeres are the structural and functional units of cardiac and skeletal muscle. A sarcomere extends from one Z line to another Z line. Within the sarcomere are alternating light and dark bands. The central dark band (A band) contains mainly myosin filaments, with some overlap with actin filaments. The lighter I band contains only actin filaments and straddles the Z band; therefore, it takes two sarcomeres to complete an I band. An H zone is found in the middle of the A band and represents the region where only myosin filaments are found. In the center of the H zone is a thin, dark band, the M band or line, that is produced by linkages between the myosin filaments. Z bands consist of short elements that interconnect and provide the thin actin filaments from two adjoining sarcomeres with an anchoring point.

The sarcoplasmic reticulum, which is comparable to the smooth ER, is composed of longitudinal tubules that run parallel to the muscle fiber and surround each myofibril. This network ends in enlarged, saclike regions called the lateral sacs or terminal cisternae. These sacs store calcium to be released during muscle contraction. A second system of tubules consists of the *transverse* or *T* tubules, which are extensions of the plasma membrane and run perpendicular to the muscle fiber. The hollow portion or lumen of the transverse tubule is continuous with the extracellular fluid compartment. Action potentials, which are rapidly conducted over the surface of the muscle fiber, are in turn propagated by the T tubules and into the sarcoplasmic reticulum. As the action potential moves through the lateral sacs, the sacs release calcium, initiating muscle contraction. The membrane of the sarcoplasmic reticulum also has an active transport mechanism for pumping calcium ions back into the reticulum. This prevents interactions between calcium ions and the actin and myosin myofilaments after cessation of a muscle contraction.

Skeletal Muscle Contraction. Muscle contraction involves the sliding of the thick myosin and thin actin filaments over each other to produce shortening of the muscle fiber, while the actual length of the individual thick and thin filaments remains unchanged. The thick myosin filaments consist of a thin tail, which provides the structural backbone for the filament, and a globular head that forms cross-bridges with the thin actin

Unit One: Mechanisms of Disease

■ FIGURE 1-17 ■ (A) Connective tissue components of a skeletal muscle. (B) Structure of the myofibril and the relationship between actin and myosin myofilaments. (C) Sarcoplasmic reticulum and system of transverse tubules.

filaments (Fig. 1-18). Myosin molecules are bundled together side by side in the thick filaments such that one half have their heads toward one end of the filament and their tails toward the other end; the other half are arranged in the opposite manner. Each globular myosin head contains a binding site able to bind to a complementary site on the actin molecule. In addition to the binding site for actin, each myosin head has a separate active site that catalyzes the breakdown of ATP to provide the energy needed to activate the myosin head so it can form a crossbridge with actin. After contraction, myosin also binds ATP, thus breaking the linkage between actin and myosin.

The thin filaments are composed mainly of actin, a globular protein lined up in two rows that coil around each other to form a long helical strand. Associated with each actin filament are two regulatory proteins, tropomyosin and troponin. Tropomyosin, which lies in grooves of the actin strand, provides the site for attachment of the globular heads of the myosin filament. In the noncontracted state, troponin covers the tropomyosin binding sites and prevents formation of cross-bridges between the actin and myosin. During an action potential, calcium ions

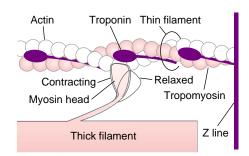


FIGURE 1-18 Molecular structure of the thin actin filament and the thicker myosin filament of striated muscle. The thin filament is a double-stranded helix of actin molecules with tropomyosin and troponin molecules lying along the grooves of the actin strands. During muscle contraction, the ATP-activated heads of the thick myosin filament swivel into position, much like the oars on a boat, form a cross-bridge with a reactive site on tropomyosin, and then pull the actin filament forward. During muscle relaxation, the troponin molecules cover the reactive sites on tropomyosin.

released from the sarcoplasmic reticulum diffuse to the adjacent myofibrils, where they bind to troponin. The binding of calcium to troponin uncovers the tropomyosin binding sites such that the myosin heads can attach and form cross-bridges.

Muscle contraction begins with activation of the crossbridges from the myosin filaments and uncovering of the tropomyosin binding sites on the actin filament. When activated by ATP, the heads of the myosin heads swivel in a fixed arc, much like the oars of a boat, as they become attached to the actin filament. During contraction, each myosin head undergoes its own cycle of movement, forming a bridge attachment and releasing it, and moving to another site where the same sequence of movement occurs. This pulls the thin and thick filaments past each other. Energy from ATP is used to break the actin and myosin cross-bridges, stopping the muscle contraction. With the breaking of the linkage between actin and myosin, the concentration of calcium around the myofibrils decreases as calcium is actively transported into the sarcoplasmic reticulum by a membrane pump that uses energy derived from ATP.

Smooth Muscle

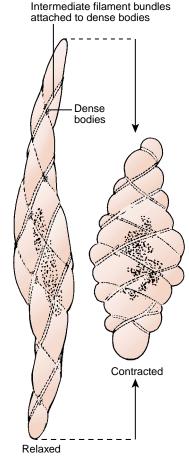
Smooth muscle is often called *involuntary muscle* because its activity arises spontaneously or through activity of the autonomic nervous system. Smooth muscle contractions are slower and more sustained than skeletal or cardiac muscle contractions.

Organization and Structure. Smooth muscle cells are spindle shaped and smaller than skeletal muscle fibers. Each smooth muscle cell has one centrally positioned nucleus. Z bands or M lines are not present in smooth muscle fibers, and the cross-striations are absent because the bundles of filaments are not parallel but crisscross obliquely through the cell. Instead, the actin filaments are attached to structures called *dense bodies*. Some of the dense bodies are attached to the cell membrane, and others are dispersed in the cell and linked together by structural proteins (Fig. 1-19).

The lack of Z lines and regular overlapping of the contractile elements provides a greater range of tension development. This is important in hollow organs that undergo changes in volume, with consequent changes in the length of the smooth muscle fibers in their walls. Even with the distention of a hollow organ, the smooth muscle fiber retains some ability to develop tension, whereas such distention would stretch skeletal muscle beyond the area where the thick and thin filaments overlap.

Smooth muscle usually is arranged in sheets or bundles. In hollow organs, such as the intestines, the bundles are organized into the two-layered muscularis externa, consisting of an outer, longitudinal layer and an inner, circular layer. A thinner muscularis mucosae often lies between the muscularis externa and the endothelium. In blood vessels, the bundles are arranged circularly or helically around the vessel wall.

Smooth Muscle Contraction. As with cardiac and skeletal muscle, smooth muscle contraction is initiated by an increase in intracellular calcium. However, smooth muscle differs from skeletal muscle in the way its cross-bridges are formed. The sarcoplasmic reticulum of smooth muscle is less developed than in skeletal muscle, and no transverse tubules are present. Thus, smooth muscle relies heavily on the entrance



■ FIGURE 1-19 ■ Structure of smooth muscle showing the dense bodies. In smooth muscle, the force of contraction is transmitted to the cell membrane by bundles of intermediate fibers. (Cormack D.H. [1993]. *Essential histology* [p. 229]. Philadelphia: J.B. Lippincott)

of extracellular calcium for muscle contraction. This dependence on movement of extracellular calcium across the cell membrane during muscle contraction is the basis for the action of calcium-blocking drugs used in the treatment of cardiovascular disease.

Smooth muscle also lacks the calcium-binding regulatory protein troponin, which is found in skeletal and cardiac muscle. Instead, it relies on another cytoplasmic protein called *calmodulin*. The calcium–calmodulin complex binds to and activates the myosin-containing thick filaments, which interact with actin.

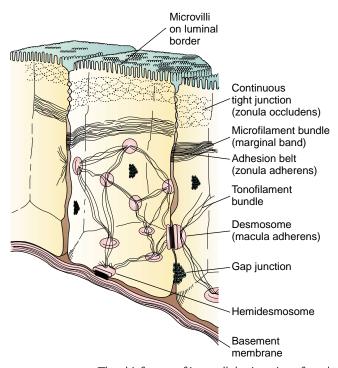
Types of Smooth Muscle. Smooth muscle may be divided into two broad categories according to the mode of activation: multiunit and single-unit smooth muscle. In *multiunit* smooth muscle, each unit operates almost independently of the others and is often enervated by a single nerve, such as occurs in skeletal muscle. It has little or no inherent activity and depends on the autonomic nervous system for its activation. Smooth muscle of this type is found in the iris, in the walls of the vas deferens, and attached to hairs in the skin. The fibers in *single-unit* smooth muscle are in close contact with each

other and can contract spontaneously without nerve or hormonal stimulation. Normally, most muscle fibers contract synchronously, thus the term *single-unit* smooth muscle. Some single-unit smooth muscle, such as that found in the gastrointestinal tract, is self-excitable. This is usually associated with a basic slow-wave rhythm transmitted from cell to cell by nexuses (*i.e.*, gap junctions) formed by the fusion of adjacent cell membranes. The cause of this slow wave is unknown. The intensity of contraction increases with the frequency of the action potential. Certain hormones, other agents, and local factors can modify smooth muscle activity by depolarizing or hyperpolarizing the membrane. The smooth muscle of the uterus and small-diameter blood vessels is also single-unit smooth muscle.

Cell Junctions and Cell-to-Cell Adhesion

Cell junctions occur at many points in cell-to-cell contact, but they are particularly plentiful and important in epithelial tissue. Three basic types of intercellular junctions are observed: tight junctions, adhering junctions, and gap junctions (Fig. 1-20). Often, the cells in epithelial tissue are joined by all three types of junctions. *Continuous tight* or occluding junctions (*i.e.*, zona occludens), which are found only in epithelial tissue, seal the surface membranes of adjacent cells together. This type of intercellular junction prevents materials such as macromolecules present in the intestinal contents from entering the intercellular space.

Adhering junctions represent a site of strong adhesion between cells. The primary role of adhering junctions may be that of preventing cell separation. Adhering junctions are not restricted to epithelial tissue; they provide adherence between ad-



■ FIGURE 1-20 ■ The chief types of intercellular junctions found in epithelial tissue. (From Cormack D.H. [1993]. *Essential histology*. Philadelphia: J.B. Lippincott) jacent cardiac muscle cells as well. Adhering junctions are found as continuous, beltlike adhesive junctions (*i.e.*, zonula adherens), or scattered, spotlike adhesive junctions called *desmosomes* (*i.e.*, macula adherens). A special feature of the adhesion belt junction is that it provides a site for anchorage of microfilaments to the cell membrane. In epithelial desmosomes, bundles of keratin-containing intermediate filaments (*i.e.*, tonofilaments) are anchored to the junction on the cytoplasmic area of the cell membrane.

Gap junctions, or *nexus junctions*, involve the close adherence of adjoining cell membranes with the formation of channels that link the cytoplasm of the two cells. Gap junctions are not unique to epithelial tissue; they play an essential role in many types of cell-to-cell communication. Because they are low-resistance channels, gap junctions are important in cell-to-cell conduction of electrical signals (*e.g.*, between cells in sheets of smooth muscle or between adjacent cardiac muscle cells, where they function as electrical synapses). These multiple communication channels also enable ions and small molecules to pass directly from one cell to another.

Hemidesmosomes are another type of junction. They are found at the base of epithelial cells and help attach the epithelial cell to the underlying connective tissue. They resemble half a desmosome, thus their name.

In summary, body cells are organized into four basic tissue types: epithelial, connective, muscle, and nervous. The epithelium covers and lines the body surfaces and forms the functional components of glandular structures. Epithelial tissue is classified into three types according to the shape of the cells and the number of layers that are present: simple, stratified, and pseudo-stratified. The cells in epithelial tissue are held together by three types of intercellular junctions: tight, adhering, and gap. They are attached to the underlying tissue by hemidesmosomes.

Connective tissue supports and connects body structures; it forms the bones and skeletal system, the joint structures, the blood cells, and the intercellular substances. Adult connective tissue can be divided into four types: loose or areolar, reticular, adipose, and dense (regular and irregular).

Muscle tissue is a specialized tissue designed for contractility. Three types of muscle tissue exist: skeletal, cardiac, and smooth. Actin and myosin filaments interact to produce muscle shortening, a process activated by the presence of calcium. In skeletal muscle, calcium is released from the sarcoplasmic reticulum in response to an action potential. Smooth muscle is often called involuntary muscle because it contracts spontaneously or through activity of the autonomic nervous system. It differs from skeletal muscle in that its sarcoplasmic reticulum is less defined and it depends on the entry of extracellular calcium ions for muscle contraction.

The cells of tissues are joined by cell junctions. There are three basic types of cell junctions: tight junctions, which prevent materials from entering the intercellular space and are found only in epithelial tissue; adhering junctions, which hold cells together and are found in epithelial as well as cardiac tissue; and gap junctions, which contain channels that link the cytoplasm of two cells and are important in cell-to-cell conduction of electrical impulses.

REVIEW QUESTIONS

State why the nucleus is called the "control center" of the cell, and explain the relationships among DNA, genes, and chromosomes.

List the cellular organelles and state their functions.

Explain how the composition of the lipid bilayer structure of the cell membrane controls access to the interior of the cell, and describe how ions, nutrients, water, and other substances cross the cell membrane.

Relate the function of ATP to cell metabolism and compare the processes involved in the aerobic and anaerobic metabolic pathways that generate ATP.

Trace the pathway for cell communication, beginning at the receptor and ending with effector response, and explain why the process is often referred to as "signal transduction."

Two conditions are necessary for a membrane potential to occur by diffusion: the membrane must be selectively permeable to a single type of ion, and the concentration of the diffusible ion must be greater on one side of the membrane than the other. Relate this concept to the role that potassium and sodium ions play in the generation of an action potential.

Describe the characteristics of the simple epithelium, stratified epithelium, and glandular epithelium; loose and dense connective tissue; skeletal and smooth muscle; and the neurons and supporting cells of the nervous system.

Characterize the function of the intracellular adhesions and junctions.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

BIBLIOGRAPHY

- Alberts B., Bray D., Lewis J., et al. (1994). *Molecular biology of the cell* (3rd ed.). New York and London: Garland Publishing.
- Ashcroft F.M. (2000). *Ion channels and disease*. San Diego: Academic Press. Cotran R.S., Kumar V., Collins T. (Eds.). (1999). *Robbins' pathologic basis of*
- disease (6th ed.). Philadelphia: W. B. Saunders.
- Guyton A.C., Hall J.E. (2000). *Textbook of medical physiology* (10th ed.). Philadelphia: W.B Saunders.
- Joachim F. (1998). How the ribosome works. American Scientist 86(5), 428–439.

Kerr J.B. (1999). Atlas of functional histology. London: Mosby.

- Mathews C.K., van Holde K.E., Ahern K.G. (2000). *Biochemistry*. San Francisco: Benjamin/Cummings.
- Moore K.L., Persaud T.V.N. (1998). The developing human: Clinically oriented embryology (6th ed.). Philadelphia: W. B. Saunders.
- Nelson D.L., Cox M.M. (2000). Lehninger's principles of biochemistry (3rd ed.). New York: Worth Publishers.
- Tortora G.J., Grabowski S.R. (2000). *Principles of anatomy and physiology*. New York: John Wiley & Sons.

CHAPTER

Tissue Adaptation and Injury

Cellular Adaptation Atrophy **Hypertrophy Hyperplasia Metaplasia Dysplasia** Intracellular Accumulations **Cell Injury and Death Causes of Cell Injury Injury From Physical Agents Radiation Injury Chemical Injury Injury From Biologic Agents** Injury From Nutritional Imbalances Mechanisms of Cell Injury Free Radical Injury Hypoxic Cell Injury Impaired Calcium Homeostasis **Reversible Cell Injury and Cell Death Reversible Cell Injury** Cell Death

hen confronted with stresses that endanger its normal structure and function, the cell undergoes adaptive changes that permit survival and maintenance of function. It is only when the stress is overwhelming or adaptation is ineffective that cell injury and death occur.

CELLULAR ADAPTATION

Cells adapt to changes in the internal environment, just as the total organism adapts to changes in the external environment. Cells may adapt by undergoing changes in size, number, and type. These changes, occurring singly or in combination, may

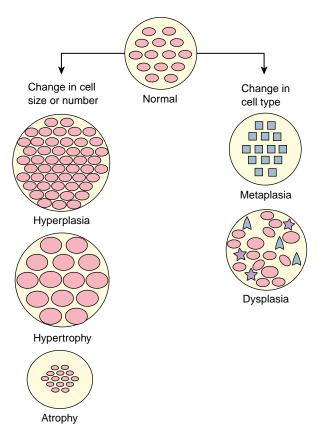
lead to atrophy, hypertrophy, hyperplasia, metaplasia, and dysplasia (Fig. 2-1). Adaptive cellular responses also include intracellular accumulations and storage of products in abnormal amounts.^{1,2}

There are numerous molecular mechanisms mediating cellular adaptation, including factors produced by other cells or by the cells themselves. These mechanisms depend largely on signals transmitted by chemical messengers that exert their effects by altering gene function. In general, the genes expressed in all cells fall into two categories: "housekeeping" genes that are necessary for normal function of a cell, and genes that determine the differentiating characteristics of a particular cell type. In many adaptive cellular responses, the expression of the differentiation genes is altered, whereas that of the housekeeping genes remains unaffected.¹ Thus, a cell is able to change size or form without compromising its housekeeping function. Once the stimulus for adaptation is removed, the effect on expression of the differentiating genes is removed and the cell resumes its previous state of specialized function. Whether adaptive cellular changes are normal or abnormal depends on whether the response was mediated by an appropriate stimulus. Normal adaptive responses occur in response to need and an appropriate stimulus. After the need has been removed, the adaptive response ceases.

Atrophy

When confronted with a decrease in work demands or adverse environmental conditions, most cells are able to revert to a smaller size and a lower and more efficient level of functioning that is compatible with survival. This decrease in cell size is called *atrophy*. Cell size, particularly in muscle tissue, is related to workload. As the workload of a cell diminishes, oxygen consumption and protein synthesis decrease. Cells that are atrophied reduce their oxygen consumption and other cellular functions by decreasing the number and size of their organelles and other structures. There are fewer mitochondria, myofilaments, and endoplasmic reticulum structures. When a sufficient number of cells are involved, the entire tissue or muscle atrophies.

The general causes of atrophy can be grouped into five categories: (1) disuse, (2) denervation, (3) loss of endocrine stimulation, (4) inadequate nutrition, and (5) ischemia or a



■ FIGURE 2-1 ■ Adaptive tissue cell responses (large circles) involving a change in number (hyperplasia), cell size (hypertrophy and atrophy), cell type (metaplasia), or size, shape, and organization (dysplasia).

decrease in blood flow. Disuse atrophy occurs when there is a reduction in skeletal muscle use. An extreme example of disuse atrophy is seen in the muscles of extremities that have been encased in plaster casts. Because atrophy is adaptive and reversible, muscle size is restored after the cast is removed and muscle use is resumed. Denervation atrophy is a form of disuse atrophy that occurs in the muscles of paralyzed limbs. Lack of endocrine stimulation produces a form of disuse atrophy. In women, the loss of estrogen stimulation during menopause results in atrophic changes in the reproductive organs. With malnutrition and decreased blood flow, cells decrease their size and energy requirements as a means of survival.

KEY CONCEPTS

CELLULAR ADAPTATIONS

- Cells are able to adapt to changes in work demands or threats to survival by changing their size (atrophy and hypertrophy), number (hyperplasia), and form (metaplasia).
- Normal cellular adaptation occurs in response to an appropriate stimulus and ceases once the need for adaptation has ceased.

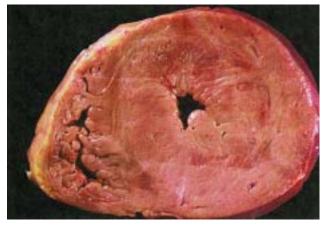
Hypertrophy

Hypertrophy represents an increase in cell size and with it an increase in the amount of functioning tissue mass. It results from an increased workload imposed on an organ or body part and is commonly seen in cardiac and skeletal muscle tissue, which cannot adapt to an increase in workload through mitotic division and formation of more cells. Hypertrophy involves an increase in the functional components of the cell that allows it to achieve equilibrium between demand and functional capacity. For example, as muscle cells hypertrophy, additional actin and myosin filaments, cell enzymes, and adenosine triphosphate (ATP) are synthesized.

Hypertrophy may occur as the result of normal physiologic or abnormal pathologic conditions. The increase in muscle mass associated with exercise is an example of physiologic hypertrophy. Pathologic hypertrophy occurs as the result of disease conditions and may be adaptive or compensatory. Examples of adaptive hypertrophy are the thickening of the urinary bladder from long-continued obstruction of urinary outflow and the myocardial hypertrophy that results from valvular heart disease or hypertension. Compensatory hypertrophy is the enlargement of a remaining organ or tissue after a portion has been surgically removed or rendered inactive. For instance, if one kidney is removed, the remaining kidney enlarges to compensate for the loss.

The precise signal for hypertrophy is unknown. It may be related to ATP depletion, mechanical forces such as stretching of the muscle fibers, activation of cell degradation products, or hormonal factors.¹ Whatever the mechanism, a limit is eventually reached beyond which further enlargement of the tissue mass is no longer able to compensate for the increased work demands. The limiting factors for continued hypertrophy might be related to limitations in blood flow. For example, in hypertension the increased workload required to pump blood against an elevated arterial pressure results in a progressive increase in left ventricular muscle mass (Fig. 2-2).

There has been recent interest in the signaling pathways that control the arrangement of contractile elements in myocardial hypertrophy. Research suggests that certain signal molecules



■ FIGURE 2-2 ■ Myocardial hypertrophy. Cross-section of the heart in a patient with long-standing hypertension. (From Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 9]. Philadelphia: Lippincott Williams & Wilkins)

can alter gene expression controlling the size and assembly of the contractile proteins in hypertrophied myocardial cells. For example, the hypertrophied myocardial cells of well-trained athletes have proportional increases in width and length. This is in contrast to the hypertrophy that develops in dilated cardiomyopathy, in which the hypertrophied cells have a relatively greater increase in length than width. In pressure overload, as occurs with hypertension, the hypertrophied cells have greater width than length.³ It is anticipated that further elucidation of the signal pathways that determine the adaptive and nonadaptive features of cardiac hypertrophy will lead to new targets for treatment.

Hyperplasia

26

Hyperplasia refers to an increase in the number of cells in an organ or tissue. It occurs in tissues with cells that are capable of mitotic division, such as the epidermis, intestinal epithelium, and glandular tissue. Nerve cells and skeletal and cardiac muscle do not divide and therefore have no capacity for hyperplastic growth. There is evidence that hyperplasia involves activation of genes controlling cell proliferation and the presence of intracellular messengers that control cell replication and growth. As with other normal adaptive cellular responses, hyperplasia is a controlled process that occurs in response to an appropriate stimulus and ceases after the stimulus has been removed.

The stimuli that induce hyperplasia may be physiologic or nonphysiologic. Physiologic hyperplasia can occur as the result of hormonal stimulation, increased functional demands, or as a compensatory mechanism. Breast and uterine enlargement during pregnancy are examples of a physiologic hyperplasia that results from estrogen stimulation. An increased demand for parathyroid hormone, as occurs in chronic renal failure, results in hyperplasia of the parathyroid gland. The regeneration of the liver that occurs after partial hepatectomy (i.e., partial removal of the liver) is an example of compensatory hyperplasia. Hyperplasia is also an important response of connective tissue in wound healing, during which proliferating fibroblasts and blood vessels contribute to wound repair. Although hypertrophy and hyperplasia are two distinct processes, they may occur together and are often triggered by the same mechanism.¹ For example, the pregnant uterus undergoes both hypertrophy and hyperplasia as the result of estrogen stimulation.

Most forms of nonphysiologic hyperplasia are due to excessive hormonal stimulation or the effects of growth factors on target tissues.² Excessive estrogen production can cause endometrial hyperplasia and abnormal menstrual bleeding (see Chapter 34). Benign prostatic hyperplasia, which is a common disorder of men older than 50 years of age, is thought to be related to the synergistic action of estrogen and androgens (see Chapter 33). Skin warts are an example of hyperplasia caused by growth factors produced by the human papillomaviruses.

Metaplasia

Metaplasia represents a reversible change in which one adult cell type (epithelial or mesenchymal) is replaced by another adult cell type. Metaplasia is thought to involve the reprogramming of undifferentiated stem cells that are present in the tissue undergoing the metaplastic changes.

Metaplasia usually occurs in response to chronic irritation and inflammation and allows for substitution of cells that are better able to survive under circumstances in which a more fragile cell type might succumb. However, the conversion of cell types never oversteps the boundaries of the primary groups of tissue (e.g., one type of epithelial cell may be converted to another type of epithelial cell, but not to a connective tissue cell). An example of metaplasia is the adaptive substitution of stratified squamous epithelial cells for the ciliated columnar epithelial cells in the trachea and large airways of a habitual cigarette smoker. A vitamin A deficiency also induces squamous metaplasia of the respiratory tract. Although the squamous epithelium is better able to survive in these situations, the protective function that the ciliated epithelium provides for the respiratory tract is lost. In addition, continued exposure to the influences that cause metaplasia may predispose to cancerous transformation of the metaplastic epithelium.

Dysplasia

Dysplasia is characterized by deranged cell growth of a specific tissue that results in cells that vary in size, shape, and appearance. Minor degrees of dysplasia are associated with chronic irritation or inflammation. The pattern is most frequently encountered in metaplastic squamous epithelium of the respiratory tract and uterine cervix. Although dysplasia is abnormal, it is adaptive in that it is potentially reversible after the irritating cause has been removed. Dysplasia is strongly implicated as a precursor of cancer. In cancers of the respiratory tract and the uterine cervix, dysplastic changes have been found adjacent to the foci of cancerous transformation. Through the use of the Papanicolaou (Pap) smear, it has been documented that cancer of the uterine cervix develops in a series of incremental epithelial changes ranging from severe dysplasia to invasive cancer. However, dysplasia is an adaptive process and as such does not necessarily lead to cancer. In many cases, the dysplastic cells revert to their former structure and function.

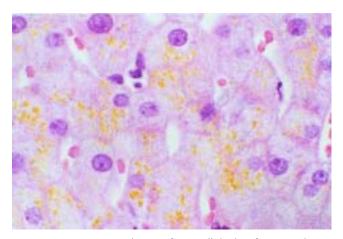
Intracellular Accumulations

Intracellular accumulations represent buildup of substances that cells cannot immediately use or dispose of. The substances may accumulate in the cytoplasm (frequently in the lysosomes) or in the nucleus. In some cases the accumulation may be an abnormal substance that the cell has produced, and in other cases the cell may be storing exogenous materials or products of pathologic processes occurring elsewhere in the body. These substances can be grouped into three categories: (1) normal body substances, such as lipids, proteins, carbohydrates, melanin, and bilirubin, that are present in abnormally large amounts; (2) abnormal endogenous products, such as those resulting from inborn errors of metabolism; and (3) exogenous products, such as environmental agents and pigments that cannot be broken down by the cell.² These substances may accumulate transiently or permanently, and they may be harmless or, in some cases, may be toxic.

The accumulation of normal cellular constituents occurs when a substance is produced at a rate that exceeds its metabolism or removal. An example of this type of process is fatty changes in the liver caused by intracellular accumulation of triglycerides. Liver cells normally contain some fat, which is either oxidized and used for energy or converted to triglycerides. This fat is derived from free fatty acids released from adipose tissue. Abnormal accumulation occurs when the delivery of free fatty acids to the liver is increased, as in starvation and diabetes mellitus, or when the intrahepatic metabolism of lipids is disturbed, as in alcoholism.

Intracellular accumulation can result from genetic disorders that disrupt the metabolism of selected substances. A normal enzyme may be replaced with an abnormal one, resulting in the formation of a substance that cannot be used or eliminated from the cell, or an enzyme may be missing, so that an intermediate product accumulates in the cell. For example, there are at least 10 genetic disorders that affect glycogen metabolism, most of which lead to the accumulation of intracellular glycogen stores. In the most common form of this disorder, von Gierke's disease, large amounts of glycogen accumulate in the liver and kidneys because of a deficiency of the enzyme glucose-6-phosphatase. Without this enzyme, glycogen cannot be broken down to form glucose. The disorder leads not only to an accumulation of glycogen but to a reduction in blood glucose levels. In Tay-Sachs disease, another genetic disorder, abnormal glycolipids accumulate in the brain and other tissues, causing motor and mental deterioration beginning at approximately 6 months of age, followed by death at 2 to 3 years of age. In a similar manner, other enzyme defects lead to the accumulation of other substances.

Pigments are colored substances that may accumulate in cells. They can be endogenous (*i.e.*, arising from within the body) or exogenous (*i.e.*, arising from outside the body). Icterus, also called *jaundice*, is a yellow discoloration of tissue caused by the retention of bilirubin, an endogenous bile pigment. This condition may result from increased bilirubin production from red blood cell destruction, obstruction of bile passage into the intestine, or toxic diseases that affect the liver's ability to remove bilirubin from the blood. Lipofuscin is a yellow-brown pigment that results from the accumulation of the indigestible residues produced during normal turnover of cell structures (Fig. 2-3). The accumulation of lipofuscin increases with age and is sometimes referred to as the *wear-and-tear pig*-



■ FIGURE 2-3 ■ Accumulation of intracellular lipofuscin. A photomicrograph of the liver of an 80-year-old man shows golden cytoplasmic granules, which represent lysosomal storage of lipofuscin. (From Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 13]. Philadelphia: Lippincott Williams & Wilkins)

ment. It is more common in heart, nerve, and liver cells than other tissues and is seen most often in conditions associated with atrophy of an organ.

One of the most common exogenous pigments is carbon in the form of coal dust. In coal miners or persons exposed to heavily polluted environments, the accumulation of carbon dust blackens the lung tissue and may cause serious lung disease. The formation of a blue lead line along the margins of the gum is one of the diagnostic features of lead poisoning. Tattoos are the result of insoluble pigments introduced into the skin, where they are engulfed by macrophages and persist for a lifetime.

The significance of intracellular accumulations depends on the cause and severity of the condition. Many accumulations, such as lipofuscin and mild fatty change, have no effect on cell function. Some conditions, such as the hyperbilirubinemia that causes jaundice, are reversible. Other disorders, such as glycogen storage diseases, produce accumulations that result in organ dysfunction and other alterations in physiologic function.

In summary, cells adapt to changes in their environment and in their work demands by changing their size, number, and characteristics. These adaptive changes are consistent with the needs of the cell and occur in response to an appropriate stimulus. The changes are usually reversed after the stimulus has been withdrawn.

When confronted with a decrease in work demands or adverse environmental conditions, cells atrophy or reduce their size and revert to a lower and more efficient level of functioning. Hypertrophy results from an increase in work demands and is characterized by an increase in tissue size brought about by an increase in cell size and functional components in the cell. An increase in the number of cells in an organ or tissue that is still capable of mitotic division is called hyperplasia. Metaplasia occurs in response to chronic irritation and represents the substitution of cells of a type that are better able to survive under circumstances in which a more fragile cell type might succumb. Dysplasia is characterized by deranged cell growth of a specific tissue that results in cells that vary in size, shape, and appearance. It is a precursor of cancer.

Under some circumstances, cells may accumulate abnormal amounts of various substances. If the accumulation reflects a correctable systemic disorder, such as the hyperbilirubinemia that causes jaundice, the accumulation is reversible. If the disorder cannot be corrected, as often occurs in many inborn errors of metabolism, the cells become overloaded, causing cell injury and death.

CELL INJURY AND DEATH

Cells can be injured in many ways. The extent to which any injurious agent can cause cell injury and death depends in large measure on the intensity and duration of the injury and the type of cell that is involved. Cell injury is usually reversible to a certain point, after which irreversible cell injury and death occur. Whether a specific stress causes irreversible or reversible cell injury depends on the severity of the insult and on vari-

KEY CONCEPTS

CELL INJURY

- Cells can be damaged in a number of ways, including physical trauma, extremes of temperature, electrical injury, exposure to damaging chemicals, radiation damage, injury from biologic agents, and nutritional factors.
- Most injurious agents exert their damaging effects through uncontrolled free radical production, impaired oxygen delivery or utilization, or the destructive effects of uncontrolled intracellular calcium release.
- Cell injury can be reversible, allowing the cell to recover, or it can be irreversible, causing cell death and necrosis.
- In contrast to necrosis, which results from tissue injury, apoptosis is a normal physiologic process designed to remove injured or worn-out cells.

ables such as blood supply, nutritional status, and regenerative capacity. Cell injury and death are ongoing processes, and in the healthy state, they are balanced by cell renewal.

Causes of Cell Injury

Cell damage can occur in many ways. For purposes of discussion, the ways by which cells are injured have been grouped into five categories: (1) injury from physical agents, (2) radiation injury, (3) chemical injury, (4) injury from biologic agents, and (5) injury from nutritional imbalances.

Injury From Physical Agents

Physical agents responsible for cell and tissue injury include mechanical forces, extremes of temperature, and electrical forces. They are common causes of injuries due to environmental exposure, occupational and transportation accidents, and physical violence and assault.

Mechanical Forces. Injury or trauma caused by mechanical forces occurs as the result of body impact with another object. The body or the object can be in motion or, as sometimes happens, both can be in motion at the time of impact. These types of injuries split and tear tissue, fracture bones, injure blood vessels, and disrupt blood flow.

Extremes of Temperature. Extremes of heat and cold cause damage to the cell, its organelles, and its enzyme systems. Exposure to low-intensity heat (43° to 46°C), such as occurs with partial-thickness burns and severe heat stroke, causes cell injury by inducing vascular injury, accelerating cell metabolism, inactivating temperature-sensitive enzymes, and disrupting the cell membrane. With more intense heat, coagulation of blood vessels and tissue proteins occurs. Exposure to cold increases blood viscosity and induces vasoconstriction by direct

action on blood vessels and through reflex activity of the sympathetic nervous system. The resultant decrease in blood flow may lead to hypoxic tissue injury, depending on the degree and duration of cold exposure. Injury from freezing probably results from a combination of ice crystal formation and vasoconstriction. The decreased blood flow leads to capillary stasis and arteriolar and capillary thrombosis.

Electrical Injuries. Electrical injuries can affect the body through extensive tissue injury and disruption of neural and cardiac impulses. The effect of electricity on the body is mainly determined by its voltage, the type of current (*i.e.*, direct or alternating), its amperage, the resistance of the intervening tissue, the pathway of the current, and the duration of exposure.^{2,4}

Lightning and high-voltage wires that carry several thousand volts produce the most severe damage.² Alternating current (AC) is usually more dangerous than direct current (DC) because it causes violent muscle contractions, preventing the person from releasing the electrical source and sometimes resulting in fractures and dislocations. In electrical injuries, the body acts as a conductor of the electrical current. The current enters the body from an electrical source, such as an exposed wire, and passes through the body and exits to another conductor, such as the moisture on the ground or a piece of metal the person is holding. The pathway that a current takes is critical because the electrical energy disrupts impulses in excitable tissues. Current flow through the brain may interrupt impulses from respiratory centers in the brain stem, and current flow through the chest may cause fatal cardiac arrhythmias.

The resistance to the flow of current in electrical circuits transforms electrical energy into heat. This is why the elements in electrical heating devices are made of highly resistive metals. Much of the tissue damage produced by electrical injuries is caused by heat production in tissues that have the highest electrical resistance. Resistance to electrical current varies from the greatest to the least in bone, fat, tendons, skin, muscles, blood, and nerves. The most severe tissue injury usually occurs at the skin sites where the current enters and leaves the body. After electricity has penetrated the skin, it passes rapidly through the body along the lines of least resistancethrough body fluids and nerves. Degeneration of vessel walls may occur, and thrombi may form as current flows along the blood vessels. This can cause extensive muscle and deep tissue injury. Thick, dry skin is more resistant to the flow of electricity than thin, wet skin. It is generally believed that the greater the skin resistance, the greater is the amount of local skin burn, and the less the resistance, the greater are the deep and systemic effects.

Radiation Injury

Electromagnetic radiation comprises a wide spectrum of wavepropagated energy, ranging from ionizing gamma rays to radiofrequency waves. A photon is a particle of radiation energy. Radiation energy above the ultraviolet (UV) range is called *ionizing radiation* because the photons have enough energy to knock electrons off atoms and molecules. *Nonionizing radiation* refers to radiation energy at frequencies below that of visible light. *UV radiation* represents the portion of the spectrum of electromagnetic radiation just above the visible range. It contains increasingly energetic rays that are powerful enough to disrupt intracellular bonds and cause sunburn. *lonizing Radiation.* Ionizing radiation affects cells by causing ionization of molecules and atoms in the cell, by directly hitting the target molecules in the cell, or by producing free radicals that interact with critical cell components.^{1,2,5} It can immediately kill cells, interrupt cell replication, or cause a variety of genetic mutations, which may or may not be lethal. Most radiation injury is caused by localized irradiation that is used in treatment of cancer (see Chapter 5). Except for unusual circumstances, such as the use of high-dose irradiation that precedes bone marrow transplantation, exposure to whole-body irradiation is rare.

The injurious effects of ionizing radiation vary with the dose, dose rate (a single dose can cause greater injury than divided or fractionated doses), and the differential sensitivity of the exposed tissue to radiation injury. Because of the effect on deoxyribonucleic acid (DNA) synthesis and interference with mitosis, rapidly dividing cells of the bone marrow and intestine are much more vulnerable to radiation injury than are tissues such as bone and skeletal muscle. Over time, occupational and accidental exposure to ionizing radiation can result in increased risk for the development of various types of cancers, including skin cancers, leukemia, osteogenic sarcomas, and lung cancer.

Many of the manifestations of radiation therapy result from acute cell injury, dose-dependent changes in the blood vessels that supply the irradiated tissues, and fibrotic tissue replacement. The cell's initial response to radiation injury involves swelling, disruption of the mitochondria and other organelles, alterations in the cell membrane, and marked changes in the nucleus. The endothelial cells in blood vessels are particularly sensitive to irradiation. During the immediate postirradiation period, only vessel dilatation takes place (e.g., the initial erythema of the skin after radiation therapy). Later or with higher levels of radiation, destructive changes occur in small blood vessels such as the capillaries and venules. Acute reversible necrosis is represented by such disorders as radiation cystitis, dermatitis, and diarrhea from enteritis. More persistent damage can be attributed to acute necrosis of tissue cells that are not capable of regeneration and chronic ischemia. Chronic effects of radiation damage are characterized by fibrosis and scarring of tissues and organs in the irradiated area (e.g., interstitial fibrosis of the heart and lungs after irradiation of the chest). Because the radiation delivered in radiation therapy inevitably travels through the skin, radiation dermatitis is common. There may be necrosis of the skin, impaired wound healing, and chronic radiation dermatitis.

Ultraviolet Radiation. Ultraviolet radiation causes sunburn and increases the risk of skin cancers (see Chapter 44). The degree of risk depends on the type of UV rays, the intensity of exposure, and the amount of protective melanin pigment in the skin. Skin damage induced by UV radiation is believed to be caused by reactive oxygen species and by damage to melanin-producing processes in the skin. UV radiation also damages DNA, resulting in the formation of pyrimidine dimers (*i.e.*, the insertion of two identical pyrimidine bases into replicating DNA instead of one). Other forms of DNA damage include the production of single-stranded breaks and formation of DNA-protein cross-links. Normally, errors that occur during DNA replication are repaired by enzymes that remove the faulty section of DNA and repair the damage. The importance of the DNA repair in protecting against UV radiation injury is

evidenced by the vulnerability of persons who lack the enzymes needed to repair UV-induced DNA damage. In a genetic disorder called *xeroderma pigmentosum*, an enzyme needed to repair sunlight-induced DNA damage is lacking. This autosomal recessive disorder is characterized by extreme photosensitivity and a greatly increased risk of skin cancer in sunexposed skin.²

Nonionizing Radiation. Nonionizing radiation includes infrared light, ultrasound, microwaves, and laser energy. Unlike ionizing radiation, which can directly break chemical bonds, nonionizing radiation exerts its effects by causing vibration and rotation of atoms and molecules. All of this vibrational and rotational energy is eventually converted to thermal energy. Lowfrequency nonionizing radiation is used widely in radar, television, industrial operations (e.g., heating, welding, melting of metals, processing of wood and plastic), household appliances (e.g., microwave ovens), and medical applications (e.g., diathermy). Isolated cases of skin burns and thermal injury to deeper tissues have occurred in industrial settings and from improperly used household microwave ovens. Injury from these sources is mainly thermal and, because of the deep penetration of the infrared or microwave rays, tends to involve dermal and subcutaneous tissue injury.

Chemical Injury

Chemicals capable of damaging cells are everywhere around us. Air and water pollution contain chemicals capable of tissue injury, as do tobacco smoke and some processed or preserved foods. Some of the most damaging chemicals exist in our environment, including gases such as carbon monoxide, insecticides, and trace metals such as lead.

Lead is a particularly toxic metal. Small amounts accumulate to reach toxic levels.⁶ There are innumerable sources of lead in the environment, including flaking paint, lead-contaminated dust and soil, lead-contaminated root vegetables, lead water pipes or soldered joints, pottery glazes, and newsprint. Children are exposed to lead through ingestion of peeling lead paint, by breathing dust from lead paint (e.g., during remodeling), or from playing in contaminated soil.⁷ Lead crosses the placenta, exposing the fetus to lead levels comparable to those of the mother. The toxicity of lead is related to its multiple biochemical effects.² It has the ability to inactivate enzymes, compete with calcium for incorporation into bone, and interfere with nerve transmission and brain development. The major targets are the red blood cells, the gastrointestinal tract, the kidneys, and the nervous system. Some of the manifestations of lead toxicity include anemia, acute abdominal pain, signs of kidney damage, and cognitive deficits and neuropathies resulting from demyelination of cerebral and cerebellar white matter and death of cortical nerve cells.

Chemical agents can injure the cell membrane and other cell structures, block enzymatic pathways, coagulate cell proteins, and disrupt the osmotic and ionic balance of the cell. Corrosive substances such as strong acids and bases destroy cells as the substances come into contact with the body. Other chemicals may injure cells in the process of metabolism or elimination. For example, carbon tetrachloride (CCl_4) causes little damage until it is metabolized by liver enzymes to a highly reactive free radical (CCl_3). Carbon tetrachloride is extremely toxic to liver cells.

Drugs. Many drugs—alcohol, prescription drugs, over-thecounter drugs, and street drugs—are capable of directly or indirectly damaging tissues. Ethyl alcohol can harm the gastric mucosa, liver (see Chapter 28), developing fetus (see Chapter 4), and other organs. Antineoplastic (anticancer) and immunosuppressant drugs can directly injure cells. Other drugs produce metabolic end-products that are toxic to cells. Acetaminophen, a commonly used analgesic drug, is detoxified in the liver, where small amounts of the drug are converted to a highly toxic metabolite. This metabolite is detoxified by a metabolic pathway that uses a substance (*i.e.*, glutathione) normally present in the liver. When large amounts of the drug are ingested, this pathway becomes overwhelmed and toxic metabolites accumulate, causing massive liver necrosis.

Injury From Biologic Agents

Biologic agents differ from other injurious agents in that they are able to replicate and can continue to produce their injurious effects. These agents range from submicroscopic viruses to the larger parasites. Biologic agents injure cells by diverse mechanisms. Viruses enter the cell and become incorporated into its DNA synthetic machinery. Certain bacteria elaborate exotoxins that interfere with cellular production of ATP. Other bacteria, such as the gram-negative bacilli, release endotoxins that cause cell injury and increased capillary permeability.

Injury From Nutritional Imbalances

Nutritional excesses and nutritional deficiencies predispose cells to injury. Obesity and diets high in saturated fats are thought to predispose persons to atherosclerosis. The body requires more than 60 organic and inorganic substances in amounts ranging from micrograms to grams. These nutrients include minerals, vitamins, certain fatty acids, and specific amino acids. Dietary deficiencies can occur in the form of starvation, in which there is a deficiency of all nutrients and vitamins, or because of a selective deficiency of a single nutrient or vitamin. Iron-deficiency anemia, scurvy, beriberi, and pellagra are examples of injury caused by the lack of specific vitamins or minerals. The protein and calorie deficiencies that occur with starvation cause widespread tissue damage.

Mechanisms of Cell Injury

The mechanisms by which injurious agents cause cell injury and death are complex. Some agents, such as heat, produce direct cell injury; other factors, such as genetic derangement, produce their effects indirectly through metabolic disturbances and altered immune responses. There seem to be at least three major mechanisms whereby most injurious agents exert their effects: free radical formation, hypoxia and ATP depletion, and disruption of intracellular calcium homeostasis.

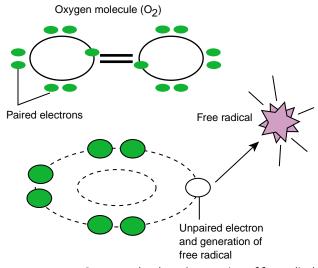
Free Radical Injury

Many injurious agents exert their damaging effects through a reactive chemical species called a *free radical*.^{2,8-10} Free radical injury is rapidly emerging as a final common pathway for tissue damage by many injurious agents.

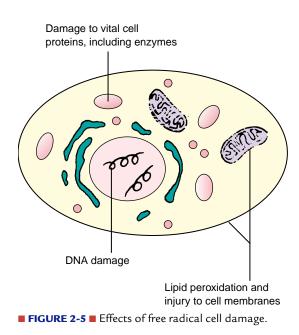
In most atoms, the outer electron orbits are filled with paired electrons moving in opposite directions to balance their spins. A free radical is a highly reactive chemical species arising from an atom that has a single unpaired electron in an outer orbit (Fig. 2-4). In this state, the radical is highly unstable and can enter into reactions with cellular constituents, particularly key molecules in cell membranes and nucleic acids. Moreover, free radicals can establish chain reactions, sometimes thousands of events long, as the molecules they react with in turn form free radicals. Chain reactions may branch, causing even greater damage. Uncontrolled free radical production causes damage to cell membranes, cross-linking of cell proteins, inactivation of enzyme systems, or damage to the nucleic acids that make up DNA.

Free radical formation is a by-product of many normal cellular reactions in the body, including energy generation, breakdown of lipids and proteins, and inflammatory processes. For example, free radical generation is the main mechanism for killing microbes in phagocytic white blood cells. Molecular oxygen (O_2), with its two unpaired outer electrons, is the most common source of free radicals. During the course of normal cell metabolism, cells process energy-producing oxygen into water; in some reactions, a superoxide radical is formed. Lipid oxidation (*i.e.*, peroxidation) is another source of free radicals. Exogenous sources of free radicals include tobacco smoke, certain pollutants and organic solvents, hyperoxic environments, pesticides, and radiation. Some of these compounds and certain medications are metabolized to free radical intermediates that cause oxidative damage to target tissues.

Although the effects of these reactive free radicals are wide ranging, three types of effects are particularly important in cell injury: lipid peroxidation, oxidative modification of proteins, and DNA effects (Fig. 2-5). Destruction of the phospholipids in cell membranes, including the outer plasma membrane and those of the intracellular organelles, results in loss of membrane integrity. Free radical attack on cell proteins, particularly those of critical enzymes, can interrupt vital processes throughout the cell. DNA is an important target of the hydroxyl free radical. Damage can involve single-stranded breaks in DNA, modification of base pairs, and cross-links between strands. In most cases, various DNA repair pathways can repair the damage. However, if the damage is extensive, the cell dies. The effects of free radical-mediated DNA changes have also been implicated in aging and malignant transformation of cells.



■ FIGURE 2-4 ■ Oxygen molecule and generation of free radical.



Under normal conditions, most cells have chemical mechanisms that protect them from the injurious effects of free radicals. These mechanisms commonly break down when the cell is deprived of oxygen or exposed to certain chemical agents, radiation, or other injurious agents. Free radical formation is a particular threat to tissues in which the blood flow has been interrupted and then restored. During the period of interrupted flow, the intracellular mechanisms that control free radicals are inactivated or damaged. When blood flow is restored, the cell is suddenly confronted with an excess of free radicals that it cannot control.

Scientists continue to investigate the use of free radical scavengers to protect against cell injury during periods when protective cellular mechanisms are impaired. Defenses against free radicals include vitamin E, vitamin C, and β -carotene.¹¹ Vitamin E is the major lipid-soluble antioxidant present in all cellular membranes. Vitamin C is an important water-soluble cytosolic chain-breaking antioxidant; it acts directly with superoxide and singlet oxygen radicals. β -carotene, a pigment found in most plants, reacts with singlet oxygen and can also function as an antioxidant.

Hypoxic Cell Injury

Hypoxia deprives the cell of oxygen and interrupts oxidative metabolism and the generation of ATP. The actual time necessary to produce irreversible cell damage depends on the degree of oxygen deprivation and the metabolic needs of the cell. Well-differentiated cells, such as those in the heart, brain, and kidneys, require large amounts of oxygen to provide energy for their special functions. For example, brain cells begin to undergo permanent damage after 4 to 6 minutes of oxygen deprivation. A thin margin can exist between the time involved in reversible and irreversible cell damage. One study found that the epithelial cells of the proximal tubule of the kidney in the rat could survive 20 but not 30 minutes of ischemia.¹²

Hypoxia can result from an inadequate amount of oxygen in the air, respiratory disease, ischemia (*i.e.*, decreased blood flow caused by circulatory disorders), anemia, edema, or inability of the cells to use oxygen. Ischemia is characterized by impaired oxygen delivery and impaired removal of metabolic end-products such as lactic acid. In contrast to pure hypoxia, which affects the oxygen content of the blood and affects all of the cells in the body, ischemia commonly affects blood flow through small numbers of blood vessels and produces local tissue injury. In cases of edema, the distance for diffusion of oxygen may become a limiting factor. In hypermetabolic states, the cells may require more oxygen than can be supplied by normal respiratory function and oxygen transport. Hypoxia also serves as the ultimate cause of cell death in other injuries. For example, toxins from certain microorganisms can interfere with cellular use of oxygen, and a physical agent such as cold can cause severe vasoconstriction and impair blood flow.

Hypoxia literally causes a power failure in the cell, with widespread effects on the cell's functional and structural components. As oxygen tension in the cell falls, oxidative metabolism ceases, and the cell reverts to anaerobic metabolism, using its limited glycogen stores in an attempt to maintain vital cell functions. Cellular pH falls as lactic acid accumulates in the cell. This reduction in pH can have profound effects on intracellular structures. The nuclear chromatin clumps and myelin figures, which derive from destructive changes in cell membranes and intracellular structures, are seen in the cytoplasm and extracellular spaces.

One of the earliest effects of reduced ATP is acute cellular swelling caused by failure of the energy-dependent sodium/ potassium (Na^+/K^+) ATPase membrane pump, which extrudes sodium from and returns potassium to the cell. With impaired function of this pump, intracellular potassium levels decrease, and sodium and water accumulate in the cell. The movement of fluid and ions into the cell is associated with dilatation of the endoplasmic reticulum, increased membrane permeability, and decreased mitochondrial function.² To this point, the cellular changes caused by ischemia are reversible if oxygenation is restored. However, if the oxygen supply is not restored there is a continued loss of essential enzymes, proteins, and ribonucleic acid through the hyperpermeable membrane of the cell. Injury to the lysosomal membranes results in leakage of destructive lysosomal enzymes into the cytoplasm and enzymatic digestion of cell components. Leakage of intracellular enzymes through the permeable cell membrane into the extracellular fluid is used as an important clinical indicator of cell injury and death. These enzymes enter the blood and can be measured by laboratory tests.

Impaired Calcium Homeostasis

Calcium functions as a messenger for the release of many intracellular enzymes. Normally, intracellular calcium levels are kept extremely low compared with extracellular levels. These low intracellular levels are maintained by energy-dependent, membrane-associated calcium/magnesium (Ca²⁺/Mg²⁺) ATPase exchange systems.² Ischemia and certain toxins lead to an increase in cytosolic calcium because of increased influx across the cell membrane and the release of calcium stored in the mitochondria and endoplasmic reticulum. The increased calcium level activates a number of enzymes with potentially damaging effects. The enzymes include the phospholipases responsible for damaging the cell membrane, proteases that damage the cytoskeleton and membrane proteins, ATPases that break down ATP and hasten its depletion, and endonucleases that fragment chromatin.

Reversible Cell Injury and Cell Death

The mechanisms of cell injury can produce sublethal and reversible cellular damage or lead to irreversible injury with cell destruction or death (Fig. 2-6). Cell destruction and removal can involve one of two mechanisms: apoptosis, which is designed to remove injured or worn-out cells, or cell death and necrosis, which occurs in irreversibly damaged cells.

Reversible Cell Injury

Reversible cell injury, although impairing cell function, does not result in cell death. Two patterns of reversible cell injury can be observed under the microscope: cellular swelling and fatty change. Cellular swelling occurs with impairment of the energy-dependent Na⁺/K⁺ ATPase membrane pump, usually as the result of hypoxic cell injury.

Fatty changes are linked to intracellular accumulation of fat. When fatty changes occur, small vacuoles of fat disperse throughout the cytoplasm. The process is usually more ominous than cellular swelling, and although it is reversible, it usually indicates severe injury. These fatty changes may occur because normal cells are presented with an increased fat load or because injured cells are unable to metabolize the fat properly. In obese persons, fatty infiltrates often occur within and between the cells of the liver and heart because of an increased fat load. Pathways for fat metabolism may be impaired during cell injury, and fat may accumulate in the cell as production exceeds use and export. The liver, where most fats are synthesized and metabolized, is particularly susceptible to fatty change, but fatty changes may also occur in the kidney, the heart, and other organs.

Cell Death

In each cell line, the control of cell number is regulated by a balance of cell proliferation and cell death. Cell death can involve apoptosis or necrosis. Apoptotic cell death involves controlled cell destruction and is involved in normal cell deletion and renewal. For example, blood cells that undergo constant renewal from progenitor cells in the bone marrow are removed by apoptotic cell death. Necrotic cell death is a pathologic form of cell death resulting from cell injury. It is characterized by cell swelling, rupture of the cell membrane, and inflammation.

Apoptosis. Apoptosis, from Greek *apo* for "apart" and *ptosis* for "fallen," means fallen apart. Apoptotic cell death, which is equated with cell suicide, eliminates cells that are worn out, have been produced in excess, have developed improperly, or have genetic damage. In normal cell turnover, this process provides the space needed for cell replacement. The process, which was first described in 1972, has become one of the most vigorously investigated processes in biology.¹³ Apoptosis is thought to be involved in several physiologic and pathologic processes. Current research is focusing on the genetic control mechanisms of apoptosis in an attempt to understand the pathogenesis of many disease states such as cancer and autoimmune disease.

Apoptotic cell death is characterized by controlled autodigestion of cell components. Cells appear to initiate their own death through the activation of endogenous enzymes. This results in cell shrinkage brought about by disruption of the cytoskeleton, condensation of the cytoplasmic organelles, disruption and clumping of nuclear DNA, and a distinctive wrinkling of the cell membrane.² As the cell shrinks, the nucleus breaks into spheres, and the cell eventually divides into membranecovered fragments. During the process, membrane changes occur, signaling surrounding phagocytic cells to engulf the cell fragments and complete the degradation process (Fig. 2-7).

Apoptosis is thought to be responsible for several normal physiologic processes, including programmed destruction of cells during embryonic development, hormone-dependent involution of tissues, death of immune cells, cell death by cytotoxic T cells, and cell death in proliferating cell populations. During embryogenesis, in the development of a number of organs such as the heart, which begins as a single pulsating tube and is gradually modified to become a four-chambered pump, apoptotic cell death allows the next stage of organ development. It also separates the webbed fingers and toes of the developing embryo (Fig. 2-8). The control of immune cell numbers and destruction of autoreactive T cells in the thymus have been credited to apoptosis. Cytotoxic T cells and natural killer cells are thought to destroy target cells by inducing apop-

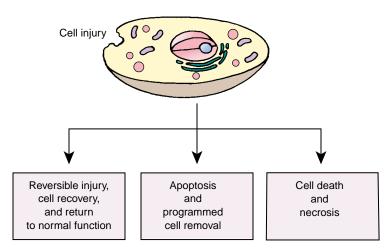
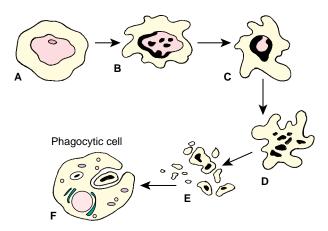


FIGURE 2-6 Outcomes of cell injury: reversible cell injury, apoptosis and programmed cell removal, cell death and necrosis.

33



■ FIGURE 2-7 ■ Apoptotic cell removal: (A) shrinking of the cell structures, (B and C) condensation and fragmentation of the nuclear chromatin, (D and E) separation of nuclear fragments and cytoplasmic organelles into apoptotic bodies, and (F) engulfment of apoptotic fragments by phagocytic cell.

totic cell death. Apoptotic cell death occurs in the hormonedependent involution of endometrial cells during the menstrual cycle and in the regression of breast tissue after weaning from breast-feeding.

Apoptosis appears to be linked to several pathologic processes. For example, suppression of apoptosis may be a determinant in the growth of cancers. Apoptosis is also thought to be involved in the cell death associated with certain viral infections, such as hepatitis B and C, and in cell death caused by a variety of injurious agents, such as mild thermal injury and radiation injury. Apoptosis may also be involved in neurodegenerative disorders such as Alzheimer's disease, Parkinson's disease, and amyotrophic lateral sclerosis (ALS). The loss of cells in these disorders does not induce inflammation; although the initiating event is unknown, apoptosis appears to be the mechanism of cell death.¹⁴

Several mechanisms appear to be involved in initiating cell death by apoptosis. As in the case of endometrial changes that occur during the menstrual cycle, the process can be triggered by the addition or withdrawal of hormones. In hepatitis B and C, the virus seems to sensitize the hepatocytes to apoptosis.¹⁵ Certain oncogenes and suppressor genes involved in the development of cancer seem to play an active role in stimulation or suppression of apoptosis. Injured cells may induce apoptotic cell death through increased cytoplasmic calcium, which leads to activation of nuclear enzymes that break down DNA. In some instances, gene transcription and protein synthesis, the events that produce new cells, may be the initiating factors. In other cases, cell surface signaling or receptor activation appears to be the influencing force.

Necrosis. Necrosis refers to cell death in an organ or tissue that is still part of a living person.² Necrosis differs from apoptosis in that it involves unregulated enzymatic digestion of cell components, loss of cell membrane integrity with uncontrolled release of the products of cell death into the intracellular space, and initiation of the inflammatory response. In contrast to apoptosis, which functions in removing cells so they can be replaced by new cells, necrosis often interferes with cell replacement and tissue regeneration.

With necrotic cell death, there are marked changes in the appearance of the cytoplasmic contents and the nucleus. These changes often are not visible, even under the microscope, for hours after cell death. The dissolution of the necrotic cell or

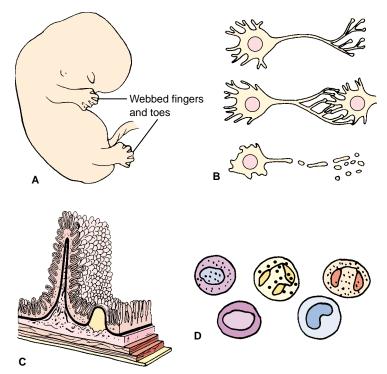


FIGURE 2-8 Examples of apoptosis: (A) separation of webbed fingers and toes in embryo, (B) development of neural connections, (C) removal of cells from intestinal villa, and (D) removal of senescent blood cells.

tissue can follow several paths. The cell can undergo liquefaction (*i.e.*, liquefaction necrosis); it can be transformed to a gray, firm mass (*i.e.*, coagulation necrosis); or it can be converted to a cheesy material by infiltration of fatlike substances (i.e., caseous necrosis). Liquefaction necrosis occurs when some of the cells die but their catalytic enzymes are not destroyed. An example of liquefaction necrosis is the softening of the center of an abscess with discharge of its contents. During coagulation necrosis, acidosis develops and denatures the enzymatic and structural proteins of the cell. This type of necrosis is characteristic of hypoxic injury and is seen in infarcted areas. Infarction (i.e., tissue death) occurs when an artery supplying an organ or part of the body becomes occluded and no other source of blood supply exists. As a rule, the infarct's shape is conical and corresponds to the distribution of the artery and its branches. An artery may be occluded by an embolus, a thrombus, disease of the arterial wall, or pressure from outside the vessel. Caseous necrosis (*i.e.*, soft, cheeselike center) is a distinctive form of coagulation necrosis. It is most commonly associated with tubercular lesions and is thought to result from immune mechanisms.

Gangrene. The term *gangrene* is applied when a considerable mass of tissue undergoes necrosis. Gangrene may be classified as dry or moist. In dry gangrene, the part becomes dry and shrinks, the skin wrinkles, and its color changes to dark brown or black. The spread of dry gangrene is slow, and its symptoms are not as marked as those of wet gangrene. The irritation caused by the dead tissue produces a line of inflammatory reaction (*i.e.*, line of demarcation) between the dead tissue of the gangrenous area and the healthy tissue (Fig. 2-9). Dry gangrene usually results from interference with arterial blood supply to a part without interference with venous return and is a form of coagulation necrosis.

In moist or wet gangrene, the area is cold, swollen, and pulseless. The skin is moist, black, and under tension. Blebs form on the surface, liquefaction occurs, and a foul odor is caused by bacterial action. There is no line of demarcation between the normal and diseased tissues, and the spread of tissue damage is rapid. Systemic symptoms are usually severe, and death may occur unless the condition can be arrested. Moist or



FIGURE 2-9 Gangrenous toes. (Biomedical Communications Group, Southern Illinois University School of Medicine, Springfield, IL)

wet gangrene primarily results from interference with venous return from the part. Bacterial invasion plays an important role in the development of wet gangrene and is responsible for many of its prominent symptoms. Dry gangrene is confined almost exclusively to the extremities, but moist gangrene may affect the internal organs or the extremities. If bacteria invade the necrotic tissue, dry gangrene may be converted to wet gangrene.

Gas gangrene is a special type of gangrene that results from infection of devitalized tissues by one of several *Clostridium* bacteria. These anaerobic and spore-forming organisms are widespread in nature, particularly in soil; gas gangrene is prone to occur in trauma and compound fractures in which dirt and debris are embedded. Some species have been isolated in the stomach, gallbladder, intestine, vagina, and skin of healthy persons. The bacteria produce toxins that dissolve the cell membranes, causing death of muscle cells, massive spreading edema, hemolysis of red blood cells, hemolytic anemia, hemo-globinuria, and renal toxicity.¹⁶ Characteristic of this disorder are the bubbles of hydrogen sulfide gas that form in the muscle. Gas gangrene is a serious and potentially fatal disease. Because the organism is anaerobic, oxygen is sometimes administered in a hyperbaric chamber.

In summary, cell injury can be caused by a number of agents, including physical agents, chemicals, radiation, and biologic agents. Among the physical agents that generate cell injury are mechanical forces that produce tissue trauma, extremes of temperature, electricity, radiation, and nutritional disorders. Chemical agents can cause cell injury through several mechanisms: they can block enzymatic pathways, cause coagulation of tissues, or disrupt the osmotic or ionic balance of the cell. Ionizing radiation affects cells by causing ionization of molecules and atoms in the cell, by directly hitting the target molecules in the cell, or by producing free radicals that interact with critical cell components. Biologic agents differ from other injurious agents in that they are able to replicate and continue to produce injury. Among the nutritional factors that contribute to cell injury are excesses and deficiencies of nutrients, vitamins, and minerals.

Injurious agents exert their effects largely through generation of free radicals, production of cell hypoxia, or unregulated intracellular calcium levels. Partially reduced oxygen species called *free radicals* are important mediators of cell injury in many pathologic conditions. They are an important cause of cell injury in hypoxia and after exposure to radiation and certain chemical agents. Lack of oxygen underlies the pathogenesis of cell injury in hypoxia and ischemia. Hypoxia can result from inadequate oxygen in the air, cardiorespiratory disease, anemia, or the inability of the cells to use oxygen. Increased intracellular calcium activates a number of enzymes with potentially damaging effects.

Injurious agents may produce sublethal and reversible cellular damage or may lead to irreversible cell injury and death. Cell death can involve two mechanisms: apoptosis or necrosis. Apoptosis involves controlled cell destruction and is the means by which the body removes and replaces cells that have been produced in excess, developed improperly, have genetic damage, or are worn out. Necrosis refers to cell death that is characterized by cell swelling, rupture of the cell membrane, and inflammation.

Chapter 2: Tissue Adaptation and Injury

Explain how cell changes that occur with atrophy, hypertrophy, hyperplasia, metaplasia, and dysplasia are adaptive in nature.

Describe the mechanisms whereby physical agents such as blunt trauma, electrical forces, extremes of temperature, and ionizing and nonionizing radiation produce cell injury.

Explain how the injurious effects of biologic agents differ from those produced by physical and chemical agents.

Define *free radical* and relate free radical formation to cell injury and death.

Explain the rationale for the different outcomes in brain tissues of someone who is hypoxic because of severe anemia and someone with an ischemic stroke.

Differentiate cell death associated with necrosis and apoptosis.

Cite the reasons for the changes that occur with the wet and dry forms of gangrene.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

 Rubin E., Farber J.L. (Eds.). (1999). *Pathology* (3rd ed., pp. 6–13, 87–103, 329–330, 338–341). Philadelphia: Lippincott Williams & Wilkins.

- Kumar V., Cotran R.S., Collins T. (Eds.). (2003). Basic Pathology (7th ed., pp. 4–25, 193, 222, 287). Philadelphia: W.B. Saunders.
- Hunter J.J., Chien K.R. (1999). Signaling pathways in cardiac hypertrophy and failure. New England Journal of Medicine 341(17), 1276–1283.
- Goodwin C.W. (1996). Electrical injury. In Bennett J.C., Plum F. (Eds.), *Cecil textbook of medicine* (20th ed., pp. 64–67). Philadelphia: W.B. Saunders.
- Upton A.C. (1996). Radiation injury. In Bennett J.C., Plum F. (Eds.), Cecil textbook of medicine (20th ed., pp. 59–64). Philadelphia: W.B. Saunders.
- Landrigan P.J., Todd A.C. (1994). Lead poisoning. Western Journal of Medicine 161, 153–156.
- Piomelli S. (2000). Lead poisoning. In Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.), *Nelson textbook of pediatrics* (16th ed., pp. 2056–2059). Philadelphia: W.B. Saunders.
- McCord J.M. (2000). The evolution of free radicals and oxidative stress. American Journal of Medicine 108, 652–659.
- 9. Kerr M.E., Bender C.M., Monti E.J. (1996). An introduction to oxygen free radicals. *Heart and Lung* 25, 200–209.
- Betteridge D.J. (2000). What is oxidative stress? *Metabolism* 49(2), 3-8.
- 11. Machlin L.J., Bendich A. (1987). Free radical tissue damage: Protective role of antioxidant nutrients. *FASEB Journal* 1(6), 441–445.
- Vogt M.T., Farber E. (1968). On the molecular pathology of ischemic renal cell death: Reversible and irreversible cellular and mitochondrial metabolic alterations. *American Journal of Pathology* 53, 1–26.
- Skikumar P., Dong Z., Mikhailov V., et. al. (1999). Apoptosis: Definitions, mechanisms, and relevance to disease. *American Journal* of Medicine 107, 490–505.
- 14. Thompson C.B. (1995). Apoptosis in the pathogenesis and treatment of disease. *Science* 267, 1456–1462.
- Rust C., Gores G.J. (2000). Apoptosis and liver disease. American Journal of Medicine 108, 568–575.
- Corry M., Montoya L. (1990). Gas gangrene: Certain diagnosis or certain death. *Critical Care Nursing* 9(10), 30–38.

CHAPTER

Genetic Control of Cell Function

Genetic Control of Cell Function Gene Structure Genetic Code Protein Synthesis Messenger RNA Transfer RNA Ribosomal RNA Regulation of Gene Expression Gene Mutations Chromosomes **Cell Division** Chromosome Structure **Patterns of Inheritance Definitions Genetic Imprinting** Mendel's Laws Pedigree Gene Mapping and Technology **Genomic Mapping Linkage Studies Dosage Studies** Hybridization Studies **Recombinant DNA Technology** Gene Therapy **DNA Fingerprinting**

enetic information is stored in the structure of *deoxyribonucleic acid* (DNA). DNA is an extremely stable macromolecule found in the nucleus of each cell. Because of the stable structure of DNA, the genetic information can survive the many processes of reduction division, in which the gametes (*i.e.*, ovum and sperm) are formed, and the fertilization process. This stability is also maintained throughout the many mitotic cell divisions involved in the formation of a new organism from the single-celled fertilized ovum called the *zygote*. The term *gene* is used to describe a part of the DNA molecule that contains the information needed to code for the types of proteins and enzymes needed for the day-to-day function of the cells in the body. In addition, a gene is the unit of heredity passed from generation to generation. For example, genes control the type and quantity of hormones that a cell produces, the antigens and receptors that are present on the cell membrane, and the synthesis of enzymes needed for metabolism. Of the estimated 35,000 to 140,000 genes that humans possess, more than 5000 have been identified and more than 2300 have been localized to a particular chromosome. With few exceptions, each gene provides the instructions for the synthesis of a single protein. This chapter includes discussions of genetic regulation of cell function, chromosomal structure, patterns of inheritance, and gene technology.

GENETIC CONTROL OF CELL FUNCTION

The genetic information needed for protein synthesis is encoded in the DNA contained in the cell nucleus. A second type of nucleic acid, *ribonucleic acid* (RNA), is involved in the actual synthesis of cellular enzymes and proteins. Cells contain several types of RNA: messenger RNA, transfer RNA, and ribosomal RNA. *Messenger RNA* (mRNA) contains the transcribed instructions for protein synthesis obtained from the DNA molecule and carries them into the cytoplasm. Transcription is followed by translation, the synthesis of proteins according to the instructions carried by mRNA. *Ribosomal RNA* (rRNA) provides the machinery needed for protein synthesis. *Transfer RNA* (tRNA) reads the instructions and delivers the appropriate amino acids to the ribosome, where they are incorporated into the protein being synthesized.

The mechanism for genetic control of cell function is illustrated in Figure 3-1. The nuclei of all the cells in an organism contain the same accumulation of genes derived from the gametes of the two parents. This means that liver cells contain the same genetic information as skin and muscle cells. For this to be true, the molecular code must be duplicated before each succeeding cell division, or mitosis. In theory, although this has not yet been achieved in humans, any of the highly differentiated cells of an organism could be used to produce a complete, genetically identical organism, or *clone*. Each particular cell type

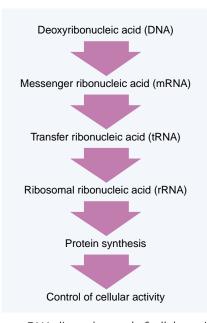
KEY CONCEPTS

FUNCTION OF DNA IN CONTROLLING CELL FUNCTION

- The information needed for the control of cell structure and function is embedded in the genetic information encoded in the stable DNA molecule.
- Although every cell in the body contains the same genetic information, each cell type uses only a portion of the information, depending on its structure and function.
- The production of the proteins that control cell function is accomplished by (1) the transcription of the DNA code for assembly of the protein onto messenger RNA, (2) the translation of the code from messenger RNA and assembly of the protein by ribosomal RNA in the cytoplasm, and (3) the delivery of the amino acids needed for protein synthesis to ribosomal RNA by transfer RNA.

in a tissue uses only part of the information stored in the genetic code. Although information required for the development and differentiation of the other cell types is still present, it is repressed.

Besides nuclear DNA, part of the DNA of a cell resides in the mitochondria. Mitochondrial DNA is inherited from the mother by her offspring (*i.e.*, matrilineal inheritance). Several



■ FIGURE 3-1 ■ DNA-directed control of cellular activity through synthesis of cellular proteins. Messenger RNA carries the transcribed message, which directs protein synthesis, from the nucleus to the cytoplasm. Transfer RNA selects the appropriate amino acids and carries them to ribosomal RNA where assembly of the proteins takes place.

genetic disorders are attributed to defects in mitochondrial DNA. Leber's hereditary optic neuropathy was the first human disease attributed to mutation in mitochondrial DNA.

Gene Structure

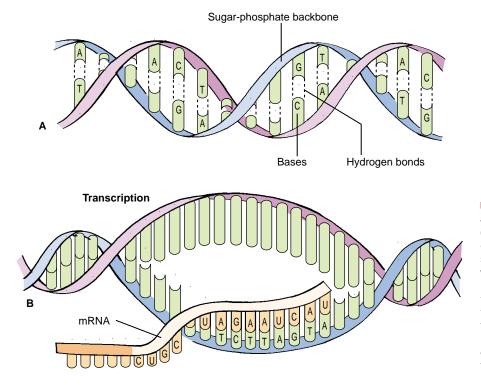
The structure that stores the genetic information in the nucleus is a long, double-stranded, helical molecule of DNA. DNA is composed of nucleotides, which consist of phosphoric acid, a five-carbon sugar called *deoxyribose*, and one of four nitrogenous bases. These nitrogenous bases carry the genetic information and are divided into two groups: the purine bases, adenine and guanine, which have two nitrogen ring structures, and the pyrimidine bases, thymine and cytosine, which have one ring. The backbone of DNA consists of alternating groups of sugar and phosphoric acid; the paired bases project inward from the sides of the sugar molecule. DNA resembles a spiral staircase, with the paired bases representing the steps (Fig. 3-2). A precise complementary pairing of purine and pyrimidine bases occurs in the double-stranded DNA molecule. Adenine is paired with thymine, and guanine is paired with cytosine. Each nucleotide in a pair is on one strand of the DNA molecule, with the bases on opposite DNA strands bound together by hydrogen bonds that are extremely stable under normal conditions. Enzymes called DNA helicases separate the two strands so that the genetic information can be duplicated or transcribed.

Several hundred to almost one million base pairs can represent a gene; the size is proportional to the protein product it encodes. Of the two DNA strands, only one is used in transcribing the information for the cell's polypeptide-building machinery. The genetic information of one strand is meaningful and is used as a template for transcription; the complementary code of the other strand does not make sense and is ignored. However, both strands are involved in DNA duplication. Before cell division, the two strands of the helix separate and a complementary molecule is duplicated next to each original strand. Two strands become four strands. During cell division, the newly duplicated double-stranded molecules are separated and placed in each daughter cell by the mechanics of mitosis. As a result, each of the daughter cells again contains the meaningful strand and the complementary strand joined in the form of a double helix. This type of DNA replication has been termed semiconservative as opposed to conservative (Fig. 3-3).

The DNA molecule is combined with several types of protein and small amounts of RNA into a complex known as *chromatin*. Chromatin is the readily stainable portion of the cell nucleus. Some DNA proteins form binding sites for repressor molecules and hormones that regulate genetic transcription; others may block genetic transcription by preventing access of nucleotides to the surface of the DNA molecule. A specific group of proteins called *histones* are thought to control the folding of the DNA strands.

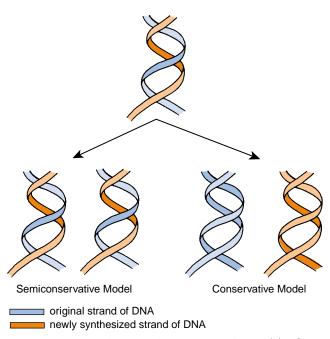
Genetic Code

The four bases—guanine, adenine, cytosine, and thymine (uracil is substituted for thymine in RNA)—make up the alphabet of the genetic code. A sequence of three of these bases forms the fundamental triplet code used in transmitting the genetic information needed for protein synthesis. This triplet code is called



■ FIGURE 3-2 ■ The DNA double helix and transcription of messenger RNA (mRNA). The top panel (**A**) shows the sequence of four bases (adenine [A], cytosine [C], guanine [G], and thymine [T]), which determines the specificity of genetic information. The bases face inward from the sugar-phosphate backbone and form pairs (*dashed lines*) with complementary bases on the opposing strand. In the bottom panel (**B**), transcription creates a complementary mRNA copy from one of the DNA strands in the double helix.

a *codon* (Table 3-1). An example is the nucleotide sequence GCU (guanine, cytosine, and uracil), which is the triplet RNA code for the amino acid alanine. The genetic code is a universal language used by most living cells (*i.e.*, the code for the amino acid tryptophan is the same in a bacterium, a plant, and



■ FIGURE 3-3 ■ Semiconservative vs conservative models of DNA replication as proposed by Meselson and Stahl in 1958. In semiconservative DNA replication, the two original strands of DNA unwind and a complementary strand is formed along each original strand.

a human being). *Stop codes*, which signal the end of a protein molecule, are also present. Mathematically, the four bases can be arranged in 64 different combinations. Sixty-one of the triplets correspond to particular amino acids, and three are stop signals. Only 20 amino acids are used in protein synthesis in humans. Several triplets code for the same amino acid; therefore, the genetic code is said to be redundant or *degenerate*. For example, AUG is a part of the initiation or start signal and the codon for the amino acid methionine. Codons that specify the same amino acid are called *synonyms*. Synonyms usually have the same first two bases but differ in the third base.

Protein Synthesis

Although DNA determines the type of biochemical product that the cell synthesizes, the transmission and decoding of information needed for protein synthesis are carried out by RNA, the formation of which is directed by DNA. The general structure of RNA differs from DNA in three respects: RNA is a single-stranded, rather than a double-stranded, molecule; the sugar in each nucleotide of RNA is ribose, instead of deoxyribose; and the pyrimidine base thymine in DNA is replaced by uracil in RNA. Three types of RNA are known: mRNA, tRNA, and rRNA. All three types are synthesized in the nucleus by RNA polymerase enzymes that take directions from DNA. Because the ribose sugars found in RNA are more susceptible to degradation than the sugars in DNA, the types of RNA molecules in the cytoplasm can be altered rapidly in response to extracellular signals.

Messenger RNA

Messenger RNA is the template for protein synthesis. It is a long molecule containing several hundred to several thousand nucleotides. Each group of three nucleotides forms a codon that

TABLE 3-1	Tri	plet C	odes	for			
Amino Acids							
Amino Acid	RNA (Codons					
Alanine	GCU	GCC	GCA	GCG			
Arginine	CGU	CGC	CGA	CGG	AGA	AGG	
Asparagine	AAU	AAC					
Aspartic acid	GAU	GAC					
Cysteine	UGU	UGC					
Glutamic acid	GAA	GAG					
Glutamine	CAA	CAG					
Glycine	GGU	GGC	GGA	GGG			
Histidine	CAU	CAC					
Isoleucine	AUU	AUC	AUA				
Leucine	CUU	CUC	CUA	CUG	UUA	UUG	
Lysine	AAA	AAG					
Methionine	AUG						
Phenylalanine	UUU	UUC					
Proline	CCU	CCC	CCA	CCG			
Serine	UCU	UCC	UCA	UCG	AGC	AGU	
Threonine	ACU	ACC	ACA	ACG			
Tryptophan	UGG						
Tyrosine	UAU	UAC					
Valine	GUU	GUC	GUA	GUG			
Start (CI)	AUG						
Stop (CT)	UAA	UAG	UGA				

(Guyton A. [2000]. *Textbook of medical physiology* [10th ed., p. 28]. Philadelphia: W.B. Saunders)

is exactly complementary to the triplet of nucleotides of the DNA molecule. Messenger RNA is formed by a process called *transcription*, in which the weak hydrogen bonds of the DNA are broken so that free RNA nucleotides can pair with their exposed DNA counterparts on the meaningful strand of the DNA molecule (see Fig. 3-2). As with the base pairing of the DNA strands, complementary RNA bases pair with the DNA bases. In RNA, uracil replaces thymine and pairs with adenine.

During transcription, a specialized nuclear enzyme, called RNA polymerase, recognizes the beginning or start sequence of a gene. The RNA polymerase attaches to the double-stranded DNA and proceeds to copy the meaningful strand into a single strand of RNA as it travels along the length of the gene. On reaching the stop signal, the enzyme leaves the gene and releases the RNA strand. The RNA strand then is processed. Processing involves the addition of certain nucleic acids at the ends of the RNA strand and cutting and splicing of certain internal sequences. Splicing often involves the removal of stretches of RNA. Because of the splicing process, the final mRNA sequence is different from the original DNA template. RNA sequences that are retained are called *exons*, and those excised are called introns. The functions of the introns are unknown. They are thought to be involved in the activation or deactivation of genes during various stages of development.

Splicing permits a cell to produce a variety of mRNA molecules from a single gene. By varying the splicing segments of the initial mRNA, different mRNA molecules are formed. For example, in a muscle cell, the original tropomyosin mRNA is spliced in as many as 10 different ways, yielding distinctly different protein products. This permits different proteins to be expressed from a single gene and reduces how much DNA must be contained in the genome.

Transfer RNA

The clover-shaped tRNA molecule contains only 80 nucleotides, making it the smallest RNA molecule. Its function is to deliver the activated form of amino acids to protein molecules in the ribosomes. At least 20 different types of tRNA are known, each of which recognizes and binds to only one type of amino acid. Each tRNA molecule has two recognition sites: the first is complementary for the mRNA codon and the second is for the amino acid itself. Each type of tRNA carries its own specific amino acid to the ribosomes, where protein synthesis is taking place; there it recognizes the appropriate codon on the mRNA and delivers the amino acid to the newly forming protein molecule.

Ribosomal RNA

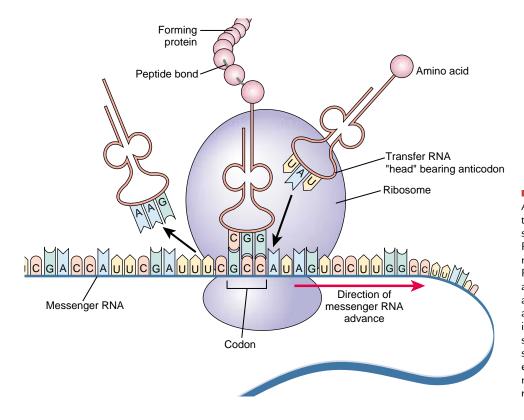
The ribosome is the physical structure in the cytoplasm where protein synthesis takes place. Ribosomal RNA forms 60% of the ribosome, with the remainder of the ribosome composed of the structural proteins and enzymes needed for protein synthesis. As with the other types of RNA, rRNA is synthesized in the nucleus. Unlike other RNAs, ribosomal RNA is produced in a specialized nuclear structure called the *nucleolus*. The formed rRNA combines with ribosomal proteins in the nucleus to produce the ribosome, which is then transported into the cytoplasm. On reaching the cytoplasm, most ribosomes become attached to the endoplasmic reticulum and begin the task of protein synthesis.

Proteins are made from a standard set of amino acids, which are joined end to end to form the long polypeptide chains of protein molecules. Each polypeptide chain may have as many as 100 to more than 300 amino acids in it. The process of protein synthesis is called *translation* because the genetic code is translated into the production language needed for protein assembly. Besides rRNA, translation requires the coordinated actions of mRNA and tRNA. Each of the 20 different tRNA molecules transports its specific amino acid to the ribosome for incorporation into the developing protein molecule. Messenger RNA provides the information needed for placing the amino acids in their proper order for each specific type of protein. During protein synthesis, mRNA contacts and passes through the ribosome, which "reads" the directions for protein synthesis in much the same way that a tape is read as it passes through a tape player (Fig. 3-4). As mRNA passes through the ribosome, tRNA delivers the appropriate amino acids for attachment to the growing polypeptide chain. The long mRNA molecule usually travels through and directs protein synthesis in more than one ribosome at a time. After the first part of the mRNA is read by the first ribosome, it moves onto a second and a third. As a result, ribosomes that are actively involved in protein synthesis are often found in clusters called *polyribosomes*.

Regulation of Gene Expression

Although all cells contain the same genes, not all genes are active all of the time, nor are the same genes active in all cell types. On the contrary, only a small, select group of genes is active in directing protein synthesis in the cell, and this group varies from one cell type to another. For the differentiation

Dr Murtadha Alshareifi e-Library

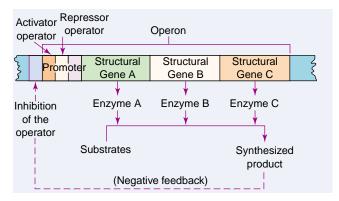


■ FIGURE 3-4 ■ Protein synthesis. A messenger RNA strand is moving through a small ribosomal RNA subunit in the cytoplasm. Transfer RNA transports amino acids to the mRNA strand and recognizes the RNA codon calling for the amino acid by base pairing (through its anticodon). The ribosome then adds the amino acid to the growing polypeptide chain. The ribosome moves along the mRNA strand and is read sequently. As each amino acid is bound to the next by a peptide bond, its tRNA is released.

process of cells to occur in the various organs and tissues of the body, protein synthesis in some cells must be different from that in others. To adapt to an ever-changing environment, certain cells may need to produce varying amounts and types of proteins. Certain enzymes, such as carbonic anhydrase, are synthesized by all cells for the fundamental metabolic processes on which life depends.

The degree to which a gene or particular group of genes is active is called *gene expression*. A phenomenon termed *induction* is an important process by which gene expression is increased. Except in early embryonic development, induction is promoted by some external influence. *Gene repression* is a process whereby a regulatory gene acts to reduce or prevent gene expression. Some genes are normally dormant but can be activated by inducer substances; other genes are naturally active and can be inhibited by repressor substances. Genetic mechanisms for the control of protein synthesis are better understood in microorganisms than in humans. However, it can be assumed that the same general principles apply.

The mechanism that has been most extensively studied is the one by which the synthesis of particular proteins can be turned on and off. For example, in the bacterium *Escherichia coli* grown in a nutrient medium containing the disaccharide lactose, the enzyme galactosidase can be isolated. The galactosidase catalyzes the splitting of lactose into a molecule of glucose and a molecule of galactose. This is necessary if lactose is to be metabolized by *E. coli*. However, if the *E. coli* is grown in a medium that does not contain lactose, very little of the enzyme is produced. From these and other studies, it is theorized that the synthesis of a particular protein, such as galactosidase, requires a series of reactions, each of which is catalyzed by a specific enzyme. At least two types of genes control protein synthesis: *structural genes* that specify the amino acid sequence of a polypeptide chain and *regulator genes* that serve a regulatory function without stipulating the structure of protein molecules. The regulation of protein synthesis is controlled by a sequence of genes, called an *operon*, on adjacent sites on the same chromosome (Fig. 3-5). An operon consists of a set of structural genes that code for enzymes used in the synthesis of a particular product and a promoter site that binds RNA polymerase and initiates transcription of the structural genes. The function of the operon is further regulated by activator and repressor opera-



■ FIGURE 3-5 ■ Function of the operon to control biosynthesis. The synthesized product exerts negative feedback to inhibit function of the operon, in this way automatically controlling the concentration of the product itself. (Guyton A., Hall J.E. [2000]. *Textbook of medical physiology* [10th ed., p. 31]. Philadelphia: W.B. Saunders with permission from Elsevier Science)

tors, which induce or repress the function of the promoter. The activator and repressor sites commonly monitor levels of the synthesized product and regulate the activity of the operon through a negative feedback mechanism. When product levels decrease, the function of the operon is activated, and when levels increase, its function is repressed. Regulatory genes found elsewhere in the genetic complex can exert control over an operon through activator or repressor substances. Not all genes are subject to induction and repression.

Gene Mutations

Rarely, accidental errors in duplication of DNA occur. These errors are called *mutations*. Mutations result from the substitution of one base pair for another, the loss or addition of one or more base pairs, or rearrangements of base pairs. Many of these mutations occur spontaneously; others occur because of environmental agents, chemicals, and radiation. Mutations may arise in somatic cells or in germ cells (ovum or sperm). Only those DNA changes that occur in germ cells can be inherited. A somatic mutation affects a cell line that differentiates into one or more of the many tissues of the body and is not transmissible to the next generation. Somatic mutations that do not have an impact on the health or functioning of a person are called *polymorphisms*. Occasionally, a person is born with one brown eye and one blue eye because of a somatic mutation. The change or loss of gene information is just as likely to affect the fundamental processes of cell function or organ differentiation. Such somatic mutations in the early embryonic period can result in embryonic death or congenital malformations. Somatic mutations are important causes of cancer and other tumors in which cell differentiation and growth get out of control. Each year, hundreds of thousands of random changes occur in the DNA molecule because of environmental events or metabolic accidents. Fortunately, fewer than 1 in 1000 base pair changes result in serious mutations. Most of these defects are corrected by DNA repair mechanisms. Several mechanisms exist, and each depends on specific enzymes such as DNA repair nucleases. Fishermen, farmers, and others who are excessively exposed to the ultraviolet radiation of sunlight have an increased risk for development of skin cancer as a result of potential radiation damage to the genetic structure of the skin-forming cells.

In summary, genes are the fundamental unit of information storage in the cell. They determine the types of proteins and enzymes made by the cell and therefore control inheritance and day-to-day cell function. Genes store information in a stable macromolecule called DNA. Genes transmit information contained in the DNA molecule as a triplet code. The arrangement of the nitrogenous bases of the four nucleotides (i.e., adenine, guanine, thymine [or uracil in RNA], and cytosine) forms the code. The transfer of stored information into production of cell products is accomplished through a second type of macromolecule called RNA. Messenger RNA transcribes the instructions for product synthesis from the DNA molecule and carries it into the cell's cytoplasm, where ribosomal RNA uses the information to direct product synthesis. Transfer RNA acts as a carrier system for delivering the appropriate amino acids to the ribosomes, where the synthesis of cell products occurs. Although all cells contain the same

genes, only a small, select group of genes is active in a given cell type. In all cells, some genetic information is repressed, whereas other information is expressed. Gene mutations represent accidental errors in duplication, rearrangement, or deletion of parts of the genetic code. Fortunately, most mutations are corrected by DNA repair mechanisms in the cell.

CHROMOSOMES

Most genetic information of a cell is organized, stored, and retrieved in small intracellular structures called *chromosomes*. Although the chromosomes are visible only in dividing cells, they retain their integrity between cell divisions. The chromosomes are arranged in pairs; one member of the pair is inherited from the father, the other from the mother. Each species has a characteristic number of chromosomes. In the human, 46 single or 23 pairs of chromosomes are present. Of the 23 pairs of human chromosomes, there are 22 pairs called *autosomes* that are alike in males and females. Each of the 22 pairs of autosomes has the same appearance in all individuals, and each has been given a numeric designation for classification purposes (Fig. 3-6).

The sex chromosomes make up the 23rd pair of chromosomes. Two sex chromosomes determine the sex of a person. All males have an X and Y chromosome (i.e., an X chromosome from the mother and a Y chromosome from the father); all females have two X chromosomes (*i.e.*, one from each parent). Only one X chromosome in the female is active in controlling the expression of genetic traits; however, both X chromosomes are activated during gametogenesis (i.e., formation of the germ cell or ovum). In the female, the active X chromosome is invisible, but the inactive X chromosome can be demonstrated with appropriate nuclear staining. This inactive chromatin mass is seen as the Barr body in epithelial cells or as the drumstick body in the chromatin of neutrophils. The genetic sex of a child can be determined by microscopic study of cell or tissue samples. The total number of X chromosomes is equal to the number of Barr bodies plus one (*i.e.*, an inactive plus an active X chromosome). For example, the cells of a normal female have one Barr body and therefore a total of two X chromosomes. A normal male has no Barr bodies. Males with Klinefelter's syndrome (one Y, an inactive X, plus an active X chromosome) exhibit one Barr body. In the female, whether the X chromosome derived from the mother or that derived from the father is active is determined within a few days after conception; the selection is random for each postmitotic cell line. This is called the Lyon principle, after Mary Lyon, the British geneticist who described it.

Cell Division

There are two types of cell division: mitosis and meiosis. *Meiosis* is limited to replicating germ cells and takes place only once in a cell line. It results in the formation of gametes or reproductive cells (*i.e.*, ovum and sperm), each of which has only a single set of 23 chromosomes. Meiosis is typically divided into two distinct phases, meiotic divisions I and II. Similar to mitosis, cells

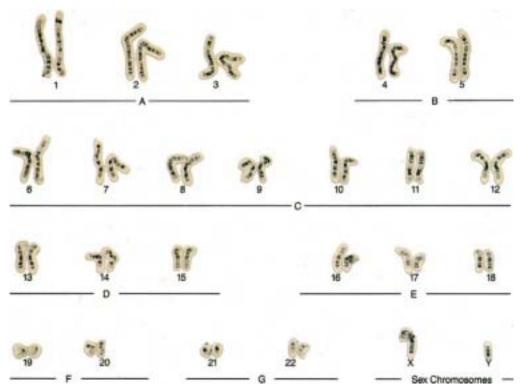


FIGURE 3-6 Karyotype of normal human boy. (Courtesy of the Prenatal Diagnostic and Imaging Center, Sacramento, CA. Frederick W. Hansen, MD, Medical Director)

about to undergo the first meiotic division replicate their DNA during interphase. During metaphase I, homologous chromosomes pair up, forming a synapsis or tetrad (two chromatids per chromosome). They are sometimes called *bivalents*. The X and Y chromosomes are not homologs and do not form bivalents. While in metaphase I, an interchange of chromatid segments can occur. This process is called *crossing over* (Fig. 3-7). Crossing over allows for new combinations of genes, increasing genetic variability. After telophase I, each of the two daughter cells contains one member of each homologous pair of chromosomes and a sex chromosome (23 double-stranded chromosomes).

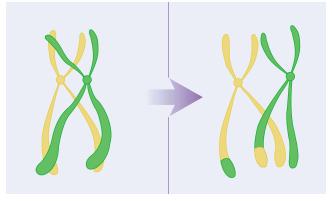
KEY CONCEPTS

CHROMOSOME STRUCTURE

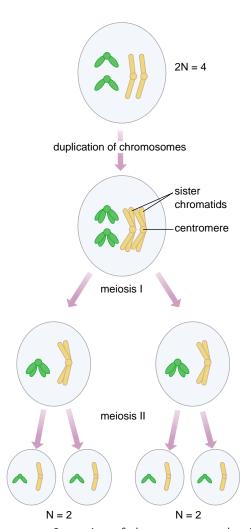
- The DNA that stores genetic material is organized into 23 pairs of chromosomes. There are 22 pairs of autosomes, which are alike for males and females, and one pair of sex chromosomes, with XX pairing in females and XY pairing in males.
- Mitosis refers to the duplication of chromosomes in somatic cell lines, in which each daughter cell receives a pair of 23 chromosomes.
- Meiosis is limited to replicating germ cells and results in the formation of a single set of 23 chromosomes.

No DNA synthesis occurs before meiotic division II. During anaphase II, the 23 double-stranded chromosomes (two chromatids) of each of the two daughter cells from meiosis I divide at their centromeres. Each subsequent daughter cell receives 23 single-stranded chromatids. Thus, a total of four daughter cells are formed by a meiotic division of one cell (Fig. 3-8).

Meiosis, which occurs only in the gamete-producing cells found in either testes or ovaries, has a different outcome in males and females. In males, meiosis (spermatogenesis) results in four viable daughter cells called *spermatids* that differentiate into sperm cells. In females, gamete formation or oogenesis is quite different. After the first meiotic division of a primary oocyte, a secondary oocyte and another structure called a *polar body* are formed. This small polar body contains little cyto-



■ FIGURE 3-7 ■ Crossing over of DNA at the time of meiosis.

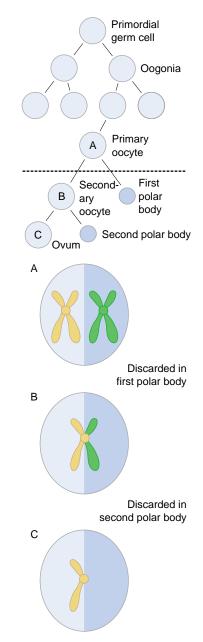


■ FIGURE 3-8 ■ Separation of chromosomes at the time of meiosis.

plasm, but it may undergo a second meiotic division, resulting in two polar bodies (Fig. 3-9). The secondary oocyte undergoes its second meiotic division, producing one mature oocyte and another polar body. Four viable sperm cells are produced during spermatogenesis, but only one ovum from oogenesis.

Chromosome Structure

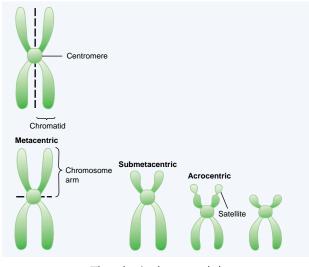
Cytogenetics is the study of the structure and numeric characteristics of the cell's chromosomes. Chromosome studies can be done on any tissue or cell that grows and divides in culture. Lymphocytes from venous blood are frequently used for this purpose. After the cells have been cultured, a drug called *colchicine* is used to arrest mitosis in metaphase. A chromosome spread is prepared by fixing and spreading the chromosomes on a slide. Subsequently, appropriate staining techniques show the chromosomal banding patterns so they can be identified. The chromosomes are photographed, and the photomicrograph of each chromosome is cut out and arranged in pairs according to a standard classification system. The completed picture is called a *karyotype*, and the procedure for preparing the picture is called *karyotyping*. A uniform system of chromosome classification was originally formulated at the 1971 Paris Chro-



■ FIGURE 3-9 ■ Essential stages of meiosis in a female, with discarding of the first and second polar bodies and formation of the ovum with the haploid number of chromosomes. The *dotted line* indicates reduction division. (Adapted from Cormack D.H. [1993]. *Essential histology*. Philadelphia: J.B. Lippincott)

mosome Conference and was later revised to describe the chromosomes as seen in more elongated prophase and prometaphase preparations.

In the metaphase spread, each chromosome takes the form of chromatids to form an "X" or "wishbone" pattern. Human chromosomes are divided into three types according to the position of the centromere (*i.e.*, central constriction; Fig. 3-10). If the centromere is in the center and the arms are of approximately the same length, the chromosome is said to be metacentric; if it is off center and the arms are of clearly different lengths, it is *submetacentric*; and if it is near one end, it is *acrocentric*. The short arm of the chromosome is designated

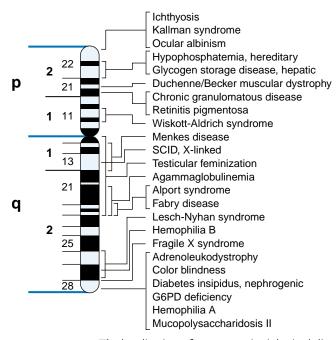


■ FIGURE 3-10 ■ Three basic shapes and the component parts of human metaphase chromosomes. The relative size of the satellite on the acrocentric is exaggerated for visibility. (Adapted from Cormack D.H. [1993]. *Essential histology*. Philadelphia: J.B. Lippincott)

as "p" for "petite," and the long arm is designated as "q" for no other reason than it is the next letter of the alphabet. Arms of the chromosome are indicated by the chromosome number followed by the p or q designation. Chromosomes 13, 14, 15, 21, and 22 have small masses of chromatin called *satellites* attached to their short arms by narrow stalks. At the ends of each chromosome are special DNA sequences called *telomeres*. Telomeres allow the end of the DNA molecule to be replicated completely.

The banding patterns of a chromosome are used in describing the position of a gene. Regions on the chromosomes are numbered from the centromere outward. The regions are further divided into bands and subbands, which are also numbered (Fig. 3-11). These numbers are used in designating the position of a gene on a chromosome. For example, Xp22.2 refers to subband 2, band 2, region 2 of the short arm (p) of the X chromosome.

In summary, the genetic information in a cell is organized, stored, and retrieved as small cellular structures called chromosomes. Forty-six chromosomes arranged in 23 pairs are present in the human being. Twenty-two of these pairs are autosomes. The 23rd pair is the sex chromosomes, which determine the sex of a person. Two types of cell division occur, meiosis and mitosis. Meiosis is limited to replicating germ cells and results in the formation of gametes or reproductive cells (ovum and sperm), each of which has only a single set of 23 chromosomes. Mitotic division occurs in somatic cells and results in the formation of 23 pairs of chromosomes. A karyotype is a photograph of a person's chromosomes. It is prepared by special laboratory techniques in which body cells are cultured, fixed, and then stained to demonstrate identifiable banding patterns. A photomicrograph is then made. Often the individual chromosomes are cut out and regrouped according to chromosome number.



■ FIGURE 3-11 ■ The localization of representative inherited diseases on the X chromosome. Notice the nomenclature of arms (P,Q), regions (1,2), bands (*e.g.*, 11,22). (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 260]. Philadelphia: Lippincott Williams & Wilkins)

PATTERNS OF INHERITANCE

The characteristics inherited from a person's parents are inscribed in gene pairs located along the length of the chromosomes. Alternate forms of the same gene are possible (*i.e.*, one inherited from the mother and the other from the father), and each may produce a different aspect of a trait.

KEY CONCEPTS

TRANSMISSION OF GENETIC INFORMATION

- The transmission of information from one generation to the next is vested in genetic material transferred from each parent at the time of conception.
- Alleles are the alternate forms of a gene (one from each parent), and the locus is the position that they occupy on the chromosome.
- The genotype of a person represents the sum total of the genetic information in the cells and the phenotype the physical manifestations of that information.
- Penetrance is the percentage in a population with a particular genotype in which that genotype is phenotypically manifested, whereas expressivity is the manner in which the gene is expressed.

Genetics has its own set of definitions. The *genotype* of a person is the genetic information stored in the base sequence triplet code. The *phenotype* refers to the recognizable traits, physical or biochemical, associated with a specific genotype. Often, the genotype is not evident by available detection methods. More than one genotype may have the same phenotype. Some brown-eyed persons are carriers of the code for blue eyes, and other brown-eyed persons are not. Phenotypically, these two types of brown-eyed persons are the same, but genotypically they are different.

When it comes to a genetic disorder, not all persons with a mutant gene are affected to the same extent. *Expressivity* refers to the manner in which the gene is expressed in the phenotype, which can range from mild to severe. *Penetrance* represents the ability of a gene to express its function. Seventy-five percent penetrance means 75% of persons of a particular genotype demonstrate a recognizable phenotype. Syndactyly (webbed fingers) and blue sclera are genetic mutations that often do not exhibit 100% penetrance.

The position of a gene on a chromosome is called its *locus*, and alternate forms of a gene at the same locus are called *alleles*. When only one pair of genes is involved in the transmission of information, the term *single-gene trait* is used. Single-gene traits follow the Mendelian laws of inheritance.

Polygenic inheritance involves multiple genes at different loci, with each gene exerting a small additive effect in determining a trait. Most human traits are determined by multiple pairs of genes, many with alternate codes, accounting for some of the dissimilar forms that occur with certain genetic disorders. Polygenic traits are predictable but less so than single-gene traits. *Multifactorial* inheritance is similar to polygenic inheritance in that multiple alleles at different loci affect the outcome; the difference is that multifactorial inheritance includes environmental effects on the genes.

Many other gene–gene interactions are known. These include *epistasis*, in which one gene masks the phenotypic effects of another nonallelic gene; *multiple alleles*, in which more than one allele affects the same trait (*e.g.*, ABO blood types); *complementary genes*, in which each gene is mutually dependent on the other; and *collaborative genes*, in which two different genes influencing the same trait interact to produce a phenotype neither gene alone could produce.

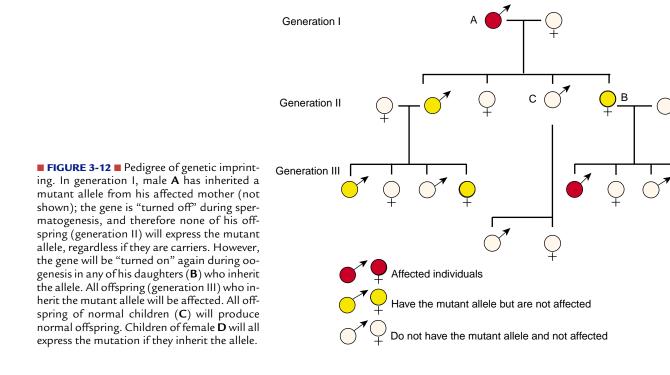
Genetic Imprinting

Besides autosomal and sex-linked genes and mitochondrial inheritance, it was found that certain genes exhibit a "parent of origin" type of transmission, in which the parental genomes do not always contribute equally in the development of an individual (Fig. 3-12). Although rare, it is estimated that approximately 100 genes exhibit genomic, or genetic, imprinting. Evidence suggests there is a genetic conflict over the developing embryo: the male genome attempts to establish larger offspring, whereas the female prefers smaller offspring to conserve her energy for the current and subsequent pregnancies.

An example of a disorder involving parenteral imprinting is uniparental disomy. This occurs when two chromosomes of the same number are inherited from one parent. Normally, this is not a problem, except in cases where a chromosome has been imprinted by a parent. If an allele is inactivated by imprinting, the offspring will have only one working copy of the chromosome, resulting in possible problems.

Mendel's Laws

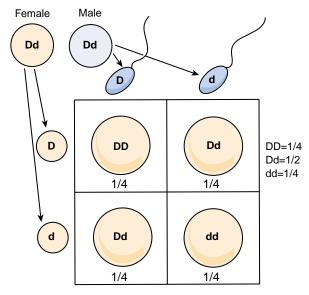
The main feature of inheritance is predictability: given certain conditions, the likelihood of the occurrence or recurrence of a specific trait is remarkably predictable. The units of inheritance are the genes, and the pattern of single-gene expression often



can be predicted using Mendel's laws of genetic transmission. Techniques and discoveries since Gregor Mendel's original work was published in 1865 have led to some modification of his original laws.

Mendel discovered the basic pattern of inheritance by conducting carefully planned experiments with simple garden peas. Experimenting with several phenotypic traits in peas, Mendel proposed that inherited traits are transmitted from parents to offspring by means of independently inherited factors now known as genes—and that these factors are transmitted as recessive and dominant traits. Mendel labeled dominant factors (his round peas) "A" and recessive factors (his wrinkled peas) "a." Geneticists continue to use capital letters to designate dominant traits and lowercase letters to identify recessive traits. The possible combinations that can occur with transmission of single-gene dominant and recessive traits can be described by constructing a figure called a *Punnett square* using capital and lowercase letters (Fig. 3-13).

The observable traits of single-gene inheritance are inherited by the offspring from the parents. During maturation, the primordial germ cells (*i.e.*, sperm and ovum) of both parents undergo meiosis, or reduction division, in which the number of chromosomes is divided in half (from 46 to 23). At this time, the two alleles from a gene locus separate so that each germ cell receives only one allele from each pair (*i.e.*, Mendel's first law). According to Mendel's second law, the alleles from the different gene loci segregate independently and recombine randomly in the zygote. Persons in whom the two alleles of a given pair are the same (AA or aa) are called *homozygotes*. *Heterozygotes* have different alleles (Aa) at a gene locus. A *recessive trait* is one expressed only in a homozygous pairing; a *dominant trait* is one expressed in either a homozygous or a heterozygous pairing. All persons with a dominant allele (depending on the



■ FIGURE 3-13 ■ The Punnett square showing all possible combinations for transmission of a single gene trait (dimpled cheeks). The example shown is when both parents are heterozygous (dD) for the trait. The alleles carried by the mother are on the left and those carried by the father are on the top. The D allele is dominant and the d allele is recessive. The DD and Dd offspring have dimples and the dd offspring does not.

penetrance of the genes) manifest that trait. A *carrier* is a person who is heterozygous for a recessive trait and does not manifest the trait. For example, the genes for blond hair are recessive and those for brown hair are dominant. Therefore, only persons with a genotype having two alleles for blond hair would be blond; persons with either one or two brown alleles would have dark hair.

Pedigree

A pedigree is a graphic method for portraying a family history of an inherited trait. It is constructed from a carefully obtained family history and is useful for tracing the pattern of inheritance for a particular trait.

In summary, inheritance represents the likelihood of the occurrence or recurrence of a specific genetic trait. The genotype refers to information stored in the genetic code of a person, whereas the phenotype represents the recognizable traits, physical and biochemical, associated with the genotype. Expressivity refers to the expression of a gene in the phenotype, and penetrance is the ability of a gene to express its function. The point on the DNA molecule that controls the inheritance of a particular trait is called a gene locus. Alternate forms of a gene at one gene locus are called alleles. The alleles at a gene locus may carry recessive or dominant traits. A recessive trait is one expressed only when there are two copies (homozygous) of the recessive alleles. Dominant traits are expressed with either homozygous or heterozygous pairing of the alleles. A pedigree is a graphic method for portraying a family history of an inherited trait.

GENE MAPPING AND TECHNOLOGY

The genome is the gene complement of an organism. Genomic mapping is the assignment of genes to specific chromosomes or parts of the chromosome. The Human Genome Project, which started in 1990, is an international project to identify and localize the estimated 35,000 to 140,000 genes in the human genome. This is a phenomenal undertaking because there are approximately 3.2 billion base pairs in the human genome. The national effort toward genomic mapping is jointly coordinated by the National Institutes of Health and the Department of Energy. Organizers of the U.S. Human Genome Project hope to have the completed map by the year 2005. It is also anticipated that the project will reveal the chemical basis for as many as 4000 genetic diseases. It is expected to provide tests for screening and diagnosing genetic disorders and the basis for new treatments. To date, chromosomes 14, 20, 21 and 22 have been completely mapped.

Although the human genome project has received most of the publicity, genome sequencing also is underway in a variety of animals, such as the mouse and fruit fly, and other organisms, from bacteria to plants. These and other endeavors will be of benefit in animal husbandry and agriculture and the study of genetic similarities among various species. The human genome project has created an anticipation of future benefits from its undertakings as well as ethical concerns that must be dealt with by both scientists and the public.

Genomic Mapping

Two types of genomic maps exist: genetic maps and physical maps. Genetic maps are like highway maps. They use linkage studies (*e.g.*, dosage, hybridization) to estimate the distances between chromosomal landmarks (*i.e.*, gene markers). Physical maps are similar to a surveyor's map. They measure the actual physical distance between chromosomal elements in biochemical units, the smallest being the nucleotide base.

Genetic maps and physical maps have been refined over the decades. The earliest mapping efforts localized genes on the X chromosome. The initial assignment of a gene to a particular chromosome was made in 1911 for the color blindness gene inherited from the mother (*i.e.*, following the X-linked pattern of inheritance). In 1968, the specific location of the Duffy blood group on the long arm of chromosome 1 was determined. The locations of more than 2300 expressed human genes have been mapped to a specific chromosome and most of them to a specific region on the chromosome. However, genetic mapping is continuing so rapidly that these numbers are constantly being updated. Documentation of gene assignments to specific human chromosomes is updated almost daily in the Online Mendelian Inheritance in Man (http://www.ncbi.nlm.nih.gov/Omim), an encyclopedia of expressed gene loci. Another source is the Human Genome Project, which is the central database for mapped genes and an international repository for most mapping information. Many methods have been used for developing genetic maps. The most important ones are family linkage studies, gene dosage methods, and hybridization studies. Often, the specific assignment of a gene is made using information from several mapping techniques.

Most of the genome mapping has been accomplished by a method called *transcript mapping*. The two-part process begins with isolating mRNAs immediately after they are transcribed. Next, the complementary DNA molecule is prepared. Although transcript mapping may provide knowledge of the gene sequence, it does not automatically mean that the function of the genetic material has been determined.

Linkage Studies

Mendel's laws were often insufficient to explain the transmission of several well-known traits, such as color blindness and hemophilia A. In some families it was noted that these two conditions were transmitted together.

Linkage studies assume that genes occur in a linear array along the chromosomes. During meiosis, the paired chromosomes of the diploid germ cell exchange genetic material because of the crossing-over phenomenon (see Fig. 3-7). This exchange usually involves more than one gene; large blocks of genes (representing large portions of the chromosome) usually are exchanged. Although the point at which the block separates from another occurs randomly, the closer together two genes are on the same chromosome, the greater the chance is that they will be passed on together to the offspring. When two inherited traits occur together at a rate significantly greater than would occur by chance, they are said to be *linked*.

Several methods take advantage of the crossing over and recombination of genes to map a particular gene. In one method, any gene that is already assigned to a chromosome can be used as a marker to assign other linked genes. For example, it was found that an extra long chromosome 1 and the Duffy blood group were inherited as a dominant trait, placing the position of the blood group gene close to the extra material on chromosome 1. Color blindness has been linked to classic hemophilia A (i.e., lack of factor VIII) in some pedigrees; hemophilia A has been linked to glucose-6-phosphate dehydrogenase deficiency in others; and color blindness has been linked to glucose-6-phosphate dehydrogenase deficiency in still others. Because the gene for color blindness is found on the X chromosome, all three genes must be located in a small section of the X chromosome. Linkage analysis can be used clinically to identify affected persons in a family with a known genetic defect. Males, because they have one X and one Y chromosome, are said to be hemizygous for sex-linked traits. Females can be homozygous (normal or mutant) or heterozygous for sex-linked traits. Heterozygous females are known as carriers for X-linked defects.

One autosomal recessive disorder that has been successfully diagnosed prenatally by linkage studies using amniocentesis is congenital adrenal hyperplasia (due to 21-hydroxylase deficiency), which is linked to an immune response gene (human leukocyte antigen [HLA] type). Postnatally, linkage studies have been used in diagnosing hemochromatosis, which is closely linked to another HLA type. Persons with this disorder are unable to metabolize iron, and it accumulates in the liver and other organs. It cannot be diagnosed by conventional means until irreversible damage has been done. Given a family history of the disorder, HLA typing can determine if the gene is present, and if it is present, dietary restriction of iron intake may be used to prevent organ damage.

Dosage Studies

Dosage studies involve measuring enzyme activity. Autosomal genes are normally arranged in pairs, and normally both are expressed. If both alleles are present and both are expressed, the activity of the enzyme should be 100%. If one member of the gene pair is missing, only 50% of the enzyme activity is present, reflecting the activity of the remaining normal allele.

Hybridization Studies

A recent biologic discovery revealed that two somatic cells from different species, when grown together in the same culture, occasionally fuse to form a new hybrid cell. Two types of hybridization methods are used in genomic studies: somatic cell hybridization and in situ hybridization.

Somatic cell hybridization involves the fusion of human somatic cells with those of a different species (typically, the mouse) to yield a cell containing the chromosomes of both species. Because these hybrid cells are unstable, they begin to lose chromosomes of both species during subsequent cell divisions. This makes it possible to obtain cells with different partial combinations of human chromosomes. By studying the enzymes that these cells produce, it is possible to determine that an enzyme is produced only when a certain chromosome is present; the coding for that enzyme must be located on that chromosome.

In situ hybridization involves the use of a specific sequence of DNA or RNA to locate genes that do not express themselves in cell culture. DNA and RNA can be chemically tagged with radioactive or fluorescent markers. These chemically tagged DNA or RNA sequences are used as probes to determine gene location. The probe is added to a chromosome spread after the DNA strands have been separated. If the probe matches the complementary DNA of a chromosome segment, it hybridizes and remains at the precise location (thus the term *in situ*) on a chromosome. Radioactive or fluorescent markers are used to determine the location of the probe.

Recombinant DNA Technology

During the past several decades, genetic engineering has provided the methods for manipulating nucleic acids and recombining genes (recombinant DNA) into hybrid molecules that can be inserted into unicellular organisms and reproduced many times over. Each hybrid molecule produces a genetically identical population, called a *clone*, that reflects its common ancestor.

The techniques of gene isolation and cloning rely on the fact that the genes of all organisms, from bacteria through mammals, are based on similar molecular organization. Gene cloning requires cutting a DNA molecule apart, modifying and reassembling its fragments, and producing copies of the modified DNA, its mRNA, and its gene product. The DNA molecule is cut apart by using a bacterial enzyme, called a *restriction enzyme*, that binds to DNA wherever a particular short sequence of base pairs is found and cleaves the molecule at a specific nucleotide site. In this way, a long DNA molecule can be broken into smaller, discrete fragments with the intent that one fragment contains the gene of interest. More than 100 restriction enzymes are commercially available that cut DNA at different recognition sites.

The selected gene fragment is replicated through insertion into a unicellular organism, such as a bacterium. To do this, a cloning vector such as a bacterial virus or a small DNA circle that is found in most bacteria, called a *plasmid*, is used. Viral and plasmid vectors replicate autonomously in the host bacterial cell. During gene cloning, a bacterial vector and the DNA fragment are mixed and joined by a special enzyme called a DNA ligase. The recombinant vectors formed are then introduced into a suitable culture of bacteria, and the bacteria are allowed to replicate and express the recombinant vector gene. Sometimes, mRNA taken from a tissue that expresses a high level of the gene is used to produce a complementary DNA molecule that can be used in the cloning process. Because the fragments of the entire DNA molecule are used in the cloning process, additional steps are taken to identify and separate the clone that contains the gene of interest.

In terms of biologic research and technology, cloning makes it possible to identify the DNA sequence in a gene and produce the protein product encoded by a gene. The specific nucleotide sequence of a cloned DNA fragment can often be identified by analyzing the amino acid sequence and mRNA codons of its protein product. Short sequences of base pairs can be synthesized, radioactively labeled, and subsequently used to identify their complementary sequence. In this way, identifying normal and abnormal gene structures is possible. Proteins that formerly were available only in small amounts can now be made in large quantities once their respective genes have been isolated. For example, genes encoding for insulin and growth hormone have been cloned to produce these hormones for pharmacologic use.

Gene Therapy

Although quite different from inserting genetic material into a unicellular organism such as bacteria, techniques are available for inserting genes into the genome of intact multicellular plants and animals. Promising delivery vehicles for these genes are the adenoviruses. These viruses are ideal vehicles because their DNA does not become integrated into the host genome; however, repeated inoculations are often needed because the body's immune system usually targets cells expressing adenovirus proteins. Sterically stable liposomes also show promise as DNA delivery mechanisms. This type of therapy is one of the more promising methods for the treatment of genetic disorders, certain cancers, cystic fibrosis, and many infectious diseases.

Two main approaches are used in gene therapy: transferred genes either replace defective genes or selectively inhibit deleterious genes. Cloned DNA sequences or ribosomes usually are the compounds used in gene therapy. However, the introduction of the cloned gene into the multicellular organism can influence only the few cells that get the gene. An answer to this problem would be the insertion of the gene into a sperm or ovum; after fertilization, the gene would be replicated in all of the differentiating cell types. Even so, techniques for cell insertion are limited. Not only are moral and ethical issues involved, but these techniques cannot direct the inserted DNA to attach to a particular chromosome or supplant an existing gene by knocking it out of its place.

DNA Fingerprinting

The technique of DNA fingerprinting is based in part on those techniques used in recombinant DNA technology and those originally used in medical genetics to detect slight variations in the genomes of different individuals. Using restrictive endonucleases, DNA is cleaved at specific regions. The DNA fragments are separated according to size by electrophoresis (*i.e.*, Southern blot) and transferred to a nylon membrane. The fragments are then broken apart and subsequently annealed with a series of radioactive probes specific for regions in each fragment. An autoradiograph reveals the DNA fragments on the membrane. When used in forensic pathology, this procedure is undertaken on specimens from the suspect and the forensic specimen. Banding patterns are then analyzed to see if they match. With conventional methods of analysis of blood and serum enzymes, a 1 in 100 to 1000 chance exists that the two specimens match because of chance. With DNA fingerprinting, these odds are 1 in 100,000 to 1 million.

In summary, the genome is the gene complement of an organism. Genomic mapping is a method used to assign genes to particular chromosomes or parts of a chromosome. The most important ones used are family linkage studies, gene dosage methods, and hybridization studies. Often the specific assignment of a gene is determined by using information from several mapping techniques. Linkage studies assign a chromosome location to genes based on their close association with other genes of known location. Recombinant DNA studies involve the extraction of specific types of messenger RNA used in synthesis of complementary DNA strands. The complementary DNA strands, labeled with a radioisotope, bind with the genes for which they are complementary and are used as gene probes. Now underway is an international project to identify and localize all 35,000 to 140,000 genes in the human genome. Genetic engineering has provided the methods for manipulating nucleic acids and recombining genes (recombinant DNA) into hybrid molecules that can be inserted into unicellular organisms and reproduced many times over. As a result, proteins that formerly were available only in small amounts can now be made in large quantities once their respective genes have been isolated. DNA fingerprinting, which relies on recombinant DNA technologies and those of genetic mapping, is often used in forensic investigations.

REVIEW QUESTIONS

- Explain the mechanisms by which genes control cell function and the characteristics of the next generation.
- Define gene locus and allele.
- Explain the difference between a person's genotype and phenotype.
- Describe the concepts of *induction* and *repression* as they apply to gene function.
- Describe the pathogenesis of gene mutation.

Explain how gene expressivity and penetrance determine the effects of a mutant gene that codes for the production of an essential enzyme.

Differentiate between autosomes and sex chromosomes and between meiosis and mitosis.

Explain the significance of the Barr body.

Construct a hypothetical pedigree for a recessive and dominant trait according to Mendel's law.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

BIBLIOGRAPHY

- Alberts B., Johnson A., Lewis J., et al. (2002). *Molecular biology of the cell* (4th ed.). New York: Garland Publishing.
- Carlson B.M. (1999). Human embryology and developmental biology (2nd ed., pp. 2–23, 128–145). St. Louis: Mosby.
- Guyton A.C., Hall J.E. (2000). Textbook of medical physiology (10th ed., pp. 24–37). Philadelphia: W.B. Saunders.
- Hawley R.S., Mori C.A. (1999). *The human genome: A user's guide.* San Diego: Harcourt Academic Press.
- Human Genome Project. (2000). Columbia encyclopedia (6th ed.). [On-line]. Available: http://www.bartleby.com/65/hu/_HumanG.html. Accessed October 22, 2000.
- The International RH Mapping Consortium. (2000). A new gene map of the human genome. [On-line]. Available: http://www.ornl.gov/ TechResources/Human_Genome. Accessed March 2003.
- Jain H.K. (1999). Genetics: Principles, concepts and implications. Enfield, NH: Science Publishers.
- Moore K.L., Persaud T.V.N. (1998). *The developing human: Clinically orientated embryology* (6th ed., pp. 17–46). Philadelphia: W.B. Saunders.
- Nussbaum, R.L., McInnes R.R., Willard H.F. (2001). Thompson & Thompson genetics in medicine. Philadelphia: W.B. Saunders.
- Ostrer H. (1998). Non-Mendelian genetics in humans. New York: Oxford University Press.
- Sadler T.W. (2000). Langman's medical embryology (8th ed., pp. 3-30, 112-135). Philadelphia: Lippincott Williams & Wilkins.
- Sapienza C. (1990). Parenteral imprinting of genes. Scientific American 263(4), 52–60.
- Shapiro L.J. (2000). Molecular basis of genetic disorders. In Behrman R.E., Kliegman R.M., Jenson H. (Eds.), *Nelson's textbook of pediatrics* (16th ed., pp. 313–317). Philadelphia: W.B. Saunders.
- Snustad D.P., Simmons J.M. (Eds.). (2000). *Principles of genetics* (2nd ed., pp. 3–21, 52–71, 91–115, 665–669). New York: John Wiley & Sons.

Genetic and Congenital Disorders

Genetic and Chromosomal Disorders Single-Gene Disorders Autosomal Dominant Disorders Autosomal Recessive Disorders X-Linked Disorders Multifactorial Inheritance Disorders Chromosomal Disorders Alterations in Chromosome Duplication Alterations in Chromosome Number Alterations in Chromosome Structure Disorders Due to Environmental Influences Period of Vulnerability Teratogenic Agents Radiation Chemicals and Drugs

Infectious Agents

CHAPTER

Genetic and congenital defects are important at all levels of health care because they affect all age groups and can involve almost any of the body tissues and organs. Congenital defects, sometimes called *birth defects*, develop during prenatal life and usually are apparent at birth or shortly thereafter. Spina bifida and cleft lip, for example, are apparent at birth, but other malformations, such as kidney and heart defects, may be present at birth but may not become apparent until they begin to produce symptoms. Not all genetic disorders are congenital, and many are not apparent until later in life.

Birth defects, which affect more than 150,000 infants each year, are the leading cause of infant death.¹ Birth defects may be caused by genetic factors (*i.e.*, single-gene or multifactorial inheritance or chromosomal aberrations), or they may be caused by environmental factors that occurred during embryonic or fetal development (*i.e.*, maternal disease, infections, or drugs taken during pregnancy). In rare cases, congenital defects may be the result of intrauterine factors such as fetal crowding, positioning, or entanglement of fetal parts with the amnion. This chapter provides an overview of genetic and congenital disorders and is divided into three parts: (1) genetic and chromosomal disorders, (2) disorders caused by environmental agents, and (3) diagnosis and counseling.

GENETIC AND CHROMOSOMAL DISORDERS

Genetic disorders involve a permanent change (or mutation) in the genome. A genetic disorder can involve a single-gene trait, multifactorial inheritance, or a chromosome disorder.

Single-Gene Disorders

Single-gene disorders are caused by a single defective or mutant gene. The defective gene may be present on an autosome or the X chromosome and it may affect only one member of an autosomal gene pair (matched with a normal gene) or both members of the pair. Single-gene defects follow the mendelian patterns of inheritance (see Chapter 3) and are often called

KEY CONCEPTS

GENETIC AND CHROMOSOMAL DISORDERS

- Genetic disorders are inherited as autosomal dominant disorders, in which each child has a 50% chance of inheriting the disorder, and as autosomal recessive disorders, in which each child has a 25% chance of being affected, a 50% chance of being a carrier, and a 25% chance of being unaffected.
- Sex-linked disorders almost always are associated with the X chromosome and are predominantly recessive.
- Chromosomal disorders reflect events that occur at the time of meiosis and result from defective movement of an entire chromosome or from breakage of a chromosome with loss or translocation of genetic material.

mendelian disorders. At last count, there were more than 8000 single-gene disorders, many of which have been mapped to a specific chromosome.²

The genes on each chromosome are arranged in pairs and in strict order, with each gene occupying a specific location or locus. The two members of a gene pair, one inherited from the mother and the other from the father, are called *alleles*. If the members of a gene pair are identical (*i.e.*, code the exact same gene product), the person is *homozygous*, and if the two members are different, the person is *heterozygous*. The genetic composition of a person is called a *genotype*, whereas the *phenotype* is the observable expression of a genotype in terms of morphologic, biochemical, or molecular traits. If the trait is expressed in the heterozygote (one member of the gene pair codes for the trait), it is said to be *dominant*; if it is expressed only in the homozygote (both members of the gene pair code for the trait), it is *recessive*.

Although gene expression usually follows a dominant or recessive pattern, it is possible for both alleles (members) of a gene pair to be fully expressed in the heterozygote, a condition called *codominance*. Many genes have only one normal version, called a *wild-type* allele. Other genes have more than one normal allele (alternate forms) at the same locus. This is called *polymorphism*. Blood group inheritance (*e.g.*, AO, BO, AB) is an example of codominance and polymorphism.

A single mutant gene may be expressed in many different parts of the body. Marfan's syndrome is a defect in connective tissue that has widespread effects involving skeletal, eye, and cardiovascular structures. In other single-gene disorders, the same defect can be caused by mutations at several different loci. Childhood deafness can result from 16 different types of autosomal recessive mutations.

Single-gene disorders are characterized by their patterns of transmission, which usually are obtained through a family genetic history. The patterns of inheritance depend on whether the phenotype is dominant or recessive, and whether the gene is located on an autosomal or sex chromosome (see Chapter 3). Disorders of autosomal inheritance include autosomal dominant and autosomal recessive traits. Among the approximate 8000 single-gene disorders, more than half are autosomal dominant. Autosomal recessive phenotypes are less common, accounting for approximately one third of single-gene disorders.3 Currently, all sex-linked genetic disorders are thought to be X-linked, and most are recessive. The only mutations affecting the Y-linked genes are involved in spermatogenesis and male fertility and thus are not transmitted. A few additional genes with homologs on the X chromosome have been mapped to the Y chromosome, but to date, no disorders resulting from mutations in these genes have been described.

Virtually all single-gene disorders lead to formation of an abnormal protein or decreased production of a gene product. The defect can result in defective or decreased amounts of an enzyme, defects in receptor proteins and their function, alterations in nonenzyme proteins, or mutations resulting in unusual reactions to drugs. Table 4-1 lists some of the common single-gene disorders and their manifestations.

Autosomal Dominant Disorders

In autosomal dominant disorders, a single mutant allele from an affected parent is transmitted to an offspring regardless of sex. The affected parent has a 50% chance of transmitting the disorder to each offspring (Fig. 4-1). The unaffected relatives of the parent or unaffected siblings of the offspring do not transmit the disorder. In many conditions, the age of onset is delayed, and the signs and symptoms of the disorder do not appear until later in life, as in Huntington's chorea (see Chapter 37).

Autosomal dominant disorders also may manifest as a new mutation. Whether the mutation is passed on to the next generation depends on the affected person's reproductive capacity. Many new autosomal dominant mutations are accompanied by reduced reproductive capacity; therefore, the defect is not perpetuated in future generations. If an autosomal defect is accompanied by a total inability to reproduce, essentially all new cases of the disorder will be due to new mutations. If the defect does not affect reproductive capacity, it is more likely to be inherited.

Although there is a 50% chance of inheriting a dominant genetic disorder from an affected parent, there can be wide variation in gene penetration and expression. When a person inherits a dominant mutant gene but fails to express it, the trait is described as having *reduced penetrance*. Penetrance is expressed in mathematical terms; a 50% penetrance indicates that a person who inherits the defective gene has a 50% chance of expressing the disorder. The person who has a mutant gene but does not express it is an important exception to the rule that unaffected persons do not transmit an autosomal dominant trait. These persons can transmit the gene to their descendants and so produce a skipped generation. Autosomal dominant disorders also can display variable expressivity, meaning that they can be expressed differently among individuals. For example, polydactyly or the presence of more than the usual number of digits may be expressed in the fingers or the toes.

The gene products of autosomal dominant disorders usually are regulatory proteins involved in rate-limiting components of complex metabolic pathways or key components of structural proteins such as collagen.^{4,5} Two disorders of autosomal inheritance, Marfan's syndrome and neurofibromatosis (NF), are described in this chapter.

Marfan's Syndrome. Marfan's syndrome is a connective tissue disorder that is manifested by changes in the skeleton, eves, and cardiovascular system. There is a wide range of variation in expression of the disorder. Persons may have abnormalities of one or all three systems. The skeletal deformities, which are the most obvious features of the disorder, include a long, thin body with exceptionally long extremities and long, tapering fingers, sometimes called arachnodactyly or spider fingers (Fig. 4-2), hyperextensible joints, and a variety of spinal deformities including kyphoscoliosis. Chest deformity, pectus excavatum (*i.e.*, deeply depressed sternum), or pigeon chest deformity, often is present. The most common eve disorder is bilateral dislocation of the lens caused by weakness of the suspensory ligaments. Myopia and predisposition to retinal detachment also are common, the result of increased optic globe length due to altered connective tissue support of ocular structures. However, the most life-threatening aspects of the disorder are the cardiovascular defects, which include mitral valve prolapse, progressive dilation of the aortic valve ring, and weakness of the aorta and other arteries. Dissection and rupture of the aorta often lead to premature death. The average age of death in persons with Marfan's syndrome is 30 to 40 years.⁴

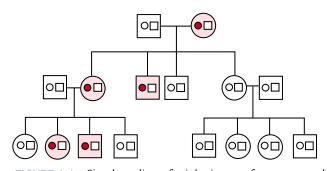
TABLE 4-1

Some Disorders of Mendelian or Single-Gene Inheritance

and Their Significance

Disorder	Significance
Autosomal Dominant	
Achondroplasia	Short-limb dwarfism
Adult polycystic kidney disease	Kidney failure
Huntington's chorea	Neurodegenerative disorder
Familial hypercholesterolemia	Premature atherosclerosis
Marfan's syndrome	Connective tissue disorder with abnormalities in skeletal, ocular, cardiovascular systems
Neurofibromatosis (NF)	Neurogenic tumors: fibromatous skin tumors, pigmented skin lesions, and ocular nodules in NF-1; bilateral acoustic neuromas in NF-2
Osteogenesis imperfecta	Molecular defects of collagen
Spherocytosis	Disorder of red blood cells
von Willebrand's disease	Bleeding disorder
Autosomal Recessive	
Color blindness	Color blindness
Cystic fibrosis	Disorder of membrane transport of ions in exocrine glands causing lung and pancreatic disease
Glycogen storage diseases	Excess accumulation of glycogen in the liver and hypoglycemia (von Gierke's disease); glycogen accumulation in striated muscle in myopathic forms
Oculocutaneous albinism	Hypopigmentation of skin, hair, eyes as result of inability to synthesize melanin
Phenylketonuria (PKU)	Lack of phenylalanine hydroxylase with hyperphenylalaninemia and impaired brain development
Sickle cell disease	Red blood cell defect
Tay-Sachs disease	Deficiency of hexosaminidase A; severe mental and physical deterioration beginning in infancy
X-Linked Recessive	
Bruton-type hypogammaglobulinemia	Immunodeficiency
Hemophilia A	Bleeding disorder
Duchenne dystrophy	Muscular dystrophy
Fragile X syndrome	Mental retardation

Neurofibromatosis. Neurofibromatosis is a condition involving neurogenic tumors that arise from Schwann cells and other elements of the peripheral nervous system.^{4,5} There are at least two genetically and clinically distinct forms of the disorder: type 1 NF (NF-1), also known as von Recklinghausen's disease, and type 2 bilateral acoustic NF (NF-2). Both of these disorders result from a genetic defect in a protein that regulates cell growth. The gene for NF-1 has been mapped to chromosome 17, and the gene for NF-2 has been mapped to chromosome 22.



■ FIGURE 4-1 ■ Simple pedigree for inheritance of an autosomal dominant trait. The small, colored circle represents the mutant gene. An affected parent with an autosomal dominant trait has a 50% chance of passing the mutant gene on to each child regardless of sex.

NF-1 is a relatively common disorder with a frequency of 1 in 3000.5 Approximately 50% of cases have a family history of autosomal dominant transmission, and the remaining 50% appear to represent a new mutation. In more than 90% of persons with NF-1, cutaneous and subcutaneous neurofibromas



■ FIGURE 4-2 ■ Long, slender fingers (arachnodactyly) in a patient with Marfan's syndrome. (Rubin E., Farber J.L. [1999]. Pathology [3rd ed., p. 242]. Philadelphia: Lippincott Williams & Wilkins)

develop in late childhood or adolescence. The cutaneous neurofibromas, which vary in number from a few to many hundreds, manifest as soft, pedunculated lesions that project from the skin. They are the most common type of lesion, often are not apparent until puberty, and are present in greatest density over the trunk (Fig. 4-3). The subcutaneous lesions grow just below the skin; they are firm and round, and may be painful. Plexiform neurofibromas involve the larger peripheral nerves. They tend to form large tumors that cause severe disfigurement of the face or an extremity. Pigmented nodules of the iris (Lisch nodules), which are specific for NF-1, usually are present after 6 years of age. They do not present any clinical problem but are useful in establishing a diagnosis.

A second major component of NF-1 is the presence of large (usually \geq 15 mm in diameter), flat cutaneous pigmentations, known as *café-au-lait spots*. They are usually a uniform light brown in whites and darker brown in African Americans and have sharply demarcated edges (Fig. 4-4). Although small single lesions may be found in normal children, larger lesions of six or more spots larger than 1.5 cm in diameter suggest NF-1. The skin pigmentations become more evident with age as the melanosomes in the epidermal cells accumulate melanin.

In addition to neurofibromatoses, persons with NF-1 have a variety of other associated lesions, the most common being skeletal lesions such as scoliosis and erosive bone defects. Persons with NF-1 also are at increased risk for development of other nervous system tumors such as meningiomas, optic gliomas, and pheochromocytomas.

NF-2 is characterized by tumors of the acoustic nerve. Most often, the disorder is asymptomatic through the first 15 years of life. The most frequently reported symptoms are headaches, hearing loss, and tinnitus (*i.e.*, ringing in the ears). There may be associated intracranial and spinal meningiomas. The condition is made worse by pregnancy, and oral contraceptives may increase the growth and symptoms of tumors. Persons with the disorder should be warned that severe disorientation may occur during diving or swimming underwater, and drowning may result. Surgery may be indicated for debulking or removal of the tumors.

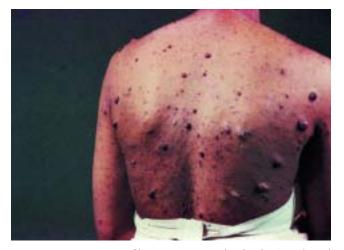


■ FIGURE 4-4 ■ Neurofibromatosis with early café-au-lait spots in a 5-year-old child. (Owen Laboratories, Inc.) (Sauer G.C., Hall J.C. [1996]. *Manual of skin diseases*. Philadelphia: Lippincott-Raven)

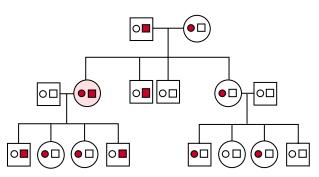
Autosomal Recessive Disorders

Autosomal recessive disorders are manifested only when both members of the gene pair are affected. In this case, both parents may be unaffected but are carriers of the defective gene. Autosomal recessive disorders affect both sexes. The occurrence risk in each pregnancy is one in four for an affected child, two in four for a carrier child, and one in four for a normal (noncarrier, unaffected) homozygous child (Fig. 4-5).

With autosomal recessive disorders, the age of onset is frequently early in life; the symptomatology tends to be more uniform than with autosomal dominant disorders; and the disorders are characteristically caused by deficiencies in enzymes, rather than abnormalities in structural proteins. In the case of a heterozygous carrier, the presence of a mutant gene usually does not produce symptoms because equal amounts of normal and defective enzymes are synthesized. The "margin of safety" ensures that cells with half their usual amount of enzyme function normally. By contrast, the inactivation of both alleles in a



■ FIGURE 4-3 ■ Neurofibromatosis on the back. (Reed and Carnick Pharmaceuticals) (Sauer G.C., Hall J.C. [1996]. *Manual of skin diseases*. Philadelphia: Lippincott-Raven)



■ FIGURE 4-5 ■ Simple pedigree for inheritance of an autosomal recessive trait. The small, colored circle and square represent a mutant gene. When both parents are carriers of a mutant gene, there is a 25% chance of having an affected child, a 50% chance of a carrier child, and a 25% chance of a nonaffected or noncarrier child, regardless of sex. All children (100%) of an affected parent are carriers.

homozygote results in complete loss of enzyme activity. Autosomal recessive disorders include almost all inborn errors of metabolism. Enzyme disorders that impair catabolic pathways result in an accumulation of dietary substances (*e.g.*, phenylketonuria [PKU]) or cellular constituents (*e.g.*, lysosomal storage diseases). Other disorders result from a defect in the enzymemediated synthesis of an essential protein (*e.g.*, the cystic fibrosis transmembrane conductance regulator in cystic fibrosis). Two examples of autosomal recessive disorders that are not covered elsewhere in this book are PKU and Tay-Sachs disease.

Phenylketonuria. Phenylketonuria is a genetically inherited enzyme defect. It is characterized by a deficiency of phenylalanine hydroxylase, the enzyme needed for conversion of phenylalanine to tyrosine. As a result of this deficiency, toxic levels of phenylalanine accumulate in the blood. Like other inborn errors of metabolism, PKU is inherited as a recessive trait and is manifested only in the homozygote. Untreated, PKU results in severe mental retardation.

PKU occurs once in approximately 10,000 births, and damage to the developing brain almost always results when high concentrations of phenylalanine and other metabolites persist in the blood.⁴ Because the symptoms of untreated PKU develop gradually and would often go undetected until irreversible mental retardation had occurred, newborn infants are routinely screened for abnormal levels of serum phenylalanine. It is important that blood samples for PKU screening be obtained at least 12 hours after birth to ensure accuracy.⁶ It also is possible to identify carriers of the trait by subjecting them to a phenylalanine test, in which a large dose of phenylalanine is administered orally and the rate at which it disappears from the bloodstream is measured.

Infants with the disorder are treated with a special diet that restricts phenylalanine intake. Dietary treatment must be started early in neonatal life to prevent brain damage. The results of dietary therapy of children with PKU have been impressive. The diet can prevent mental retardation as well as other neurodegenerative effects of untreated PKU.

Tay-Sachs Disease. Tay-Sachs disease is a variant of a class of lysosomal storage diseases, known as gangliosidoses, in which substances (gangliosides) found in membranes of nervous tissue are deposited in neurons of the central nervous system and retina because of a failure of lysosomal degradation.^{4,5} The disease is particularly prevalent among eastern European (Ashkenazi) Jews. Infants with Tay-Sachs disease appear normal at birth but begin to manifest progressive weakness, muscle flaccidity, and decreased attentiveness at approximately 6 to 10 months of age. This is followed by rapid deterioration of motor and mental function, often with development of generalized seizures. Retinal involvement leads to visual impairment and eventual blindness. Death usually occurs before 4 years of age. Although there is no cure for the disease, analysis of the blood serum for a deficiency of the lysosomal enzyme, hexosaminidase A, which is deficient in Tay-Sachs disease, allows for accurate identification of the genetic carriers for the disease.

X-Linked Disorders

Sex-linked disorders are almost always associated with the X, or female, chromosome, and the inheritance pattern is predominantly recessive. Because of a normal paired gene, female heterozygotes rarely experience the effects of a defective gene. The common pattern of inheritance is one in which an unaffected mother carries one normal and one mutant allele on the X chromosome. This means that she has a 50% chance of transmitting the defective gene to her sons, and her daughters have a 50% chance of being carriers of the mutant gene. When the affected son procreates, he transmits the defective gene to all of his daughters, who become carriers of the mutant gene. Because the genes of the Y chromosome are unaffected, the affected male does not transmit the defect to any of his sons, and they will not be carriers or transmit the disorder to their children. X-linked recessive disorders include the fragile X syndrome, glucose-6-phosphate dehydrogenase deficiency (see Chapter 13), hemophilia A (see Chapter 12), and X-linked agammaglobulinemia (see Chapter 10).

Fragile X Syndrome. Fragile X syndrome is an X-linked disorder associated with a fragile site on the X chromosome where the chromatin fails to condense during mitosis. As with other X-linked disorders, fragile X syndrome affects males more often than females. The disorder, which affects approximately 1 in 1000 male infants, is the second most common cause of mental retardation, after Down syndrome.⁷

Affected males are mentally retarded and share a common physical phenotype that includes a long face with large mandible; large, everted ears; and large testicles (macroorchidism). Hyperextensible joints, a high-arched palate, and mitral valve prolapse, which is observed in some cases, mimic a connective tissue disorder. Some physical abnormalities may be subtle or absent. The most distinctive feature, which is present in 90% of prepubertal males, is macroorchidism.^{5,8}

In 1991, the fragile X syndrome was mapped to a small area on the X chromosome (Xq27), now designated FMR-1 (fragile X, mental retardation 1) site.⁷ The mechanism by which the normal FMR-1 gene is converted to an altered, or mutant, gene capable of producing disease symptoms involves an increase in the length of the gene. A small region of the gene that contains the CCG triplet code undergoes repeated duplication, resulting in a longer gene. The longer gene is susceptible to methylation, a chemical process that results in inactivation of the gene. When the number of repeats is small (<200), the person often has few or no manifestations of the disorder, compared with those evidenced in persons with a larger number of repeats.

In fragile X families, the probability of being affected with the disorder is related to the position in the pedigree. Later generations are more likely to be affected than earlier generations. For example, brothers of transmitting males are at a 9% risk of having mental retardation, whereas grandsons of transmitting males are at a 40% risk.⁵ Approximately 20% of males who have been shown to carry the fragile X mutation are clinically and cytogenetically normal. Because male carriers transmit the trait through all their daughters (who are phenotypically normal) to affected grandchildren, they are called *transmitting males*. Approximately 50% of female carriers are affected (mentally retarded), a proportion that is higher than with other Xlinked disorders.⁵

Multifactorial Inheritance Disorders

Multifactorial inheritance disorders are caused by multiple genes and, in many cases, environmental factors. The exact number of genes contributing to multifactorial traits is not known, and these traits do not follow a clear-cut pattern of inheritance, as do single-gene disorders. Multifactorial inheritance has been described as a threshold phenomenon in which the factors contributing to the trait might be compared with water filling a glass.9 Using this analogy, one might say that expression of the disorder occurs when the glass overflows. Disorders of multifactorial inheritance can be expressed during fetal life and be present at birth, or they may be expressed later in life. Congenital disorders that are thought to arise through multifactorial inheritance include cleft lip or palate, clubfoot, congenital dislocation of the hip, congenital heart disease, pyloric stenosis, and urinary tract malformation. Environmental factors are thought to play a significant role in disorders of multifactorial inheritance that develop in adult life, such as coronary artery disease, diabetes mellitus, hypertension, cancer, and common psychiatric disorders such as manic-depressive psychoses and schizophrenia.

Although multifactorial traits cannot be predicted with the same degree of accuracy as the mendelian single-gene mutations, characteristic patterns exist. First, multifactorial congenital malformations tend to involve a single organ or tissue derived from the same embryonic developmental field. Second, the risk of recurrence in future pregnancies is for the same or a similar defect. This means that parents of a child with a cleft palate defect have an increased risk of having another child with a cleft palate, but not with spina bifida. Third, the increased risk (compared with the general population) among first-degree relatives of the affected person is 2% to 7%, and among second-degree relatives, it is approximately one-half that amount.5 The risk increases with increasing incidence of the defect among relatives. This means that the risk is greatly increased when a second child with the defect is born to a couple. The risk also increases with severity of the disorder and when the defect occurs in the sex not usually affected by the disorder.

Chromosomal Disorders

Chromosomal disorders form a major category of genetic disease, accounting for a large proportion of reproductive wastage (early gestational abortions), congenital malformations, and mental retardation. Specific chromosomal abnormalities can be linked to more than 60 identifiable syndromes that are present in 0.7% of all live births, 2% of all pregnancies in women older than 35 years of age, and 50% of all firstterm abortions.³

During cell division (*i.e.*, mitosis) in nongerm cells, the chromosomes replicate so that each cell receives a full diploid number. In germ cells, a different form of division (*i.e.*, meiosis) takes place. During meiosis, the double sets of 22 autosomes and the 2 sex chromosomes (normal diploid number) are reduced to single sets (haploid number) in each gamete. At the time of conception, the haploid number in the ovum and that in the sperm join and restore the diploid number of chromosomes. Chromosomal defects usually develop because of defective movement during meiosis or because of breakage of a chromosome with loss or translocation of genetic material.

Chromosome abnormalities are commonly described according to the shorthand description of the karyotype. In this system, the total number of chromosomes is given first, followed by the sex chromosome complement, and then the description of any abnormality. For example, a male with trisomy 21 is designated 47,XY,+21.

Alterations in Chromosome Duplication

Mosaicism is the presence in one individual of two or more cell lines characterized by distinctive karyotypes. This defect results from an accident during chromosomal duplication. Sometimes, mosaicism consists of an abnormal karyotype and a normal one, in which case the physical deformities caused by the abnormal cell line usually are less severe.

Alterations in Chromosome Number

A change in chromosome number is called *aneuploidy*. Among the causes of aneuploidy is failure of the chromosomes to separate during oogenesis or spermatogenesis. This can occur in the autosomes or the sex chromosomes and is called nondisjunction. Nondisjunction gives rise to germ cells that have an even number of chromosomes (22 or 24). The products of conception formed from this even number of chromosomes have an uneven number of chromosomes, 45 or 47. Monosomy refers to the presence of only one member of a chromosome pair. The defects associated with monosomy of the autosomes are severe and usually cause abortion. Monosomy of the X chromosome (45,X/O), or Turner's syndrome, causes less severe defects. Polysomy, or the presence of more than two chromosomes to a set, occurs when a germ cell containing more than 23 chromosomes is involved in conception. This defect has been described for the autosomes and the sex chromosomes. Trisomies of chromosomes 8, 13, 18, and 21 are the more common forms of polysomy of the autosomes. There are several forms of polysomy of the sex chromosomes in which extra X or Y chromosomes are present.

Trisomy 21. First described in 1866 by John Langon Down, trisomy 21, or Down syndrome, causes a combination of birth defects, including characteristic facial features, some degree of mental retardation, and other health problems. According to the National Down Syndrome Association, it is the most common chromosomal disorder, occurring approximately once in every 800 to 1000 births. Currently, there are approximately 350,000 people in the United States with Down syndrome.¹⁰

Approximately 95% of cases of Down syndrome are caused by nondisjunction or an error in cell division during meiosis, resulting in a trisomy of chromosome 21. Most of the remaining cases are due to a translocation in which part of chromosome 21 breaks off and attaches to another chromosome (usually chromosome 14). Although there still are only 46 chromosomes in the cell, the presence of the extra part of chromosome 21 causes the features of Down syndrome.

The risk of having a child with Down syndrome increases with maternal age: it is 1/1300 at 25 years of age, 1/365 at 35 years, and 1/30 at 45 years of age.¹¹ The reason for the correlation between maternal age and nondisjunction is unknown, but is thought to reflect some aspect of aging of the oocyte. Although males continue to produce sperm throughout their reproductive life, females are born with all the oocytes they ever will have. These oocytes may change as a result of the aging process. With increasing age, there is a greater chance of a woman having been exposed to damaging environmental agents such as drugs, chemicals, and radiation.

56 Unit One: Mechanisms of Disease

The physical features of a child with Down syndrome are distinctive, and therefore the condition usually is apparent at birth. These features include a small and rather square head. There is upward slanting of the eyes; small, low-set, and malformed ears; a fat pad at the back of the neck; an open mouth; and a large, protruding tongue (Fig. 4-6). The child's hands usually are short and stubby, with fingers that curl inward, and there usually is only a single palmar (*i.e.*, simian) crease. Hypotonia and joint laxity also are present in infants and young children. There often are accompanying congenital heart defects and an increased risk of gastrointestinal malformations. Approximately 1% of persons with trisomy 21 Down syndrome have mosaicism (i.e., cell populations with the normal chromosome number and trisomy 21); these persons may be less severely affected. Of particular concern is the much greater risk of development of acute leukemia among children with Down syndrome-10 to 20 times greater than that of other children.5 With increased life expectancy due to improved health care, it has been found that there is an increased risk of Alzheimer's disease among older persons with Down syndrome.

There are several prenatal screening tests that can be done to determine the risk of having a child with Down syndrome. The most commonly used test is the triple screen— α -fetoprotein (AFP), human chorionic gonadotropin (HCG), and unconjugated estriol. The results of these tests, which usually are done between 15 and 20 weeks of gestation, together with the woman's age often are used to determine the probability of a pregnant woman having a child with Down syndrome. These tests are able to accurately detect only approximately 60% of fetuses with Down syndrome. Some women are given falsepositive readings, and some are given false-negative readings. In 1992, fetal nuchal (back of neck) translucency, as measured by ultrasonography in the first trimester, was proposed as another screening measure; the nucha was found to be thicker in fetuses with Down syndrome.12 This test, which must be done by a highly trained professional, continues to be investigated as a screening method. The only way to accurately determine



■ **FIGURE 4-6** ■ A child with Down syndrome. (Courtesy of March of Dimes Birth Defects Foundation, White Plains, NY)

the presence of Down syndrome in the fetus is through chromosome analysis using chorionic villus sampling, amniocentesis, or percutaneous umbilical blood sampling.

Monosomy X. Monosomy X, or Turner's syndrome, describes a monosomy of the X chromosome (45,X/0) with gonadal agenesis, or absence of the ovaries. This disorder affects approximately 1 of every 2500 live births, and it has been estimated that more than 99% of fetuses with the 45,X/0 karyotype are spontaneously aborted during the first trimester.¹³ There are variations in the syndrome, with abnormalities ranging from essentially none to webbing of the neck with redundant skin folds, nonpitting lymphedema of the hands and feet, and congenital heart defects, particularly coarctation of the aorta. There also may be abnormalities in kidney development (*i.e.*, abnormal location, abnormal vascular supply, or double collecting system). There may be other abnormalities, such as changes in nail growth, high-arched palate, short fourth metacarpal, and strabismus.

Characteristically, the female with Turner's syndrome is short in stature, but her body proportions are normal. She does not menstruate and shows no signs of secondary sex characteristics. When a mosaic cell line (*i.e.*, 45,X/0 and 46,X/X or 45,X/0 and 46,X/Y) is present, the manifestations associated with the chromosomal defect tend to be less severe.

Administration of female sex hormones (*i.e.*, estrogens) is used to promote development of secondary sexual characteristics and produce additional skeletal growth in women with Turner's syndrome. Growth hormone also may be used to increase skeletal growth.¹⁴

The diagnosis of Turner's syndrome often is delayed until late childhood or early adolescence in girls who do not present with the classic features of the syndrome.¹⁵ Early diagnosis is an important aspect of treatment for Turner's syndrome. It allows for counseling about the phenotypic characteristics of the disorder; screening for cardiac, renal, thyroid, and other abnormalities; provision of emotional support for the girl and her family; and planning for growth hormone therapy, if appropriate.13 Because of the potential for delay in diagnosis, it has been recommended that girls with unexplained short stature (height below the fifth percentile), webbed neck, peripheral lymphedema, coarctation of the aorta, or delayed puberty have chromosome studies done. In addition, chromosome analysis should be considered for girls who remain above the fifth percentile but have two or more features of Turner's syndrome, including high palate, nail deformities, short fourth metacarpal, and strabismus.15

Polysomy X. Polysomy X, or Klinefelter's syndrome, is a condition of testicular dysgenesis accompanied by the presence of one or more extra X chromosomes in excess of the normal male XY complement. Most males with Klinefelter's syndrome have one extra X chromosome (XXY). In rare cases, there may be more than one extra X chromosome (XXYY). The syndrome is characterized by enlarged breasts, sparse facial and body hair, small testes, and the inability to produce sperm.¹⁵ Regardless of the number of X chromosomes present, the male phenotype is retained. Based on studies conducted in the 1970s, including one sponsored by the National Institutes of Health and Human Development that checked the chromosomes of more than 40,000 infants, it has been estimated that

Chapter 4: Genetic and Congenital Disorders

57

the XXY syndrome is one of the most common genetic abnormalities known, occurring as frequently as 1 in 500 to 1 in 1000 male births.¹⁶ Although the presence of the extra chromosome is fairly common, the syndrome with its accompanying signs and symptoms that may result from the extra chromosome is uncommon. Many men live their lives without being aware that they have an additional chromosome. For this reason, it has been suggested that the term *Klinefelter's syndrome* be replaced with XXY male.

The condition often goes undetected at birth. The infant usually has normal male genitalia, with a small penis and small, firm testicles. At puberty, the intrinsically abnormal testes do not respond to stimulation from the gonadotropins and undergo degeneration. This leads to a tall stature with abnormal body proportions in which the lower part of the body is longer than the upper part. Later in life, the body build may become heavy, with a female distribution of subcutaneous fat and variable degrees of breast enlargement. There may be deficient secondary male sex characteristics, such as a voice that remains feminine in pitch and sparse beard and pubic hair. There may be sexual dysfunction along with the complete infertility that occurs owing to the inability to produce sperm. Regular administration of testosterone, beginning at puberty, can promote more normal growth and development of secondary sexual characteristics. Although the intellect usually is normal, most XXY males have some degree of language impairment. They often learn to talk later than do other children and often have trouble with learning to read and write.

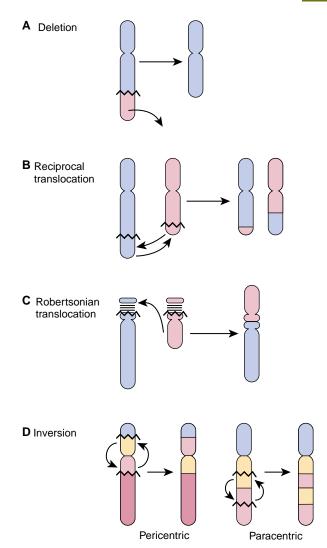
The presence of the extra X chromosome in the XXY male results from nondisjunction during meiotic division in one of the parents. The additional X chromosome (or chromosomes) is of maternal origin in approximately two thirds of cases and of paternal origin in the remaining one third.⁴ The cause of the nondisjunction is unknown. Advanced maternal age increases the risk, but only slightly.

Alterations in Chromosome Structure

Aberrations in chromosome structure occur when there is a break in one or more of the chromosomes followed by rearrangement or deletion of the chromosome parts. Among the factors believed to cause chromosome breakage are exposure to radiation sources, such as x-rays; influence of certain chemicals; extreme changes in the cellular environment; and viral infections.

Several patterns of chromosome breakage and rearrangement can occur (Fig. 4-7). There can be a *deletion* of the broken portion of the chromosome. When one chromosome is involved, the broken parts may be *inverted*. *Isochromosome formation* occurs when the centromere, or central portion, of the chromosome separates horizontally instead of vertically. *Ring formation* results when deletion is followed by uniting of the chromatids to form a ring. *Translocation* occurs when there are simultaneous breaks in two chromosomes from different pairs, with exchange of chromosome parts. With a balanced reciprocal translocation, no genetic information is lost; therefore, persons with translocations usually are normal. However, these persons are translocation carriers and may have normal and abnormal children.

A special form of translocation called a *centric fusion* or *Robertsonian translocation* involves two acrocentric chromosomes in which the centromere is near the end. Typically, the



■ FIGURE 4-7 ■ Examples of structural abnormalities of human chromosomes: (A) deletion of part of a chromosome leads to loss of genetic material and shortening of the chromosome; (B) a reciprocal translocation involves breaks in two nonhomologous chromosomes, with exchange of the acentric segment; (C) robertsonian translocation in which two nonhomologous chromosomes break near their centromeres, after which the long arms fuse to form one large metacentric chromosome; (D) inversion, which requires two breaks in a single chromosome with inversion to the opposite side of the centromere (pericentric), or paracentric if the breaks are on the same arm. (Adapted from Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 225]. Philadelphia: Lippincott Williams & Wilkins)

break occurs near the centromere affecting the short arm in one chromosome and the long arm in the other. Transfer of the chromosome fragments leads to one long and one extremely short chromosome. The short fragments commonly are lost. In this case, the person has only 45 chromosomes, but the amount of genetic material that is lost is so small that it often goes unnoticed. However, difficulty arises during meiosis; the result is gametes with an unbalanced number of chromosomes.

A rare form of Down syndrome can occur in the offspring of persons in whom there has been a translocation involving the long arm of chromosome 21q and the long arm of one of the acrocentric chromosomes (most often 14 or 22). The translocation adds to the normal long arm of chromosome 21; therefore, the person with this type of Down syndrome has 46 chromosomes, but essentially a trisomy of 21q.³

The manifestations of aberrations in chromosome structure depend to a great extent on the amount of genetic material that is lost. Many cells sustaining unrestored breaks are eliminated within the next few mitoses because of deficiencies that may in themselves be fatal. This is beneficial because it prevents the damaged cells from becoming a permanent part of the organism or, if it occurs in the gametes, from giving rise to grossly defective zygotes. Some altered chromosomes, such as those that occur with translocations, are passed on to the next generation.

In summary, genetic disorders can affect a single gene (mendelian inheritance) or several genes (polygenic inheritance). Single-gene disorders may be present on an autosome or on the X chromosome and they may be expressed as a dominant or recessive trait. In autosomal dominant disorders, a single mutant allele from an affected parent is transmitted to an offspring regardless of sex. The affected parent has a 50% chance of transmitting the disorder to each offspring. Autosomal recessive disorders are manifested only when both members of the gene pair are affected. Usually, both parents are unaffected but are carriers of the defective gene. Their chances of having an affected child are one in four; of having a carrier child, two in four; and of having a noncarrier unaffected child, one in four. Sex-linked disorders, which are associated with the X chromosome, are those in which an unaffected mother carries one normal and one mutant allele on the X chromosome. She has a 50% chance of transmitting the defective gene to her sons, and her daughters have a 50% chance of being carriers of the mutant gene. Because of a normal paired gene, female heterozygotes rarely experience the effects of a defective gene. Multifactorial inheritance disorders are caused by multiple genes and, in many cases, environmental factors.

Chromosomal disorders result from a change in chromosome number or structure. A change in chromosome number is called *aneuploidy*. *Monosomy* involves the presence of only one member of a chromosome pair; it is seen in Turner's syndrome, in which there is monosomy of the X chromosome. *Polysomy* refers to the presence of more than two chromosomes in a set. Klinefelter's syndrome involves polysomy of the X chromosome. Trisomy 21 (*i.e.*, Down syndrome) is the most common form of chromosome disorder. Alterations in chromosome structure involve deletion or addition of genetic material, which may involve a translocation of genetic material from one chromosome pair to another.

DISORDERS DUE TO ENVIRONMENTAL INFLUENCES

The developing embryo is subject to many nongenetic influences. After conception, development is influenced by the environmental factors that the embryo shares with the mother. The physiologic status of the mother—her hormone balance, her general state of health, her nutritional status, and the drugs she takes—undoubtedly influences the development of the unborn child. For example, diabetes mellitus is associated with increased risk of congenital anomalies. Smoking is associated with lower than normal neonatal weight. Alcohol, in the context of chronic alcoholism, is known to cause fetal abnormalities. Some agents cause early abortion. Measles and other infectious agents cause congenital malformations. Other agents, such as radiation, can cause chromosomal and genetic defects and produce developmental disorders.

Period of Vulnerability

The embryo's development is most easily disturbed during the period when differentiation and development of the organs are taking place. This time interval, which is often referred to as the period of *organogenesis*, extends from day 15 to day 60 after conception. Environmental influences during the first 2 weeks after fertilization may interfere with implantation and result in abortion or early resorption of the products of conception. Each organ has a critical period during which it is highly susceptible to environmental derangements (Fig. 4-8). Often, the effect is expressed at the biochemical level just before the organ begins to develop. The same agent may affect different organ systems that are developing at the same time.

Teratogenic Agents

A teratogenic agent is an environmental agent that produces abnormalities during embryonic or fetal development. It is important to remember that, in this case, the environment is that of the embryo and fetus. Maternal disease or altered metabolic state also can affect the environment of the embryo or fetus. For discussion purposes, teratogenic agents have been divided into three groups: radiation, drugs and chemical substances, and infectious agents. Chart 4-1 lists commonly identified agents in each of these groups. Theoretically, environmental agents can cause birth defects in three ways: by direct exposure of the pregnant woman and the embryo or fetus to the agent; through exposure of the soon-to-be-pregnant woman with an agent that has a slow clearance rate such that a teratogenic dose is retained during early pregnancy; or as a result of mutagenic effects of an environmental agent that occur before pregnancy, causing permanent damage to a woman's (or a man's) reproductive cells.

Radiation

Heavy doses of ionizing radiation have been shown to cause microcephaly, skeletal malformations, and mental retardation. There is no evidence that diagnostic levels of radiation cause congenital abnormalities. However, because the question of safety remains many agencies require that the day of a woman's last menstrual period be noted on all radiologic requisitions. Other institutions may require a pregnancy test before any extensive diagnostic x-ray studies are performed. Radiation is teratogenic and mutagenic, and there is the possibility of effecting inheritable changes in genetic materials. Administration of therapeutic doses of radioactive iodine (¹³¹I) during the 13th week of gestation, the time when the fetal thyroid is beginning to concentrate iodine, has been shown to interfere with thyroid development.

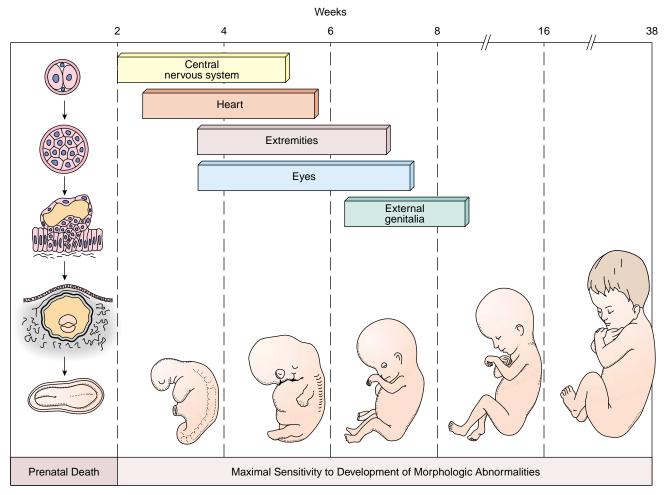


FIGURE 4-8 Sensitivity of specific organs to teratogenic agents at critical periods in embryogenesis. Exposure of adverse influences in the preimplantation and early postimplantation stages of development (*far left*) leads to prenatal death. Periods of maximal sensitivity to teratogens (*horizontal bars*) vary for different organ systems but overall are limited to the first 8 weeks of pregnancy. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 216]. Philadelphia: Lippincott Williams & Wilkins)

Chemicals and Drugs

Environmental chemicals and drugs can cross the placenta and cause damage to the developing embryo and fetus. It has been estimated that only 2% to 3% of developmental defects have a known drug or environmental origin. Some of the best-

KEY CONCEPTS

TERATOGENIC AGENTS

- Teratogenic agents such as radiation, chemicals and drugs, and infectious organisms are agents that produce abnormalities in the developing embryo.
- The stage of development of the embryo determines the susceptibility to teratogens. The period during which the embryo is most susceptible to teratogenic agents is the time during which rapid differentiation and development of body organs and tissues are taking place, usually from days 15 to 60 postconception.

documented environmental teratogens are the organic mercurials, which cause neurologic deficits and blindness. Sources of exposure to mercury include contaminated food (fish) and water.¹⁷ The precise mechanism by which chemicals and drugs exert their teratogenic effects is largely unknown. They may produce cytotoxic (cell-killing), antimetabolic, or growthinhibiting properties. Often their effects depend on the time of exposure (in terms of embryonic and fetal development) and extent of exposure (dosage).

Drugs top the list of chemical teratogens, probably because they are regularly used at elevated doses. Most drugs can cross the placenta and expose the fetus to both the pharmacologic and teratogenic effects. Factors that affect placental drug transfer and drug effects on the fetus include the rate at which the drug crosses the placenta, the duration of exposure, and the stage of placental and fetal development at the time of exposure.¹⁸ Lipid-soluble drugs tend to cross the placenta more readily and enter the fetal circulation. The molecular weight of a drug also influences the rate of transfer and the amount of drug transferred across the placenta. Drugs with a molecular weight less than 500 can cross the placenta easily, depending on lipid solubility and degree of ionization; those with a

59

Radiation		
Drugs and Chemical Substances		
Alcohol		
Anticoagulants		
Warfarin		
Anticonvulsants		
Cancer drugs		
Aminopterin Methotrexate		
6-Mercaptopurine		
Isotretinoin (Accutane)		
Propylthiouracil		
Tetracycline		
Thalidomide		
Infectious Agents		
Viruses		
Cytomegalovirus		
Herpes simplex virus		
Measles (rubella)		
Mumps		
Varicella-zoster virus (chickenpox)		
Nonviral factors		
Syphilis Toxoplasmosis		
*Not inclusive.		

molecular weight of 500 to 1000 cross the placenta with more difficulty; and those with molecular weights greater than 1000 cross very poorly.¹⁸

A number of drugs are suspected of being teratogens, but only a few have been identified with certainty.¹⁹ Perhaps the best known of these drugs is thalidomide, which has been shown to give rise to a full range of malformations, including phocomelia (i.e., short, flipper-like appendages) of all four extremities. Other drugs known to cause fetal abnormalities are the antimetabolites that are used in the treatment of cancer, the anticoagulant drug warfarin, several of the anticonvulsant drugs, ethyl alcohol, and cocaine. Some drugs affect a single developing structure; for example, propylthiouracil can impair thyroid development and tetracycline can interfere with the mineralization phase of tooth development. More recently, vitamin A and its derivatives (the retinoids) have been targeted for concern because of their teratogenic potential. Concern over the teratogenic effects of vitamin A derivatives became evident with the introduction of the acne drug isotretinoin (Accutane). Fetal abnormalities such as cleft palate, heart defects, retinal and optic nerve abnormalities, and central nervous system malformations were observed in women ingesting therapeutic doses of the drug during the first trimester of pregnancy.²⁰ There also is concern about the teratogenic effects when a woman consumes high doses of vitamin A, such as those contained in some dietary supplements or vitamin pills. It is currently recommended that doses greater than 10,000 IU should be avoided.21

In 1983, the U.S. Food and Drug Administration established a system for classifying drugs according to probable risks to the fetus. According to this system, drugs are put into five categories: A, B, C, D, and X. Drugs in category A are the least dangerous, and categories B, C, and D are increasingly more dangerous. Those in category X are contraindicated during pregnancy because of proven teratogenicity.²² The law does not require classification of drugs that were in use before 1983.

Because many drugs are suspected of causing fetal abnormalities, and even those that were once thought to be safe are now being viewed critically, it is recommended that women in their childbearing years avoid unnecessary use of drugs. This pertains to nonpregnant women as well as pregnant women because many developmental defects occur early in pregnancy. As happened with thalidomide, the damage to the embryo may occur before pregnancy is suspected or confirmed. Two drugs of particular importance are alcohol and cocaine.

Fetal Alcohol Syndrome. The term *fetal alcohol syndrome* (FAS) refers to a constellation of physical, behavioral, and cognitive abnormalities resulting from maternal alcohol consumption. It has been reported that 1 in 1000 infants born in the United States manifests some characteristics of the syndrome.²³ Alcohol, which is lipid soluble and has a molecular weight between 600 and 1000, passes freely across the placental barrier; concentrations of alcohol in the fetus are at least as high as in the mother. Unlike other teratogens, the harmful effects of alcohol are not restricted to the sensitive period of early gestation but extend throughout pregnancy.

Alcohol has widely variable effects on fetal development, ranging from minor abnormalities to FAS. Criteria for defining FAS were standardized by the Fetal Alcohol Study Group of the Research Society on Alcoholism in 1980,24 and modifications were proposed in 1989 by Sokol and Clarren.²⁵ The proposed criteria are prenatal or postnatal growth retardation (i.e., weight or length below the 10th percentile); central nervous system involvement, including neurologic abnormalities, developmental delays, behavioral dysfunction, intellectual impairment, and skull and brain malformation; and a characteristic face with short palpebral fissures (*i.e.*, eve openings), a thin upper lip, and an elongated, flattened midface and philtrum (*i.e.*, the groove in the middle of the upper lip). The facial features of FAS may not be as apparent in the newborn but become more prominent as the infant develops. As the children grow into adulthood, the facial features become more subtle, making diagnosis of FAS in older individuals more difficult.²⁶ Each of these defects can vary in severity, probably reflecting the timing of alcohol consumption in terms of the period of fetal development, amount of alcohol consumed, and hereditary and environmental influences. Because of problems with terminology and the diagnostic criteria, the Institute of Medicine in 1996 proposed the terms alcohol-related neurodevelopmental disorder (ARND) and alcoholrelated birth defects (ARBD) to describe conditions in which there is a history of maternal alcohol consumption.²⁷ This new terminology uses pathophysiologic diagnostic categories to describe the conditions resulting from confirmed alcohol exposure. For example, facial abnormalities, growth retardation, and central nervous system abnormalities would be classified as FAS; central nervous system and cognitive abnormalities would be classified as ARND; and birth defects as ARBD.²⁸

The mechanisms whereby alcohol exerts its teratogenic effects are unclear. Evidence suggests that the effects of alcohol observed in children with FAS are related to the timing of alcohol consumption and peak alcohol dose.

The amount of alcohol that can be safely consumed during pregnancy also is unknown. Animal studies suggest that the fetotoxic effects of alcohol are dose dependent, rather than threshold dependent. Studies suggest that even three drinks per day may be associated with a lower IQ at 4 years of age.²⁹ However, it may be that the time during which alcohol is consumed is equally important. Even small amounts of alcohol consumed during critical periods of fetal development may be teratogenic. For example, if alcohol is consumed during the period of organogenesis, a variety of skeletal and organ defects may result. When alcohol is consumed later in gestation, when the brain is undergoing rapid development, there may be behavioral and cognitive disorders in the absence of physical abnormalities. Chronic alcohol consumption throughout pregnancy may result in a variety of effects, ranging from physical abnormalities to growth retardation and compromised central nervous system functioning. Evidence suggests that short-lived high concentrations of alcohol such as those that occur with binge drinking may be particularly significant, with abnormalities being unique to the period of exposure. The recommendation of the U.S. Surgeon General is that women abstain from alcohol during pregnancy.30

Cocaine Babies. Of concern is the increasing use of cocaine by pregnant women. In 1992, approximately 45,000 women in this country used cocaine during their pregnancy.³¹ Determining exposure of infants to maternal cocaine use often is difficult. In utero exposure often is ascertained by testing maternal urine for cocaine and its metabolites and by interviewing the mother. Urine testing provides evidence only of recent cocaine use, and information from an interview may be inaccurate. Urine testing of infants provides evidence only of recent exposure to cocaine.

Among the effects of cocaine use during pregnancy is a decrease in uteroplacental blood flow, maternal hypertension, stimulation of uterine contractions, and fetal vasoconstriction. The decrease in uteroplacental blood flow is associated with an increase in preterm births, intrauterine growth retardation, microcephaly, and neurologic abnormalities.^{32,33} Furthermore, there appears to be a dose-related relationship between increasing levels of chronic cocaine abuse and impaired fetal growth and neurologic function.33 Maternal hypertension may increase the risk of abruptio placentae, particularly if it is accompanied by a decrease in uteroplacental blood flow.³¹ Fetal vasoconstriction has been suggested as the cause of fetal anomalies, particularly limb reduction defects and urogenital tract defects such as hydronephrosis, hypospadias, and undescended testicles, as well as ambiguous genitalia.^{34,35} Exposure of the fetus to cocaine also may lead to destructive lesions of the brain, including cerebral infarction and intracranial hemorrhage. Sudden infant death syndrome (SIDS) also has been more common in infants of mothers who have used cocaine during their pregnancy.36

Although the immediate effects of maternal cocaine use on infant behavior are being reported, the long-term effects are

largely unknown. Unfortunately, cocaine addiction often affects the behavior of the pregnant woman to the extent that the need to procure larger amounts of the drug overwhelms all other considerations of maternal and fetal well-being; other factors such as malnutrition, use of other drugs and teratogens, and lack of prenatal care also may contribute to fetal disorders.

Folic Acid Deficiency. Although most birth defects are related to exposure to a teratogenic agent, deficiencies of nutrients and vitamins also may be a factor. Folic acid deficiency has been implicated in the development of neural tube defects (e.g., anencephaly, spina bifida, encephalocele). Studies have shown a reduction in neural tube defects when folic acid was taken before conception and continued during the first trimester of pregnancy.^{37,38} The Public Health Service recommends that all women of childbearing age should take 400 micrograms (μg) of folic acid daily. It has been suggested that this recommendation may help to prevent as many as 50% of neural tube defects.³⁹ The Institute of Medicine Panel for Folate and Other B Vitamins and Choline has recently revised the Recommended Dietary Allowance for pregnant women to 600 µg.40 These recommendations are particularly important for women who have previously had an affected pregnancy, for couples with a close relative with the disorder, and for women with diabetes mellitus and those taking anticonvulsant drugs who are at increased risk for having infants with birth defects.

Since 1998, all enriched cereal grain products in the United States have been fortified with folic acid. To achieve an adequate intake of folic acid, pregnant women should couple a diet that contains folate-rich foods (*e.g.*, orange juice, dark, leafy green vegetables, and legumes) with sources of synthetic folic acid, such as fortified food products.⁴⁰

Infectious Agents

Many microorganisms cross the placenta and enter the fetal circulation, often producing multiple malformations. The acronym TORCH stands for toxoplasmosis, other, rubella (*i.e.*, German measles), cytomegalovirus, and herpes, which are the agents most frequently implicated in fetal anomalies.³ Other infections include varicella-zoster virus infection, listeriosis, leptospirosis, Epstein-Barr virus infection, tuberculosis, and syphilis. The TORCH screening test examines the infant's serum for the presence of antibodies to these agents. These infections tend to cause similar clinical manifestations, including microcephaly, hydrocephalus, defects of the eye, and hearing problems.

Toxoplasmosis is a protozoal infection that can be contracted by eating raw or poorly cooked meat. The domestic cat also seems to carry the organism, excreting the protozoa in its stools. It has been suggested that pregnant women should avoid contact with excrement from the family cat. The introduction of the rubella vaccine in the United States has virtually eliminated congenital rubella. The epidemiology of cytomegalovirus infection is largely unknown. Some infants are severely affected at birth, and others, although having evidence of the infection, have no symptoms. In some symptomfree infants, brain damage becomes evident over a span of several years. There also is evidence that some infants contract the infection during the first year of life, and in some of them the infection leads to retardation a year or two later. Herpes simplex type 2 infection is considered to be a genital infection and usually is transmitted through sexual contact. The infant acquires this infection in utero or in passage through the birth canal.

In summary, a teratogenic agent is one that produces abnormalities during embryonic or fetal life. It is during the early part of pregnancy (15 to 60 days after conception) that environmental agents are most apt to produce their deleterious effects on the developing embryo. A number of environmental agents can be damaging to the unborn child, including radiation, drugs and chemicals, and infectious agents. FAS is a risk for infants of women who regularly consume alcohol during pregnancy. Of recent concern is the use of cocaine by pregnant women. Because many drugs have the potential for causing fetal abnormalities, often at an early stage of pregnancy, it is recommended that women of childbearing age avoid unnecessary use of drugs. It also has been shown that folic acid deficiency can contribute to neural tube defects. The acronym TORCH stands for toxoplasmosis, other, rubella, cytomegalovirus, and herpes, which are the infectious agents most frequently implicated in fetal anomalies.

REVIEW QUESTIONS

Explain the difference between a hereditary and congenital defect.

Contrast disorders attributable to multifactorial inheritance with those caused by single-gene inheritance.

Describe two chromosomal abnormalities that demonstrate aneuploidy.

Relate maternal age and the occurrence of Down syndrome.

Cite the most susceptible period of intrauterine life for development of defects caused by environmental agents.

State the cautions that should be observed when considering the use of drugs during pregnancy.

Describe the effects of alcohol and cocaine abuse on fetal development and birth outcomes.

List four infectious agents that cause congenital defects.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- 1. March of Dimes Birth Defects Foundation. (2003). Birth defects information. [On-line]. Available: http://www.modimes.org.
- 2. Online Mendelian Inheritance in Man (OMIM[™]). (2003). National Center for Biotechnology Information, National Library of Medicine. [On-line]. Available: http://www.ncbi.nlm.nih.gov/omim

- Nussbaum R.L., McInnes R.R., Willard H.F. (2001). Thompson and Thompson genetics in medicine (6th ed., pp. 51–78, 135, 159, 359– 388). Philadelphia: W.B. Saunders.
- Rubin E., Farber J.E. (Eds.). (1999). *Pathology* (3rd ed., pp. 221–223, 236, 241–242, 252, 257–259). Philadelphia: Lippincott Williams & Wilkins.
- Cotran R.S., Kumar V., Collins T. (Eds.). (1999). Robbins pathologic basis of disease (6th ed., pp. 144–149, 155–156, 162–165, 170–171, 177–178). Philadelphia: W.B. Saunders.
- Koch R.K. (1999). Issues in newborn screening for phenylketonuria. *American Family Physician* 60, 1462–1466.
- National Institute of Child Health and Human Development. (2000). Facts about fragile X syndrome. [On-line]. Available: http://www. nichd.nih.gov/publications/pubs/Fragilex.htm.
- Warren S.T. (1997). Trinucleotide repetition and fragile X syndrome. *Hospital Practice* 31(4), 73–85, 90–98.
- 9. Riccardi V.M. (1977). *The genetic approach to human disease* (p. 92). New York: Oxford University Press.
- March of Dimes. (2003). Down syndrome. [On-line]. Available: http://www.modimes.org/HealthLibrary2/FactSheets/Down_ syndrome.htm.
- Newberger D.S. (2000). Down syndrome: Prenatal risk assessment and diagnosis. American Family Physician 62, 825–832, 837–838.
- Wald N.J., Watt H.C., Hacshaw A.K. (1999). Integrated screening for Down's syndrome based on tests performed during the first and second trimester. *New England Journal of Medicine* 341, 461–467.
- 13. Rosenfeld R.G. (2000). Turner's syndrome: A growing concern. *Pediatrics* 137, 443–444.
- Saenger P. (1996). Turner's syndrome. New England Journal of Medicine 335, 1749–1754.
- Savendahl L., Davenport M. (2000). Delayed diagnoses of Turner's syndrome: Proposed guidelines for change. *Journal of Pediatrics* 137, 455–459.
- National Institute of Child Health and Human Development. (2000). A guide for XXY males and their family. [On-line]. Available: http:// www.nichd.nih.gov/publications/ pubs/klinefelter.htm.
- Steurerwald U., Weibe P., Jorgensen P.J., et al. (2000). Maternal seafood diet, methylmercury exposure, and neonatal neurologic function. *Journal of Pediatrics* 136, 599–605.
- Katzung B.D. (2001). Basic and clinical pharmacology (8th ed., pp. 1025–1035). Stamford, CT: Appleton & Lange.
- Koren G., Pstuszak A., Ito S. (1998). Drugs in pregnancy. New England Journal of Medicine 338, 1128–1137.
- Ross S.A., McCaffery P.J., Drager U.C., et al. (2000). Retinoids in embryonal development. *Physiological Reviews* 80, 1021–1055.
- 21. Oakley G.P., Erickson J.D. (1995). Vitamin A and birth defects. *New England Journal of Medicine* 333, 1414–1415.
- U.S. Food and Drug Administration. (2000). Pregnancy categories. [On-line]. Available: http://www.fda.gov.
- March of Dimes. (1999). Leading categories of birth defects. [On-line]. Available: http://www.modimes.org/HealthLibrary2/ InfantHealthStatistics/bdtable.htm.
- 24. Rosett H.L. (1980). A clinical perspective of the fetal alcohol syndrome. *Alcoholism, Clinical and Experimental Research* 4, 162–164.
- Sokol R.J., Clarren S.K. (1989). Guidelines for use of terminology describing the impact of prenatal alcohol on the offspring. *Alcoholism, Clinical and Experimental Research* 13, 587–589.
- 26. Lewis D.D., Woods S.E. (1994). Fetal alcohol syndrome. *American Family Physician* 50, 1025–1032.
- Stratton K., Howe C., Battaglia F. (Eds.). (1996). Fetal alcohol syndrome: Diagnosis, epidemiology, prevention and treatment (pp. 4–21). Washington, DC: National Academy Press.
- American Academy of Pediatrics. (2000). Fetal alcohol syndrome and alcohol-related neurodevelopmental disorders. *Pediatrics* 106, 358–361.
- 29. Ernhart C.B., Bowden D.M., Astley S.J. (1987). Alcohol teratogenicity in the human: A detailed assessment of specificity, critical period, and threshold. *American Journal of Obstetrics and Gynecology* 156, 33–39.

Chapter 4: Genetic and Congenital Disorders

- Surgeon General's advisory on alcohol and pregnancy. (1981). FDA Drug Bulletin 2, 10.
- March of Dimes. (2003). Cocaine use during pregnancy. [On-line]. Available: http://www.modimes.org/HealthLibrary2/FactSheets/ Cocaine_use_during_pregnancy.htm.
- 32. Volpe J.J. (1972). Effect of cocaine use on the fetus. *New England Journal of Medicine* 327, 399–407.
- Chiriboga C.A., Brust C.M., Bateman D., et al. (1999). Dose-response effect of fetal cocaine exposure on newborn neurologic function. *Pediatrics* 103, 79–85.
- 34. MacGregor S.N., Keith L.G., Chasnoff I.J., et al. (1987). Cocaine use during pregnancy: Adverse outcome. *American Journal of Obstetrics and Gynecology* 157, 686–690.
- 35. Chasnoff I.J., Chisum G.M., Kaplan W.E. (1988). Maternal cocaine use and genitourinary malformations. *Teratology* 37, 201–204.

- Riley J.B., Brodsky N.L., Porat R. (1988). Risk of SIDS in infants with in utero cocaine exposure: A prospective study [Abstract]. *Pediatric Research* 23, 454A.
- 37. Committee on Genetics. (1993). Folic acid for the prevention of neural tube defects. *Pediatrics* 92, 493–494.
- Centers for Disease Control and Prevention. (1992). Recommendations for use of folic acid to reduce the number of cases of spina bifida and other neural tube defects. *Morbidity and Mortality Weekly Report* 41, 1–8.
- Scholl T.O., Johnson W.G. (2000). Folic acid: Influence on outcome of pregnancy. *American Journal of Clinical Nutrition* 71 (Suppl.), 12958–1303S.
- Bailey L.B. (2000). New standard for dietary folate intake in pregnant women. *American Journal of Clinical Nutrition* 71 (Suppl.), 1304S–1307S.

CHAPTER

Alterations in Cell Growth and Replication: Neoplasia

Concepts of Cell Growth The Cell Cycle **Cell Proliferation Cell Differentiation Characteristics of Benign and Malignant Neoplasms** Terminology **Benign Neoplasms** Malignant Neoplasms **Cancer Cell Characteristics** Invasion and Metastasis **Tumor Growth Carcinogenesis and Causes of Cancer Oncogenesis: The Molecular Basis of Cancer Tumor Cell Transformation** Heredity **Carcinogenic Agents Chemical Carcinogens Radiation Carcinogenesis Oncogenic Viruses Immunologic Defects Clinical Features Diagnosis and Staging** The Pap Smear **Biopsy Tumor Markers** Staging and Grading of Tumors **Cancer Treatment** Surgery **Radiation Therapy** Systemic Cancer Therapy **Childhood Cancers Diagnosis and Treatment** Adult Survivors of Childhood Cancer

ancer is the second leading cause of death in the United States after cardiovascular disease. The disease affects all age groups, causing more death in children 3 to 15 years of age than any other disease. The American Cancer Society estimates that 1.3 million Americans will develop cancer in 2003, and that one in two males and one in three females will have cancer during their lifetime. It also is estimated that approximately 556,500 Americans will die in the year from neoplastic diseases.¹ As age-adjusted cancer mortality rates increase and heart disease mortality decreases, it is predicted that cancer will become the leading cause of death in a few decades.² Trends in cancer survival demonstrate that relative 5-year survival rates have improved since the early 1960s. It is estimated that approximately 59% of people who develop cancer each year will be alive 5 years later.

CONCEPTS OF CELL GROWTH

Cancers result from a process of altered cell differentiation and growth. The resulting tissue is called *neoplasia*. The term *neoplasm* comes from a Greek word meaning *new formation*. Unlike the tissue growth that occurs with hypertrophy and hyperplasia, the growth of a neoplasm is uncoordinated and relatively autonomous in that it lacks normal regulatory controls over cell growth and division. Neoplasms tend to increase in size and continue to grow after the stimulus has ceased or the needs of the organism have been met.

Cancer is not a single disease. The term describes almost all forms of malignant neoplasia. Cancer can originate in almost any organ, with the prostate being the most common site in men and the breast in women. The ability of cancer to be cured varies considerably and depends on the type of cancer and the extent of the disease at diagnosis. Cancers such as acute lymphocytic leukemia, Hodgkin's disease, testicular cancer, and osteosarcoma, which only a few decades ago had poor prognoses, are today cured in many cases. However, lung cancer, which is the leading cause of death in men and women in the United States, is resistant to therapy, and although some progress has been made in its treatment, mortality rates remain high.

Tissue renewal and repair involves cell proliferation and differentiation. *Proliferation*, or the process of cell division, is an inherent adaptive mechanism for replacing body cells when old cells die or additional cells are needed. *Differentiation* is the process of specialization whereby new cells acquire the structure and function of the cells they replace. In adult tissues, the size of a population of cells is determined by the rates of cell proliferation, differentiation, and death by apoptosis.³ Apoptosis, which is discussed in Chapter 2, is a form of programmed cell death designed to eliminate senescent cells or unwanted cells. A balance of cellular signals that regulate cell proliferation, differentiation, and apoptosis regulates the size of cell populations.

The Cell Cycle

The cell cycle is the interval between each cell division. It regulates the duplication of genetic information and appropriately aligns the duplicated chromosomes to be received by the daughter cells. In addition, pauses or checkpoints in the cell cycle determine the accuracy with which deoxyribonucleic acid (DNA) is duplicated. These checkpoints allow for any defects to be edited and repaired, thereby assuring that the daughter cells receive the full complement of genetic information, identical to that of the parent cell.³

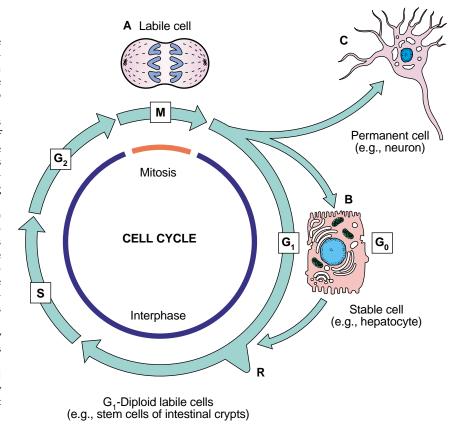
The cell cycle is divided into four distinct phases referred to as G_1 , S, G_2 , and M (Fig. 5-1). G_1 (*gap 1*), is the postmitotic phase during which DNA synthesis ceases while ribonucleic acid (RNA) and protein synthesis and cell growth take place. This is the phase during which cells pursue their own specialized type of function. Some cells such as neurons become terminally differentiated after mitosis and remain in G_1 . Continually dividing cells, such as the crypt cells in the intestinal mucosa, pass through restriction point (R) in G_1 that commits them to progression to the synthesis (S) phase and a new round of cell division. During the S phase, DNA synthesis occurs, giving rise to two separate sets of chromosomes, one for each daughter cell. G_2 (*gap 2*) is the premitotic gap and is similar to G_1 in that DNA synthesis ceases while RNA and protein synthesis continues. The *M phase* is the phase of cellular division or mitosis. Stable cells, such as hepatocytes, enter a quiescent period in the cell cycle, the G_0 gap. These quiescent cells reenter the cell cycle in response to extracellular nutrients, growth factors, hormones, and other signals, such as blood loss or tissue injury, that signal for cell renewal.^{4,5}

The duration of the phases of the cell cycle vary depending on the cell type, the frequency with which the cells divide, and host characteristics, such as the presence of appropriate growth factors. Very rapidly dividing cells can complete the cell cycle in less than 8 hours, whereas others can take longer than 1 year. Most of this variability occurs in the G_0 and G_1 phases. The duration of the S phase (10 to 20 hours), the G_2 phase (2 to 10 hours), and the M phase (0.5 to 1 hour) appears to be relatively constant.⁵

Cell Proliferation

Cell proliferation is the process by which cells divide and reproduce. Cell division provides the body with the means for replacing cells that have a limited life span such as skin and blood cells, increasing tissue mass during periods of growth, and providing for tissue repair and wound healing. In normal

■ FIGURE 5-1 ■ The cell cycle. (A) Labile cells (e.g., intestinal crypt cells) undergo continuous replication and the interval between two consecutive mitoses is designated the cell cycle. After division, the cells enter a gap (G_1) during which DNA synthesis ceases and RNA and protein synthesis takes place as the cell develops its own specialized type of function. Cells that continue in the cell cycle pass the restriction point (R), which commits them to a new round of cell division and continuation to the synthesis (S) phase during which all the chromosomes are replicated. The S phase is followed by the short gap (G_2) during which DNA synthesis ceases and protein synthesis continues. The M phase is the period of mitosis. After each cycle, one daughter cell will become committed to differentiation and the other will continue cycling. (B) Some cell types, such as hepatocytes, are stable. After cell mitosis, the cells take up their specialized functions (G₀) and not reenter the cell cycle unless stimulated by the loss of other cells. (C) Permanent cells (neurons) become terminally differentiated after mitosis and cannot reenter the cell cycle. (Rubin E., Farber J.L. [1999]. Pathology [3rd ed., p. 85]. Philadelphia: Lippincott Williams & Wilkins)



tissue, cell proliferation is regulated so that the number of cells actively dividing is equivalent to the number dying or being shed.

In terms of cell proliferation, the 200 or more cell types of the body can be divided into 3 large groups: the welldifferentiated neurons and cells of skeletal and cardiac muscle that are unable to divide and reproduce; the parent, or progenitor cells, that continue to divide and reproduce, such as blood cells, skin cells, and liver cells; and the undifferentiated stem cells that can be triggered to enter the cell cycle and produce large numbers of progenitor cells when the need arises. The rates of reproduction of these cells vary greatly. White blood cells and cells that line the gastrointestinal tract live several days and must be replaced constantly. In most tissues, the rate of cell reproduction is greatly increased when tissue is injured or lost. For example, bleeding stimulates the rapid reproduction of the blood-forming cells of the bone marrow. In some types of tissue, the genetic program for cell replication normally is repressed but can be resumed under certain conditions. For example, the liver has extensive regenerative capabilities under certain conditions.

Cell Differentiation

Cell differentiation is the process whereby proliferating cells are transformed into different and more specialized cell types. This process leads to a fully differentiated, adult cell that has achieved its specific set of structural, functional, and life expectancy characteristics. For example, a red blood cell is programmed to develop into a concave disk that functions as a vehicle for oxygen transport and lives approximately 120 days.

All of the different cell types of the body originate from a single cell—the fertilized ovum. As the embryonic cells increase in number, they engage in an orderly process of differentiation that is necessary for the development of all the various organs of the body. The process of differentiation is regulated by a combination of internal programming that in-

KEY CONCEPTS

CELL PROLIFERATION AND GROWTH

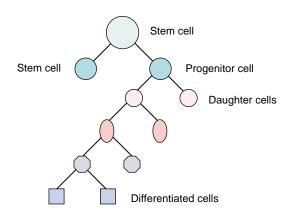
- Tissue growth and repair involve cell proliferation and differentiation.
- Cell proliferation is the process whereby tissues acquire new or replacement cells through cell division.
- Cell differentiation is the orderly process in which proliferating cells are transformed into different and more specialized types. It determines the microscopic characteristics of the cell, how the cell functions, and how long it will live.
- Cells that are fully differentiated are no longer capable of cell division.

volves the expression of specific genes and external stimuli provided by neighboring cells, exposure to substances in the maternal circulation, and a variety of growth factors, nutrients, oxygen, and ions.⁶ What makes the cells of one organ different from those of another organ is the type of gene that is expressed. Although all cells have the same complement of genes, only a small number of these genes are expressed in postnatal life. When cells, such as those of the developing embryo, differentiate and give rise to committed cells of a particular tissue type, the appropriate genes are maintained in an active state while the remainder are inactive. Normally, the rate of cell reproduction and the process of cell differentiation are precisely controlled in prenatal and postnatal life so that both of these mechanisms cease once the appropriate numbers and types of cells are formed.

The process of differentiation occurs in orderly steps; with each progressive step, increased specialization is exchanged for a loss of ability to develop different cell characteristics and different cell lines. The more highly specialized a cell becomes, the more likely it is to lose its ability to undergo mitosis. Neurons, which are the most highly specialized cells in the body, lose their ability to divide and reproduce once development of the nervous system is complete. More important, there are no reserve or parent cells to direct their replacement. However, appropriate numbers of these cell types are generated in the embryo such that loss of a certain percentage of cells does not affect the total cell population. Although these cells never divide and are not replaced if lost, they exist in sufficient numbers to carry out their specific functions. In other, less specialized tissues, such as the skin and mucosal lining, cell renewal continues throughout life.

Even in the continuously renewing cell populations, highly specialized cells are similarly unable to divide. An alternative mechanism provides for their replacement. There are progenitor cells of the same lineage that have not yet differentiated to the extent that they have lost their ability to divide. These cells are sufficiently differentiated that their daughter cells are limited to the same cell line, but they are insufficiently differentiated to preclude the potential for active proliferation. As a result, these parent or progenitor cells are able to provide large numbers of replacement cells. However, the progenitor cells have limited capacity for self-renewal and they become restricted to producing a single type of cell.

Another type of cell, called a stem cell, remains incompletely differentiated throughout life. Stem cells are reserve cells that remain quiescent until there is a need for cell replenishment, in which case they divide, thereby producing other stem cells and cells that can carry out the functions of the differentiated cell (Fig. 5-2). There are several types of stem cells, some of which include the muscle satellite cell, the epidermal stem cell, the spermatogonium, and the basal cell of the olfactory epithelium. These stem cells are unipotent in that they give rise to only one type of differentiated cell. Oligopotent stem cells can produce a small number of cells, and pluripotent stem cells, such as those involved in hematopoiesis, give rise to numerous cell types.⁴ Stem cells are the primary cellular component of bone marrow transplantation, in which the stem cells in the transplanted marrow re-establish the recipient's blood production and immune system. Peripheral blood stem cell transplantation is a transplantation procedure that by-



■ FIGURE 5-2 ■ Mechanism of stem cell-mediated cell replacement. Division of a stem cell with an unlimited potential for proliferation results in one daughter cell, which retains the characteristics of a stem cell, and a second daughter cell, which differentiates into a progenitor or parent cell, with a limited potential for differentiation and proliferation. As the daughter cells of the progenitor cell proliferate, they become more differentiated, until reaching the stage where they are fully differentiated and no longer able to divide.

passes the need for bone marrow infusion by infusing stem cells that have been separated and removed from the donor blood.

In summary, the term *neoplasm* refers to an abnormal mass of tissue in which the growth exceeds and is uncoordinated with that of the normal tissues. Unlike normal cellular adaptive processes such as hypertrophy and hyperplasia, neoplasms do not obey the laws of normal cell growth. They serve no useful purpose, they do not occur in response to an appropriate stimulus, and they continue to grow at the expense of the host.

Cell proliferation is the process whereby cells divide and bear offspring; it normally is regulated so that the number of cells that are actively dividing is equal to the number dying or being shed. The process of cell growth and division is called the *cell cycle*. It is divided into four phases: G_1 , the postmitotic phase, during which DNA synthesis ceases while RNA and protein synthesis and cell growth take place; S, the phase during which DNA synthesis occurs, giving rise to two separate sets of chromosomes; G_2 , the premitotic phase, during which RNA and protein synthesis continues; and M, the phase of cell mitosis or cell division. The G_0 phase is a resting or quiescent phase in which nondividing cells reside.

Cell differentiation is the process whereby cells are transformed into different and more specialized cell types as they proliferate. It determines the structure, function, and life span of a cell. There are three types of cells: well-differentiated cells that are no longer able to divide, progenitor or parent cells that continue to divide and bear offspring, and undifferentiated stem cells that can be recruited to become progenitor cells when the need arises. As a cell line becomes more differentiated, it becomes more highly specialized in its function and less able to divide.

CHARACTERISTICS OF BENIGN AND MALIGNANT NEOPLASMS

Neoplasms are composed of two types of tissue: parenchymal tissue and the stroma or supporting tissue. The *parenchymal cells* represent the functional components of an organ. The *supporting tissue* consists of the connective tissue, blood vessels, and lymph structure. The parenchymal cells of a tumor determine its behavior and are the component for which a tumor is named. The supporting tissue carries the blood vessels and provides support for tumor survival and growth.

Terminology

By definition, a *tumor* is a swelling that can be caused by a number of conditions, including inflammation and trauma. Although they are not synonymous, the terms *tumor* and *neoplasm* often are used interchangeably. Neoplasms usually are classified as benign or malignant. Neoplasms that contain well-differentiated cells that are clustered together in a single mass are considered to be *benign*. These tumors usually do not cause death unless their location or size interferes with vital functions. In contrast, malignant neoplasms are less well differentiated and have the ability to break loose, enter the circulatory or lymphatic systems, and form secondary malignant tumors at other sites. *Malignant neoplasms* usually cause suffering and death if untreated or uncontrolled.

Tumors usually are named by adding the suffix *-oma* to the parenchymal tissue type from which the growth originated. Thus, a benign tumor of glandular epithelial tissue is called an *adenoma*, and a benign tumor of bone tissue is called an *osteoma*. The term *carcinoma* is used to designate a malignant tumor of epithelial tissue origin. In the case of a malignant adenoma, the term *adenocarcinoma* is used. Malignant tumors of mesenchymal origin are called *sarcomas* (*e.g.*, osteosarcoma). *Papillomas* are benign microscopic or macroscopic fingerlike projections that grow on any surface. A *polyp* is a growth that projects from a mucosal surface, such as the intestine. Although the term usually implies a benign neoplasm, some malignant tumors also appear as polyps.³ Oncology is the study of tumors and their treatment. Table 5-1 lists the names of selected benign and malignant tumors according to tissue types.

Benign and malignant neoplasms usually are differentiated by their (1) cell characteristics, (2) manner of growth, (3) rate of growth, (4) potential for metastasizing or spreading to other parts of the body, (5) ability to produce generalized effects, (6) tendency to cause tissue destruction, and (7) capacity to cause death. The characteristics of benign and malignant neoplasms are summarized in Table 5-2.

Benign Neoplasms

Benign tumors are characterized by a slow, progressive rate of growth that may come to a standstill or regress, an expansive manner of growth, the presence of a well-defined fibrous capsule, and failure to metastasize to distant sites. Benign tumors are composed of well-differentiated cells that resemble the cells of the tissue of origin. For example, the cells of a uterine

TABLE 5-1 Names of Selected Benign and Malignant Tumors

According to Tissue Types

Epithelial Surface Papilloma Squ	Jamous cell carcinoma enocarcinoma
Surface Papilloma Squ	
Glandular Adenoma Ade	enocarcinoma
Connective	
Fibrous Fibroma Fibr	rosarcoma
Adipose Lipoma Lipo	osarcoma
Cartilage Chondroma Cho	ondrosarcoma
Bone Osteoma Oste	eosarcoma
5	mangiosarcoma
	nphangiosarcoma
Lymph tissue Lym	nphosarcoma
Muscle	
Smooth Leiomyoma Leio	omyosarcoma
Striated Rhabdomyoma Rha	abdomyosarcoma
Neural Tissue	
Nerve cell Neuroma Neu	uroblastoma
Glial tissue Glioma (benign) Glio	oblastoma, astrocytoma, medulloblastoma, oligodendroglioma
Nerve sheaths Neurilemmoma Neu	urilemmal sarcoma
Meninges Meningioma Mer	ningeal sarcoma
Hematologic	
	elocytic leukemia
	throcytic leukemia
	Itiple myeloma
	nphocytic leukemia or lymphoma
	nocytic leukemia
Endothelial Tissue	
Blood vessels Hemangioma Hem	mangiosarcoma
Lymph vessels Lymphangioma Lym	nphangiosarcoma

Characteristics of Benign and Malignant Neoplasms TABLE 5-2

Characteristics	Benign	Malignant
Cell characteristics	Well-differentiated cells that resemble normal cells of the tissue from which the tumor originated	Cells are undifferentiated and often bear little resemblance to the normal cells of the tissue from which they arose
Mode of growth	Tumor grows by expansion and does not infiltrate the surrounding tissues; usually encapsulated	Grows at the periphery and sends out processes that infil- trate and destroy the surrounding tissues
Rate of growth	Rate of growth usually is slow	Rate of growth is variable and depends on level of differ- entiation; the more anaplastic the tumor, the more rapid the rate of growth
Metastasis	Does not spread by metastasis	Gains access to the blood and lymph channels and metas- tasizes to other areas of the body
General effects	Usually is a localized phenomenon that does not cause generalized effects unless its location interferes with vital functions	Often causes generalized effects such as anemia, weak- ness, and weight loss
Tissue destruction	Usually does not cause tissue damage unless its location interferes with blood flow	Often causes extensive tissue damage as the tumor out- grows its blood supply or encroaches on blood flow to the area; also may produce substances that cause cell damage
Ability to cause death	Usually does not cause death unless its location interferes with vital functions	Usually causes death unless growth can be controlled

leiomyoma resemble uterine smooth muscle cells. For unknown reasons, benign tumors seem to have lost the ability to suppress the genetic program for cell replication but retain the program for normal cell differentiation. Benign tumors grow by expansion and are enclosed in a fibrous capsule. This is in sharp contrast to malignant neoplasms, which grow by infiltrating the surrounding tissue (Fig. 5-3). The capsule is responsible for a sharp line of demarcation between the benign tumor and the adjacent tissues, a factor that facilitates surgical removal. The formation of the capsule is thought to represent the reaction of the surrounding tissues to the tumor.⁶

Benign tumors do not undergo degenerative changes as readily as malignant tumors, and they usually do not cause death unless they interfere with vital functions because of their location. For instance, a benign tumor growing in the cranial cavity can eventually cause death by compressing brain structures. Benign tumors also can cause disturbances in the function of adjacent or distant structures by producing pressure on tissues, blood vessels, or nerves. Some benign tumors are also known for their ability to cause alterations in body function through abnormal elaboration of hormones.

KEY CONCEPTS

BENIGN AND MALIGNANT NEOPLASMS

- A tumor is a new growth or neoplasm.
- Benign neoplasms are well-differentiated tumors that resemble the tissues of origin but have lost the ability to control cell proliferation. They grow by expansion, are enclosed in a fibrous capsule, and do not cause death unless their location is such that it interrupts vital body functions.
- Malignant neoplasms are less well-differentiated tumors that have lost the ability to control both cell proliferation and differentiation. They grow in a crablike manner to invade surrounding tissues, have cells that break loose and travel to distant sites to form metastases, and inevitably cause suffering and death unless their growth can be controlled through treatment.





■ FIGURE 5-3 ■ Photographs of a benign encapsulated fibroadenoma of the breast (top) and a bronchogenic carcinoma of the lung (bottom). The fibroadenoma has sharply defined edges, but the bronchogenic carcinoma is diffuse and infiltrates the surrounding tissues.

Malignant Neoplasms

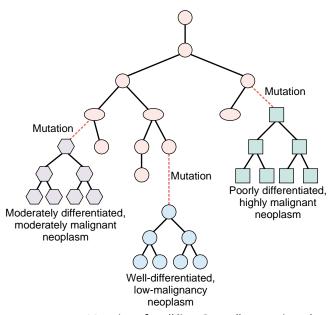
Malignant neoplasms tend to grow rapidly, spread widely, and kill regardless of their original location. Because of their rapid rate of growth, malignant tumors tend to compress blood vessels and outgrow their blood supply, causing ischemia and tissue necrosis; rob normal tissues of essential nutrients; and liberate enzymes and toxins that destroy tumor tissue and normal tissue. The destructive nature of malignant tumors is related to their lack of cell differentiation, cell characteristics, rate of growth, and ability to spread and metastasize.

There are two categories of cancer—solid tumors and hematologic cancers. Solid tumors initially are confined to a specific tissue or organ. As the growth of a solid tumor progresses, cells are shed from the original tumor mass and travel through the blood and lymph system to produce metastasis in distant sites. Hematologic cancers involve the blood-forming cells that naturally migrate to the blood and lymph systems, thereby making them disseminated diseases from the beginning.

Cancer Cell Characteristics

Cancer cells, unlike normal cells, fail to undergo normal cell proliferation and differentiation. It is thought that cancer cells develop from mutations that occur during the differentiation process (Fig. 5-4). When the mutation occurs early in the process, the resulting tumor is poorly differentiated and highly malignant; when it occurs later in the process, better differentiated and less malignant tumors result.

The term *anaplasia* is used to describe the lack of cell differentiation in cancerous tissue. Undifferentiated cancer cells are altered in appearance and nuclear size and shape from the cells in the tissue from which the cancer originated. In descending the scale of differentiation, enzymes and specialized pathways of metabolism are lost and cells undergo functional simplification.³ Highly anaplastic cancer cells, whatever their tissue of origin, begin to resemble each other more than they do their tissue of origin. For example, when examined under the



■ FIGURE 5-4 ■ Mutation of a cell line. Generally, mutations that occur early in the differentiation process result in poorly differentiated neoplasms and those that appear late in the differentiation process result in relatively well-differentiated neoplasms.

microscope, cancerous tissue that originates in the liver does not have the appearance of normal liver tissue. Some cancers display only slight anaplasia, and others display marked anaplasia.

Because cancer cells lack differentiation, they do not function properly, nor do they die in the time frame of normal cells. For example, in some types of leukemia the lymphocytes do not follow the normal developmental process. They do not differentiate fully, acquire the ability to destroy bacteria, or die on schedule. Instead, these long-lived, defective cells continue to grow, crowding the normal developing blood cells, thereby affecting the development of other cell lineages, such as the erythrocytes, platelets, and other white blood cells. This results in reduced numbers of mature, effectively functioning cells, producing white blood cells that cannot effectively fight infection, erythrocytes that cannot effectively transport oxygen to tissues, or platelets that cannot participate in the clotting system.

Alterations in cell differentiation also are accompanied by changes in cell characteristics and cell function that distinguish cancer cells from their fully differentiated normal counterparts. These changes include (1) alterations in contact inhibition, (2) loss of cohesiveness and adhesion, (3) impaired cell-to-cell communication, (4) expression of altered tissue antigens, and (5) elaboration of degradative enzymes that participate in invasion and metastatic spread.

Contact inhibition is the cessation of growth after a cell comes in contact with another cell. Contact inhibition usually switches off cell growth by blocking the synthesis of DNA, RNA, and protein. In wound healing, contact inhibition causes fibrous tissue growth to cease at the point where the edges of the wound come together. However, cancer cells tend to grow rampantly without regard for other tissue. The reduced tendency of cancer cells to stick together (*i.e., cohesiveness* and *adhesiveness*) permits shedding of the tumor's surface cells; these cells appear in the surrounding body fluids or secretions and can often be detected using the *Papanicolaou* (Pap) test.

Chemical messengers carry out cell-to-cell communication. These messengers bind to specific cell surface receptors and serve to control cell growth and modulate cell behavior. Impaired cell-to-cell communication may interfere with the formation of intercellular connections and responsiveness to membrane-derived signals.

Cancer cells express a number of cell surface molecules or antigens that are immunologically identified as foreign. These *tissue antigens* are coded by the genes of a cell. Many transformed cancer cells revert to earlier stages of gene expression and produce antigens that are immunologically distinct from the antigens that are expressed by cells of the well-differentiated tissue from which the cancer originated. Some cancers express fetal antigens that are not produced by comparable cells in the adult. Tumor antigens may be used clinically as markers to indicate the presence or progressive growth of a cancer.

Most cancers synthesize and secrete enzymes (*i.e.*, proteases and glycosidases) that break down proteins involved in ensuring intracellular organization and cell-to-cell cohesion. The degradation of the extracellular matrix by these enzymes facilitates invasiveness of the tumor. The production of degradative enzymes, such as fibrinolysins, contributes to the breakdown of the intercellular matrix, which leads to changes in the organization of the cell's cytoskeleton and affects cell-to-cell adhesion, cellular migration, and cellular communication. Cancers of nonendocrine tissues may assume hormone synthesis to produce so-called *ectopic hormones* (discussed with paraneoplastic syndromes in the section on general effects).

Invasion and Metastasis

Unlike benign tumors, which grow by expansion and usually are surrounded by a capsule, malignant tumors grow by extensive infiltration and invasion of the surrounding tissues. The word *cancer* is derived from the Latin word meaning *crablike* because cancerous growth spreads by sending crablike projections into the surrounding tissues. The lack of a sharp line of demarcation separating them from the surrounding tissue makes the complete surgical removal of malignant tumors more difficult than removal of benign tumors. *Seeding* of cancer cells into body cavities occurs when a tumor erodes into these spaces. Most often, the peritoneal cavity is involved, but other spaces such as the pleural cavity, pericardial cavity, and joint spaces may be involved. Seeding into the peritoneal cavity is particularly common with ovarian cancers.

The term *metastasis* is used to describe the development of a secondary tumor in a location distant from the primary tumor. Metastatic tumors retain many of the characteristics of the primary tumor from which they were derived. Because of this, it usually is possible to determine the site of the primary tumor from the cellular characteristics of the metastatic tumor. Some tumors tend to metastasize early in their developmental course, but others do not metastasize until later. Occasionally, the metastatic tumor is far advanced before the primary tumor becomes clinically detectable. For example, malignant tumors of the kidney may go undetected and be asymptomatic, even when a metastatic lesion is found in the lung.

Metastasis occurs by way of the lymph channels (*i.e.*, lymphatic spread) and the blood vessels (*i.e.*, hematogenic spread).^{3,7} In many types of cancer, the first evidence of disseminated disease is the presence of tumor cells in the lymph nodes that drain the tumor area. When metastasis occurs by way of the lymphatic channels, the tumor cells lodge first in the regional lymph nodes that received drainage from the tumor site. Once in the lymph node, the cells may die because of the lack of a proper environment, grow into a discernible mass, or remain dormant for unknown reasons. Because the lymphatic channels empty into the venous system, cancer cells that survive may eventually break loose and gain access to the venous system.

With hematologic spread, the blood-borne cancer cells typically follow the venous flow that drains the site of the neoplasm. Before entering the general circulation, venous blood from the gastrointestinal tract, pancreas, and spleen is routed through the portal vein to the liver. Thus, the liver is a common site for metastatic spread for cancers that originate in these organs. Although the site of hematologic spread is generally related to vascular drainage of the primary tumor, some tumors metastasize to distant and unrelated sites. One explanation is that cells of different tumors tend to metastasize to specific target organs that provide substances such as hormones or growth factors that are needed for their survival. For example, prostatic cancer preferentially spreads to bone, bronchiogenic cancer spreads to the adrenal glands and brain, and neuroblastoma spreads to the liver and bones.

The selective nature of hematologic spread indicates that metastasis is a finely orchestrated, multistep process, and only a small, select clone of cancer cells has the right combination of gene products to perform all of the steps needed for establishment of a secondary tumor. It has been estimated that fewer than 1 in 10,000 tumor cells that leave a primary tumor survives to start a secondary tumor.⁸ To metastasize, a cancer cell must be able to break loose from the primary tumor, invade the surrounding extracellular matrix, gain access to a blood vessel, survive its passage in the bloodstream, emerge from the bloodstream at a favorable location, invade the surrounding tissue, and begin to grow (Fig. 5-5).

Considerable evidence suggests that cancer cells capable of metastasis secrete enzymes that break down the surrounding extracellular matrix, allowing them to move through the degraded matrix and gain access to a blood vessel. Once in the circulation, the tumor cells are vulnerable to destruction by host immune cells. Some tumor cells gain protection from the antitumor host cells by aggregating and adhering to circulating blood components, particularly platelets, to form tumor emboli. Tumor cells that survive their travel in the circulation must be able to halt their passage by adhering to the vessel wall. After that, they must be able to exit the vessel, move through the extracellular matrix of the target tissue, and subsequently establish growth of a secondary tumor.

Once in the target tissue, the process of tumor development depends on the establishment of blood vessels and specific growth factors that promote proliferation of the tumor cells. Tumor cells secrete tumor angiogenesis factor, which enables the development of new blood vessels in the tumor.⁷ The presence of stimulatory or inhibitor growth factors correlates with the site-specific pattern of metastasis. For example, a potent growth-stimulating factor has been isolated from lung tissue,

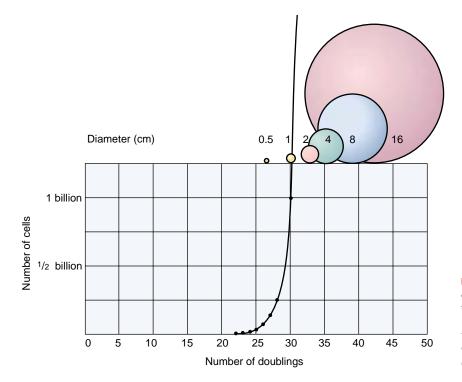
and stromal cells in bone have been shown to produce a factor that stimulates growth of prostatic cancer cells.⁷

Tumor Growth

The rate of tissue growth in normal and cancerous tissue depends on three factors: (1) the number of cells that are actively dividing or moving through the cell cycle, (2) the duration of the cell cycle, and (3) the number of cells that are being lost compared with the number of new cells being produced. One of the reasons cancerous tumors often seem to grow so rapidly relates to the size of the cell pool that is actively engaged in cycling. It has been shown that the cell cycle time of cancerous tissue cells is not necessarily shorter than that of normal cells; rather, cancer cells do not die on schedule. Also, the growth factors that allow cells to enter the resting or G_0 phase when they are not needed for cell replacement are lacking. Thus, a greater percentage of cells are actively engaged in moving through the cell cycle than occurs in normal tissue.

The ratio of dividing cells to resting cells in a tissue mass is called the *growth fraction*. The doubling time is the length of time it takes for the total mass of cells in a tumor to double. As the growth fraction increases, the doubling time decreases. When normal tissues reach their adult size, an equilibrium between cell birth and cell death is reached. However, cancer cells continue to divide until limitations in blood supply and nutrients inhibit their growth. As this happens, the doubling time for cancer cells decreases. If tumor growth is plotted against time on a semilogarithmic scale, the initial growth rate is exponential and then tends to decrease or flatten with time. This characterization of tumor growth is called the *Gompertzian model*.⁵

A tumor usually is undetectable until it has doubled 30 times and contains more than 1 billion (10⁹) cells. At this



■ FIGURE 5-6 ■ Growth curve of a hypothetical tumor on arithmetic coordinates. Notice the number of doubling times before the tumor reaches an appreciable size. (Adapted from Collins V.P., et al. [1956]. Observations of growth rates of human tumors. *Am J Roent Rad Ther Nuclear Med* 76, 988.)

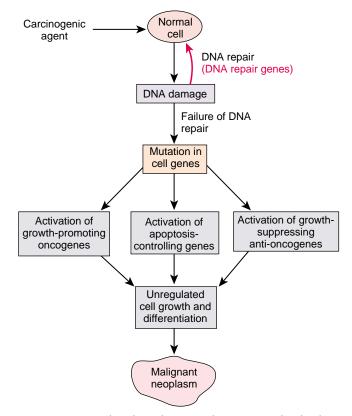
point, it is approximately 1 cm in size (Fig. 5-6). After 35 doublings, the mass contains more than 1 trillion (10^{12}) cells, which is a sufficient number to kill the host.

CARCINOGENESIS AND CAUSES OF CANCER

Carcinogenesis, or the development of cancer, is a multistep process that involves both the molecular aspects of cell transformation and the overall growth and spread of the tumor mass. Because cancer is not a single disease, it is reasonable to assume that it does not have a single cause. More likely, cancer occurs because of interactions between multiple risk factors or repeated exposure to a single carcinogenic (cancer-producing) agent. Among the risk factors that have been linked to cancer are heredity, chemical and environmental carcinogens, cancercausing viruses, and immunologic defects.

Oncogenesis: The Molecular Basis of Cancer

The term *oncogenesis* refers to a genetic mechanism whereby normal cells are transformed into cancer cells. There are three kinds of genes that control cell growth and replication: *protooncogenes, anti-oncogenes,* and genes that control programmed cell death or *apoptosis.*³ In addition to these three classes of genes, a fourth category of genes, those that regulate repair of damaged DNA, is implicated in the process of oncogenesis (Fig. 5-7). The DNA repair genes affect cell proliferation and survival indirectly through their ability to repair nonlethal damage in other genes, including proto-oncogenes, anti-oncogenes, and the genes that control apoptosis.³ These



■ FIGURE 5-7 ■ Flowchart depicting the stages in the development of a malignant neoplasm resulting from exposure to an oncogenic agent that produces DNA damage. When DNA repair genes are present (red arrow), the DNA is repaired and gene mutation does not occur.

Chapter 5: Alterations in Cell Growth and Replication: Neoplasia

genes have been implicated as the principal targets of genetic damage occurring during the development of cancer cells.^{9,10} Such genetic damage may be acquired by the action of chemicals, radiation, or viruses, or it may be inherited in the germ line. Significantly, it appears that the acquisition of a single gene mutation is not sufficient to transform normal cells into cancer cells. Instead, cancerous transformation appears to require the activation of many independently mutated genes.

Proto-oncogenes have essential roles in regulating the growth and proliferation of normal cells. Proto-oncogene products may act as growth factors, as receptors for growth factors, or as second messengers that transmit growth factor signals. The involvement of these genes in the cancer process is attributable to a somatic mutation that takes place in a specific target tissue, converting its proto-oncogenes into oncogenes.

Cancer suppressor genes, or anti-oncogenes, inhibit the proliferation of cells in a tumor. When this type of gene is inactivated, a genetic signal that normally inhibits proliferation is removed, thereby causing the cell to begin unregulated growth. Several human tumor suppressor genes have been identified.³ Of particular interest in this group is the p53 gene, located on the short arm of chromosome 17, that codes for a protein that is pivotal in growth regulation and functions as a suppressor of tumor growth. Mutation of the p53 gene has been implicated in the development of lung, breast, and colon cancer—the three leading causes of cancer death.³ The p53 gene also appears to initiate apoptosis of radiation- and chemotherapydamaged tumor cells. Thus, tumors that retain normal p53 function are more likely to respond to such therapy than are tumors that carry a defective p53 gene.³

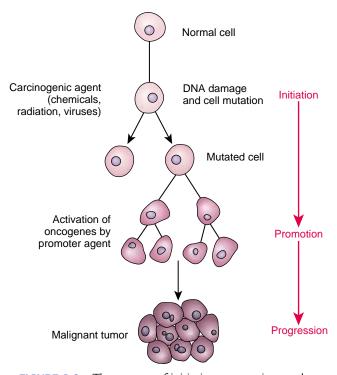
Tumor Cell Transformation

The transformation of normal cells to cancer cells by carcinogenic agents is a multistep process that can be divided into three stages: (1) initiation, (2) promotion, and (3) progression³ (Fig. 5-8). *Initiation* involves the exposure of cells to appropriate doses of a carcinogenic agent that makes them sus-

KEY CONCEPTS

ONCOGENESIS

- Normal cell growth is controlled by growthpromoting proto-oncogenes and growth-suppressing anti-oncogenes. Normally, cell growth is genetically controlled so that potentially malignant cells are targeted for elimination by tumor-suppressing genes.
- Oncogenesis is a genetic process whereby normal cells are transformed into cancer cells. It involves mutations in the normal growth-regulating genes.
- The transformation of normal cells into cancer cells is multifactorial, involving the inheritance of cancer susceptibility genes and environmental factors such as chemicals, radiation, and viruses.



■ FIGURE 5-8 ■ The process of initiation, promotion, and progression in the clonal evolution of malignant tumors. Initiation involves the exposure of cells to appropriate doses of a carcinogenic agent; promotion, the unregulated and accelerated growth of the mutated cells; and progression, the acquisition of malignant characteristics by the tumor cells.

ceptible to malignant transformation. The carcinogenic agents can be chemical, physical, or biologic, and produce irreversible changes in the genome of a previously normal cell. Because the effects of initiating agents are irreversible, multiple divided doses may achieve the same effects as single exposures of the same comparable dose or small amounts of highly carcinogenic substances. The most susceptible cells for mutagenic alterations in the genome are the cells that are actively synthesizing DNA.¹¹

Promotion involves the induction of unregulated accelerated growth in already initiated cells by various chemicals and growth factors. Promotion is reversible if the promoter substance is removed. Cells that have been irreversibly initiated may be promoted even after long latency periods. The latency period varies with the type of agent, the dosage, and the characteristics of the target cells. Many chemical carcinogens are called *complete carcinogens* because they can initiate and promote neoplastic transformation. *Progression* is the process whereby tumor cells acquire malignant phenotypic changes that promote invasiveness, metastatic competence, a tendency for autonomous growth, and increased karyotypic instability.¹¹

Heredity

A hereditary predisposition has been observed in approximately 50 types of cancer. For example, breast cancer occurs more frequently in women whose grandmothers, mothers, aunts, and sisters also have experienced a breast malignancy.

74 Unit One: Mechanisms of Disease

The genetic predisposition for development of cancer has been documented for a number of cancerous and precancerous lesions that follow mendelian inheritance patterns. Cancer is found in approximately 10% of persons having one affected first-degree relative, in approximately 15% of persons having two affected family members, and in 30% of persons having three affected family members. The risk increases to approximately 50% in women 65 years of age who have multiple family members with breast cancer. Two oncogenes, called BRAC1 (breast carcinoma 1) and BRAC2 (breast carcinoma 2), have been implicated in a genetic susceptibility to breast cancer.⁶

Several cancers exhibit an autosomal dominant inheritance pattern. In approximately 40% of cases, retinoblastoma (a malignant eye tumor that occurs in children) is inherited as an autosomal dominant trait; the remaining cases are nonhereditary. The penetrance of the genetic trait is high; in carriers of the dominant retinoblastoma gene, the penetrance for this gene is 95% for at least one tumor, and the affected person may be unilaterally or bilaterally affected.¹² Familial adenomatous polyposis of the colon also follows an autosomal dominant inheritance pattern. In people who inherit this gene, hundreds of adenomatous polyps may develop, some of which inevitably become malignant.³

Carcinogenic Agents

A carcinogen is an agent capable of causing cancer. The role of environmental agents in causation of cancer was first noted in 1775 by Sir Percivall Pott, who related the high incidence of scrotal cancer in chimney sweeps to their exposure to coal soot.¹³ In 1915, a group of Japanese investigators conducted the first experiments in which a chemical agent was used to produce cancer.¹³ These investigators found that a cancerous growth developed when they painted a rabbit's ear with coal tar. Coal tar has since been found to contain potent polycyclic aromatic hydrocarbons. Since then, many carcinogenic agents have been identified (Chart 5-1).

Chemical Carcinogens

More than six million chemicals have been identified. It is estimated that less than 1000 of these have been extensively examined for their carcinogenic potential.¹⁴ Some have been found to cause cancers in animals, and others are known to cause cancers in humans. These agents include both natural (*e.g.*, aflatoxin B₁) and artificial products (*e.g.*, vinyl chloride).

Chemical carcinogens can be divided into two groups: direct-reacting agents, which do not require activation in the body to become carcinogenic, and indirect-reacting agents, called *procarcinogens* or *initiators*, which become active only after metabolic conversion. Direct- and indirect-acting initiators form highly reactive species (*i.e.*, electrophiles and free radicals) that bind with the nucleophilic residues on DNA, RNA, or cellular proteins. The action of these reactive species tends to cause cell mutation or alteration in synthesis of cell enzymes and structural proteins in a manner that alters cell replication and interferes with cell regulatory controls. The carcinogenicity of some chemicals, called *promoters*, is augmented by agents that by themselves have little or no cancercausing ability. It is believed that promoters exert their effect by changing the expression of genetic material in a cell, in-

CHART 5-1 Chemical and Environmental Agents Known to be Carcinogenic in Humans

Polycyclic Hydrocarbons

Soots, tars, and oils Cigarette smoke

Industrial Agents

 Aniline and azo dyes

 Arsenic compounds

 Asbestos

 β-Naphthylamine

 Benzene

 Benzo[a]pyrene

 Carbon tetrachloride

 Insecticides, fungicides

 Nickel and chromium compounds

 Polychlorinated biphenyls

 Vinyl chloride

Food and Drugs

Smoked foods Nitrosamines Aflatoxin B₁ Diethylstilbestrol Anticancer drugs (*e.g.,* alkylating agents, cyclophosphamide, chlorambucil, nitrosourea)

creasing DNA synthesis, enhancing gene amplification (*i.e.*, number of gene copies that are made), and altering intercellular communication.

The exposure to many chemical carcinogens is associated with lifestyle risk factors, such as smoking, dietary factors, and alcohol consumption. Cigarette smoke contains both procarcinogens and promoters. It is directly associated with lung and laryngeal cancer¹⁵ and has been linked with cancers of the esophagus, pancreas, kidney, uterine cervix, and bladder. Chewing tobacco or tobacco products increases the risk of cancers of the oral cavity and esophagus. It has been estimated that 30% of current cancer deaths in the United States are related to tobacco. Not only is the smoker at risk, but others passively exposed to cigarette smoke are at risk. Environmental tobacco smoke has been classified as a "group A" carcinogen based on the U.S. Environmental Protection Agency's system of carcinogen classification. It also is estimated that between 20% and 60% of the deaths that occur each year from nonsmokingrelated lung cancers may be caused by environmental tobacco smoke.16

There is strong evidence that certain elements in the diet contain chemicals that contribute to cancer risk. For example, benzo[a]pyrene and other polycyclic hydrocarbons are converted to carcinogens when foods are fried in fat that has been reused multiple times. Nitrosamines, which are powerful carcinogens, may be formed from nitrites that are derived from nitrates added to vegetables and foods as a preservative. Among the most potent of the procarcinogens are the polycyclic hydrocarbons. The polycyclic hydrocarbons are of particular interest because they are produced in the combustion of tobacco and are present in cigarette smoke. They also are produced from animal fat in the process of broiling meats and are present in smoked meats and fish. Formation of these nitrosamines may be inhibited by the presence of antioxidants such as vitamin C in the stomach. Cancer of the colon has been associated with high dietary intake of fat, protein, and beef and low intake of dietary fiber. The carcinogenic factors associated with a high-fat diet have yet to be confirmed. However, some studies have shown a relationship between high levels of fecally excreted bile acids and colon cancer. A high-fat diet increases the flow of primary bile acids. These acids are converted to secondary bile acids in the presence of anaerobic bacteria in the colon. These acids are thought to be tumor promoters, rather than initiators.⁷

Alcohol modifies the metabolism of carcinogens in the liver and esophagus.¹⁷ It is believed to influence the transport of carcinogens, increasing the contact between an externally induced carcinogen and the stem cells that line the upper oral cavity and esophagus. The carcinogenic effect of cigarette smoke can be enhanced by concomitant consumption of alcohol; persons who smoke and drink considerable amounts of alcohol are at increased risk for the development of cancer of the oral cavity and esophagus.

The effects of carcinogenic agents usually are dose dependent—the larger the dose or the longer the duration of exposure, the greater the risk that cancer will develop. Some chemical carcinogens may act in concert with other carcinogenic influences, such as viruses or radiation, to induce neoplasia. There usually is a time delay ranging from 5 to 30 years from the time of chemical carcinogen exposure to the development of overt cancer. This is unfortunate because many people may have been exposed to the agent and its carcinogenic effects before the association was recognized. This occurred with the use of diethylstilbestrol, which was widely used in the United States from the mid-1940s to 1970 to prevent miscarriages. But it was not until the late 1960s that many cases of vaginal adenosis and adenocarcinoma in young women were found to be the result of their exposure in utero to diethylstilbestrol.¹⁸

Radiation Oncogenesis

The effects of *ionizing radiation* in carcinogenesis have been well documented in atomic bomb survivors, in patients diagnostically exposed, and in industrial workers, scientists, and physicians who were exposed during employment. Malignant epitheliomas of the skin and leukemia were significantly elevated in these populations.⁶ Between 1950 and 1970, the death rate from leukemia alone in the most heavily exposed population groups of the atomic bomb survivors in Hiroshima and Nagasaki was 147 per 100,000 persons, 30 times the expected rate.¹⁹

The type of cancer that developed depended on the dose of radiation, the gender of the person, and the age at which exposure occurred. For instance, approximately 25 to 30 years after total body or trunk irradiation, there were increased incidences of leukemia and cancers of the breast, lung, stomach, thyroid, salivary gland, gastrointestinal system, and lymphoid tissues. The length of time between exposure and the onset of cancer is related to the age of the individual. For example, children exposed to ionizing radiation in utero have an increased risk for developing leukemias and childhood tumors, particularly 2 to 3 years after birth. This latency period for leukemia extends to 5 to 10 years if the child was exposed after birth and to 20 years for certain solid tumors.⁶ As another example, the latency period for the development of thyroid cancer in infants and small children who received radiation to the head and neck to decrease the size of the tonsils or thymus was as long as 35 years after exposure.

The association between sunlight and the development of skin cancer has been reported for more than 100 years. *Ultraviolet radiation* emits relatively low-energy rays that do not deeply penetrate the skin (see Chapter 44). As with other carcinogens, the effects of ultraviolet radiation usually are additive, and there usually is a long delay between the time of exposure and the time that cancer can be detected.

Oncogenic Viruses

An oncogenic virus is one that can induce cancer. Viruses, which are small particles containing genetic (DNA or RNA) material, enter a host cell and become incorporated into its chromosomal DNA or take control of the cell's machinery for the purpose of producing viral proteins. A large number of DNA and RNA viruses (*i.e.*, retroviruses) have been shown to be oncogenic in animals. However, only a few viruses have been linked to cancer in humans.³ Among the recognized oncogenic viruses in humans are the human papilloma virus (HPV), Epstein-Barr virus (EBV), hepatitis B virus (HBV), and human T-cell leukemia virus-1 (HTLV-1).⁶ Herpes simplex type 2 also has been associated with cervical cancer, but the evidence supporting its role as a carcinogenic influence is less clear.

Three DNA viruses have been implicated in human cancers: HPV, EBV, and HBV. The transforming DNA viruses form stable associations with the human genome using genes that allow them to complete their replication cycle and be expressed in transformed cells. There is strong evidence to suggest that the DNA viruses act in concert with other factors to cause cancer.

There are more than 60 genetically different types of HPV. Some types (*i.e.*, types 1, 2, 4, 7) have been shown to cause benign squamous papillomas (*i.e.*, warts). HPVs also have been implicated in squamous cell carcinoma of the cervix and anogenital region. HPV types 16 and 18 and, less commonly, types 31, 33, 35, and 51, have been found in approximately 85% of squamous cell carcinomas of the cervix and presumed precursors (*i.e.*, severe cervical dysplasia and carcinoma in situ).³

EBV is a member of the herpesvirus family. It has been implicated in the pathogenesis of four human cancers: Burkitt's lymphoma, nasopharyngeal cancer, B-cell lymphomas in immunosuppressed individuals such as those with acquired immunodeficiency syndrome (AIDS), and in some cases of Hodgkin's lymphoma. Burkitt's lymphoma, a tumor of B lymphocytes, is endemic in parts of East Africa and occurs sporadically in other areas worldwide. In persons with normal immune function, the EBV-driven, B-cell proliferation is readily controlled, and the person becomes asymptomatic or experiences a self-limited episode of infectious mononucleosis (see Chapter 11). In regions of the world where Burkitt's lymphoma is endemic, concurrent malaria or other infections cause impaired immune function, allowing sustained B-lymphocyte proliferation. The incidence of nasopharyngeal cancer is high in some areas of China, particularly southern China, and in the Cantonese population in Singapore.

HBV is the etiologic agent in the development of hepatitis B, cirrhosis, and hepatocellular carcinoma. There is a significant correlation between elevated rates of hepatocellular carcinoma worldwide and the prevalence of HBV carriers. Other etiologic factors also may contribute to the development of liver cancer. Ingestion of aflatoxin and infection with HCV, the hepatitis C virus, have been implicated. The precise mechanism by which HBV induces hepatocellular cancer has not been determined, although it has been suggested that it may be the result of prolonged HBV-induced liver damage and regeneration.

HTLV-1, a retrovirus, is associated with a form of T-cell leukemia that is endemic in certain parts of Japan and some areas of the Caribbean and Africa and is found sporadically elsewhere, including the United States and Europe.³ Similar to the AIDS virus, HTLV-1 is attracted to the CD4⁺T cells, and this subset of T cells is therefore the major target for cancerous transformation. The virus requires transmission of infected T cells by way of sexual intercourse, infected blood, or breast milk. A second type of HTLV virus, HTLV-2, has been isolated from individuals with an unusual form of hairy cell leukemia. Hairy cell leukemia is usually the result of an alteration in the B-lymphocyte lineage.²⁰

Immunologic Defects

There is growing evidence for the immune system's participation in resistance against the progression and spread of cancer. The central concept, known as the *immune surveillance hypothesis*, which was first proposed by Paul Ehrlich in 1909, postulates that the immune system plays a central role in resistance against the development of tumors.²¹ In addition to cancer-host interactions as a mechanism of cancer development, immunologic mechanisms provide a means for the detection, classification, and prognostic evaluation of cancers and as a potential method of treatment. *Immunotherapy* (discussed later in this chapter) is a cancer treatment modality designed to heighten the patient's general immune responses to increase tumor destruction.

It has been suggested that the development of cancer might be associated with impairment or decline in the surveillance capacity of the immune system. For example, increases in cancer incidence have been observed in people with immunodeficiency diseases and in those with organ transplants who are receiving immunosuppressant drugs. The incidence of cancer also is increased in the elderly, in whom there is a known decrease in immune activity. The association of Kaposi's sarcoma with AIDS further emphasizes the role of the immune system in preventing malignant cell proliferation.¹³

It has been shown that most tumor cells have molecular configurations that can be specifically recognized by immune T cells or by antibodies and thus are termed tumor antigens. The most relevant tumor antigens fall into two categories: unique tumor-specific antigens found only on tumor cells, and tumor-associated antigens found on tumor cells and on normal cells. Quantitative and qualitative differences permit the use of these tumor-associated antigens to distinguish cancer cells from normal cells.²²

Virtually all of the components of the immune system have the potential for eradicating cancer cells, including T lymphocytes, B lymphocytes, antibodies, macrophages, and natural killer (NK) cells (see Chapter 9). The T-cell response is undoubtedly one of the most important host responses for controlling the growth of antigenic tumor cells; it is responsible for the direct killing of tumor cells and for activation of other components of the immune system. The T-cell response to cancer cells reflects the function of two subsets of T cells: the CD4⁺ helper T cells and CD8⁺ cytotoxic T cells. The finding of tumor-reactive antibodies in the serum of people with cancer supports the role of the B cell as a member of the immune surveillance team. Antibodies can destroy cancer cells through complement-mediated mechanisms or through antibodydependent cellular cytotoxicity, in which the antibody binds the cancer cell to another effector cell, such as the NK cell, that does the actual killing of the cancer cell. NK cells do not require antigen recognition and can lyse a wide variety of target cells. The cytotoxic activity of NK cells can be augmented by the lymphokines, IL-2, and interferon, and NK activity can be amplified by immune T-cell responses.²³ Macrophages are important in tumor immunity as antigen-presenting cells to initiate the immune response and as potential effector cells to participate in tumor cell lysis.

In summary, neoplasms may be benign or malignant. Benign and malignant tumors differ in terms of cell characteristics, manner of growth, rate of growth, potential for metastasis, ability to produce generalized effects, tendency to cause tissue destruction, and capacity to cause death. The growth of a benign tumor is restricted to the site of origin, and the tumor usually does not cause death unless it interferes with vital functions. However, cancers or malignant neoplasms grow wildly and without organization, spread to distant parts of the body, and cause death unless their growth is inhibited or stopped by treatment.

There are two types of cancer: solid tumors and hematologic tumors. Solid tumors initially are confined to a specific organ or tissue, but hematologic cancers are disseminated from the onset. Cancer is a disorder of cell proliferation and differentiation. Cancer cells often are poorly differentiated compared with normal cells, and they display abnormal membrane characteristics, have abnormal antigens, produce abnormal biochemical products, and have abnormal karyotypes. All cancers result from nonlethal genetic changes that transform a normal cell into a cancer cell. The spread of cancer occurs through three pathways: direct invasion and extension, seeding of cancer cells in body cavities, and metastatic spread through vascular or lymphatic pathways. Only a proportionately small clone of cancer cells is capable of metastasis. To metastasize, a cancer cell must be able to break loose from the primary tumor, invade the surrounding extracellular matrix, gain access to a blood vessel, survive its passage in the bloodstream, emerge from the bloodstream at a favorable location, invade the surrounding tissue, and begin to grow.

Because cancer is not a single disease, it is reasonable to assume that it does not have a single cause. Multiple factors probably interact at the genetic level to transform normal cells into cancer cells. This transformation process is called *oncogenesis*. Four kinds of genes control normal cell growth and replication: growth-promoting regulatory genes (*i.e.*, proto-oncogenes) and growth-inhibiting regulatory genes (*i.e.*, anti-oncogenes), genes that control apoptosis, and gene-repair genes. These genes are implicated as principal targets of the genetic damage that occurs during the development of a cancer cell. Such genetic damage may be acquired by the action of chemicals (*i.e.*, chemical carcinogens), radiation, or viruses, or it may be inherited in the cell line.

CLINICAL FEATURES

TABLE 5-3

There probably is no single body function left unaffected by the presence of cancer (Table 5-3). Because tumor cells replace normally functioning parenchymal tissue, the initial manifestations of cancer usually reflect the primary site of involvement. For example, cancer of the lung initially produces impairment of respiratory function; as the tumor grows and metastasizes, other body structures become affected.

Cancer disrupts tissue integrity. As cancers grow, they compress and erode blood vessels, causing ulceration and necrosis along with frank bleeding and sometimes hemorrhage. One of the early warning signals of colorectal cancer is blood in the stool. Cancer cells also may produce enzymes and metabolic toxins that are destructive to the surrounding tissues. Usually, tissue damaged by cancerous growth does not heal normally. Instead, the damaged area persists and often continues to grow; a sore that does not heal is another warning signal of cancer. Cancer has no regard for normal anatomic boundaries; as it grows, it invades and compresses adjacent structures. For example, abdominal cancer often compresses the viscera and causes bowel obstruction. Cancer may obstruct lymph flow and penetrate serous cavities, causing pleural effusion and ascites. In its late stages, cancer often causes pain (see Chapter 39). Pain is probably one of the most dreaded aspects of cancer, and pain management is one of the major treatment concerns for persons with incurable cancers.

Abnormalities in energy, carbohydrate, lipid, and protein regulation are common manifestations during progressive tumor growth. Many cancers are associated with weight loss and wasting of body fat and lean protein, a condition called cancer cachexia. Although anorexia, reduced food intake, and abnormalities of taste are common in people with cancer and often are accentuated by treatment methods, the extent of weight loss and protein wasting cannot be explained in terms of diminished food intake alone. There also is a disparity between the size of the tumor and the severity of cachexia, which supports the existence of other mediators in the development of cachexia. Cachexia is thought to be the result of tumor-derived or host-derived factors that cause anorexia directly by acting on satiety centers in the hypothalamus or indirectly by injuring tissues that subsequently release anorexigenic substances.

Cachectin was the first identified cytokine associated with wasting. Cachectin was later found to be identical to tumor necrosis factor (TNF), a cytokine secreted primarily from macrophages in response to tumor cell growth or gram-negative bacterial infections.²⁴ TNF causes anorexia by suppressing satiety centers and by suppressing the synthesis of lipoprotein lipase, an enzyme that facilitates the release of fatty acids from lipoproteins so they can be used by tissues. TNF is an endogenous pyrogen that induces fever by its actions on cells in the hypothalamic regulatory regions of the brain. TNF also induces a number of inflammatory responses, activates the coagulation system, suppresses bone marrow stem cell division, acts on hepatocytes to increase the synthesis of specific serum proteins in response to inflammatory stimuli, and mediates endotoxic shock secondary to trauma, burns, and sepsis.²⁵ The role of TNF and its full impact on cancer cachexia are uncertain. It has been suggested by some that the cytokine may be an endogenous antineoplastic agent. Interleukin (IL)-1, another cytokine secreted from macrophages, shares with TNF the ability to initiate cachexia.

	1
Overall Effect	Related Tumor Action
Altered function of the involved tissue	Destruction and replacement of parenchymal tissue by neoplastic growth
Bleeding and hemorrhage	Compression of blood vessels, with ischemia and necrosis of tissue; or tumor may outgrow its blood supply
Ulceration, necrosis, and infection of tumor area	Ischemia associated with rapid growth, with subsequent bacterial invasion
Obstruction of hollow viscera or communication pathways	Expansive growth of tumor with compression and invasion of tissues
Effusion in serous cavities	Impaired lymph flow from the serous cavity or erosion of tumor into the cavity
Increased risk of vascular thrombosis	Abnormal production of coagulation factors by the tumor, obstruction of venous channels, and immobility
Anemia	Bleeding and depression of red blood cell production
Bone destruction	Metastatic invasion of bony structures
Hypercalcemia	Destruction of bone due to metastasis or production by the tumor of parathyroid-like hormone
Pain	Liberation of pain mediators by the tumor, compression, or ischemia of structures
Cachexia, weakness, wasting of tissues	Catabolic effect of the tumor on body metabolism along with selective trapping of nutrients by rapidly growing tumor cells
Inappropriate hormone production	Production by the tumor of hormones or hormone-like substances that are not
(e.g., ADH or ACTH secretion by cancers such	regulated by normal feedback mechanisms
as bronchogenic carcinoma)	

General Effects on Body Function Associated With Cancer Growth

Unit One: Mechanisms of Disease

In addition to signs and symptoms at the sites of primary and metastatic disease, cancer can produce manifestations in sites that are not directly affected by the disease. Such manifestations are collectively referred to as *paraneoplastic syndromes*. Some of these manifestations are caused by the elaboration of hormones by cancer cells, and others result from the production of circulating factors that produce nonmetastatic hematopoietic, neurologic, and dermatologic syndromes. For example, cancers may produce procoagulation factors that contribute to an increased risk of venous thrombosis. It is estimated that approximately 10% of persons with cancer are affected by these syndromes.3 The three most common endocrine syndromes associated with cancer are the syndrome of inappropriate antidiuretic hormone secretion (see Chapter 6), Cushing's syndrome caused by ectopic adrenocorticotropic hormone (now called *corticotropin*) production (see Chapter 31), and hypercalcemia (see Chapter 6). Hypercalcemia of malignancy does not appear to be related to parathyroid hormone (PTH) but to PTH-related protein, which shares several biologic actions with PTH. It also can be caused by osteolytic processes induced by cancer such as multiple myeloma or bony metastases from other cancers. The paraneoplastic syndromes may be the earliest indication that a person has cancer; they also may signal early recurrence of the disease in previously treated patients.

Diagnosis and Staging

The methods used in the diagnosis and staging of cancer are determined largely by the location and type of cancer suspected. A number of procedures are used in the diagnosis of cancer, including x-ray studies, endoscopic examinations, urine and stool tests, blood tests for tumor markers, bone marrow aspirations, ultrasound imaging, magnetic resonance imaging (MRI), and computed tomography (CT) scan.

The Pap Smear

The Pap smear is an example of the type of test called exfoliative cytology. It consists of a microscopic examination of a properly prepared slide by a cytotechnologist or pathologist for the purpose of detecting the presence of abnormal cells. The usefulness of exfoliative cytology relies on the fact that the cancer cells lack the cohesive properties and intercellular junctions that are characteristic of normal tissue; without these characteristics, cancer cells tend to exfoliate and become mixed with secretions surrounding the tumor growth. The American Cancer Society recommends that the test be done annually to detect cervical cancer in women who are or have been sexually active and who have reached 18 years of age. After three consecutive normal findings, the test may be performed less frequently at the discretion of the physician.²⁶ Exfoliative cytology also can be performed on other body secretions, including nipple drainage, pleural or peritoneal fluid, and gastric washings.

Biopsy

Tissue biopsy is the removal of a tissue specimen for microscopic study. Biopsies are obtained in a number of ways, including needle aspiration (*i.e.*, fine, percutaneous, or core needle); by endoscopic methods, such as bronchoscopy or cystoscopy, which involve the passage of an endoscope through an orifice and into the involved structure; or by laparoscopic methods. In some instances, a surgical incision is made from which biopsy specimens are obtained. Excisional biopsies are those in which all of the tumor is removed. The tumors usually are small, solid, palpable masses. If the tumor is too large to be completely removed, a wedge of tissue from the mass can be excised for examination. Tissue diagnosis is of critical importance in designing the treatment plan should cancer cells be found.27

Tumor Markers

Tumor markers are antigens that are expressed on the surface of tumor cells or substances released from normal cells in response to the presence of tumor. Some substances, such as hormones and enzymes, are produced normally by the tissue involved but become overexpressed as a result of cancer. Other tumor markers, such as oncofetal proteins, are produced during fetal development and are induced to re-appear later in life as a result of benign and malignant neoplasms. Tumor markers are used for screening, diagnosis, establishing prognosis, monitoring treatment, and detecting recurrent disease.

As diagnostic tools, tumor markers have limitations. The value of a marker depends on its sensitivity, specificity, proportionality, and feasibility.²⁸ Sensitivity implies that the marker is apparent early in the development of the tumor and has few false-negative results. Specificity indicates that the marker is specific for the specific cancer and is not elevated in other disease conditions (i.e., has few false-positive results). Proportionality means that the level of marker accurately reflects the growth of the tumor, such that higher levels reflect a larger growth. Feasibility implies that the methods are readily available, easy to use, and that the cost is not prohibitive. Nearly all markers can be elevated in benign conditions, and most are not elevated in the early stages of malignancy. Thus, tumor markers have limited value as screening tests. Extremely elevated levels of a tumor marker can indicate a poor prognosis or the need for more aggressive treatment. Perhaps the greatest value of tumor markers is in monitoring therapy in people with widespread cancer. Nearly all markers show an association with the clinical course of the disease. The levels of most markers decline with successful treatment and increase with recurrence of the tumor.

The markers that have been most useful in practice have been human chorionic gonadotropin (hCG), CA 125, prostatespecific antigen (PSA), prostatic acid phosphatase (PAP), α -fetoprotein (AFP), and carcinoembryonic antigen (CEA). HCG is a hormone normally produced by the placenta. It is used as a marker for diagnosing, prescribing treatment, and following the disease course in persons with high-risk gestational trophoblastic tumors. PSA and PAP are used as markers in prostate cancer, and CA 125 is used as a marker in ovarian cancer.

Some cancers express oncofetal antigens, which are differentiation antigens normally present only during embryonal development.³ The two that have proven the most useful as tumor markers are AFP and CEA. AFP is synthesized by the fetal liver, yolk sac, and gastrointestinal tract and is the major serum protein in the fetus. Elevated levels are encountered in people with primary liver cancers and have also been observed in some testicular, ovarian, pancreatic, and stomach cancers. CEA normally is produced by embryonic tissue in the gut, pancreas, and liver and is elaborated by a number of different cancers. Depending on the serum level adopted for significant elevation, CEA is elevated in approximately 60% to 90% of colorectal carcinomas, 50% to 80% of pancreatic cancers, and 25% to 50% of gastric and breast tumors.³ As with most other tumor markers, elevated levels of CEA and AFP are found in other, noncancerous conditions, and elevated levels of both depend on tumor size so that neither is useful as an early test for cancer.

Staging and Grading of Tumors

The two basic methods for classifying cancers are grading according to the histologic or cellular characteristics of the tumor and staging according to the clinical spread of the disease. Both methods are used to determine the course of the disease and aid in selecting an appropriate treatment or management plan. Grading of tumors involves the microscopic examination of cancer cells to determine their level of differentiation and the number of mitoses. Cancers are classified as grades I, II, III, and IV with increasing anaplasia or lack of differentiation. Staging of cancers uses methods to determine the progress and spread of the disease. Surgery may be used to determine tumor size and lymph node involvement.

The clinical staging of cancer is intended to provide a means by which information related to the progress of the disease, the methods and success of treatment modalities, and the prognosis can be communicated to others. The TNM system, which has evolved from the work of the International Union Against Cancer (IUAC) and the American Joint Committee on Cancer Staging and End Stage Reporting (AJCCS), is used by many cancer facilities. This system, which is briefly described in Chart 5-2, classifies the disease into stages using three tumor components: T stands for the extent of the primary tumor, N refers to the involvement of the regional lymph nodes, and M describes the extent of the metastatic involvement. The time of staging is indicated as cTNM, clinical-diagnostic staging; pTNM, postsurgical resection-pathologic staging; sTNM, surgical-evaluative staging; rTNM, retreatment staging; and aTNM, autopsy staging.29

CHART	5-2 TNM Classification System
T (tumo	or)
Tx T0 Tis T1-4	Tumor cannot be adequately assessed No evidence of primary tumor Carcinoma in situ Progressive increase in tumor size or involvement
N (nod	es)
Nx N0 N 1–3	Regional lymph nodes cannot be assessed No evidence of regional node metastasis Increasing involvement of regional lymph nodes
M (met	astasis)
Mx M0 M1	Not assessed No distant metastasis Distant metastasis present, specify sites

Cancer Treatment

The goals of cancer treatment methods fall into three categories: curative, controlling, and palliative. The most common modalities are surgery, radiation, chemotherapy, hormonal therapy, and biotherapy. The treatment of cancer involves the use of a carefully planned program that combines the benefits of multiple treatment modalities and the expertise of an interdisciplinary team of specialists including medical, surgical, and radiation oncologists; clinical nurse specialists; nurse practitioners; pharmacists; and a variety of ancillary personnel.

Surgery

Surgery is used for diagnosis, the staging of cancer, tumor removal, and palliation (*i.e.*, relief of symptoms) when a cure cannot be achieved. The type of surgery to be used is determined by the extent of the disease, the location and structures involved, the tumor growth rate and invasiveness, the surgical risk to the patient, and the quality of life the patient will experience after the surgery. If the tumor is small and has welldefined margins, the entire tumor often can be removed. However, if the tumor is large or involves vital tissues, surgical removal may be difficult if not impossible.

Surgical techniques have expanded to include electrosurgery, cryosurgery, chemosurgery, cytoreductive surgery, and laser surgery. Electrosurgery uses the cutting and coagulating effects of high-frequency current applied by needle, blade, or electrodes. Once considered a palliative procedure, it now is being used as an alternative treatment for certain cancers of the skin, oral cavity, and rectum. Cryosurgery involves the instillation of liquid nitrogen into the tumor through a probe. It is used in treating cancers of the oral cavity, brain, and prostate. Chemosurgery is used in skin cancers. It involves the use of a corrosive paste in combination with multiple frozen sections to ensure complete removal of the tumor. Laser surgery uses a laser beam to resect a tumor. It has been used effectively in retinal and vocal cord surgery.

Radiation Therapy

Radiation can be used singly as the primary method of treatment, as preoperative or postoperative treatment, with chemotherapy, or with chemotherapy and surgery. It can also be used as a palliative treatment to reduce symptoms in persons with advanced cancers. It is effective in reducing the pain associated with bone metastasis and, in some cases, improves mobility. Radiation also is used to treat several oncologic emergencies such as superior vena cava syndrome, spinal cord compression, bronchial obstruction, and hemorrhage.³⁰

Radiation therapy exerts its effects by direct or indirect ionization. Indirect ionization produced by x-rays or gamma rays causes cellular damage when these rays are absorbed into tissue and give up their energy by producing fast-moving electrons. These electrons interact with free or loosely bonded electrons of the absorber cells and subsequently produce free radicals that interact with critical cell components (see Chapter 3). It can immediately kill cells, delay or halt cell cycle progression, or at dose levels commonly used in radiation therapy, cause damage to the cell nucleus that results in cell death after replication. Cell damage can be sublethal, in which case a single break in the strand can repair itself before the next radiation insult. Double-stranded breaks in DNA are generally 80

believed to be the primary damage that leads to radiation death of cells. The result of unrepaired DNA is that cells may continue to function until they undergo cell mitosis, at which time the genetic damage from the irradiation may result in death of the cell. The clinical significance is that the rapidly proliferating and poorly differentiated cells of a cancerous tumor are more likely to be injured by radiation therapy than are the slower proliferating cells of normal tissue. However, to some extent radiation is injurious to all rapidly proliferating cells, including those of the bone marrow and the mucosal lining of the gastrointestinal tract. This results in many of the common adverse effects of radiation therapy, including infection, bleeding, and anemia due to loss of blood cells and nausea and vomiting due to loss of gastrointestinal cells. In addition to its lethal effects, radiation also produces sublethal injury. Recovery from sublethal doses of radiation occurs in the interval between the first dose of radiation and subsequent doses. This is why large total doses of radiation can be tolerated when they are divided into multiple smaller fractionated doses. Normal tissue is usually able to recover from radiation damage more readily than is cancerous tissue.

Systemic Cancer Therapy

The use of chemotherapy drugs, hormones, antihormones, and biotherapy has become a highly specialized and increasingly effective means of treating cancers. The therapies rely on systemic agents that are distributed throughout the body. Gene therapy, although investigational, may provide a foundation for the development of more effective treatments in the future.

Bone marrow transplantation and peripheral blood stem cell transplantation are two treatment approaches for leukemias, certain solid tumors, and other cancers previously thought to be incurable.

Chemotherapy. Since the early 1960s, cancer chemotherapy has evolved as a major treatment modality. More than 30 different chemotherapeutic drugs are used alone or in various combinations. Administering higher doses of multiple drugs may be used as a strategy to achieve cure or optimal palliation; however, the adverse drug interactions and side effects can be unpredictable and intense. Chemotherapeutic drugs may be the primary form of treatment, or they may be used as adjuncts to other treatments. Chemotherapy is the primary treatment for most hematologic and some solid tumors, including choriocarcinoma, testicular cancer, acute and chronic leukemia, Burkitt's lymphoma, Hodgkin's disease, and multiple myeloma.

Cancer chemotherapeutic drugs exert their effects through several mechanisms. At the cellular level, they exert their lethal action by creating adverse conditions that prevent cell growth and replication. These mechanisms include disrupting production of essential enzymes; inhibiting DNA, RNA, and protein synthesis; and preventing cell mitosis.^{6,31}

For most chemotherapy drugs, the relationship between tumor cell survival and drug dose is exponential, with the number of cells surviving being proportional to drug dose, and the number of cells at risk for exposure being proportional to the destructive action of the drug. Chemotherapeutic drugs are most effective in treating tumors that have a high growth fraction because of their ability to kill rapidly dividing cells. Exponential killing implies that a proportion or percentage of tumor cells are killed, rather than an absolute number. This proportion is a constant percentage of the total number of cells. For this reason, multiple courses of treatment are needed if the tumor is to be eradicated.⁵

The anticancer drugs may be classified as either cell cycle specific or cell cycle nonspecific. Drugs are cell cycle specific if they exert their action during a specific phase of the cell cycle. For example, methotrexate, an antimetabolite, acts by interfering with DNA synthesis and thereby interrupts the S phase of the cell cycle. Drugs that are cell cycle nonspecific affect cancer cells through all the phases of the cell cycle. The alkylating agents, which are cell cycle nonspecific, act by disrupting DNA when the cells are in the resting state and when they are dividing. The site of action of various cancer drugs varies. Chemotherapeutic drugs that have similar structures and effects on cell function usually are grouped together, and these drugs usually have similar toxic and side effects. Because they differ in their mechanisms of action, combinations of cell cycle-specific and cell cycle-nonspecific agents often are used to treat cancer.

Combination chemotherapy has been found to be more effective than treatment with a single drug. With this method, several drugs with different mechanisms of action, metabolic pathways, times of onset of action and recovery, side effects, and onset of side effects are used. Drugs used in combinations are individually effective against the tumor and synergistic with each other. The regimens for combination therapy often are referred to by acronyms. Two well-known combinations are CHOP (cyclophosphamide, doxorubicin, Oncovin [vincristine], and prednisone), used in the treatment of Hodgkin's disease, and CMF (cyclophosphamide, methotrexate, and 5-fluorouracil), used in the treatment of breast cancer. The maximum possible drug doses usually are used to ensure the maximum cell killing. Routes of administration and dosage schedules are carefully designed to ensure optimal delivery of the active forms of the drugs to a tumor during the sensitive phase of the cell cycle.

Hormone and Antihormone Therapy. Hormone therapy consists of administration of hormones or hormone-blocking drugs. It is used for cancers that are responsive to or dependent on hormones for growth. The actions of hormones depend on the presence of specific receptors in the tumor. Among the tumors known to be responsive to hormonal manipulations are those of the breast, prostate, adrenal gland, and uterine endometrium. Hormones commonly used for cancer treatment include estrogens (e.g., diethylstilbestrol, estradiol), androgens (*e.g.*, testosterone), and progestins (*e.g.*, hydroxyprogesterone). Hormone therapy also involves use of the adrenal corticosteroid hormones such as prednisone, dexamethasone, and methylprednisolone. These compounds inhibit mitosis and are cytotoxic to cells of lymphocytic origin. Hormones are cell cycle nonspecific and are thought to alter the synthesis of RNA and proteins by binding to receptor sites. The side effects of hormonal treatment are directly related to the normal action of the hormones. Because dosages of these drugs are usually higher than those that normally occur in the body, the normal actions of the hormone are accentuated.³¹

Hormone-blocking drugs include the antiestrogen drugs tamoxifen and leuprolide (*i.e.*, a gonadotropin-releasing hormone analog that blocks both estrogens and androgens) and the antiadrenal drug aminoglutethimide. **Biotherapy**. Biotherapy involves the use of biologic response modifiers (BRMs) that change the person's own biologic response to cancer. The BRMs are products normally produced in the body that serve as regulators and messengers of normal cellular function. Although biotherapy relies heavily on immune mechanisms, it is not limited to them. Three major mechanisms by which biotherapy exerts its effects are: (1) modification of host responses, (2) direct destruction of cancer cells by suppressing tumor growth or killing the tumor cell, and (3) modification of tumor cell biology.

Immunotherapy techniques include active and passive immunotherapy. Active immunotherapy involves nonspecific techniques, such as bacille Calmette-Guérin (BCG) and levamisole, and specific techniques, such as purified or recombinant antigens. Passive immunotherapy is divided into nonspecific techniques such as lymphokine-activated killer (LAK) cells and cytokine therapy, specific techniques such as antibody therapy, and combined techniques that include LAK cells and antibodies. Active immunotherapy focuses on stimulating immune response. BCG is an attenuated strain of the bacterium that causes bovine tuberculosis. BCG acts as a nonspecific stimulant of the immune system. A second method involves the use of vaccines made from the patient's own tumor (autologous) or from pooled tumor-associated antigens (allogeneic) that have been obtained from a number of tumors. Active immunotherapy has been studied as treatment for melanoma, renal cell carcinoma, and leukemia.32

Adoptive immunotherapy is a technique that uses lymphokine-activated NK cells or tumor-specific T-cell immunity as a means of eradicating cancer cells. Originally, only LAK cells were used. These NK cells are grown in culture supported by IL-2. Because NK cells are nonspecific in their function, LAK cells attack both normal and tumor cells. The technique of adoptive therapy has been expanded to the production of tumor-specific T cells. These cells are derived from a person's own *tumor-infiltrating lymphocytes* (TILs) that have been expanded in the laboratory so that a large amount of cells are available for reinfusion. Because the TILs are tumor specific, they do not attack normal host cells.

Four types of biologic response modifiers are being used or investigated: interferon therapy, interleukin therapy, monoclonal antibodies, and hematopoietic growth factors. Some agents, such as the interferons, have more than one biologic action, including antiviral, immunomodulatory, and antiproliferative actions. The *interferons* are endogenous polypeptides that are synthesized by a number of cells in response to a variety of cellular or viral stimuli. The three major types of interferons are alpha (α), beta (β) and gamma (γ), each group differing in terms of their cell surface receptors. The exact physiologic roles of each of the interferons remain unclear. They appear to inhibit viral replication and also may be involved in inhibiting tumor protein synthesis and in prolonging the cell cycle and increasing the percentage of cells in the G₀ phase. Interferons stimulate NK cells and T-lymphocyte killer cells. Although 17 interleukins have been identified, only one, IL-2, has been approved by the U.S. Food and Drug Administration (FDA) for the treatment of metastatic renal cell carcinoma. IL-2 has been found to reduce tumor size in a minority of patients with metastatic renal cancer and melanoma.32

Monoclonal antibodies (MoAbs) are highly specific antibodies derived from cloned cells or hybridomas. Scientists were able to produce large quantities of these MoAbs that were specific for tumor cells. Several MoAbs have been approved: muromonab-CD3 (OKT-3), which targets the CD3 receptor of human T cells, for the treatment of acute allograft rejection in renal transplant recipients; satumomab pendetide, used in the detection of colorectal and ovarian cancers,³³ and rituximab, an anti-CD20 MoAb used in the treatment of B-cell malignant lymphomas.³⁴

Hematopoietic growth factors are growth and maturation factors that include the colony-stimulating factors (CSFs). The CSFs are factors that control the production of neutrophils and monocytes/macrophages, erythropoietin, and thrombopoietin (see Chapter 11).³⁴

In summary, cancer compresses and erodes blood vessels; obstructs lymph flow; penetrates serous cavities; and compresses visceral structures. It produces chemical mediators, such as TNF, that produce pain, sap energy reserves, and cause weight loss and tissue wasting. Paraneoplastic syndromes arise from the ability of cancers to elaborate hormones and other chemical messengers that produce nonmetastatic endocrine, hematopoietic, neurologic, and dermatologic syndromes.

The methods used in the diagnosis of cancer vary with the type of cancer and its location. Because many cancers are curable if diagnosed early, health care practices designed to promote early detection are important. Pap smears, tissue biopsies, and tumor markers are used to detect the presence of cancer cells and in diagnosis. There are two basic methods of classifying tumors: (1) grading according to the histologic or tissue characteristics, and (2) clinical staging according to spread of the disease. Histologic studies are done in the laboratory using cells or tissue specimens. The TNM system for clinical staging of cancer uses tumor size, lymph node involvement, and presence of metastasis.

Treatment plans that use more than one type of therapy, often in combination, are providing cures for a number of cancers that a few decades ago had a poor prognosis and are increasing the life expectancy in other types of cancer. Surgical procedures are more precise as a result of improved diagnostic equipment and new techniques, such as laser surgery. Radiation equipment and radioactive sources permit greater and more controlled destruction of cancer cells while causing less damage to normal tissues. Chemotherapy involves the use of drugs that exert their effects at the cellular level to prevent cell replication. Successes with immunotherapy techniques offer hope that the body's own defenses can be used in fighting cancer.

CHILDHOOD CANCERS

In the United States, cancer is the second leading cause of death in children 1 to 14 years of age.¹ Between 1974 and 1991, children younger than 14 years of age exhibited a 1% average yearly increase in the incidence of all malignant neoplasms, with a 1.6% average increase in the incidence of acute lymphocytic leukemia and a greater than 2% increase for astroglial tumors, rhabdomyosarcomas, germ cell tumors, and

osteosarcomas.³⁵ The spectrum of cancers that affect children differs markedly from those that affect adults. Although most adult cancers are of epithelial cell origin (*e.g.*, lung cancer, breast cancer, colorectal cancers), childhood cancers usually involve the hematopoietic system, nervous system, or connective tissue. Chart 5-3 lists the most common forms of solid childhood cancers.

As with adult cancers, there probably is no one cause of childhood cancer. However, many forms of childhood cancer repeat in families and may result from polygenic or single-gene inheritance, chromosomal aberrations (*e.g.*, translocations, deletions, insertions, inversions, duplications), exposure to mutagenic environmental agents, or a combination of these factors (see Chapter 4). If cancer develops in one child, the risk of cancer in siblings is approximately twice that of the general population, and if the disease develops in two children, the risk is even greater.

Heritable forms of cancer tend to have an earlier age of onset, a higher frequency of multifocal lesions in a single organ, and bilateral involvement of paired organs or multiple primary tumors. The two-hit hypothesis has been used as one explanation of heritable cancers.³ The first "hit" or mutation occurs prezygotically (*i.e.*, in germ cells before conception) and is present in the genetic material of all somatic cells. Cancer subsequently develops in one or several somatic cell lines that undergo a second mutation.

Children with heritable disorders are at increased risk for developing certain forms of cancer. For example, Down syndrome is associated with increased risk of leukemia; primary immunodeficiency disorders (see Chapter 10) are associated with lymphoma, leukemia, and brain cancer; and xeroderma pigmentosum is associated with basal and squamous cell carcinoma and melanoma.

Diagnosis and Treatment

The early diagnosis of childhood cancers often is overlooked because the signs and symptoms often are similar to those of common childhood diseases and because cancer occurs less frequently in children than in adults.³⁶ Symptoms of prolonged fever, unexplained weight loss, and growing masses (especially in association with weight loss) should be viewed as warning signs of cancer in children. Diagnosis of childhood cancers involves many of the same methods that are used in adults. Accurate disease staging is especially beneficial in childhood cancers, in which the potential benefits of treatment must be carefully weighed against potential long-term effects.

CHART 5-3 Common Solid Tumors of Childhood

Brain and nervous system tumors Medulloblastoma Glioma Neuroblastoma Wilms' tumor Rhabdomyosarcoma and embryonal sarcoma Retinoblastoma Osteosarcoma Ewing's sarcoma

Adult Survivors of Childhood Cancer

With improvement in treatment methods, the number of children who survive childhood cancer is continuing to increase.³⁷ Unfortunately, therapy may produce late sequelae, such as impaired growth, neurologic dysfunction, hormonal dysfunction, cardiomyopathy, pulmonary fibrosis, and risk of second malignancies. Although cures for large numbers of children have been possible only since the 1970s, much already is known about the potential for delayed effects.

Children reaching adulthood after cancer therapy may have reduced physical stature because of the therapy they received, particularly radiation, which retards the growth of normal tissues along with cancer tissue. The younger the age and the higher the radiation dose, the greater the deviation from normal growth. There also is concern that central nervous system radiation as a prophylactic measure in childhood leukemia has an effect on cognition and learning. Children younger than 6 years of age at the time of radiation and those receiving the highest radiation doses are most likely to have subsequent cognitive difficulties.

Delayed sexual maturation in both boys and girls can result from irradiation of the gonads. Delayed sexual maturation also is related to the treatment of children with alkylating agents. Cranial irradiation may result in premature menarche in girls, with subsequent early closure of the epiphysis and a reduction in final growth achieved. Data related to fertility and health of the offspring of childhood cancer survivors are just becoming available.

Vital organs such as the heart and lungs may be affected by cancer treatment. Children who received anthracyclines (*i.e.*, doxorubicin or daunorubicin) may be at risk for cardiomyopathy and congestive heart failure developing. Pulmonary irradiation may cause lung dysfunction and restrictive lung disease. Drugs such as bleomycin, methotrexate, and busulfan also can cause lung disease.

For survivors of childhood cancers, the risk of second cancers is reported to range from 3% to 12%. There is a special risk of second cancers in children with the retinoblastoma gene. Because of this risk, children who have been treated for cancer should be followed up routinely.

In summary, although most adult cancers are of epithelial cell origin, most childhood cancers usually involve the hematopoietic system, nervous system, or connective tissue. Heritable forms of cancer tend to have an earlier age of onset, a higher frequency of multifocal lesions in a single organ, and bilateral involvement of paired organs or multiple primary tumors. The early diagnosis of childhood cancers often is overlooked because the signs and symptoms often are similar to those of other childhood diseases. With improvement in treatment methods, the number of children who survive childhood cancer is continuing to increase. As these children approach adulthood, there is continued concern that the life-saving therapy they received during childhood may produce late sequelae, such as impaired growth, neurologic dysfunction, hormonal dysfunction, cardiomyopathy, pulmonary fibrosis, and risk of second malignancies.

REVIEW QUESTIONS

Use the cell cycle to explain the difference in regenerative capabilities of cells in self-renewing tissues such as the skin, stable cells such as liver cells, and permanent cells such as neurons.

Define *neoplasm* and explain how neoplastic growth differs from the normal adaptive changes seen in atrophy, hypertrophy, and hyperplasia.

Relate the properties of cell differentiation to the development of a cancer cell line and the behavior of the tumor.

Trace the pathway for hematologic spread of a metastatic cancer cell and explain why some cancers preferentially metastasize to certain tissues in the body.

Use the concepts of growth fraction and doubling time to explain the number of cancer cells that would be present in a breast cancer lesion at the time it can be detected by breast self exam.

Describe the role of proto-oncogenes and anti-oncogenes in the transformation of a normal cell line to a cancer cell line.

- Compare methods used in histological grading of tumors and clinical staging of cancers.
- Cite the early warning signs of cancer in children.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- Jemal A., Murray T., Samuels A., et al. (2003). Cancer statistics, 2000. CA: A Cancer Journal for Clinicians 53, 5–26.
- 2. Li F.P. (1996). Hereditary cancer susceptibility. Cancer 78, 553-557.
- Kumar V., Cotran R.S., Robbins S.L. (Eds.). (2003). Neoplasia. In Basic pathology (7th ed., pp. 165–209). Philadelphia: W.B. Saunders.
- Lee W.M.F., Dang C.V. (2000). Control of cell growth and differentiation. In Hoffman R., Benz E.K., Shattil S.J., et al. (Eds.), *Hematology: Basic principles and practice* (3rd ed., pp. 57–71). New York: Churchill Livingstone.
- Buick R.N. (1994). Cellular basis of chemotherapy. In Dorr R.T., Von Hoff D.D. (Eds.), *Cancer chemotherapy handbook* (pp. 3–14). Norwalk, CT: Appleton & Lange.
- 6. Ruddon R.W. (Ed.). (1995). *Cancer biology* (pp. 3–60, 141–276). New York and Oxford: Oxford University Press.
- Fidler I.J. (1997). Molecular biology of cancer: Invasion and metastasis. In DeVita V.T. Jr., Hellman S., Rosenberg S.A. (Eds.), *Cancer: Principles and practice of oncology* (5th ed., pp. 135–148). Philadelphia: Lippincott-Raven.
- Liotta L.A. (1992). Cancer cell invasion and metastasis. Scientific American 266 (2), 54–63.
- Caudell K.A., Cuaron L.J., Gallucci, B.B. (1996). Cancer biology: Molecular and cellular aspects. In McCorkle R., Grant M., Frank-Stromborg M., et al. (Eds.), *Cancer nursing: A comprehensive textbook* (2nd ed., pp. 150–170). Philadelphia: W.B. Saunders.
- Levine A.J. (1996). Tumor suppressor genes. In Pusztai L., Lewis C.E., Yap E. (Eds.), Cell proliferation in cancer: Regulatory mechanisms of neoplastic cell growth (pp. 86–104). Oxford: Oxford University Press.
- Pusztai L., Cooper K. (1996). Introduction: Cell proliferation and carcinogenesis. In Pusztai L., Lewis C.E., Yap E. (Eds.), Cell proliferation in cancer: Regulatory mechanisms of neoplastic cell growth (pp. 3–24). Oxford: Oxford University Press.

- 12. Knudson A.G. (1974). Heredity and human cancer. American Journal of Pathology 77, 77–84.
- Rubin E., Farber J.L. (1999). Pathology (3rd ed., pp. 155–212, 434). Philadelphia: Lippincott Williams & Wilkins.
- 14. Stellman J.M., Stellman S.D. (1996). Cancer and the workplace. CA: A Cancer Journal for Clinicians 46, 70–92.
- Frank-Stromborg M., Heusinkveld K.B., Rohan K. (1996). Evaluating cancer risks and preventive oncology. In McCorkle R., Grant M., Frank-Stromborg M., et al. (Eds.), *Cancer nursing: A comprehensive textbook* (2nd ed., pp. 213–264). Philadelphia: W.B. Saunders.
- Bartecchi C.E., MacKenzie T.D., Schrier R. (1994). The human costs of tobacco use. New England Journal of Medicine 330, 907–912.
- 17. McMillan S. (1992). Carcinogenesis. Seminars in Oncology Nursing 8, 10–19.
- Poskanzer D.C., Herbst A. (1977). Epidemiology of vaginal adenosis and adenocarcinoma associated with exposure to stilbestrol in utero. *Cancer* 39, 1892–1895.
- Jablon S., Kato H. (1972). Studies of the mortality of A-bomb survivors: 5. Radiation dose and mortality, 1950–1970. *Radiation Research* 50, 649–698.
- Howley P.M. (1996). Viral carcinogenesis. In Pusztai L., Lewis C.E., Yap E. (Eds.), Cell proliferation in cancer: Regulatory mechanisms of neoplastic cell growth (pp. 38–58). Oxford: Oxford University Press.
- 21. Burnett F.M. (1967). Immunologic aspects of malignant disease. Lancet 1, 1171.
- Beverley P. (1996). Tumor immunology. In Roitt I., Brostoff J., Male D. (Eds.), *Immunology* (4th ed., pp. 20.1–20.8). London: Mosby.
- Greenberg P.D. (1997). Mechanisms of tumor immunology. In Stites D.P., Terr A.I., Parslow T.G. (Eds.), *Medical immunology* (9th ed., pp. 631–639). Stamford, CT: Appleton & Lange.
- 24. Beutler B. (1993). Cytokines and cancer cachexia. *Hospital Practice* 28(4), 45–52.
- Oppenheim J.J., Ruscetti F.W. (1997). Cytokines. In Stites D.P., Terr A.I., Parslow T.G. (Eds.), *Medical immunology* (9th ed., pp. 147–168). Stamford, CT: Appleton & Lange.
- 26. American Cancer Society. (1993). American Cancer Society facts and figures: 1993. Atlanta: Author.
- Weintraub F.N., Neumark D.E. (1996). Surgical oncology. In McCorkle R., Grant M., Frank-Stromborg M., et al. (Eds.), *Cancer nursing: A comprehensive textbook* (2nd ed., pp. 315–330). Philadelphia: W.B. Saunders.
- Collins M.C. (1990). Tumor markers and screening tools in cancer detection. Nursing Clinics of North America 25, 283–290.
- American Joint Committee on Cancer. (1997). AJCC cancer staging manual/American Joint Committee on Cancer (5th ed.). Philadelphia: Lippincott-Raven.
- Hilderley L.J., Dow K.H. (1996). Radiation oncology. In McCorkle R., Grant M., Frank-Stromborg M., et al. (Eds.), *Cancer nursing: A comprehensive textbook* (2nd ed., pp. 331–358). Philadelphia: W.B. Saunders.
- Guy J.L., Ingram B.A. (1996). Medical oncology: The agents. In McCorkle R., Grant M., Frank-Stromborg M., et al. (Eds.), *Cancer nursing: A comprehensive textbook* (2nd ed., pp. 359–394). Philadelphia: W.B. Saunders.
- Royal R.E., Steinberg S.M., Krause P.S., et al. (1996). Correlates of response to IL-2 therapy in patients treated for metastatic renal cancer and melanoma. *The Cancer Journal* 2 (2), 91–98.
- Farrell M.M. (1996). Biotherapy and the oncology nurse. Seminars in Oncology Nursing 12, 82–88.
- McLaughlin P., Grillo-Lopez A.J., Link B.K., et al. (1998). Rituximab chimeric anti-CD20 monoclonal antibody therapy for relapsed indolent lymphoma: Half of patients respond to a four-dose treatment program. *Journal of Clinical Oncology* 16, 2825–2833.
- Gurney J.G., Davis S., Severson P.K., et al. (1996). Trends in cancer incidence among children in the U.S. *Cancer* 78, 532–541.
- Crist W.M. (2000). Neoplastic diseases and tumors. In Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.), *Nelson textbook of pediatrics* (16th ed., pp. 1531–1543). Philadelphia: W.B. Saunders.
- Ward J.D. (2001). Pediatric cancer survivors. Nurse Practitioner 26(2), 18–37.

CHAPTER

Alterations in Fluids, Electrolytes, and Acid-Base Balance

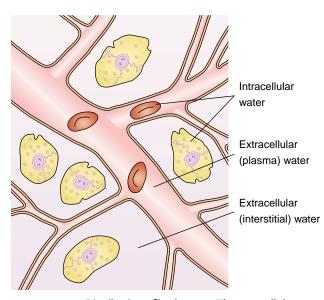
Composition and Compartmental Distribution of Body Fluids Introductory Concepts **Diffusion and Osmosis Compartmental Distribution of Body Fluids** Intracellular Fluid Volume **Extracellular Fluid Volume** Capillary/Interstitial Fluid Exchange Edema **Third-Space Accumulation Sodium and Water Balance Regulation of Sodium and Water Balance Regulation of Sodium Balance Regulation of Water Balance Alterations in Isotonic Fluid Volume** Isotonic Fluid Volume Deficit Isotonic Fluid Volume Excess **Alterations in Sodium Concentration** Hyponatremia Hypernatremia **Potassium Balance Regulation of Potassium Balance** Hypokalemia Hyperkalemia **Calcium and Magnesium Balance Calcium Balance Regulation of Serum Calcium** Hypocalcemia Hypercalcemia Magnesium Balance **Regulation of Magnesium** Hypomagnesemia Hypermagnesemia

Acid-Base Balance Introductory Concepts Acid-Base Chemistry Metabolic Acid and Bicarbonate Production Regulation of pH Laboratory Tests Alterations in Acid-Base Balance Metabolic Versus Respiratory Acid-Base Disorders Primary Versus Compensatory Mechanisms Metabolic Acidosis Metabolic Alkalosis Respiratory Alkalosis

Fluids and electrolytes are present in body cells, in the tissue spaces between the cells, and in the blood that fills the vascular compartment. Body fluids serve to transport gases, nutrients, and wastes; help to generate the electrical activity needed to power body functions; take part in the transforming of food into energy; and otherwise maintain the overall function of the body. Although fluid volume and composition remain relatively constant in the presence of a wide range of changes in intake and output, conditions such as environmental stresses and disease can increase fluid loss, impair its intake, and otherwise interfere with mechanisms that regulate fluid volume, composition, and distribution.

COMPOSITION AND COMPARTMENTAL DISTRIBUTION OF BODY FLUIDS

Body fluids are distributed between the intracellular fluid (ICF) and extracellular fluid (ECF) compartments. The ICF compartment consists of fluid contained within all of the billions of cells in the body. It is the larger of the two compartments, containing approximately two thirds of the body water in healthy adults. The remaining one third of body water is in the ECF compartment, which contains all the fluids outside the cells, including that in the interstitial or tissue spaces and blood vessels (Fig. 6-1). The ECF, including the plasma and interstitial fluids,



■ FIGURE 6-1 ■ Distribution of body water. The extracellular space includes the vascular compartment (plasma water) and the interstitial spaces.

contain large amounts of sodium and chloride, moderate amounts of bicarbonate, but only small quantities of potassium. In contrast to the ECF fluid, the ICF contains small amounts of sodium, chloride, and bicarbonate and large amounts of potassium (Table 6-1). It is the ECF levels of electrolytes in the blood or blood serum that are measured clinically. Although blood levels usually are representative of the total body levels of an electrolyte, this is not always the case, particularly with potassium, which is approximately 28 times more concentrated inside the cell than outside.

The cell membrane serves as the primary barrier to the movement of substances between the ECF and ICF compartments. Lipid-soluble substances such as gases (*i.e.*, oxygen and carbon dioxide), which dissolve in the lipid bilayer of the cell membrane, pass directly through the membrane. Many ions, such as sodium (Na⁺) and potassium (K⁺), rely on transport

TABLE 6-1Concentrations ofExtracellular and IntracellularElectrolytes in Adults		
Electrolyte	Extracellular Concentration*	Intracellular Concentration*
Sodium	135–145 mEq/L	10–14 mEq/L
Potassium	3.5–5.0 mEq/L	140–150 mEq/L
Chloride	98–106 mEq/L	3–4 mEq/L
Bicarbonate	24–31 mEq/L	7–10 mEq/L
Calcium	8.5–10.5 mg/dL	<1 mEq/L
Phosphate/ phosphorus	2.5–4.5 mg/dL	4 mEq/kg [†]
Magnesium	1.8–3.0 mg/dL	40 mEq/kg†

*Values may vary among laboratories, depending on the method of analysis used.

*Values vary among various tissues and with nutritional status.

mechanisms such as the Na⁺/K⁺-ATPase pump that is located in the cell membrane for movement across the membrane (see Chapter 1). Water crosses the cell membrane by osmosis using special protein channels.

Introductory Concepts

The electrolytes in body fluids are substances that dissociate in solution to form charged particles, or *ions*. For example, a sodium chloride (NaCl) molecule dissociates to form a positively charged Na⁺ ion and a negatively charged Cl⁻ ion. Because of their attraction forces, positively charged cations are always accompanied by negatively charged anions. The distribution of electrolytes between body compartments is influenced by their electrical charge. However, one cation may be exchanged for another, providing it carries the same charge. For example, a positively charged H⁺ ion may be exchanged for a positively charged K⁺ and a negatively charged HCO₃⁻ ion may be exchanged for a negatively charged Cl⁻ anion.

Diffusion and Osmosis

Diffusion is the movement of charged or uncharged particles along a concentration gradient. All molecules and ions, including water and dissolved molecules, are in constant random motion. It is the motion of these particles, each colliding with one another, that supplies the energy for diffusion. Because there are more molecules in constant motion in a concentrated solution, particles move from an area of higher concentration to one of lower concentration.

Osmosis is the movement of water across a semipermeable membrane (*i.e.*, one that is permeable to water but impermeable to most solutes). As with solute particles, water diffuses down its concentration gradient, moving from the side of the membrane with the lesser number of particles and greater concentration of water to the side with the greater number of particles and lesser concentration of water (Fig. 6-2). As water moves across the semipermeable membrane, it generates a pressure, called the *osmotic pressure*. The osmotic pressure represents the pressure (measured in millimeters of mercury [mm Hg]) needed to oppose the movement of water across the membrane.

The osmotic activity that nondiffusible particles exert in pulling water from one side of the semipermeable membrane to the other is measured by a unit called an *osmole*. The osmole is derived from the gram molecular weight of a substance (*i.e.*, 1 gram molecular weight of a nondiffusible and nonionizable substance is equal to 1 osmole). In the clinical setting, osmotic activity usually is expressed in milliosmoles (one thousandth of an osmole) per liter. Each nondiffusible particle, large or small, is equally effective in its ability to pull water through a semipermeable membrane. Thus, it is the number, rather than the size, of the nondiffusible particles that determines the osmotic activity of a solution.

The osmotic activity of a solution may be expressed in terms of either its osmolarity or osmolality. *Osmolarity* refers to the osmolar concentration in 1 L of solution (mOsm/L) and *osmolality* to the osmolar concentration in 1 kg of water (mOsm/kg of H_2O). Osmolarity is usually used when referring to fluids outside the body and osmolality for describing fluids inside the body. Because 1 L of water weighs 1 kg, the terms *osmolarity* and *osmolality* are often used interchangeably.

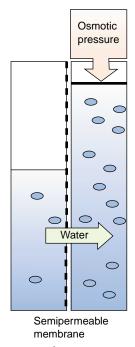


FIGURE 6-2 Movement of water across a semipermeable membrane. Water moves from the side that has fewer nondiffusible particles to the side that has more. The osmotic pressure is equal to the hydrostatic pressure needed to oppose water movement across the membrane.

Serum osmolality, which is largely determined by sodium and its attendant anions (chloride and bicarbonate), normally ranges between 275 and 295 mOsm/kg. Blood urea nitrogen (BUN) and glucose, which also are osmotically active, account for less than 5% of the total osmotic pressure in the ECF compartment. However, this can change, such as when blood glucose levels are elevated in persons with diabetes mellitus or when BUN levels change rapidly in persons with renal failure.

Tonicity. A change in water content causes cells to swell or shrink. The term *tonicity* refers to the tension or effect that the effective osmotic pressure of a solution with impermeable solutes exerts on cell size because of water movement across the cell membrane. An effective osmole is one that exerts an osmotic force and cannot permeate the cell membrane, whereas an ineffective osmole is one that exerts an osmotic force but crosses the cell membrane. Tonicity is determined solely by effective solutes such as glucose that cannot penetrate the cell membrane, thereby producing an osmotic force that pulls water into or out of the cell and causing it to change size.

Solutions to which body cells are exposed can be classified as isotonic, hypotonic, or hypertonic, depending on whether they cause cells to swell or shrink (Fig. 6-3). Cells placed in an isotonic solution, which has the same effective osmolality as the ICF (*i.e.*, 280 mOsm/L), neither shrink nor swell. An example of an isotonic solution is 0.9% sodium chloride. When cells are placed in a hypotonic solution, which has a lower effective osmolality than the ICF, they swell as water moves into the cell; when they are placed in a hypertonic solution, which has a greater effective osmolality than ICF, they shrink as water is pulled out of the cell.

Measurement Units

Laboratory measurements of electrolytes in body fluids are expressed as a concentration or amount of solute in a given volume of fluid, such as milligrams per deciliter (mg/dL), milliequivalents per liter (mEq/L), or millimoles per liter (mmol/L).

The use of *milligrams (mg) per deciliter* expresses the weight of the solute in one tenth of a liter (dL). The concentration of electrolytes, such as calcium, phosphate, and magnesium, is often expressed in mg/dL.

The *milliequivalent* is used to express the charge equivalency for a given weight of an electrolyte: 1 mEq of sodium has the same number of charges as 1 mEq of chloride, regardless of molecular weight. The number of milliequivalents of an electrolyte in a liter of solution can be derived from the following equation:

mEq =
$$\frac{\text{mg}/100 \text{ mL} \times 10 \times \text{valence}}{\text{atomic weight}}$$

The Système Internationale (SI) units express electrolyte concentration in *millimoles per liter* (mmol/L). A millimole is one thousandth of a mole, or the molecular weight of a substance expressed in milligrams. The number of millimoles of an electrolyte in a liter of solution can be calculated using the following equation:

$$mmol/L = \frac{mEq/L}{valence}$$

Compartmental Distribution of Body Fluids

Body water is distributed between the ICF and ECF compartments. In the adult, the fluid in the ICF compartment constitutes approximately 40% of body weight.¹ The fluid in the ECF compartment is further divided into two major subdivisions: the plasma compartment, which constitutes approximately 4%

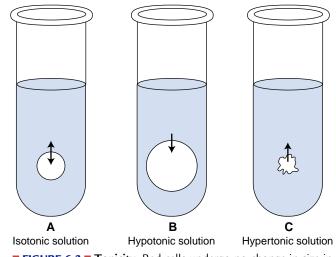


FIGURE 6-3 Tonicity. Red cells undergo no change in size in isotonic solutions (**A**). They increase in size in hypotonic solutions (**B**) and decrease in size in hypertonic solutions (**C**).

of body weight, and the interstitial fluid compartment, which constitutes approximately 15% of body weight (Fig. 6-4).

A third, usually minor, subdivision of the ECF compartment is the transcellular compartment. It includes the cerebrospinal fluid and fluid contained in the various body spaces, such as the peritoneal, pleural, and pericardial cavities; the joint spaces; and the gastrointestinal tract. Normally, only approximately 1% of ECF is in the transcellular space. This amount can increase considerably in conditions such as ascites, in which large amounts of fluid are sequestered in the peritoneal cavity. When the transcellular fluid compartment becomes considerably enlarged, it is referred to as a *third space*, because this fluid is not readily available for exchange with the rest of the ECF.

Intracellular Fluid Volume

The intracellular fluid volume is regulated by proteins and organic compounds in the ICF and by solutes that move between the ECF and ICF. The membrane in most cells is freely permeable to water; therefore, water moves between the ECF and ICF as a result of osmosis. In contrast, osmotically active proteins and other organic compounds cannot pass through the membrane. Water entry into the cell is regulated by these osmotically active substances as well as by solutes such as sodium and potassium that pass through the cell membrane. Many of the intracellular proteins are negatively charged and attract positively charged ions such as the K⁺ ion, accounting for its higher concentration in the ICF. The Na⁺ ion, which has a greater concentration in the ECF, tends to enter the cell by diffusion. The Na⁺ ion is osmotically active, and its entry would, if left unchecked, pull water into the cell until it ruptured. The reason this does not occur is because the Na⁺/K⁺-ATPase membrane pump continuously removes three Na⁺ ions from the cell for every two K⁺ ions that are moved back into the cell (see Chapter 1). Situations that impair the function of the Na⁺/K⁺-ATPase pump, such as hypoxia, cause cells to swell because of an accumulation of Na+ ions.

Intracellular volume is also affected by the concentration of osmotically active substances in the extracellular fluid that cannot cross the cell membrane. In diabetes mellitus, for example, glucose cannot enter the cell and its increased concentration in the ECF pulls water out of the cell.

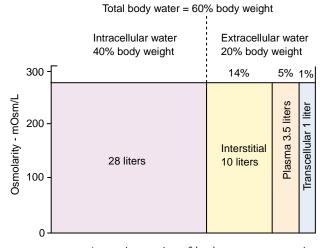


FIGURE 6-4 Approximate size of body compartments in a 70-kg adult.

Extracellular Fluid Volume

The ECF is divided between the vascular and interstitial fluid compartments. The vascular compartment contains blood, which is essential to the transport of substances such as electrolytes, gases, nutrients, and waste products throughout the body. The fluid in the interstitial compartment acts as a transport vehicle for gases, nutrients, wastes, and other materials that move between the vascular compartment and body cells. The interstitial fluid compartment also provides a reservoir from which vascular volume can be maintained during periods of hemorrhage or loss of vascular volume. A tissue gel, which is a spongelike material composed of large quantities of mucopolysaccharides, fills the tissue spaces and aids in even distribution of interstitial fluid. Normally, most of the fluid in the interstitium is in gel form. The tissue gel is supported by collagen fibers that hold the gel in place. The tissue gel, which has a firmer consistency than water, opposes the outflow of water from the capillaries and prevents the accumulation of free water in the interstitial spaces.

Capillary/Interstitial Fluid Exchange

The transfer of water between the vascular and interstitial compartments occurs at the capillary level. Four forces control the movement of water between the capillary and interstitial spaces: (1) the capillary filtration pressure, which pushes water out of the capillary into the interstitial spaces; (2) the capillary colloidal osmotic pressure, which pulls water back into the capillary; (3) the interstitial hydrostatic pressure, which opposes the movement of water out of the capillary; and (4) the tissue colloidal osmotic pressure, which pulls water out of the capillary into the interstitial spaces (Fig. 6-5). Normally, the combination of these four forces is such that only a small excess of fluid remains in the interstitial compartment. This excess fluid is removed from the interstitium by the lymphatic system and returned to the systemic circulation.

Capillary filtration refers to the movement of water through capillary pores because of a mechanical, rather than an osmotic, force. The capillary filtration pressure, sometimes called the *capillary hydrostatic pressure*, is the pressure pushing water out of the capillary into the interstitial spaces. It reflects the

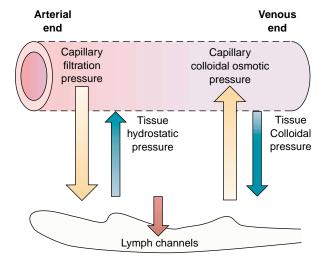


FIGURE 6-5 Exchange of fluid at the capillary level.

Unit One: Mechanisms of Disease

arterial and venous pressures, the precapillary (arterioles) and postcapillary (venules) resistances, and the force of gravity.² A rise in arterial or venous pressure increases capillary pressure. A decrease in arterial resistance or increase in venous resistance increases capillary pressure, and an increase in arterial resistance or decrease in venous resistance decreases capillary pressure. The force of gravity increases capillary pressure in the dependent parts of the body. In a person who is standing absolutely still, the weight of blood in the vascular column causes an increase of 1 mm Hg in pressure for every 13.6 mm of distance from the heart.² This pressure results from the weight of water and is therefore called *hydrostatic pressure*. In the adult who is standing absolutely still, the pressure in the veins of feet can reach 90 mm Hg. This pressure is then transmitted to the capillaries.

The *capillary colloidal osmotic pressure* is the osmotic pressure generated by the plasma proteins that are too large to pass through the pores of the capillary wall. The term *colloidal osmotic pressure* differentiates this type of osmotic pressure from the osmotic pressure that develops at the cell membrane from the presence of electrolytes and nonelectrolytes. Because plasma proteins do not normally penetrate the capillary pores and because their concentration is greater in the plasma than in the interstitial fluids, it is capillary colloidal osmotic pressure that pulls fluids back into the capillary.

The interstitial fluid pressure and the tissue colloidal osmotic pressure contribute to movement of water into and out of the interstitial spaces. The interstitial fluid pressure opposes the outward movement of water from the capillary into the interstitial spaces. The tissue colloidal osmotic pressure pulls water out of the capillary into the tissue spaces. It reflects the small amount of plasma proteins that normally escape from the capillary to enter the interstitial spaces.

Edema

Edema can be defined as palpable swelling produced by expansion of the interstitial fluid volume. Edema does not become evident until the interstitial fluid volume has been increased by 2.5 to 3 L.³

The physiologic mechanisms that contribute to edema formation include factors that: (1) increase the capillary filtration pressure, (2) decrease the capillary colloidal osmotic pressure, (3) increase capillary permeability, or (4) produce obstruction to lymph flow. The causes of edema are summarized in Chart 6-1.

Increased Capillary Filtration Pressure. As the capillary filtration pressure rises, the movement of vascular fluid into the interstitial spaces increases. Among the factors that increase capillary pressure are: (1) a decrease in the resistance to flow through the precapillary sphincters; (2) an increase in venous pressure or resistance to outflow at the postcapillary sphincters, and (3) capillary distention caused by increased vascular volume.

Edema can be either localized or generalized. The localized edema that occurs with urticaria (*i.e.*, hives) or other allergic or inflammatory conditions results from the release of histamine and other inflammatory mediators that cause dilation of the precapillary sphincters and arterioles that supply the swollen lesions. Thrombophlebitis obstructs venous flow, producing an elevation of venous pressure and edema of the affected part, usually one of the lower extremities.

CHART 6-1 Causes of Edema

Increased Capillary Pressure

Increased Capillary Pressure
Increased vascular volume
Heart failure
Kidney disease
Premenstrual sodium retention
Pregnancy Environmental heat stress
Venous obstruction
Liver disease with portal vein obstruction
Acute pulmonary edema
Venous thrombosis (thrombophlebitis)
Decreased arteriolar resistance
Calcium channel-blocking drug responses
Decreased Colloidal Osmotic Pressure
Increased loss of plasma proteins Protein-losing kidney diseases Extensive burns Decreased production of plasma proteins Liver disease
Starvation, malnutrition
Increased Capillary Permeability
Inflammation Allergic reactions (<i>e.g.,</i> hives, angioneurotic edema) Malignancy (<i>e.g.,</i> ascites, pleural effusion) Tissue injury and burns
Obstruction of Lymphatic Flow

Malignant obstruction of lymphatic structures Surgical removal of lymph nodes

Generalized edema is common in conditions such as congestive heart failure that produce fluid retention and venous congestion. In right-sided heart failure, blood dams up throughout the entire venous system, causing organ congestion and edema of the dependent extremities. Decreased sodium and water excretion by the kidneys leads to an increase in ECF volume with an increase in capillary volume and pressure with subsequent movement of fluid into the tissue spaces. The swelling of hands and feet that occurs in healthy persons during hot weather results from vasodilation of superficial blood vessels along with sodium and water retention.

Because of the effects of gravity, edema resulting from increased capillary pressure commonly causes fluid to accumulate in the dependent parts of the body, a condition referred to as *dependent edema*. For example, edema of the ankles and feet becomes more pronounced during prolonged periods of standing.

Decreased Capillary Colloidal Osmotic Pressure. Plasma proteins exert the osmotic force needed to pull fluid back into the capillary from the tissue spaces. The plasma proteins constitute a mixture of proteins, including albumin, globulins, and fibrinogen. Albumin, the smallest of the plasma proteins, has a molecular weight of 69,000; globulins have molecular weights of approximately 140,000; and fibrinogen has a molecular weight of 400,000.² Because of its lower molecular weight, 1 g of albumin has approximately twice as many osmotically active molecules as 1 g of globulin and almost six times as many osmotically active molecules as 1 g of fibrinogen. In addition, the concentration of albumin (approximately 4.5 g/dL) is greater than that of the globulins (2.5 g/dL) and fibrinogen (0.3 mg/dL).

Edema caused by decreased capillary colloidal osmotic pressure usually is the result of inadequate production or abnormal loss of plasma proteins, mainly albumin. The plasma proteins are synthesized in the liver. In persons with severe liver failure, impaired synthesis of albumin results in a decrease in colloidal osmotic pressure. In starvation and malnutrition, edema develops because there is a lack of the amino acids needed in plasma protein synthesis.

The most common site of plasma protein loss is the kidney. In kidney diseases such as nephrosis, the glomerular capillaries become permeable to the plasma proteins, particularly albumin, which is the smallest of the proteins. When this happens, large amounts of albumin are filtered out of the blood and lost in the urine. An excessive loss of plasma proteins also occurs when large areas of skin are injured or destroyed. Edema is a common problem during the early stages of a burn, resulting from capillary injury and loss of plasma proteins.

Because the plasma proteins are evenly distributed throughout the body and are not affected by the force of gravity, edema caused by decreased capillary colloidal osmotic pressure tends to affect tissues in nondependent as well as dependent parts of the body. There is swelling of the face as well as the legs and feet.

Increased Capillary Permeability. When the capillary pores become enlarged or the integrity of the capillary wall is damaged, capillary permeability is increased. When this happens, plasma proteins and other osmotically active particles leak into the interstitial spaces, increasing the tissue colloidal osmotic pressure and thereby contributing to the accumulation of interstitial fluid. Among the conditions that increase capillary permeability are burn injury, capillary congestion, inflammation, and immune responses.

Obstruction of Lymph Flow. Osmotically active plasma proteins and other large particles that cannot be reabsorbed through the pores in the capillary membrane rely on the lymphatic system for movement back into the circulatory system. Edema caused by impaired lymph flow is commonly referred to as *lymphedema*. Malignant involvement of lymph structures and removal of lymph nodes at the time of cancer surgery are common causes of lymphedema. Another cause of lymphedema is infection involving the lymphatic channels and lymph nodes.

Manifestations. The effects of edema are determined largely by its location. Edema of the brain, larynx, or lungs is an acute, life-threatening condition. Although not life threatening, edema may interfere with movement by limiting joint motion. Swelling of the ankles and feet often is insidious in onset and may or may not be associated with disease. At the tissue level, edema increases the distance for diffusion of oxygen, nutrients, and wastes. Edematous tissues usually are more susceptible to

injury and the development of ischemic tissue damage, including pressure ulcers. Edema can also compress blood vessels. For example, the skin of a severely swollen finger can act as a tourniquet, shutting off the blood flow to the finger. Edema can also be disfiguring, causing psychological effects and disturbances in self-concept.

Assessment and Treatment. Methods for assessing edema include daily weight, visual assessment, measurement of the affected part, and application of finger pressure to assess for pitting edema. Daily weight performed at the same time each day with the same amount of clothing provides a useful index of water gain (1 L of water weighs 2.2 pounds) attributable to edema. Visual inspection and measurement of the circumference of an extremity can also be used to assess the degree of swelling. This is particularly useful when swelling is caused by thrombophlebitis. Pitting edema occurs when the accumulation of interstitial fluid exceeds the absorptive capacity of the tissue gel. In this form of edema, the tissue water becomes mobile and can be translocated with pressure exerted by a finger. Finger pressure can be used to assess the degree of pitting edema. If an indentation remains after the finger has been removed, pitting edema is identified. It is evaluated on a scale of +1 (minimal) to +4 (severe) (Fig. 6-6).

Treatment of edema usually is directed toward maintaining life when the swelling involves vital structures, correcting or controlling the cause, and preventing tissue injury. Diuretic therapy commonly is used to treat edema. Edema of the lower extremities may respond to simple measures such as elevating the feet.

Elastic support stockings and sleeves increase interstitial fluid pressure and resistance to outward movement of fluid from the capillary into the tissue spaces. These support devices typically are prescribed for patients with conditions such as lymphatic or venous obstruction and are most efficient if applied before the tissue spaces have filled with fluid, such as in the morning, before the effects of gravity have caused fluid to move into the ankles.



■ FIGURE 6-6 ■ 3 + pitting edema of the left foot. (Used with permission from Bates B. [1999]. *Bates' guide to physical examination and history taking* [7th ed., p. 472]. Philadelphia: Lippincott Williams & Wilkins)

Third-Space Accumulation

Third spacing represents the loss or trapping of ECF in the transcellular space. The serous cavities are part of the transcellular compartment (i.e., third space) located in strategic body areas where there is continual movement of body structures-the pericardial sac, the peritoneal cavity, and the pleural cavity. The exchange of ECF among the capillaries, the interstitial spaces, and the transcellular space of the serous cavity uses the same mechanisms as capillaries elsewhere in the body. The serous cavities are closely linked with lymphatic drainage systems. The milking action of the moving structures, such as the lungs, continually forces fluid and plasma proteins back into the circulation, keeping these cavities empty. Any obstruction to lymph flow causes fluid accumulation in the serous cavities. As with edema fluid, third-space fluids represent an accumulation or trapping of body fluids that contribute to body weight but not to fluid reserve or function.

The prefix *hydro-* may be used to indicate the presence of excessive fluid, as in *hydrothorax*, which means excessive fluid in the pleural cavity. The accumulation of fluid in the peritoneal cavity is called *ascites*. The transudation of fluid into the serous cavities is also referred to as *effusion*. Effusion can contain blood, plasma proteins, inflammatory cells (*i.e.*, pus), and ECF.

In summary, body fluids are distributed between the ICF and ECF compartments of the body. Two thirds of body fluids are contained in the body cells of the ICF compartment, and one third is contained in the vascular compartment, interstitial spaces, and third-space areas of the ECF compartment. Electrolytes and nonelectrolytes move by diffusion across cell membranes that separate the ICF and ECF compartments. Water moves by osmosis across semipermeable membranes, moving from the side of the membrane that has the lesser number of particles and greater concentration of water to the side that has the greater number of particles and lesser concentration of water. The osmotic tension or effect that a solution exerts on cell volume in terms of causing the cell to swell or shrink is called *tonicity*.

Intracellular volume is regulated by the large numbers of proteins and other inorganic solutes that cannot cross the cell's membrane and solutes such as sodium, potassium, and glucose that selectively move between the ICF and ECF dependent upon concentration gradients and transport mechanisms. ECF volume, which is distributed between the vascular and interstitial compartments, is regulated by the elimination of sodium and water by the kidney.

Edema represents an increase in interstitial fluid volume. The physiologic mechanisms that predispose to edema formation are increased capillary filtration pressure, decreased capillary colloidal osmotic pressure, increased capillary permeability, and obstruction of lymphatic flow. The effect that edema exerts on body function is determined by its location; cerebral edema can be a life-threatening situation, but swollen feet can be a normal discomfort that accompanies hot weather. Fluid can also accumulate in the transcellular compartment—the joint spaces, pericardial sac, the peritoneal cavity, and the pleural cavity. Because this fluid is not easily exchanged with the rest of the ECF, it is often referred to as third-space fluid.

SODIUM AND WATER BALANCE

The movement of body fluids between the ICF and ECF compartments occurs at the cell membrane and depends on regulation of ECF water and sodium. Water provides approximately 90% to 93% of the volume of body fluids and sodium salts approximately 90% to 95% of the ECF solutes. Normally, equivalent changes in sodium and water are such that the volume and osmolality of the ECF is maintained within a normal range. Because it is the concentration of sodium (in milligrams per liter) that controls ECF osmolality, changes in sodium are usually accompanied by proportionate changes in water volume.

Alterations of sodium and water balance can be divided into two main categories: (1) isotonic contraction or expansion of ECF volume and (2) hypotonic dilution (hyponatremia) or hypertonic concentration (hypernatremia) of sodium brought about by changes in extracellular water (Fig. 6-7). Isotonic disorders usually are confined to the ECF compartment producing a contraction (fluid volume deficit) or expansion (fluid volume excess) of the interstitial and vascular fluids. Disorders of sodium concentration produce a change in the osmolality of the ECF with movement of water from the ECF compartment into the ICF compartment (hyponatremia) or from the ICF compartment into the ECF compartment (hypernatremia) (Fig. 6-8).

Regulation of Sodium and Water Balance

Regulation of Sodium Balance

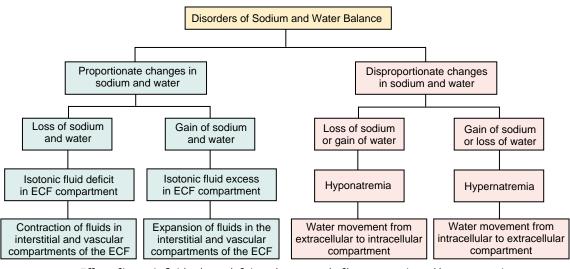
Sodium is the most abundant cation in the body, averaging approximately 60 mEq/kg of body weight.² Most of the body's sodium is in the ECF compartment (135 to 145 mEq/L), with only a small amount (10 to 14 mEq/L) located in the ICF compartment.

Sodium functions mainly in regulating extracellular fluid volume, including that in the vascular compartment. As the major cation in the ECF compartment, Na⁺ and its attendant anions (Cl⁻ and HCO₃⁻) account for most of the osmotic activity in the ECF. Because sodium is part of the sodium bicarbonate molecule, it is important in regulating acid-base balance. As a current-carrying ion, sodium contributes to the function of the nervous system and other excitable tissue.

Gains and Losses. Sodium normally enters the body through the gastrointestinal tract. Sodium intake normally is derived from dietary sources. Other sources of sodium are intravenous saline infusions and medications that contain sodium.

Sodium leaves the body through the kidney, gastrointestinal tract, and skin. Most sodium losses occur through the kidney. The kidneys are extremely efficient in regulating sodium output, and when sodium intake is limited or conservation of sodium is needed, the kidneys are able to reabsorb almost all the sodium that has been filtered by the glomerulus. This results in an essentially sodium-free urine. Conversely, urinary losses of sodium increase as intake increases.

Usually less than 10% of sodium intake is lost through the gastrointestinal tract and skin. Sodium losses increase with conditions such as vomiting, diarrhea, fistula drainage, and



■ FIGURE 6-7 ■ Effect of isotonic fluid volume deficit and excess and of hyponatremia and hypernatremia on ECF and ICF volume.

gastrointestinal suction that remove sodium from the upper gastrointestinal tract. Irrigation of gastrointestinal tubes with distilled water removes sodium from the gastrointestinal tract, as do repeated tap water enemas. Sweat losses, which usually are negligible, can increase greatly during exercise and periods of exposure to a hot environment. Loss of skin integrity, such as occurs in extensive burns, also leads to excessive skin losses of sodium. **Mechanisms of Sodium Regulation.** The kidney is the main regulator of sodium. The kidney monitors arterial pressure and retains sodium when the arterial pressure is decreased and eliminates it when the arterial pressure is increased. The rate at which the kidney excretes or conserves sodium is coordinated by the sympathetic nervous system and the renin-angiotensin-aldosterone system. The sympathetic nervous system responds to changes in arterial pressure and blood volume by adjusting

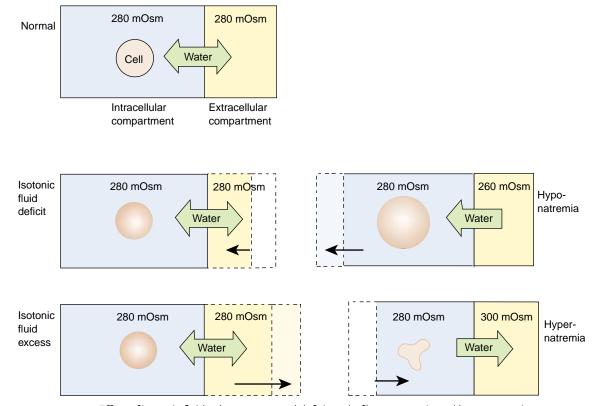


FIGURE 6-8 Effect of isotonic fluid volume excess and deficit and of hyponatremia and hypernatremia on extracellular and intracellular fluid volume.

the glomerular filtration rate and the rate at which sodium is filtered from the blood. Sympathetic activity also regulates tubular reabsorption of sodium and renin release. The reninangiotensin-aldosterone system exerts its action through angiotensin II and aldosterone (see Chapter 16). Angiotensin II acts directly on the renal tubules to increase sodium reabsorption. It also acts to constrict renal blood vessels, thereby decreasing the glomerular filtration rate and slowing renal blood flow so that less sodium is filtered and more is reabsorbed. Angiotensin II is also a powerful regulator of aldosterone, a hormone secreted by the adrenal cortex. Aldosterone acts at the level of the cortical collecting tubules of the kidneys to increase sodium reabsorption while increasing potassium elimination.

Regulation of Water Balance

Total body water (TBW) varies with gender and weight. These differences can be explained by differences in body fat, which is essentially water free. In men, body water approximates 60% of body weight during young adulthood and decreases to approximately 50% in old age. In young women, it is approximately 50% and in elderly women, approximately 40%.³ Obesity produces further decreases in body water, sometimes reducing these levels to values as low as 30% to 40% of body weight in adults (Fig. 6-9).

Infants have a high TBW content. TBW constitutes approximately 75% to 80% of body weight in full-term infants and is even greater in premature infants. In addition to having proportionately more body water than adults, infants have relatively more water in their ECF compartment. Infants have more than half of their TBW in the ECF compartment, whereas adults have only approximately a third.⁴ The greater extracellular water content of an infant can be explained in terms of its higher metabolic rate, larger surface area in relation to its body mass, and its inability to concentrate its urine because of immature

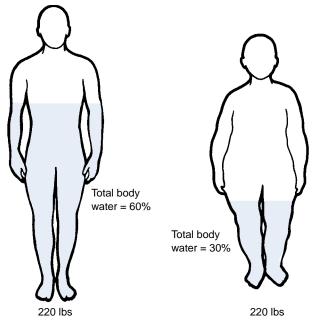


FIGURE 6-9 Body composition of a lean and an obese individual. (Adapted with permission from Statland H. [1963]. *Fluids and electrolytes in practice*. [3rd ed.]. Philadelphia: J.B. Lippincott)

kidney structures. Because ECF is more easily lost from the body, infants are more vulnerable to fluid deficit than are older children and adults. As an infant grows older, TBW decreases, and by the second year of life, the percentages and distribution of body water approach those of an adult.⁵

Gains and Losses. Regardless of age, all healthy persons require approximately 100 mL of water per 100 calories metabolized for dissolving and eliminating metabolic wastes. This means that a person who expends 1800 calories for energy requires approximately 1800 mL of water for metabolic purposes. The metabolic rate increases with fever; it rises approximately 12% for every 1°C (7% for every 1°F) increase in body temperature.² Fever also increases the respiratory rate, resulting in additional loss of water vapor through the lungs.

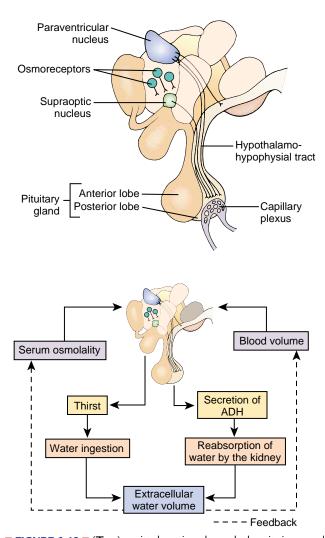
The main source of water gain is through oral intake and metabolism of nutrients. Water, including that obtained from liquids and solid foods, is absorbed from the gastrointestinal tract. Metabolic processes also generate a small amount of water. The amount of water gained from these processes varies from 150 to 300 mL/day, depending on metabolic rate.

Normally, the largest loss of water occurs through the kidneys, with lesser amounts being lost through the skin, lungs, and gastrointestinal tract. Even when oral or parenteral fluids are withheld, the kidneys continue to produce urine as a means of ridding the body of metabolic wastes. The urine output that is required to eliminate these wastes is called the *obligatory urine output*. The obligatory urine loss is approximately 300 to 500 mL/day. Water losses that occur through the skin and lungs are referred to as *insensible water losses* because they occur without a person's awareness. The gains and losses of body water are summarized in Table 6-2.

Mechanisms of Regulation. There are two main physiologic mechanisms that assist in regulating body water: thirst and antidiuretic hormone (ADH). Thirst is primarily a regulator of water intake and ADH a regulator of water output. Both mechanisms respond to changes in extracellular osmolality and volume (Fig. 6-10).

Thirst. Thirst is controlled by the thirst center in the hypothalamus. There are two stimuli for true thirst based on water need: (1) cellular dehydration caused by an increase in extracellular osmolality and (2) a decrease in blood volume, which may or may not be associated with a decrease in serum osmolality. Sensory neurons, called *osmoreceptors*, which are located in or

TABLE 6-2 Sources of Body Water Gains and Losses in the Adult			
Gains		Losses	
Oral intake		Urine	1500 mL
As water In food	1000 mL 1300 mL	Insensible losses Lungs	300 mL
Water of	200 mL	Skin	500 mL
oxidation		Feces	200 mL
Total	2500 mL	Total	2500 mL



■ FIGURE 6-10 ■ (Top) sagittal section through the pituitary and anterior hypothalamus. Antidiuretic hormone (ADH) is formed primarily in the supraoptic nucleus and to a lesser extent in the paraventricular nucleus of the hypothalamus. It is then transported down the hypothalamohypophysial tract and stored in secretory granules in the posterior pituitary, where it can be released into the blood. (**Bottom**) pathways for regulation of extracellular water volume by thirst and antidiuretic hormone.

near the thirst center in the hypothalamus, respond to changes in extracellular osmolality by stimulating the sensation of thirst (Fig. 6-10). Thirst normally develops when there is as little as a 1% to 2% change in serum osmolality.⁶ Stretch receptors in the vascular system that are sensitive to changes in arterial blood pressure and central blood volume also aid in the regulation of thirst. A third important stimulus for thirst is angiotensin II, which becomes increased in response to low blood volume and low blood pressure.

Dryness of the mouth produces a sensation of thirst that is not necessarily associated with the body's hydration status. Thirst sensation also occurs in those who breathe through their mouths, such as smokers and persons with chronic respiratory disease or hyperventilation syndrome.

Hypodipsia represents a decrease in the ability to sense thirst. There is evidence that thirst is decreased and water in-

take reduced in elderly persons, despite higher serum sodium and osmolality levels.^{7,8} The inability to perceive and respond to thirst is compounded in elderly persons who have had a stroke and may be further influenced by confusion and sensory disturbances.

Polydipsia, or excessive thirst, can be classified into three categories: (1) symptomatic or true thirst, (2) inappropriate or false thirst that occurs despite normal levels of body water and serum osmolality, and (3) compulsive water drinking. Symptomatic thirst develops when there is a loss of body water and resolves after the loss has been replaced. Among the most common causes of symptomatic thirst are water losses associated with diarrhea, vomiting, diabetes mellitus, and diabetes insipidus. Inappropriate or excessive thirst may persist despite adequate hydration. It is a common complaint in persons with renal failure and congestive heart failure. Although the cause of thirst in these persons is unclear, it may result from increased angiotensin levels. Thirst is also a common complaint in persons with dry mouth caused by decreased salivary function or treatment with drugs with an anticholinergic action (e.g., antihistamines, atropine) that lead to decreased salivary flow.

Psychogenic polydipsia involves compulsive water drinking and is usually seen in persons with psychiatric disorders, most commonly schizophrenia. Persons with the disorder drink large amounts of water and excrete large amounts of urine. The cause of excessive water drinking in these persons is uncertain. It has been suggested that the compulsive water drinking may share the same pathology as the psychosis because persons with the disorder often increase their water drinking during periods of exacerbation of their psychotic symptoms.⁹ The condition may be compounded by antipsychotic medications that increase ADH levels and interfere with water excretion by the kidneys. Cigarette smoking, which is common among persons with psychiatric disorders, also stimulates ADH secretion. Excessive water ingestion coupled with impaired water excretion (or rapid ingestion at a rate that exceeds renal excretion) in persons with psychogenic polydipsia can lead to water intoxication (see Hyponatremia). Treatment consists of water restriction and behavioral measures aimed at decreasing water consumption.10

Antidiuretic Hormone. The reabsorption of water by the kidneys is regulated by ADH, also known as vasopressin. ADH is synthesized by cells in the supraoptic and paraventricular nuclei of the hypothalamus; transported along a neural pathway (*i.e.*, hypothalamohypophysial tract) to the neurohypophysis (*i.e.*, posterior pituitary); and then released into the circulation (see Fig. 6-10).¹¹

As with thirst, ADH levels are controlled by extracellular volume and osmolality. Osmoreceptors in the hypothalamus sense changes in extracellular osmolality and stimulate the production and release of ADH. A small increase in serum osmolality of 1% is sufficient to cause ADH release. Likewise, stretch receptors (baroreceptors) that are sensitive to changes in blood pressure and central blood volume aid in the regulation of ADH release. A blood volume decrease of 5% to 10% produces a maximal increase in ADH levels.¹² As with many other homeostatic mechanisms, acute conditions produce greater changes in ADH levels than do chronic conditions; long-term changes in blood volume or blood pressure may exist without affecting ADH levels. An abnormal increase in ADH synthesis and release occurs in a number of stress situations. Severe pain, nausea, trauma, surgery, certain anesthetic agents, and some narcotics (*e.g.*, morphine and meperidine) increase ADH levels. Nausea is a potent stimulus of ADH secretion; it can increase ADH levels 10 to 1000 times those required for maximal diuresis.¹³ Among the drugs that affect ADH are nicotine, which stimulates its release, and alcohol, which inhibits it. Two important conditions alter ADH levels: diabetes insipidus and inappropriate secretion of ADH.

Diabetes insipidus (DI) is caused by a deficiency of or a decreased response to ADH. Persons with DI are unable to concentrate their urine during periods of water restriction; they excrete large volumes of urine, usually 3 to 20 L/day, depending on the degree of ADH deficiency or renal insensitivity to ADH. This large urine output is accompanied by excessive thirst. As long as the thirst mechanism is normal and fluid is readily available, there is little or no alteration in the fluid levels of persons with DI. The danger arises when the condition develops in someone who is unable to communicate the need for water or is unable to secure the needed water. In such cases, inadequate fluid intake rapidly leads to hypertonic dehydration and increased serum osmolality.

There are two types of DI: central or neurogenic and nephrogenic DI. Neurogenic DI occurs because of a defect in the synthesis or release of ADH and nephrogenic DI occurs because the kidneys do not respond to ADH.¹⁴ In neurogenic DI, loss of 75% to 80% of ADH-secretory neurons is necessary before polyuria becomes evident. Most persons with neurogenic DI have an incomplete form of the disorder and retain some ability to concentrate their urine. Temporary neurogenic DI may follow head injury or surgery near the hypophysial tract. Nephrogenic DI is characterized by impairment of the urineconcentrating ability of the kidney and free-water conservation. It may occur as a genetic trait that affects the ADH receptors in the kidney, as a side effect of drugs such as lithium,¹⁵ or as the result of electrolyte disorders such as potassium depletion or chronic hypercalcemia.

The syndrome of inappropriate ADH (SIADH) results from a failure of the negative feedback system that regulates the release and inhibition of ADH.¹⁶ In persons with this syndrome, ADH secretion continues even when serum osmolality is decreased, causing marked water retention and dilutional hyponatremia.

SIADH may occur as a transient condition, such as in a stress situation, or as a chronic condition, resulting from disorders such as a lung tumor. Stimuli, such as surgery, pain, stress, and temperature changes, are capable of stimulating ADH release through the central nervous system (CNS). Drugs induce SIADH in different ways; some drugs are thought to increase hypothalamic production and release of ADH, and others are believed to act directly on the renal tubules to enhance the action of ADH. More chronic forms of SIADH may be the result of lung tumors, chest lesions, and CNS disorders. Tumors, particularly bronchogenic carcinoma and cancers of the lymphoid tissues, prostate, and pancreas, are known to produce and release ADH independent of normal hypothalamic control mechanisms. Other intrathoracic conditions, such as advanced tuberculosis, severe pneumonia, and positive-pressure breathing, can also cause SIADH. The suggested mechanism for SIADH in positive-pressure ventilation is activation of baroreceptors (e.g., aortic baroreceptors, cardiopulmonary receptors) that respond to marked changes in intrathoracic pressure. Human immunodeficiency virus (HIV) infection is emerging as a new cause of SIADH. It has been reported that as many as 35% of persons with acquired immunodeficiency syndrome (AIDS) who are admitted to the acute care setting will have SIADH related to *Pneumocystis carinii* pneumonia, central nervous system infections, or malignancies.¹⁷

The manifestations of SIADH are those of dilutional hyponatremia. Urine output decreases despite adequate or increased fluid intake. Urine osmolality is high, and serum osmolality is low. Hematocrit and the serum sodium and BUN levels are all decreased because of the expansion of the ECF volume. The severity of symptoms usually is proportional to the extent of sodium depletion and water intoxication.

Alterations in Isotonic Fluid Volume

Isotonic fluid volume disorders represent an expansion or contraction of the ECF brought about by proportionate changes in both sodium and water.

Isotonic Fluid Volume Deficit

Fluid volume deficit is characterized by a decrease in ECF, including circulating blood volume. The term *isotonic fluid volume deficit* is used to differentiate fluid deficit in which there are proportionate losses in sodium and water from water deficit and the hyperosmolar state associated with hypernatremia. Unless other fluid and electrolyte imbalances are present, the concentration of serum electrolytes remains essentially unchanged. When the effective circulating blood volume is compromised, the condition is often referred to as *hypovolemia*.

Causes. Isotonic fluid volume deficit results when water and electrolytes are lost in isotonic proportions. It is almost always caused by a loss of body fluids and is often accompanied by a

KEY CONCEPTS

SODIUM AND WATER BALANCE

- It is the amount of water and its effect on sodium concentration in the ECF that serves to regulate the distribution of fluid between the ICF and the ECF compartments.
- Isotonic changes in body fluids that result from proportionate gains or losses of sodium and water are largely confined to the ECF compartment. Many of the manifestations of isotonic fluid deficit or excess reflect changes in vascular and interstitial fluid volume.
- Hyponatremia or hypernatremia brought about by disproportionate losses or gains in sodium or water exert their effects on the ICF compartment, causing water to move in or out of body cells. Many of the manifestations of changes in sodium concentration reflect changes in the intracellular volume of cells, particularly those in the nervous system.

decrease in fluid intake. It can occur because of a loss of gastrointestinal fluids, polyuria, or sweating caused by fever and exercise. Third-space losses cause sequestering of ECF in the serous cavities, extracellular spaces in injured tissues, or lumen of the gut.

In a single day, 8 to 10 L of ECF is secreted into the gastrointestinal tract. Most of it is reabsorbed in the ileum and proximal colon, and only about 150 to 200 mL per day is eliminated in the feces. Vomiting and diarrhea interrupt the reabsorption process and, in some situations, lead to increased secretion of fluid into the intestinal tract. Gastrointestinal suction, fistulas, and drainage tubes can remove large amounts of fluid from the gastrointestinal tract.

Excess sodium and water losses can also occur through the kidney. Certain forms of kidney disease are characterized by salt wasting caused by impaired sodium reabsorption. Fluid volume deficit also can result from osmotic diuresis or injudicious use of diuretic therapy. Glucose in the urine filtrate prevents reabsorption of water by the renal tubules, causing a loss of sodium and water. In Addison's disease, a condition of chronic adrenocortical insufficiency, there is unregulated loss of sodium in the urine with a resultant loss of ECF volume (see Chapter 31).

The skin acts as an exchange surface for heat and as a vapor barrier to prevent water from leaving the body. Body surface losses of sodium and water increase when there is excessive sweating or when large areas of skin have been damaged. Hot weather and fever increase sweating. The respiratory rate and sweating usually are increased as body temperature rises. Burns are another cause of excess fluid loss.

Manifestations. The manifestations of fluid volume deficit reflect a decrease in ECF volume. They include thirst, signs of water conservation by the kidney, loss of body weight, im-

paired temperature regulation, and signs of reduced interstitial and vascular volume (Table 6-3).

Thirst is a common symptom of fluid deficit, although it is not always present during early stages of isotonic fluid deficit. It develops as the effective circulating volume decreases to a point sufficient to stimulate the thirst mechanism. Urine output decreases and urine osmolality and specific gravity increase as ADH levels rise because of a decrease in vascular volume. Although there is an isotonic loss of sodium and water from the vascular compartment, other substances such as hematocrit and BUN become more concentrated.

A loss in fluid volume is accompanied by a decrease in body weight. One liter of water weighs 1 kg (2.2 lb). A mild ECF deficit exists when weight loss equals 2% of body weight.⁵ In a person who weighs 68 kg (150 lb.), this percentage of weight loss equals 1.4 L of water. A moderate deficit equates to a 5% loss in weight and a severe deficit to an 8% or greater loss in weight.⁵ To be accurate, weight must be measured at the same time each day with the person wearing the same amount of clothing. Because the fluid is trapped within the body in persons with third-space losses, their body weight may not decrease.

The fluid content of body tissues decreases as fluid is removed from the interstitial spaces. The eyes assume a sunken appearance and feel softer than normal as the fluid content in the anterior chamber of the eye is decreased. Fluids add resiliency to the skin and underlying tissues that is referred to as *skin* or *tissue turgor*.⁵ Pinching a fold of skin between the thumb and forefinger provides a means for assessing tissue turgor. The skin should immediately return to its original configuration when the fingers are released. A loss of 3% to 5% of body water in children causes the resiliency of the skin to be lost. Decreased tissue turgor is less predictive of fluid deficit in older persons (>65 years) because of the loss of tissue elasticity. In

TABLE 6-3 Manifestations of Isotonic Fluid Volume Deficit and Excess		
Fluid Volume Deficit	Fluid Volume Excess	
Acute Weight Loss (% body weight) Mild fluid volume deficit: 2% Moderate fluid volume deficit: 5% Severe fluid volume deficit: >8%	Acute Weight Gain (% body weight) Mild fluid volume excess: 2% Moderate fluid volume excess: 5% Severe fluid volume excess: >8%	
Signs of Compensatory Mechanisms Increased thirst Increased ADH: oliguria and high urine-specific gravity		
Decreased Interstitial Fluid Volume Decreased skin and tissue turgor Dry mucous membranes Sunken and soft eyeballs Depressed fontanel in infants	Increased Interstitial Fluid Volume Dependent and generalized edema	
Decreased Vascular Volume Postural hypotension Weak rapid pulse Decreased vein filling Hypotension and shock (severe deficit)	Increased Vascular Volume Full and bounding pulse Venous distention Pulmonary edema (severe excess) Shortness of breath Crackle Dyspnea Cough	

infants fluid deficit may be evidenced by depression of the anterior fontanel because of a decrease in cerebrospinal fluid.

Arterial and venous volumes decline during periods of fluid deficit, as does filling of the capillary circulation. As the volume in the arterial system declines, the blood pressure decreases, the heart rate increases, and the pulse becomes weak and thready. Postural hypotension (a drop in blood pressure upon standing) is an early sign of fluid deficit. On the venous side of the circulation, the veins become less prominent, and venous refill time increases. Body temperature may be subnormal because of decreased metabolism.⁵ When volume depletion becomes severe, signs of hypovolemic shock and vascular collapse appear (see Chapter 18).

Diagnosis and Treatment. Treatment of fluid volume deficit consists of fluid replacement and measures to correct the underlying cause. Usually, isotonic electrolyte solutions are used for fluid replacement. Acute hypovolemia and hypovolemic shock can cause renal damage; therefore, prompt assessment of the degree of fluid deficit and adequate measures to resolve the deficit and treat the underlying cause are essential.

Isotonic Fluid Volume Excess

Fluid volume excess represents an isotonic expansion of the ECF compartment with increases in both interstitial and vascular volumes. Although increased fluid volume is usually the result of a disease condition, this is not always true. For example, a compensatory isotonic expansion of body fluids can occur in healthy persons during hot weather as a mechanism for increasing body heat loss.

Causes. Isotonic fluid volume excess almost always results from an increase in total body sodium that is accompanied by a proportionate increase in body water. Although it can occur as the result of excessive sodium intake, it is most commonly caused by a decrease in sodium and water elimination by the kidney. Among the causes of decreased sodium and water elimination are disorders of renal function, heart failure, liver failure, and corticosteroid excess.

Heart failure produces a decrease in renal blood flow and a compensatory increase in sodium and water retention (Chapter 18). Persons with severe congestive heart failure maintain a precarious balance between sodium and water intake and output. Even small increases in sodium intake can precipitate a state of fluid volume excess and a worsening of heart failure. Liver failure impairs aldosterone metabolism and alters renal perfusion, leading to increased salt and water retention. Corticosteroid hormones increase sodium reabsorption by the kidney. Persons taking corticosteroid medications and those with Cushing's syndrome (see Chapter 31) often have problems with sodium retention.

Manifestations. Isotonic fluid volume excess is characterized by an increase in interstitial and vascular fluids. It is manifested by weight gain over a short period of time. A mild fluid volume excess represents a 2% weight gain; moderate fluid volume excess, a 5% weight gain; and severe fluid volume excess, a weight gain of 8% or more (Table 6-3).⁵ The presence of edema is characteristic of isotonic fluid volume excess. When the excess fluid accumulates gradually, as often happens in debilitating diseases and starvation, edema fluid may mask the loss of tissue

mass. As the vascular volume increases, the neck veins become distended, the pulse becomes full and bounding, and the central venous pressure becomes elevated. The BUN and hematocrit may be decreased as a result of the expanded plasma volume. When excess fluid accumulates in the lungs (*i.e.*, pulmonary edema), there is shortness of breath, complaints of difficult breathing, respiratory crackles, and a productive cough (see Chapter 18). Ascites and pleural effusion may occur with severe fluid volume excess.

Diagnosis and Treatment. The treatment of fluid volume excess focuses on providing a more favorable balance between sodium intake and output. A sodium-restricted diet is often prescribed. Diuretic therapy may be used to increase sodium elimination.

Alterations in Sodium Concentration

The normal serum sodium ranges from 135 to 145 mEq/L (135 to 145 mmol/L). Serum sodium values, expressed in mEq/L, reflect the concentration or dilution of sodium by water, rather than its absolute value. Because sodium and its attendant anions account for 90% to 95% of the osmolality of the ECF (normal range, 275 to 295 mOsm/kg), changes in serum sodium generally are accompanied by changes in serum osmolality.

Hyponatremia

Hyponatremia represents a serum sodium concentration below 135 mEq/L (135 mmol/L). Because of the effects of other osmotically active particles in the ECF, such as glucose, hyponatremia may be associated with high or low tonicity.¹⁷⁻¹⁹

Hypertonic (translocational) hyponatremia results from an osmotic shift of water from ICF to the ECF, such as occurs with hyperglycemia. In this case, the sodium in the ECF becomes diluted as water moves out of cells in response to the osmotic effects of the elevated blood glucose level.

Hypotonic (dilutional) is by far the most common form of hyponatremia. It is caused by water retention and characterized by a decrease in serum osmolality. Dilutional hyponatremia can present as a hypervolemic, euvolemic, or hypovolemic condition. Hypervolemic hyponatremia involves an increase in ECF volume and is seen when hyponatremic conditions are accompanied by edema-forming disorders such as congestive heart failure, cirrhosis, and advanced kidney disease.

Euvolemic hyponatremia represents a retention of water with dilution of sodium while maintaining the ECF volume within a normal range. It is usually the result of inappropriate thirst or SIADH. Hypovolemic hyponatremia occurs when water is lost along with sodium but to a lesser extent. It occurs with diuretic use, excessive sweating in hot weather, and vomiting and diarrhea.

Causes. The most common causes of acute dilutional hyponatremia in adults are drug therapy (diuretics and drugs that increase ADH levels), inappropriate fluid replacement during heat exposure or after heavy exercise, SIADH, and polydipsia in persons with psychotic disorder.

Among the causes of hypovolemic hyponatremia are excessive sweating in hot weather, particularly during heavy exercise, which leads to loss of salt and water; hyponatremia develops when water, rather than electrolyte-containing liquids, is used to replace fluids lost in sweating. Another potential cause of hypovolemic hyponatremia is the loss of sodium from the gastrointestinal tract caused by repeated tap water enemas or frequent gastrointestinal irrigations with distilled water. Isoosmotic fluid loss, such as occurs in vomiting or diarrhea, does not usually lower serum sodium levels unless these losses are replaced with disproportionate amounts of orally ingested or parenterally administered water. Gastrointestinal fluid loss and ingestion of excessively diluted formula are common causes of acute hyponatremia in infants and children.

Hypovolemic hyponatremia is a common complication of adrenal insufficiency and is attributable to the effects of aldosterone and cortisol deficiency (see Chapter 31). A lack of aldosterone increases renal losses of sodium, and a cortisol deficiency leads to increased release of ADH with water retention.

The risk of euvolemic hyponatremia is increased during the postoperative period. During this time ADH levels are often high, producing an increase in water reabsorption by the kidney (see SIADH). Although these elevated levels usually resolve in about 72 hours, they can persist for as long as 5 days. The hyponatremia becomes exaggerated when electrolyte-free fluids (*e.g.*, 5% glucose in water) are used for fluid replacement. *Manifestations.* The manifestations of hyponatremia will vary dependent upon the serum osmolality, the ECF fluid volume status, the rapidity of onset, and the severity of the sodium dilution (Table 6-4). The signs and symptoms may be acute, as in severe water intoxication, or more insidious in onset and less severe, as in chronic hyponatremia. Because of water movement, hypotonic hyponatremia causes intracellular hypoosmolality, which is responsible for many of the clinical manifestations of the disorder.¹⁹ Fingerprint edema is a sign of excess intracellular water. This phenomenon is demonstrated by pressing the finger firmly over the bony surface of the sternum for 15 to 30 seconds. When excess intracellular water is present, a fingerprint similar to that observed when pressing on a piece of modeling clay is seen.

Muscle cramps, weakness, and fatigue reflect the hypoosmolality of skeletal muscle cells and are often early signs of hyponatremia. These effects commonly are observed in persons with hyponatremia that occurs during heavy exercise in hot weather. Gastrointestinal manifestations such as nausea and vomiting, abdominal cramps, and diarrhea may develop.

The brain and nervous system are the most seriously affected by pronounced increases in intracellular water. Symptoms include apathy, lethargy, and headache, which can progress to disorientation, confusion, gross motor weakness, depression of

TABLE 6-4 Manifestations of Hyponatremia and Hypernatremia		
Hyponatremia (Hypotonic)	Hypernatremia	
Laboratory Values Serum sodium <135 mEq/L Decreased serum osmolality Dilutional decrease in blood components, including hematocrit, blood urea nitrogen	Laboratory Values Serum sodium >145 mEq/L Increased serum osmolality Increase in blood components, including hematocrit, blood urea nitrogen Compensatory Mechanisms Increased thirst Increased ADH with oliguria and high urine-specific gravity	
Increased Intracellular Fluid Fingerprint edema	Decreased Intracellular Fluid Dry skin and mucous membranes Decreased tissue turgor Decreased salivation and lacrimation	
Hypo-osmolality and Movement of Water Into Muscle, Neural, and Gastrointestinal Tract Tissue Muscle Muscle cramps and weakness Central Nervous System Headache Apprehension, feeling of impending doom Personality changes Lethargy Stupor and coma (severe) Gastrointestinal Tract Anorexia, nausea, vomiting Abdominal cramps, diarrhea	Hyperosmolality and Movement of Water Out of Neural Tissue Headache Disorientation and agitation Decreased reflexes Seizures and coma (severe)	
	Decreased Vascular Volume Weak rapid pulse Possible impaired temperature regulation with fever Decreased blood pressure Vascular collapse (severe)	

deep tendon reflexes. Seizures and coma occur when serum sodium levels reach extremely low levels. These severe effects, which are caused by brain swelling, may be irreversible.¹⁹ If the condition develops slowly, signs and symptoms do not develop until serum sodium levels approach 125 mEq/L. The term "water intoxication" is often used to describe the neurologic effects of acute hypotonic hyponatremia.

Treatment. The treatment of hyponatremia with water excess focuses on the underlying cause. When hyponatremia is caused by water intoxication, limiting water intake or discontinuing medications that contribute to SIADH may be sufficient. The administration of saline solution orally or intravenously may be needed in severe hyponatremia caused by sodium deficiency.

Hypernatremia

Hypernatremia implies a serum sodium level above 145 mEq/L and a serum osmolality greater than 295 mOsm/kg. Because sodium is functionally an impermeable solute, it contributes to the tonicity and movement of water across cell membranes. Hypernatremia is characterized by hypertonicity of the ECF and almost always causes cellular dehydration.²⁰

Causes. Hypernatremia represents a deficit of water in relation to the body's sodium levels. It can be caused by net gain of sodium or net loss of water. Rapid ingestion or infusion of sodium with insufficient time or opportunity for water ingestion can produce a disproportionate gain in sodium. A defect in thirst or inability to obtain or drink water can interfere with water replacement.

Hypernatremia also occurs when there is an excess loss of body fluids that have a lower than normal concentration of sodium so that water is lost in excess of sodium. This can result from increased losses from the respiratory tract during fever or strenuous exercise, from watery diarrhea, or when osmotically active tube feedings are given with inadequate amounts of water. With pure water loss, each body fluid compartment loses an equal percentage of its volume. Because approximately one third of the water is in the ECF compartment, compared with the two thirds in the ICF compartment, more actual water volume is lost from the ICF than the ECF compartment.⁵

Normally, water deficit stimulates thirst and increases water intake. Therefore, hypernatremia is more likely to occur in infants and in persons who cannot express their thirst or obtain water to drink. With hypodipsia, or impaired thirst, the need for fluid intake does not activate the thirst response. Hypodipsia is particularly prevalent among the elderly. In persons with diabetes insipidus, hypernatremia can develop when thirst is impaired or access to water is impeded.

Manifestations. The clinical manifestations of hypernatremia caused by water loss are largely those of ECF fluid loss and cellular dehydration (Table 6-4). The severity of signs and symptoms is greatest when the increase in serum sodium is large and occurs rapidly. Body weight is decreased in proportion to the amount of water that has been lost. Because blood plasma is roughly 90% to 93% water, the concentrations of blood cells, hematocrit, BUN, and other solutes increase as ECF water decreases.

Thirst is an early symptom of water deficit, occurring when water losses are equal to 0.5% of body water. Urine output is decreased and urine osmolality increased because of renal water-conserving mechanisms. Body temperature frequently is elevated, and the skin becomes warm and flushed. As the vascular volume decreases, the pulse becomes rapid and thready, and the blood pressure drops. Hypernatremia produces an increase in serum osmolality and results in water being pulled out of body cells. As a result, the skin and mucous membranes become dry, and salivation and tearing of the eyes are decreased. The mouth becomes dry and sticky, and the tongue becomes rough and fissured. Swallowing is difficult. The subcutaneous tissues assume a firm, rubbery texture. Most significantly, water is pulled out of the cells in the CNS, causing decreased reflexes, agitation, headache, and restlessness. Coma and seizures may develop as hypernatremia progresses.

Treatment. The treatment of hypernatremia includes measures to treat the underlying cause of the disorder and fluid replacement therapy to treat the accompanying dehydration. Replacement fluids can be given orally or intravenously. The oral route is preferable. Oral glucose–electrolyte replacement solutions are available for the treatment of infants with diarrhea.²¹ Until recently, these solutions were used only early in diarrheal illness or as a first step in re-establishing oral intake after parenteral replacement therapy. These solutions are now widely available in grocery stores and pharmacies for use in the treatment of diarrhea and other dehydrating disorders in infants and young children.

In summary, body fluids are distributed between the ICF and ECF compartments. Regulation of fluid volume, solute concentration, and distribution between the two compartments depends on water and sodium balance. Water provides approximately 90% to 93% of fluid volume and sodium salts, approximately 90% to 95% of extracellular solutes. Body water is regulated by thirst, which controls water intake, and ADH, which controls urine concentration and renal output. Sodium is ingested in the diet and eliminated by the kidneys under the influence of the sympathetic nervous system and the renin-angiotensin-aldosterone system.

Isotonic fluid disorders result from contraction or expansion of ECF volume brought about by proportionate losses of sodium and water. *Isotonic fluid volume deficit* is characterized by a decrease in ECF volume. It causes thirst, decreased vascular volume and circulatory function, decreased urine output, and increased urine specific gravity. *Isotonic fluid volume excess* is characterized by an increase in ECF volume. It is manifested by signs of increased vascular volume and edema.

Alterations in extracellular sodium concentration are brought about by a disproportionate gain (hyponatremia) or loss (hypernatremia) of water. As the major cation in the ECF compartment, sodium controls the ECF osmolality and its effect on cell volume. Hypotonic hyponatremia is characterized by water being pulled into the cell from the extracellular compartment, causing cells to swell. It is manifested by muscle cramps and weakness; nausea, vomiting, abdominal cramps, and diarrhea; and CNS signs such as lethargy, headache, depression of deep tendon reflexes, and in severe cases seizure and coma. Hypernatremia is characterized by intracellular water being pulled into the extracellular compartment, causing cells to shrink. It is manifested by thirst and decreased urine output; dry mouth and decreased tissue turgor; signs of decreased vascular volume (tachycardia, weak and thready pulse); and CNS signs, such as decreased reflexes, agitation, headache, and in severe cases seizures and coma.

POTASSIUM BALANCE

Potassium is the second most abundant cation in the body and the major cation in the ICF compartment. Approximately 98% of body potassium is contained within body cells, with an intracellular concentration of 140 to 150 mEq/L.²² The potassium content of ECF (3.5 to 5.0 mEq/L) is considerably less. Because potassium is an intracellular ion, total body stores of potassium are related to body size and muscle mass. Approximately 65% to 75% of potassium is in muscle.²³ Thus, total body potassium declines with age, mainly as a result of a decrease in muscle mass.

As the major intracellular cation, potassium is critical to many body functions. It is involved in a wide range of body functions, including the maintenance of the osmotic integrity of cells, acid-base balance, and the kidney's ability to concentrate urine. Potassium is necessary for growth, and it contributes to the intricate chemical reactions that transform carbohydrates into energy, change glucose into glycogen, and convert amino acids to proteins.

Potassium also plays a critical role in conducting nerve impulses and the excitability of skeletal, cardiac, and smooth muscle (see Chapter 1). It does this by regulating: (1) the rest-

KEY CONCEPTS

POTASSIUM BALANCE

- Most of the body's potassium is contained in the ICF, with only a small, but vital, amount being present in the ECF.
- Two major mechanisms function in the control of serum potassium: (1) renal mechanisms that conserve or eliminate potassium, and (2) transcellular buffer systems that remove potassium from and release it into the serum as needed.
- The distribution of potassium between the ICF and ECF compartments regulates electrical membrane potentials controlling the excitability of nerve and muscle cells as well as contractility of skeletal, cardiac, and smooth muscle tissue. Many of the manifestations of potassium deficit or excess are related to potassium's effect on membrane potentials in cardiac, skeletal, and smooth muscle in the gastrointestinal tract.

ing membrane potential, (2) the opening of the sodium channels that control the flow of current during the action potential, and (3) the rate of repolarization. Changes in nerve and muscle excitability are particularly important in the heart, where alterations in serum potassium can produce serious dysrhythmias and conduction defects. Changes in serum potassium also affect skeletal muscles and the smooth muscle in blood vessels and the gastrointestinal tract.

The resting membrane potential is determined by the ratio of intracellular to extracellular potassium. A decrease in serum potassium causes the resting membrane potential to become more negative (hyperpolarization), moving further from the threshold for excitation (Fig. 6-11). Thus, it takes a greater stimulus to reach threshold and open the sodium channels that are responsible for the action potential. An increase in serum potassium has the opposite effect; it causes the resting membrane potential to become more positive (hypopolarized), moving closer to threshold. This produces an initial increase in excitability. Activation and opening of the sodium channels that control the flow of current during an action potential are also affected by potassium levels. With severe hyperkalemia, the sodium channels become inactivated, producing a net decrease in excitability. The rate of repolarization also varies with serum potassium levels.¹¹ It is more rapid in hyperkalemia and delayed in hypokalemia. The rate of repolarization is important clinically because it predisposes to conduction defects and dysrhythmias in the heart.

Regulation of Potassium Balance

Potassium intake is normally derived from dietary sources. In healthy persons, potassium balance usually can be maintained by a daily dietary intake of 50 to 100 mEq. Additional amounts of potassium are needed during periods of trauma and stress. The kidneys are the main source of potassium loss. Approximately 80% to 90% of potassium losses occur in the urine, with the remainder being lost in stools or sweat.

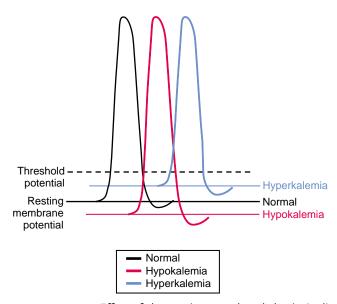


FIGURE 6-11 Effect of changes in serum hypokalemia (red) and hyperkalemia (blue) on the resting membrane potential.

Mechanisms of Regulation. Normally, the ECF concentration of potassium is precisely regulated at about 4.2 mEq/ml. The precise control is necessary because many cell functions are sensitive to even small changes in ECF potassium levels. An increase in serum potassium levels of only maybe 0.3 to 0.4 mEq/L can cause serious cardiac dysrhythmias and even death.

Serum potassium levels are largely regulated through two mechanisms: (1) renal mechanisms that conserve or eliminate potassium, and (2) a transcellular shift of potassium between the ICF and ECF compartments. Normally, it takes 6 to 8 hours to eliminate 50% of potassium intake.³ To avoid an increase in extracellular potassium levels during this time, excess potassium is temporarily shifted into red blood cells and other cells such as those of muscle, liver, and bone.

Renal Regulation. The kidney provides the major route for potassium elimination. Potassium is filtered in the glomerulus, reabsorbed along with sodium and water in the proximal tubule and with sodium and chloride in the thick ascending loop of Henle, and then secreted into the late distal and cortical collecting tubules for elimination in the urine. The latter mechanism serves to "fine-tune" the concentration of potassium in the ECF.

Aldosterone plays an essential role in regulating potassium elimination by the kidney (see Chapter 22). In the presence of aldosterone, sodium is transported back into the blood and potassium is secreted into the tubular filtrate for elimination in the urine. There is also a potassium-hydrogen exchange system in the collecting tubules of the kidney. When serum potassium levels are increased, potassium is secreted into the urine and hydrogen is reabsorbed into the blood, producing a decrease in pH and metabolic acidosis. Conversely, when potassium levels are low, potassium is reabsorbed and hydrogen is secreted into the urine, leading to metabolic alkalosis.

Extracellular-Intracellular Shifts. The movement of potassium between the ECF and ICF allows potassium to move into body cells when serum levels are high and move out when serum levels are low. Among the factors that alter potassium distribution between the ECF and ICF are insulin, β -adrenergic stimulation, serum osmolality, and acid-base disorders. Both insulin and the β -adrenergic catecholamines (*e.g.*, epinephrine) increase cellular uptake of potassium. Insulin increases the cellular uptake of potassium after a meal. The potassium content of a single meal is often as high as 50 mEq; the actions of insulin prevent a rise in serum potassium to life-threatening levels. The catecholamines, particularly epinephrine, facilitate the movement of potassium into muscle tissue during periods of physiologic stress.

Extracellular osmolality and pH also influence the movement of potassium between the ICF and ECF. Acute increases in serum osmolality cause potassium to move out of cells. When serum osmolality is increased because of the presence of impermeable solutes such as glucose (without insulin), water leaves the cell. The loss of cell water produces an increase in intracellular potassium, causing it to move out of the cell into the ECF. Acid-base disorders are often accompanied by a change in serum potassium. Both the hydrogen and potassium ions are positively charged, and both ions move freely between the ICF and ECF compartments. In metabolic acidosis, hydrogen ions move into body cells for buffering; this causes potassium to leave the cells and move into the ECF. Metabolic alkalosis has the opposite effect.

Exercise can also produce compartmental shifts in potassium. Repeated muscle contraction releases potassium into the ECF. Although the increase usually is small with modest exercise, it can be considerable during exhaustive exercise. Even the repeated clenching and unclenching of the fist during a blood draw can cause potassium to move out of cells and artificially elevate serum potassium levels.

Hypokalemia

Hypokalemia refers to a serum potassium level below 3.5 mEq/L (3.5 mmol/L). Because of transcellular shifts, temporary changes in serum K⁺ may occur as the result of movement between the ICF and ECF compartments.

Causes. The causes of potassium deficit can be grouped into three categories: (1) inadequate intake; (2) excessive losses through the kidney, skin, and gastrointestinal tract; and (3) redistribution between the ICF and ECF compartments.

Inadequate intake is a frequent cause of hypokalemia. A potassium intake of at least 10 to 30 mEq/day is needed to compensate for obligatory urine losses.^{4,23} A person on a potassium-free diet continues to lose approximately 5 to 15 mEq of potassium daily. Insufficient dietary intake may result from the inability to obtain or ingest food or from a diet that is low in potassium-containing foods.

Excessive renal losses can occur with diuretic use, metabolic alkalosis, magnesium depletion, trauma and stress, and increased levels of aldosterone. Diuretic therapy, with the exception of potassium-sparing diuretics, is the most common cause of hypokalemia. Both thiazide and loop diuretics increase the loss of potassium in the urine. The degree of hypokalemia is directly related to diuretic dose and is greater when sodium intake is higher.²³ Magnesium depletion causes renal potassium wasting. Magnesium deficiency often coexists with potassium depletion because of diuretic therapy or disease processes such as diarrhea. Importantly, the ability to correct potassium deficiency is impaired when magnesium deficiency is present.

The kidneys do not have the homeostatic mechanisms needed to conserve potassium during periods of stress or insufficient intake. After trauma and in stress situations, urinary losses of potassium are greatly increased, sometimes approaching levels of 150 to 200 mEq/L.²⁴ Renal losses of potassium are accentuated by aldosterone and cortisol. Trauma and surgery produce a stress-related increase in these hormones. Primary aldosteronism, caused by an aldosterone-secreting tumor of the adrenal cortex, can produce severe urinary losses of potassium. Cortisol binds to aldosterone receptors and exerts aldosterone-like effects on potassium elimination.

Although potassium losses from the gastrointestinal tract and the skin usually are minimal, these losses can become excessive under certain conditions. The gastrointestinal tract is one of the most common sites for acute potassium losses. Vomiting and gastrointestinal suction lead to hypokalemia, partly because of actual potassium losses and because of renal losses associated with metabolic alkalosis (see section on metabolic alkalosis under acid-base balance). Because intestinal secretions are high in potassium content, diarrhea and intestinal suction can also cause large losses of potassium. Excessive skin losses occur with loss of the protective skin surface and sweating. Burns and other types of skin injury increase the loss of potassium through wound drainage. Losses caused by sweating increase in persons who are acclimating to a hot climate, partly because increased secretion of aldosterone during heat acclimatization increases the loss of potassium in urine and sweat.

Because of the high ratio of intracellular to extracellular potassium, a redistribution of potassium from the ECF to the ICF compartment can produce a marked decrease in the serum concentration. One cause of potassium redistribution is insulin. Because insulin increases the movement of glucose and potassium into cells, potassium deficit often develops during treatment of diabetic ketoacidosis. β -adrenergic agonist drugs, such as pseudoephedrine and albuterol, can have a similar effect on potassium distribution.

Manifestations. The manifestations of hypokalemia include the effect of altered membrane potentials on cardiovascular, neuromuscular and gastrointestinal function (see Table 6-5). The signs and symptoms of potassium deficit seldom develop until the serum potassium level has fallen to less than 3.0 mEq/L. They are typically gradual in onset, so the disorder may go undetected for some time.

The most serious effects of hypokalemia are those affecting cardiovascular function. Postural hypotension is common. Most persons with a serum potassium level of less than 3.0 mEq/L demonstrate electrocardiographic (ECG) changes typical of hypokalemia. These changes include prolongation of the PR interval, depression of the ST segment, flattening of the T wave, and appearance of a prominent U wave (Fig. 6-12). Although these ECG changes usually are not serious, they may predispose to sinus bradycardia and ectopic ventricular dysrhythmias (see Chapter 17). Digitalis toxicity can be provoked in persons treated with this drug, and there is an increased risk of ventricular dysrhythmias, particularly in persons with underlying heart disease. Potassium and digitalis compounds compete for binding to sites on the Na⁺/K⁺-ATPase membrane pump. In hypokalemia, more sites are available for digitalis to bind to and exert its action. The dangers associated with digitalis toxicity are compounded in persons who are receiving diuretics that increase urinary losses of potassium.

Complaints of weakness, fatigue, and muscle cramps, particularly during exercise, are common in moderate hypokalemia (serum potassium 3.0 to 2.5 mEq/L). Muscle paralysis with life-threatening respiratory insufficiency can occur with severe hypokalemia (serum potassium <2.5 mEq/L). Leg muscles, particularly the quadriceps, are most prominently affected. Some persons report muscle tenderness and paresthesias, rather than weakness. In chronic potassium deficiency, muscle atrophy may contribute to muscle weakness.

There are numerous signs and symptoms associated with altered gastrointestinal function, including anorexia, nausea, and vomiting. Atony of the gastrointestinal smooth muscle can cause constipation, abdominal distention, and, in severe hypokalemia, paralytic ileus. When gastrointestinal symptoms occur

TABLE 6-5 Manifestations of Hypokalemia and Hyperkalemia		
Hypokalemia	Hyperkalemia	
Laboratory Values	Laboratory Values	
Serum potassium <3.5 mEq/L	Serum potassium >5.0 mEq/L	
Thirst and Urine		
Increased thirst		
Inability to concentrate urine with polyuria and		
urine with low specific gravity		
Effects of Changes in Membrane Potentials on Neural	Effects of Changes in Membrane Potentials on Neural	
and Muscle Function	and Muscle Function	
Gastrointestinal	Gastrointestinal	
Anorexia, nausea, vomiting	Nausea, vomiting	
Abdominal distention	Intestinal cramps	
Paralytic ileus (severe hypokalemia)	Diarrhea	
Neuromuscular	Neuromuscular	
Muscle weakness, flabbiness, fatigue	Weakness, dizziness	
Muscle cramps and tenderness	Muscle cramps	
Paresthesias	Paresthesias	
Paralysis (severe hypokalemia)	Paralysis (severe hyperkalemia)	
Central Nervous System	Cardiovascular	
Confusion, depression	Electrocardiogram changes	
Cardiovascular	Risk of cardiac arrest with severe hyperkalemia	
Postural hypotension		
Predisposition to digitalis toxicity		
Electrocardiogram changes		
Cardiac dysrhythmias		
Acid-Base Balance		

Metabolic alkalosis

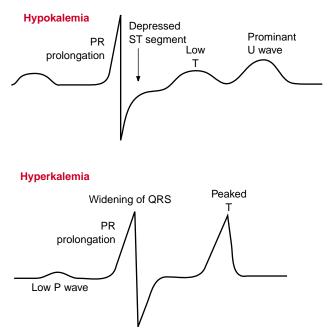


FIGURE 6-12 Electrocardiographic changes with hyperkalemia and hypokalemia.

gradually and are not severe, they often impair potassium intake and exaggerate the condition.

The renal processes that conserve potassium during hypokalemia interfere with the kidney's ability to concentrate urine. As a result, urine output and serum osmolality are increased, urine specific gravity is decreased, and polyuria, nocturia, and thirst are common. Metabolic alkalosis and renal chloride wasting are signs of severe hypokalemia.

Treatment. When possible, hypokalemia caused by potassium deficit is treated by increasing the intake of foods high in potassium content. Oral potassium supplements are prescribed for persons whose intake of potassium is insufficient in relation to losses. This is particularly true of persons who are receiving diuretic therapy and those who are taking digitalis.

Potassium may be given intravenously when the oral route is not tolerated or when rapid replacement is needed. Magnesium deficiency may impair potassium correction; in such cases, magnesium replacement is indicated.²⁵ The rapid infusion of a concentrated potassium solution can cause death from cardiac arrest. Health personnel who assume responsibility for administering potassium-containing intravenous solutions should be fully aware of all the precautions pertaining to potassium dilution and flow rate.

Hyperkalemia

Hyperkalemia refers to an increase in serum levels of potassium in excess of 5.0 mEq/L (5.0 mmol/L). It seldom occurs in healthy persons because the body is extremely effective in preventing excess potassium accumulation in the ECF.

Causes. The three major causes of potassium excess are (1) decreased renal elimination, (2) excessively rapid administration, and (3) movement of potassium from the ICF to ECF compartment.

The most common cause of hyperkalemia is decreased renal function. Chronic hyperkalemia is almost always associated with renal failure. Usually, the glomerular filtration rate must decline to less than 10 mL/minute before hyperkalemia develops. Some renal disorders, such as sickle cell nephropathy, lead nephropathy, and systemic lupus nephritis, can selectively impair tubular secretion of potassium without causing renal failure. As discussed previously, acidosis diminishes potassium elimination by the kidney. Persons with acute renal failure accompanied by lactic acidosis or ketoacidosis are at increased risk for the development of hyperkalemia. Correcting the acidosis usually helps to correct the hyperkalemia.²⁶

Aldosterone acts at the level of the distal tubular sodium/ potassium exchange system to increase potassium excretion while facilitating sodium reabsorption. A decrease in aldosterone-mediated potassium elimination can result from adrenal insufficiency (*i.e.*, Addison's disease), depression of aldosterone release caused by a decrease in renin or angiotensin II, or impaired ability of the kidneys to respond to aldosterone. Potassium-sparing diuretics (*e.g.*, spironolactone, amiloride, triamterene) can produce hyperkalemia by means of the latter mechanism. Because of their ability to decrease aldosterone levels, angiotensin-converting enzyme inhibitors also can produce an increase in serum potassium levels.

Potassium excess can result from excessive oral ingestion or intravenous administration of potassium. It is difficult to increase potassium intake to the point of causing hyperkalemia when renal function is adequate and the aldosterone sodium/ potassium exchange system is functioning. An exception to this rule is the intravenous route of administration. In some cases, severe and fatal incidents of hyperkalemia have occurred when intravenous potassium solutions were infused too rapidly. Because the kidneys control potassium elimination, the administration of potassium-containing intravenous solutions should not be initiated until urine output has been assessed and renal function has been deemed to be adequate.

The movement of potassium out of body cells into the ECF also can lead to elevated serum potassium levels. Tissue injury, such as that caused by burns and crushing injuries, causes release of intracellular potassium into the ECF compartment. The same injuries often diminish renal function, which contributes to the development of hyperkalemia. Transient hyperkalemia may be induced during extreme exercise or seizures, when muscle cells are permeable to potassium.

Manifestations. The signs and symptoms of potassium excess are closely related to the alterations in neuromuscular excitability (see Table 6-5). The neuromuscular manifestations of potassium excess usually are absent until the serum concentration exceeds 6 mEq/L. The first symptom associated with hyperkalemia typically is paresthesia. There may be complaints of generalized muscle weakness or dyspnea secondary to respiratory muscle weakness.

The most serious effect of hyperkalemia is on the heart. As potassium levels increase, disturbances in cardiac conduction occur. The earliest changes are peaked, narrow T waves and widening of the QRS complex. If serum levels continue to rise, the PR interval becomes prolonged and is followed by the disappearance of P waves (see Fig. 6-12). The heart rate may be slow. Ventricular fibrillation and cardiac arrest are terminal events. Detrimental effects of hyperkalemia on the heart

are most pronounced when the serum potassium level rises rapidly.

Treatment. The treatment of potassium excess varies with the severity of the disturbance and focuses on decreasing or curtailing intake or absorption, increasing renal excretion, and increasing cellular uptake. Decreased intake can be achieved by restricting dietary sources of potassium. The major ingredient in most salt substitutes is potassium chloride, and such substitutes should not be given to patients with renal problems.

Increasing potassium output often is more difficult. People with renal failure may require hemodialysis or peritoneal dialysis to reduce serum potassium levels. Most emergency methods focus on measures that cause serum potassium to move into the ICF compartment. An intravenous infusion of insulin and glucose is often used for this purpose.

In summary, potassium, which is the major intracellular cation, contributes to the maintenance of intracellular osmolality, is necessary for normal neuromuscular function, and influences acid-base balance. Potassium levels are influenced by dietary intake and elimination by the kidney. A transcellular shift can produce a redistribution of potassium between the ECF and ICF compartments, causing blood levels to increase or decrease.

Hypokalemia represents a serum potassium level below 3.5 mEq/L. It can result from inadequate intake, excessive losses, or redistribution between the ICF and ECF compartments. The manifestations of potassium deficit include alterations in renal, skeletal muscle, gastrointestinal, and cardiovascular function, reflecting the crucial role of potassium in cell metabolism and neuromuscular function. *Hyperkalemia* represents an increase in serum potassium in excess of 5.0 mEq/L. It seldom occurs in healthy persons because the body is extremely effective in preventing excess potassium accumulation in the ECF. The major causes of potassium excess are decreased renal elimination of potassium, excessively rapid intravenous administration of potassium, and a transcellular shift of potassium out of the cell to the ECF compartment. The most serious effect of hyperkalemia is cardiac arrest.

CALCIUM AND MAGNESIUM BALANCE

Calcium Balance

Calcium is one of the major divalent cations in the body. Approximately 99% of body calcium is found in bone, where it provides the strength and stability for the skeletal system and serves as an exchangeable source to maintain extracellular calcium levels. Most of the remaining calcium (approximately 1%) is located inside cells, and only 0.1%–0.2% is present in the ECF.

Serum calcium exists in three forms: (1) protein bound, (2) complexed, and (3) ionized. Approximately 40% of serum calcium is bound to plasma proteins, mostly albumin, and cannot diffuse or pass through the capillary wall to leave the vascular compartment (Fig. 6-13). Another 10% is complexed (*i.e.*, chelated) with substances such as citrate, phosphate, and

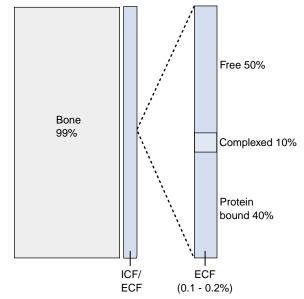
KEY CONCEPTS

CALCIUM BALANCE

- About 99% of body calcium is stored in bone; 1% is located inside cells; and 0.1% is found in the ECF.
- ECF calcium levels are made up of free (ionized), complexed, and protein-bound fractions. Only the ionized calcium ions play an essential role in neuromuscular and cardiac excitability.
- Serum calcium levels are regulated by parathyroid hormone and by renal mechanisms in which serum levels of calcium and phosphate are reciprocally regulated to prevent the damaging deposition of calcium phosphate crystals in the soft tissues of the body.

sulfate. This form is not ionized. The remaining 50% of serum calcium is present in the ionized form. It is the ionized form of calcium that is free to leave the vascular compartment and participate in cellular functions. The total serum calcium level fluctuates with changes in serum albumin and pH.

Ionized calcium serves a number of functions. It participates in many enzyme reactions; exerts an important effect on membrane potentials and neuronal excitability; is necessary for contraction in skeletal, cardiac, and smooth muscle; participates in the release of hormones, neurotransmitters, and other chemical messengers; influences cardiac contractility and automaticity by way of slow calcium channels; and is essential for blood clotting. The use of calcium channel-blocking drugs in circulatory disorders demonstrates the importance of the calcium ion



■ FIGURE 6-13 ■ Distribution of body calcium between the bone and the intracellular and extracellular fluids. The percentages of free, complexed, and protein-bound calcium in extracellular fluids are indicated.

in the normal function of the heart and blood vessels. Calcium is required for all but the first two steps of the intrinsic pathway for blood coagulation. Because of its ability to bind calcium, citrate often is used to prevent clotting in blood that is to be used for transfusions.

Regulation of Serum Calcium

Calcium enters the body through the gastrointestinal tract, is absorbed from the intestine under the influence of vitamin D, is stored in bone, and is excreted by the kidney. The major sources of calcium are milk and milk products. Only 30% to 50% of dietary calcium is absorbed from the duodenum and upper jejunum; the remainder is eliminated in the stool.

Calcium is filtered in the glomerulus of the kidney and then selectively reabsorbed back into the blood. Approximately 60% to 65% of filtered calcium is passively reabsorbed in the proximal tubule, driven by the reabsorption of sodium chloride; 15% to 20% is reabsorbed in the thick ascending loop of Henle, driven by the Na⁺/K⁺/ 2Cl⁻ cotransport system; and 5% to 10% is reabsorbed in the distal convoluted tubule (see Chapter 22). The distal convoluted tubule is an important regulatory site for controlling the amount of calcium that enters the urine. PTH and possibly vitamin D stimulate calcium reabsorption in this segment of the nephron. Other factors that may influence calcium reabsorption in the distal convoluted tubule are phosphate levels and glucose and insulin levels. Thiazide diuretics, which exert their effects in the distal convoluted tubule, enhance calcium reabsorption.

Serum calcium, which is responsible for the physiologic functions of calcium, is directly or indirectly regulated by parathyroid hormone (PTH) and vitamin D. Calcitonin, a hormone produced by C cells in the thyroid, is thought to act on the kidney and bone to remove calcium from the circulation. The regulation of serum calcium is also strongly influenced by serum phosphate levels. The role of PTH, calcitonin, and vitamin D on skeletal function is discussed in Chapter 44.

Parathyroid hormone, a major regulator of serum calcium and phosphate, is secreted by the parathyroid glands. The response to a decrease in serum calcium is prompt, occurring within seconds. The main function of PTH is to maintain the calcium concentration of the ECF (Fig. 6-14). It performs this function by promoting the release of calcium from bone, increasing the activation of vitamin D as a means of enhancing intestinal absorption of calcium, and stimulating calcium conservation by the kidney while increasing phosphate excretion (see Chapter 31).

Vitamin D, although classified as a vitamin, functions as a hormone. Vitamin D_3 , the precursor of the active form of vitamin D, is synthesized in the skin or obtained from foods in the diet, many of which are fortified with vitamin D. Vitamin D_3 is hydroxylated in the liver and is transformed to its active form in the kidney. The major action of the activated form of vitamin D is to increase the absorption of calcium from the intestine.

The extracellular concentrations of calcium and phosphate are reciprocally regulated such that calcium levels fall when phosphate levels are high and vice versa. Normal serum levels of calcium (8.5 to 10.5 mg/dL in adults) and phosphate (2.5 to 4.5 mg/dL in adults) are regulated so that the product of the two concentrations ($[Ca^{2+}] \times [PO_4^{2-}]$) is normally maintained at less than 70.²⁷ Maintenance of the calcium-phosphate prod-

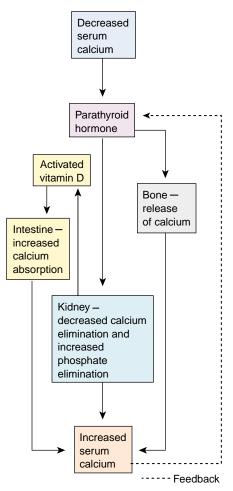


FIGURE 6-14 Regulation of serum calcium concentration by parathyroid hormone.

uct within this range is important in preventing the deposition of CaPO₄ salts in soft tissue, damaging the kidneys, blood vessels, and lungs.

Hypocalcemia

Hypocalcemia represents a serum calcium level of less than 8.5 mg/dL. Hypocalcemia occurs in many forms of critical illness and has affected as many as 70% to 90% of patients in intensive care units.²⁷

Causes. The causes of hypocalcemia can be divided into three categories: (1) impaired ability to mobilize calcium bone stores, (2) abnormal losses of calcium from the kidney, and (3) increased protein binding or chelation such that greater proportions of calcium are in the nonionized form. A pseudo-hypocalcemia is caused by hypoalbuminemia. It results in a decrease in protein-bound, rather than ionized, calcium and usually is asymptomatic.²⁸ Calcium deficit caused by dietary deficiency exerts its effects on bone stores, rather than extracellular calcium levels.

Serum calcium exists in a dynamic equilibrium with calcium in bone. The ability to mobilize calcium from bone depends on adequate levels of PTH. Decreased levels of PTH may result from primary or secondary forms of hypoparathyroidism (see Chapter 31). Suppression of PTH release may also occur when vitamin D levels are elevated. Magnesium deficiency inhibits PTH release and impairs the action of PTH on bone resorption. This form of hypocalcemia is difficult to treat with calcium supplementation alone and requires correction of the magnesium deficiency.

Phosphate elimination is impaired in renal failure. Because of the inverse relation between calcium and phosphate, serum calcium levels fall as phosphate levels rise in renal failure. Hypocalcemia and hyperphosphatemia occur when the glomerular filtration rate falls to less than 25 to 30 mL/minute (normal is 100 to 120 mL/minute).

Only the ionized form of calcium is able to leave the capillary and participate in body functions. A change in pH alters the proportion of calcium that is in the bound and ionized forms. An acid pH decreases binding of calcium to protein, causing a proportionate increase in ionized calcium, whereas total serum calcium remains unchanged. An alkaline pH has the opposite effect. As an example, hyperventilation sufficient to cause respiratory alkalosis can produce tetany because of increased protein binding of calcium. Free fatty acids increase binding of calcium to albumin, causing a reduction in ionized calcium. Elevations in free fatty acids sufficient to alter calcium binding may occur during stressful situations that cause elevations of epinephrine, glucagon, growth hormone, and adrenocorticotropin levels.

Hypocalcemia is a common finding in a patient with acute pancreatitis. Inflammation of the pancreas causes release of proteolytic and lipolytic enzymes. It is thought that the calcium ion combines with free fatty acids released by lipolysis in the pancreas, forming soaps and removing calcium from the circulation.

Manifestations. Hypocalcemia can manifest as an acute or chronic condition. The manifestations of acute hypocalcemia reflect the increased neuromuscular excitability and cardiovascular effects of a decrease in ionized calcium (see Table 6-6). Ionized calcium stabilizes neuromuscular excitability, thereby making nerve cells less sensitive to stimuli. Nerves exposed to low ionized calcium levels show decreased thresholds for excitation, repetitive responses to a single stimulus, and, in extreme cases, continuous activity. The severity of the manifestations depends on the underlying cause, rapidity of onset, accompanying electrolyte disorders, and extracellular pH. Increased neuromuscular excitability can manifest as paresthesias (i.e., tingling around the mouth and in the hands and feet) and tetany (i.e., muscle spasms of the muscles of the face, hands, and feet). Severe hypocalcemia can lead to laryngeal spasm, seizures, and even death.

The cardiovascular effects of acute hypocalcemia include hypotension, cardiac insufficiency, cardiac dysrhythmias (particularly heart block and ventricular fibrillation), and failure to have a response to drugs such as digitalis, norepinephrine, and dopamine that act through calcium-mediated mechanisms.

Chronic hypocalcemia is often accompanied by skeletal manifestations and skin changes. There may be bone pain, fragility, deformities, and fractures. The skin may be dry and

Hypocalcemia	Hypercalcemia
Laboratory Serum calcium <8.5 mg/dL	Laboratory Serum calcium >10.5 mg/dL
	Inability to Concentrate Urine and Exposure of Kidney to Increased Concentration of Calcium Polyuria Increased thirst Flank pain Signs of acute renal insufficiency Signs of kidney stones
Neural and Muscle Effects (Increased Excitability) Paresthesias, especially numbness and tingling Skeletal muscle cramps Abdominal muscle spasms and cramps Hyperactive reflexes Carpopedal spasm Tetany Laryngeal spasm	Neural and Muscle Effects (Decreased Excitability) Muscle weakness Ataxia, loss of muscle tone Lethargy Personality and behavioral changes Stupor and coma
Cardiovascular Effects Hypotension Signs of cardiac insufficiency Decreased response to drugs that act by calcium-mediated mechanisms Prolongation of the QT interval predisposes to ventricular dysrhythmias	Cardiovascular Effects Hypertension Shortening of the QT interval Atrioventricular block
Skeletal Effects (Chronic Deficiency) Osteomalacia Bone pain	Gastrointestinal Effects Anorexia Nausea, vomiting Constipation

scaling, the nails brittle, and hair dry. The development of cataracts is common. A person with chronic hypocalcemia may also present with mild diffuse brain disease mimicking depression, dementia, or psychoses.

Diagnosis and Treatment. Chvostek's and Trousseau's tests can be used to assess for an increase in neuromuscular excitability and tetany.⁵ *Chvostek's sign* is elicited by tapping the face just below the temple at the point where the facial nerve emerges. Tapping the face over the facial nerve causes spasm of the lip, nose, or face when the test result is positive. An inflated blood pressure cuff is used to test for *Trousseau's sign*. The cuff is inflated above systolic blood pressure for 3 minutes. Contraction of the fingers and hands (*i.e.*, carpopedal spasm) indicates the presence of tetany.

Acute hypocalcemia is an emergency situation, requiring prompt treatment. An intravenous infusion containing calcium is used when tetany or acute symptoms are present or anticipated because of a decrease in the serum calcium level.

Chronic hypocalcemia is treated with oral intake of calcium. One glass of milk contains approximately 300 mg of calcium. Oral calcium supplements may be used. In some cases, long-term treatment may require the use of vitamin D preparations. The active form of vitamin D is administered when the liver or kidney mechanisms needed for hormone activation are impaired.

Hypercalcemia

Hypercalcemia represents a total serum calcium concentration greater than 10.5 mg/dL. Falsely elevated levels of calcium can result from prolonged drawing of blood with an excessively tight tourniquet. Increased plasma proteins (*e.g.*, hyperalbuminemia, hyperglobulinemia) may elevate the total serum calcium but not affect the ionized calcium concentration

Causes. A serum calcium excess (*i.e.*, hypercalcemia) results when calcium movement into the circulation overwhelms the calcium regulatory hormones or the ability of the kidney to remove excess calcium ions. The most common causes of hypercalcemia are increased bone resorption caused by neoplasms or hyperparathyroidism. Hypercalcemia is a common complication of cancer, occurring in approximately 10% to 20% of persons with advanced disease.²⁹ A number of malignant tumors, including carcinoma of the lungs, have been associated with hypercalcemia. Some tumors destroy the bone, but others produce humoral agents that stimulate osteoclastic activity, increase bone resorption, or inhibit bone formation.

Less common causes of hypercalcemia are prolonged immobilization, increased intestinal absorption of calcium, and excessive doses of vitamin D. Prolonged immobilization and lack of weight bearing cause demineralization of bone and release of calcium into the bloodstream. Intestinal absorption of calcium can be increased by excessive doses of vitamin D or as a result of a condition called the *milk-alkali syndrome*. The milkalkali syndrome is caused by excessive ingestion of calcium (often in the form of milk) and absorbable antacids. Because of the advent of nonabsorbable antacids, the condition is seen less frequently than in the past, but it may occur in women who are overzealous in taking calcium preparations for osteoporosis prevention.

A variety of drugs elevate calcium levels. The use of lithium to treat bipolar disorders has caused hypercalcemia and hyperparathyroidism. The thiazide diuretics increase calcium reabsorption in the distal convoluted tubule of the kidney. Although the thiazide diuretics seldom cause hypercalcemia, they can unmask hypercalcemia from other causes such as underlying bone disorders and conditions that increase bone resorption.

Manifestations. The signs and symptoms associated with calcium excess originate from three sources: (1) changes in neural excitability, (2) alterations in smooth and cardiac muscle function, and (3) exposure of the kidneys to high concentrations of calcium (see Table 6-6).

Neural excitability is decreased in patients with hypercalcemia. There may be a dulling of consciousness, stupor, weakness, and muscle flaccidity. Behavioral changes may range from subtle alterations in personality to acute psychoses.

The heart responds to elevated levels of calcium with increased contractility and ventricular dysrhythmias. Digitalis accentuates these responses. Gastrointestinal symptoms reflect a decrease in smooth muscle activity and include constipation, anorexia, nausea, and vomiting. Pancreatitis is another potential complication of hypercalcemia and is probably related to stones in the pancreatic ducts.

High calcium concentrations in the urine impair the ability of the kidneys to concentrate urine by interfering with the action of ADH. This causes salt and water diuresis and an increased sensation of thirst. Hypercalciuria also predisposes to the development of renal calculi.

Hypercalcemic crisis describes an acute increase in the serum calcium level.²⁹ Malignant disease and hyperparathyroidism are major causes of hypercalcemic crisis. In hypercalcemic crisis, polyuria, excessive thirst, volume depletion, fever, altered levels of consciousness, azotemia (*i.e.*, nitrogenous wastes in the blood), and a disturbed mental state accompany other signs of calcium excess. Symptomatic hypercalcemia is associated with a high mortality rate; death often is caused by cardiac arrest.

Treatment. The treatment of calcium excess usually is directed toward rehydration and measures to increase urinary excretion of calcium and inhibit release of calcium from bone. Fluid replacement is needed in situations of volume depletion. The excretion of sodium is accompanied by calcium excretion. Diuretics and sodium chloride can be administered to increase urinary elimination of calcium after the ECF volume has been restored. Loop diuretics commonly are used, rather than thiazide diuretics, which increase calcium reabsorption.

The initial lowering of calcium levels is followed by measures to inhibit bone reabsorption. Drugs that are used to inhibit calcium mobilization include bisphosphonates, calcitonin, and the glucocorticosteroids. The bisphosphonates are a relatively new group of drugs that act mainly by inhibiting osteoclastic activity. Calcitonin inhibits osteoclastic activity, thereby decreasing resorption. The glucocorticosteroids inhibit bone resorption and are used to treat hypercalcemia associated with cancer.

Magnesium Balance

Magnesium is the second most abundant intracellular cation. The average adult has approximately 24 g of magnesium distributed throughout the body.³⁰ Of the total magnesium content, approximately 50% to 60% is stored in bone, 39% to 49% is contained in the body cells, and the remaining 1% is dis-

persed in the ECF.³¹ Approximately 20% to 30% of the extracellular magnesium is protein bound, and only a small fraction of intracellular magnesium (15% to 30%) is exchangeable with the ECF. The normal serum concentration of magnesium is 1.8 to 2.7 mg/dL.

Only recently has the importance of magnesium to the overall function of the body been recognized.³¹ Magnesium acts as a cofactor in many intracellular enzyme reactions, including those related to transfer of phosphate groups. It is essential to all reactions that require ATP, for every step related to replication and transcription of DNA, and for the translation of messenger RNA. It is required for cellular energy metabolism, functioning of the sodium-potassium membrane pump, membrane stabilization, nerve conduction, ion transport, and calcium channel activity. Magnesium binds to calcium receptors, and it has been suggested that alterations in magnesium levels may exert their effects through calciummediated mechanisms. Magnesium may bind competitively to calcium binding sites, producing the appropriate response; it may compete with calcium for a binding site but not exert an effect; or it may alter the distribution of calcium by interfering with its movement across the cell membrane.

Regulation of Magnesium

Magnesium is ingested in the diet, absorbed from the intestine, and excreted by the kidneys. Intestinal absorption is not closely regulated, and approximately 25% to 65% of dietary magnesium is absorbed. Magnesium is contained in all green vegetables, grains, nuts, meats, and seafood. Magnesium is also present in much of the groundwater in North America.

The kidney is the principal organ of magnesium regulation. Magnesium is a unique electrolyte in that only approximately 30% to 40% of the filtered amount is reabsorbed in the proximal tubule. The greatest quantity, approximately 50% to 70%, is reabsorbed in the thick ascending loop of Henle. The distal tubule, which reabsorbs a small amount of magnesium, is the major site of magnesium regulation. Magnesium reabsorption is decreased in the presence of increased serum levels, stimulated by PTH, and inhibited by increased calcium levels. The major driving force for magnesium absorption in the thick ascending loop of Henle is the Na⁺/K⁺/2Cl⁻ cotransport system (see Chapter 22). Inhibition of this transport system by loop diuretics lowers magnesium reabsorption.

Hypomagnesemia

Hypomagnesemia represents a serum magnesium concentration of less than 1.8 mg/dL (Table 6-7).³² It is seen in conditions that limit intake or increase intestinal or renal losses, and it is a common finding in emergency department and critical care patients.

Causes. Magnesium deficiency can result from insufficient intake, excessive losses, or movement between the ECF and ICF compartments. It can result from conditions that directly limit intake, such as malnutrition, starvation, or prolonged maintenance of magnesium-free parenteral nutrition. Other conditions, such as diarrhea, malabsorption syndromes, prolonged nasogastric suction, or laxative abuse can serve to decrease intestinal absorption. Excessive calcium intake impairs intestinal absorption of magnesium by competing for the same transport site. Another common cause of magnesium deficiency is chronic alcoholism. There are many factors that contribute to hypomagnesemia in alcoholism, including low intake and gastrointestinal losses from diarrhea.

Although the kidneys are able to defend against hypermagnesemia, they are less able to conserve magnesium and prevent hypomagnesemia. Urine losses are increased in diabetic ketoacidosis, hyperparathyroidism, and hyperaldosteronism. Some drugs increase renal losses of magnesium, including diuretics (particularly loop diuretics) and nephrotoxic drugs such as aminoglycoside antibiotics, cyclosporine, cisplatin, and amphotericin B.

A relative hypomagnesemia may also develop in conditions that promote movement of magnesium between the ECF and ICF compartments, including rapid administration of glucose, insulin-containing parenteral solutions, and alkalosis. Although transient, these conditions can cause serious alterations in body function.

Manifestations. Signs of magnesium deficiency are not usually apparent until the serum magnesium is less than 1 mEq/dL. Hypomagnesemia is characterized by an increase in neuro-muscular excitability as evidenced by muscle weakness and tremors. Other manifestations may include hyperactive deep tendon reflexes, paresthesias (*e.g.*, numbness, pricking, tingling sensation), muscle fasciculations, and tetanic muscle contractions. A positive Chvostek's or Trousseau's may be present,

TABLE 6-7 Manifestations of Hypomagnesemia and Hypermagnesemia		
Hypomagnesemia	Hypermagnesemia	
Laboratory Values	Laboratory Values	
Serum magnesium <1.8 mg/dL	Serum magnesium >2.7 mg/dL	
Neural and Muscle Effects	Neural and Muscle Effects	
Personality changes	Lethargy	
Athetoid or choreiform movements	Hyporeflexia	
Nystagmus	Confusion	
Tetany	Coma	
Cardiovascular Effects	Cardiovascular Effects	
Tachycardia	Hypotension	
Hypertension	Cardiac dysrhythmias	
Cardiac dysrhythmias	Cardiac arrest	

particularly if hypocalcemia is present. Because decreased serum magnesium increases irritability in nervous tissue, seizures may occur. Other manifestations may include ataxia, vertigo, disorientation, depression, and psychotic symptoms.

Cardiovascular manifestations include tachycardia, hypertension, and ventricular dysrhythmias. There may be ECG changes such as widening of the QRS complex, appearance of peak T waves, prolongation of the PR interval, T-wave inversion, and appearance of U waves. Ventricular dysrhythmias, particularly in the presence of digitalis, may be difficult to treat unless magnesium levels are normalized.

Magnesium deficiency often occurs in conjunction with hypocalcemia and hypokalemia, producing a number of related neurologic and cardiovascular manifestations. Hypocalcemia is typical of severe hypomagnesemia. Most persons with hypomagnesemia-related hypocalcemia have decreased PTH levels, probably as a result of impaired magnesiumdependent mechanisms that control PTH release and synthesis. Hypokalemia also is a typical feature of hypomagnesemia. It leads to a reduction in intracellular potassium and impairs the ability of the kidney to conserve potassium. When hypomagnesemia is present, hypokalemia is unresponsive to potassium replacement therapy.

Treatment. Hypomagnesemia is treated with magnesium replacement. The route of administration depends on the severity of the condition. Symptomatic, moderate to severe magnesium deficiency is treated by parenteral administration. Treatment must be continued for several days to replace stored and serum levels. In conditions of chronic intestinal or renal loss, maintenance support with oral magnesium may be required. Patients with any degree of renal failure must be carefully monitored to prevent magnesium excess. Magnesium often is used therapeutically to treat cardiac arrhythmia, myocardial infarct, angina, and pregnancy complicated by preeclampsia or eclampsia. Caution to prevent hypermagnesemia is essential.

Hypermagnesemia

Hypermagnesemia represents a serum magnesium concentration in excess of 2.7 mg/dL. Because of the ability of the normal kidney to excrete magnesium, hypermagnesemia is rare.

When hypermagnesemia does occur, it usually is related to renal insufficiency and the injudicious use of magnesiumcontaining medications such as antacids, mineral supplements, or laxatives. The elderly are particularly at risk because they have age-related reductions in renal function and tend to consume more magnesium-containing medications. Magnesium sulfate is used to treat toxemia of pregnancy and premature labor; in these cases, careful monitoring for signs of hypermagnesemia is essential.

Hypermagnesemia affects neuromuscular and cardiovascular function (see Table 6-7). The signs and symptoms occur only when serum magnesium levels exceed 4.9 mg/dL.³²

Hypermagnesemia diminishes neuromuscular transmission, causing hyporeflexia, muscle weakness, and confusion. Magnesium decreases acetylcholine release at the myoneural junction and may cause neuromuscular blockade and respiratory paralysis. Cardiovascular effects are related to the calcium channelblocking effects of magnesium. Blood pressure is decreased, and the ECG shows an increase in the PR interval, a shortening of the QT interval, T-wave abnormalities, and prolongation of the QRS and PR intervals. Hypotension caused by vasodilation and cardiac dysrhythmias can occur with moderate hypermagnesemia (\geq 5 to 10 mg/dL), and confusion and coma can occur with severe hypermagnesemia (\geq 10 mg/dL). Very severe hypermagnesemia (\geq 15 mg/dL) may cause cardiac arrest.

The treatment of hypermagnesemia includes cessation of magnesium administration. Calcium is a direct antagonist of magnesium, and intravenous administration of calcium may be used. Peritoneal dialysis or hemodialysis may be required.

In summary, calcium is one of the major divalent ions in the body. Approximately 99% of body calcium is found in bone; about 1% is in the ICF and 0.1% to 0.2% is in the ECF. The calcium in bone is in dynamic equilibrium with extracellular calcium. Of the three forms of ECF calcium (*i.e.*, protein bound, complexed, and ionized), only the ionized form can cross the cell membrane and contribute to cellular function. Ionized calcium has a number of functions. It contributes to neuromuscular function, plays a vital role in the blood clotting process, and participates in a number of enzyme reactions. Alterations in ionized calcium levels produce neural effects; neural excitability is increased in hypocalcemia and decreased in hypercalcemia.

Magnesium is the second most abundant intracellular cation. It acts as a cofactor in many enzyme reactions and affects neuromuscular function in the same manner as the calcium ion. Magnesium deficiency can result from insufficient intake, excessive losses, or movement between the ECF and ICF compartments. The manifestations of hypomagnesemia are characterized by a decrease in neuromuscular excitability as evidenced by paresthesias and hyperactive reflexes. Cardiovascular effects include tachycardia, hypertension, and ventricular dysrhythmias. Hypermagnesemia usually is related to renal insufficiency and the injudicious use of magnesiumcontaining medications such as antacids, mineral supplements, or laxatives. It diminishes neuromuscular transmission leading to hyporeflexia, muscle weakness, and confusion.

ACID-BASE BALANCE

Metabolic activities of the body require the precise regulation of acid-base balance, which is reflected by the pH of ECF. Membrane excitability, enzyme systems, and chemical reactions depend on pH being regulated within a narrow physiologic range. Many conditions, pathologic or otherwise, can alter body pH.

Introductory Concepts

Normally, the concentration of body acids and bases is regulated so that the pH of extracellular body fluids is maintained within a very narrow range of 7.35 to 7.45. This balance is maintained through mechanisms that generate, buffer, and eliminate acids and bases.

Acid-Base Chemistry

An *acid* is a molecule that can release a hydrogen ion (H^+) , and a *base* is a molecule that can accept or combine with an H^+ ion. Most of the body's acids and bases are weak acids and

bases; the most important are *carbonic acid* (H_2CO_3), which is a weak acid derived from carbon dioxide (CO_2), and *bicarbonate* (HCO_3^-), which is a weak base.

The concentration of H⁺ ions in body fluids is low compared with other ions. For example, the Na⁺ ion is present at a concentration approximately 1 million times that of the H⁺ ion. Because of its low concentration in body fluids, the H⁺ ion concentration is commonly expressed in terms of pH. Specifically, *pH* represents the negative logarithm (p) of the H⁺ ion concentration in mEq/L. A pH value of 7.0 implies a H⁺ ion concentration of 10⁻⁷ (0.0000001 mEq/L). Because the pH is inversely related to the H⁺ ion concentration, a low pH indicates a high concentration of H⁺ ions and a high pH, a low concentration.²

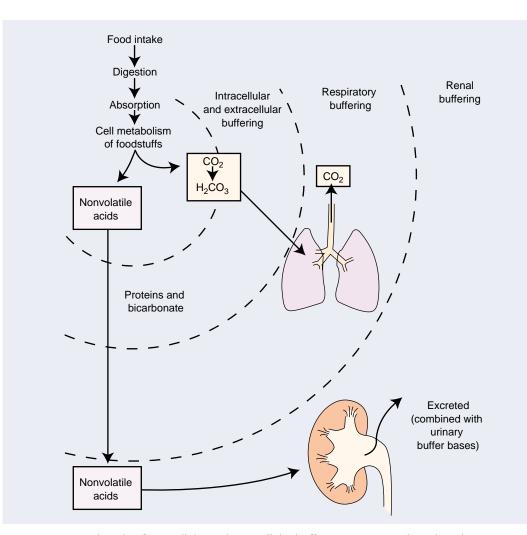
Metabolic Acid and Bicarbonate Production

Acids are continuously generated as by-products of metabolic processes. Physiologically, these acids fall into two groups: the volatile acid carbonic acid and all other nonvolatile or fixed acids (Fig. 6-15).

KEY CONCEPTS

MECHANISMS OF ACID-BASE BALANCE

- The pH is determined by the ratio of the HCO₃[−] base to CO₂ (H₂CO₃). At a pH of 7.4, the ratio is normally 20 to 1.
- The HCO₃⁻ part of the pH equation reflects the generation of nonvolatile metabolic acids that are buffered in intracellular and extracellular buffers and eliminated by the kidney.
- The H_2CO_3 or volatile acid part of the equation represents the level of dissolved CO_2 ($H_2O + CO_2 \leftrightarrow H_2CO_3$) in the blood. It is regulated by the elimination of CO_2 by the lungs.



■ FIGURE 6-15 ■ The role of intracellular and extracellular buffer, respiratory, and renal mechanisms in maintaining normal blood pH. (Rhoades R.A., Tanner G.A. [1996]. *Medical physiology* [p. 468]. Boston: Little, Brown)

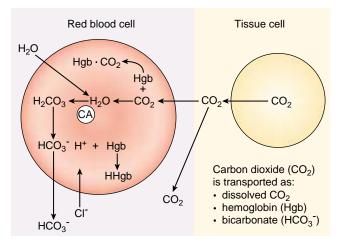
110 Unit One: Mechanisms of Disease

The difference between the two types of acids arises because H_2CO_3 is in equilibrium with the volatile CO_2 , which leaves the body by way of the lungs. Therefore, the H_2CO_3 concentration is determined by the lungs and their respiratory capacity. The *noncarbonic acids* (*e.g.*, sulfuric, hydrochloric, phosphoric) are *nonvolatile* and are not eliminated by the lungs. Instead, they are buffered by body proteins or extracellular buffers, such as HCO_3^- , and then excreted by the kidney.

Carbon Dioxide and Bicarbonate Production. Body metabolism results in the production of approximately 15,000 mmol of CO₂ each day.³³ Carbon dioxide is transported in the circulation in three forms: (1) attached to hemoglobin, (2) as dissolved CO₂ (*i.e.*, PCO₂), and as (3) HCO₃⁻ (Fig. 6-16). Collectively, dissolved CO₂ and HCO₃⁻ constitute approximately 77% of the CO₂ that is transported in the ECF; the remaining CO₂ travels attached to hemoglobin. Although CO₂ is not an acid, a small percentage of the gas combines with water in the blood-stream to form H₂CO₃ (CO₂ + H₂O \leftrightarrow H₂CO₃). The reaction between CO₂ and water is catalyzed by an enzyme, called *carbonic anhydrase*, which is present in large quantities in red blood cells, renal tubular cells, and other tissues in the body.

Because it is almost impossible to measure H_2CO_3 , dissolved CO_2 measurements are commonly substituted when calculating pH. The H_2CO_3 content of the blood can be calculated by multiplying the partial pressure of the CO_2 in the blood by its solubility coefficient, which is 0.03. This means that the concentration of H_2CO_3 in venous blood, which normally has a PCO₂ of approximately 45 mm Hg, is 1.35 mEq/L ($45 \times 0.03 = 1.35$).

Production of Metabolic Acids. The metabolism of dietary proteins is the major source of strong *inorganic acids*—sulfuric acid, hydrochloric acid, and phosphoric acid.^{33, 34} Oxidation of the sulfur-containing amino acids (*e.g.*, methionine, cysteine, cystine) results in the production of sulfuric acid. Oxidation of arginine and lysine produces hydrochloric acid, and oxidation of phosphorus-containing nucleic acids yields phosphoric acid. Incomplete oxidation of glucose results in the formation of lactic acid, and incomplete oxidation of fats results in the production of ketoacids. The major source of base is the metabolism of amino acids such as aspartate and glutamate and the



■ FIGURE 6-16 ■ Mechanisms of carbon dioxide transport.

metabolism of certain organic anions (*e.g.*, citrate, lactate, acetate). Acid production normally exceeds base production, with the net effect being the addition of approximately 1 mmol/kg body weight of nonvolatile acid to the body each day.³³ A vegetarian diet, which contains large amounts of organic anions, results in the net production of base.

Calculation of pH. The pH is calculated with the *Henderson-Hasselbalch equation* using the dissociation constant for the bicarbonate buffer system (6.1) and the HCO₃⁻ to H₂CO₃ ratio.

$$pH = 6.1 + \log \frac{HCO_3}{H_2CO_3}$$

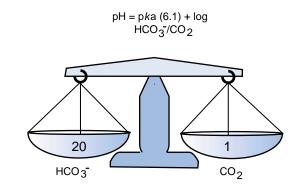
The log of 20 is 1.3. Thus, when the ratio is 20 to 1, the pH is within the normal range at 7.4 (Fig. 6-17). Because the ratio is used, a change in HCO_3^- will have little or no effect on pH, as long as there are accompanying changes in H_2CO_3 . The generation of metabolic acids and the availability of bicarbonate to buffer these acids control the HCO_3^- part of the equation. The H_2CO_3 part of the equation is regulated by the lungs and their ability to eliminate CO_2 . The kidney functions in the generation and reabsorption of HCO_3^- and contributes to control of the metabolic part of the equation.

Regulation of pH

The pH of body fluids is regulated by three major mechanisms: (1) ICF and ECF buffering systems, (2) the lungs, which control the elimination of CO_2 , and (3) the kidneys, which eliminate H⁺ and regulate the elimination of HCO_3^- .

Intracellular and Extracellular Buffer Systems. The momentby-moment regulation of pH depends on the ICF and ECF buffer systems. A *buffer system* consists of a weak acid and the base salt of that acid or of a weak base and its acid salt. In the process of preventing large changes in pH, the system trades a strong acid for a weak acid or a strong base for a weak base.

The three major buffer systems that protect the pH of body fluids are (1) proteins, (2) the HCO_3^- buffer system, and (3) the transcellular H⁺/K⁺ exchange system.² These buffer systems are immediately available to combine with excess acids or bases and prevent large changes in pH from occurring during the time it takes for the respiratory and renal mechanisms to become effective. Bone also represents an important site for the buffering of acids and bases.³³ The role of bone buffers is even greater in chronic acid-base disorders. One consequence of



■ FIGURE 6-17 ■ The Henderson-Hasselbalch equation expressed as a simple scale.

bone buffering is the release of calcium from bone and increased renal excretion of calcium. In addition to causing demineralization of bone, it also predisposes to kidney stones.

Proteins are the largest buffer system in the body. Proteins are *amphoteric*, meaning that they can function as either acids or bases. They contain many ionizable groups that can release or bind H⁺. The protein buffers are largely located within cells, and H⁺ ions and CO₂ diffuse across cell membranes for buffering by intracellular proteins. Albumin and plasma globulins are the major protein buffers in the vascular compartment.

The bicarbonate buffer system uses H_2CO_3 as its weak acid and HCO_3^- as its weak base. It substitutes the weak H_2CO_3 for a strong acid such as hydrochloric acid or the weak HCO_3^- base for a strong base such as sodium hydroxide. The HCO_3^-/H_2CO_3 buffer system is a particularly efficient system because the buffer components can be readily added or removed from the body.³⁴ Metabolism provides an ample supply of CO_2 , which can replace any H_2CO_3 that is lost when excess base is added, and CO_2 can be readily eliminated when excess acid is added. Likewise, the kidney can form new HCO_3^- when excess base is added.

The transcompartmental exchange of the K⁺ and H⁺ ions is important in the regulation of acid-base balance. Both ions are positively charged, and both ions move freely between the ICF and ECF compartments. When excess H⁺ is present in the ECF, it moves into the body cells in exchange for K⁺. Likewise, when excess K⁺ is present in the ECF, it moves into the cell in exchange for H⁺. On the average, serum K⁺ rises by approximately 0.6 mEq/L for every 0.1 unit fall in pH.33 Thus, alterations in K⁺ levels can affect acid-base balance, and changes in acid-base balance can influence K⁺ levels. Potassium shifts tend to be more pronounced in acidemia than alkalemia and are greater in metabolic acidosis than respiratory acidosis.⁵ Metabolic acidosis caused by an accumulation of nonorganic acids (e.g., hydrochloric acid that occurs in diarrhea, phosphoric acid that occurs in renal failure) produces a greater increase in K⁺ than does acidosis caused by an accumulation of organic acids (e.g., lactic acid, ketoacids).

An important implication of the K⁺ and H⁺ transcompartmental exchange is its effect on the resting membrane potential of neurons and other excitable tissue. In acidosis, increased levels of serum K⁺ cause the resting membrane potential to become less negative and in alkalosis, decreased levels cause the resting membrane potential to become more negative. Changes in neural excitability are further influenced by alterations in ionized calcium. In acidosis, the ionized portion of the extracellular calcium is increased, making neurons less excitable, and in alkalosis, the amount of ionized calcium is reduced, making neurons more excitable.

Respiratory Control Mechanisms. The respiratory system provides for the elimination of CO_2 into the air and plays a major role in acid-base regulation. The respiratory control of pH is rapid, occurring within minutes, and is maximal within 12 to 24 hours.³³ However, it does not return the pH to normal.

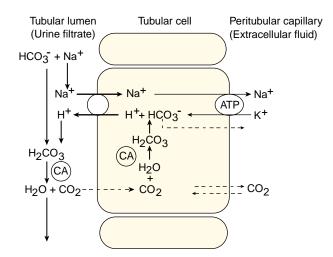
Renal Control Mechanisms. The kidneys regulate acid-base balance by excreting acidic or alkaline urine. Excreting an acidic urine reduces the amount of acid in the ECF, and excreting an alkaline urine removes base from the ECF. The renal mechanisms for regulating acid-base balance cannot adjust the pH

within minutes, as respiratory mechanisms can, but they continue to function for days until the pH has returned to normal or near-normal range.

The kidneys filter HCO_3^- in the glomerulus and then reabsorb it in the tubules as a means of maintaining ECF levels. Bicarbonate ions do not readily cross the membranes of renal tubular cells; therefore, HCO_3^- that has been filtered in the glomerulus cannot be directly reabsorbed. Instead, the $HCO_3^$ ions are reabsorbed in a special process in which they first combine with H⁺ ions that have been secreted into the tubular fluid to form H_2CO_3 , which is converted to CO_2 and water (Fig. 6-18). The resulting CO_2 can readily cross the tubular membrane and enter the tubular cell, where it combines with water, under the influence of carbonic anhydrase, to generate a new H_2CO_3 molecule. The newly formed H_2CO_3 , in turn, dissociates to form a HCO_3^- and a H⁺ ion. The HCO_3^- that is formed moves out of the tubular cell into the bloodstream and the H⁺ is secreted into the tubular fluid.

The HCO_3^- reabsorption system serves a dual purpose—it buffers the H⁺ ions that have been secreted from the blood into the tubular fluid, and it conserves the HCO_3^- ions that have been filtered from the blood into the tubular fluid. Normally, only a few of the secreted H⁺ ions remain in the tubular fluid because the secretion of H⁺ ions is roughly equivalent to the number of HCO_3^- ions that are filtered in the glomerulus. When the number of H⁺ ions secreted into the tubular fluid exceeds the amount of filtered HCO_3^- ions, the urine filtrate becomes acidic.

The elimination of H^+ is accomplished through secretion into the urine filtrate by the cells in the renal tubules. Because extremely acidic urine would be damaging to structures in the urinary tract, the pH of the urine is maintained within a range from 4.5 to 8.0. This limits the number of unbuffered H^+ ions that can be excreted by the kidney. When the number of free H^+ ions secreted into the tubular fluid threatens to cause the urine to become too acidic, the H^+ ions must be carried in



■ FIGURE 6-18 ■ Hydrogen ion (H⁺) secretion and bicarbonate ion (HCO₃⁻) reabsorption in a renal tubular cell. Carbon dioxide (CO₂) diffuses from the blood or urine filtrate into the tubular cell, where it combines with water in a carbonic anhydrase-catalyzed reaction that yields carbonic acid (H₂CO₃). The H₂CO₃ dissociates to form H⁺ and HCO₃⁻. The H⁺ is secreted into the tubular fluid in exchange for Na⁺. The Na⁺ and HCO₃⁻ enter the extracellular fluid.

some other form. This is accomplished by combining the H⁺ ions with intratubular buffers before being excreted in the urine. There are two important intratubular buffer systems: the phosphate buffer system and the ammonia buffer system.

The phosphate buffer system uses HPO_4^{2-} and $H_2PO_4^{-}$ that are present in the tubular filtrate to buffer H⁺. The combination of H⁺ with HPO_4^{2-} to form $H_2PO_4^{-}$ allows the kidneys to increase their secretion of H⁺ ions. Because they are poorly reabsorbed, the phosphates become more concentrated as they move through the tubules. This system works best when the renal tubular fluid contains a high concentration of H⁺ ions.

Another important but more complex buffer system is the *ammonia buffer system*. Renal tubular cells are able to use the amino acid glutamine to synthesize ammonia (NH₃) and secrete it into the tubular fluid. The H⁺ ions then combine with the NH₃ to form an ammonium ion (NH₄⁺). The NH₄⁺ ions combine with Cl⁻ ions (Cl⁻), which are present in the tubular fluid, to form ammonium Cl⁻ (NH₄Cl), which is then excreted in the urine.

Plasma potassium levels influence renal elimination of H^+ ions and vice versa. When plasma potassium levels fall, there is movement of K^+ ions from tubular cells into the plasma and a reciprocal movement of H^+ ions from the plasma into tubular cells. Potassium depletion also produces a reduction in Cl^- reabsorption in the distal tubule. The result is increased reabsorption of the filtered bicarbonate and development of metabolic alkalosis. An elevation in serum potassium has the opposite effect.

Aldosterone also influences H⁺ ion elimination by the kidney. It acts in the collecting duct to indirectly stimulate H⁺ ion secretion, while increasing Na⁺ ion reabsorption and K⁺ secretion. Hyperaldosteronism tends to lead to a decrease in serum K⁺ levels and increased pH and alkalosis caused by increased H⁺ ion secretion. Hypoaldosteronism has the opposite effect. It leads to increased K⁺ levels, decreased H⁺ ion secretion, and acidosis.

One of the mechanisms that the kidneys use in regulating the pH of the ECF is the conservation or elimination of $HCO_3^$ ions; in the process, it often is necessary to shuffle anions. Chloride is the most abundant anion in the ECF and can substitute for HCO_3^- when an anion shift is needed. As an example, serum HCO_3^- levels normally increase as hydrochloric acid (HCl) is secreted into the stomach after a heavy meal, causing what is called the *postprandial alkaline tide*. Later, as the Cl⁻ ion is reabsorbed in the small intestine, the pH returns to normal. *Hypochloremic alkalosis* refers to an increase in pH that is induced by a decrease in serum Cl⁻ levels and *hyperchloremic acidosis* to a decrease in pH that occurs when excess levels of Cl⁻ are present.

Laboratory Tests

Laboratory tests that are used in assessing acid-base balance include arterial blood gas measurements, carbon dioxide content and bicarbonate levels, base excess or deficit, and the anion gap.

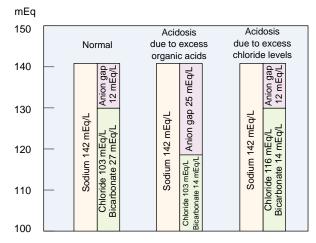
Arterial blood gases provide a means of assessing the respiratory component of acid-base balance. H_2CO_3 levels are determined from arterial CO_2 levels and the solubility coefficient for CO_2 (normal arterial PCO₂ is 38 to 42 mm Hg). Arterial blood gases are used because *venous blood gases* are highly variable, depending on metabolic demands of the various tissues that empty into the vein from where the sample is being drawn. Laboratory measurements of electrolytes include the CO_2 content and bicarbonate levels. The CO_2 content refers to the total CO_2 content of blood, including that contained in bicarbonate. It is determined by adding a strong acid to a plasma sample and measuring the amount of CO_2 generated. More than 70% of the CO_2 in the blood is in the form of bicarbonate.² The serum *bicarbonate* is then determined from the total CO_2 content of the blood.

Base excess or *deficit* is a measurement of bicarbonate excess or deficit. It describes the amount of a fixed acid or base that must be added to a blood sample to achieve a pH of 7.4 (normal \pm 3.0 mEq/L).³³ A base excess indicates metabolic alkalosis, and a base deficit indicates metabolic acidosis.

The *anion gap* describes the difference between the plasma concentration of the major measured cation (Na⁺) and the sum of the measured anions (Cl⁻ and HCO₃⁻). This difference represents the concentration of unmeasured anions, such as phosphates, sulfates, organic acids, and proteins (Fig. 6-19). Normally, the anion gap ranges between 8 and 12 mEq/L (a value of 16 mEq is normal if Na⁺ and K⁺ concentrations are used in the calculation). The anion gap is increased in conditions such as lactic acidosis and ketoacidosis that result from elevated levels of metabolic acids. A low anion gap is found in conditions that produce a fall in unmeasured anions (primarily albumin) or rise in unmeasured cations. An increase in unmeasured anions can occur in hyperkalemia, hypercalcemia, hypermagnesemia, lithium intoxication, or multiple myeloma, in which an abnormal immunoglobulin is produced.³⁴

Alterations in Acid-Base Balance

The terms *acidosis* and *alkalosis* describe the clinical conditions that arise as a result of changes in dissolved CO_2 and HCO_3^- concentrations. An alkali represents a combination of one or more alkali metals such as sodium or potassium with a highly basic ion such as a hydroxyl ion (OH⁻). Sodium bicarbonate (NaHCO₃) is the main alkali in the ECF. Although the definitions differ somewhat, the terms *alkali* and *base* are often used



■ FIGURE 6-19 ■ The anion gap in acidosis due to excess metabolic acids and excess serum chloride levels. Unmeasured anions such as phosphates, sulfates, and organic acids increase the anion gap because they replace bicarbonate. This assumes there is no change in sodium content.

KEY CONCEPTS

ACID-BASE DISORDERS

- The manifestations of acid-base disorders can be divided into three groups: (1) those due to the primary cause of the imbalance, (2) those due to the changed pH, and (3) those due to the elicited compensatory mechanisms.
- Metabolic acidosis represents a decrease in HCO₃⁻ that is caused by an excess of nonvolatile acids or loss of bicarbonate, and metabolic alkalosis represents an increase in HCO₃⁻ caused by a decrease in nonvolatile acids or increase in HCO₃⁻ intake or generation.
- Respiratory acidosis represents an increase in H₂CO₃ that is caused by respiratory conditions that interfere with the elimination of CO₂, and respiratory alkalosis represents a decrease in H₂CO₃ that is caused by excess elimination of CO₂.

interchangeably. Thus, the term *alkalosis* has come to mean the opposite of *acidosis*.

Metabolic Versus Respiratory Acid-Base Disorders

There are two types of acid-base disorders: metabolic and respiratory. Metabolic disorders produce an alteration in bicarbonate concentration and result from the addition or loss of nonvolatile acid or alkali to or from the ECF. A decrease in pH caused by a decrease in bicarbonate is called *metabolic acidosis*, and an elevated pH caused by increased bicarbonate levels is called *metabolic alkalosis*. Respiratory disorders involve changes in the CO₂, reflecting an increase or decrease in alveolar ventilation. *Respiratory acidosis* is characterized by a decrease in pH, reflecting a decrease in ventilation and an increase in CO₂. *Respiratory alkalosis* involves an increase in pH, resulting from an increase in alveolar ventilation and a decrease in CO₂.

Primary Versus Compensatory Mechanisms

Acidosis and alkalosis typically involve a *primary* or *initiating event* and a *compensatory state* that results from homeostatic mechanisms that attempt to correct or prevent large changes in pH. For example, a person may have a primary metabolic acidosis as a result of overproduction of ketoacids and respiratory alkalosis because of a compensatory increase in ventilation (see Table 6-8).

Compensatory mechanisms adjust the pH toward a more normal level without correcting the underlying cause of the disorder. Often, compensatory mechanisms are interim measures that permit survival while the body attempts to correct the primary disorder. Compensation requires the use of mechanisms that are different from those that caused the primary disorder. In other words, the lungs cannot compensate for respiratory acidosis that is caused by lung disease, nor can the kidneys compensate for metabolic acidosis that occurs because of renal failure. Compensatory mechanisms often become more effective with time; thus, there are differences between the level of pH change that occurs with acute and chronic acidbase disorders.

Metabolic Acidosis

Metabolic acidosis involves a primary deficit in base bicarbonate along with a decrease in plasma pH. In metabolic acidosis, the body compensates for the decrease in pH by increasing the respiratory rate in an effort to decrease CO₂ and H₂CO₃ levels.

Causes. Metabolic acidosis can be caused by one of four mechanisms: (1) increased production of nonvolatile metabolic acids, (2) decreased acid secretion by the kidney, (3) excessive loss of bicarbonate, or (4) an increase in chloride. Metabolic acids increase when there is an accumulation of lactic acid, overproduction of ketoacids, or ingestion of drugs (e.g., aspirin and salicylates) or chemicals (e.g., methanol or ethylene glycol) that results in production of metabolic acids or an inability of the kidneys to excrete metabolic acids or conserve bicarbonate. The anion gap is often useful in determining the cause of the metabolic acidosis (Chart 6-2). The presence of excess metabolic acids produces an increase in the anion gap as sodium bicarbonate is replaced by the sodium salt of the offending acid (e.g., sodium lactate). When acidosis results from increased chloride levels (e.g., hyperchloremic acidosis), the anion gap remains within normal levels.

Acute lactic acidosis is one of the most common types of metabolic acidosis. Lactic acid is produced by the anaerobic metabolism of glucose. Lactic acidosis develops when there is excess production of lactic acid or diminished lactic acid removal from the blood. Most cases of lactic acidosis are caused by inadequate oxygen delivery, as in shock or cardiac arrest. These conditions

TABLE 6-8S	Summary of Acid-Base Imbalances		
Acid-Base Imbala	nce Primary Disturbance	Respiratory Compensation	Renal Compensation
Metabolic acidosis	Decrease in bicarbonate	Hyperventilation to decrease PCO ₂	If no renal disease, increased H ⁺ excretion and increased HCO3 ⁻ reabsorption
Metabolic alkalosis	Increase in bicarbonate	Hypoventilation to increase PCO ₂	If no renal disease, decreased H ⁺ excretion and decreased HCO ₃ ⁻ reabsorption
Respiratory acidosis	Increase in PCO ₂	None	Increased H ⁺ excretion and increased HCO ₃ ⁻ reabsorption
Respiratory alkalosis	Decrease in PCO ₂	None	Decreased H ⁺ excretion and decreased HCO ₃ ⁻ reabsorption

CHART 6-2 The Anion Gap in Differential Diagnosis of Metabolic Acidosis

Decreased Anion Gap (<8 mEq/L)

Hypoalbuminemia (decrease in unmeasured anions) Multiple myeloma (increase in unmeasured cationic IgG paraproteins)

Increased unmeasured cations (hyperkalemia, hypercalcemia, hypermagnesemia, lithium intoxication)

Increased Anion Gap (>12 mEq/L)

Presence of unmeasured metabolic anion Diabetic ketoacidosis Alcoholic ketoacidosis Lactic acidosis Starvation Renal insufficiency Presence of drug or chemical anion Salicylate poisoning Methanol poisoning Ethylene glycol poisoning Normal Anion Gap (8–12 mEq/L)

Loss of bicarbonate Diarrhea Pancreatic fluid loss Ileostomy (unadapted) Chloride retention Renal tubular acidosis Ileal loop bladder Parenteral nutrition (arginine and lysine)

increase lactic acid production, and they impair lactic acid clearance because of poor liver perfusion. Excess lactate also is produced with vigorous exercise or grand mal seizures (*i.e.*, convulsions), during which there is a local disproportion between oxygen supply and demand in the contracting muscles.

Lactic acidosis is also associated with disorders in which tissue hypoxia does not appear to be present. It has been reported in patients with leukemia, lymphomas, or other cancers and in patients with severe liver failure. The mechanisms causing lactic acidosis in these conditions are poorly understood. Some conditions such as neoplasms may produce local increases in tissue metabolism and lactate production or may interfere with blood flow delivery to noncancerous cells.

Ketoacids (i.e., acetoacetic and β -hydroxybutyric acid), produced in the liver from fatty acids, are the source of fuel for many body tissues. An overproduction of ketoacids occurs when carbohydrate stores are inadequate or when the body cannot use available carbohydrates as a fuel. Under these conditions, fatty acids are mobilized from adipose tissue and delivered to the liver, where they are converted to ketones. Ketoacidosis develops when ketone production exceeds tissue use.

The most common cause of ketoacidosis is uncontrolled diabetes mellitus, in which an insulin deficiency leads to the release of fatty acids from adipose cells with subsequent production of excess ketoacids (see Chapter 32). Ketoacidosis may also develop as the result of fasting or food deprivation, during which the lack of carbohydrates produces a self-limited state of ketoacidosis (self-limited because the resultant decrease in insulin release suppresses the release of fatty acids for fat cells). Ketones are also formed during the oxidation of alcohol, a process that occurs in the liver. A condition called *alcoholic ketoacidosis* can develop in persons who engage in excess alcohol consumption.³⁵ It usually follows prolonged alcohol ingestion, particularly if accompanied by decreased food intake and vomiting that results in using fatty acids as an energy source.

Kidney failure is the most common cause of chronic metabolic acidosis. The kidneys normally conserve HCO_3^- and secrete H⁺ ions into the urine as a means of regulating acid-base balance. In renal failure, there is loss of glomerular and tubular function, with retention of nitrogenous wastes and metabolic acids. In a condition called *renal tubular acidosis*, glomerular function is normal, but the tubular secretion of H⁺ or reabsorption of HCO_3^- is abnormal.

Increased HCO₃⁻ losses occur with the loss of bicarbonaterich body fluids or with impaired conservation of HCO₃⁻ by the kidney. Intestinal secretions have a high HCO₃⁻ concentration. Consequently, excessive losses of HCO₃⁻ ions occur with severe diarrhea; small bowel, pancreatic, or biliary fistula drainage; ileostomy drainage; and intestinal suction. In diarrhea of microbial origin, HCO₃⁻ is secreted into the bowel to neutralize the metabolic acids produced by the microorganisms causing the diarrhea.

Hyperchloremic acidosis occurs when serum Cl⁻ ion levels are increased. Because Cl⁻ and HCO₃⁻ are anions, the serum HCO₃⁻ ion concentration decreases when there is an increase in Cl⁻ ions. Hyperchloremic acidosis can occur as the result of abnormal absorption of chloride by the kidneys or as a result of treatment with chloride-containing medications (*i.e.*, sodium chloride, amino acid-chloride hyperalimentation solutions, and ammonium chloride). The administration of intravenous sodium chloride or parenteral hyperalimentation solutions that contain an amino acid-chloride combination can cause acidosis in a similar manner.³⁶ With hyperchloremic acidosis, the anion gap is within the normal range, but the chloride levels are increased and bicarbonate levels are decreased.

Manifestations. Metabolic acidosis is characterized by a decrease in pH (<7.35) and a decrease in serum HCO_3^- levels (<24 mEq/L). Acidosis typically produces a compensatory increase in respiratory rate with a decrease in PCO_2 and H_2CO_3 (see Table 6-9).

The manifestations of metabolic acidosis fall into three categories: (1) signs and symptoms of the disorder causing the acidosis, (2) alterations in function resulting from the decreased pH, and (3) changes in body function related to recruitment of compensatory mechanisms. The signs and symptoms of metabolic acidosis usually begin to appear when the plasma $HCO_3^$ concentration falls to 20 mEq/L or less. Metabolic acidosis is seldom a primary disorder; the manifestations of the disorder are frequently superimposed on the symptoms of the contributing health problem. With diabetic ketoacidosis, which is a common cause of metabolic acidosis, there is an increase in blood and urine glucose and a characteristic smell of ketones to the breath. In metabolic acidosis that accompanies renal failure, blood urea nitrogen levels are elevated, and tests of renal function yield abnormal results.

Changes in pH have a direct effect on body function that can produce signs and symptoms common to most types of meta-

TABLE 6-9 Manifestations of Metabolic Acidosis and Alkalosis		
Metabolic Acidosis	Metabolic Alkalosis	
Laboratory Tests pH decreased Bicarbonate (primary) decreased PCO ₂ (compensatory) decreased	Laboratory Tests pH increased Bicarbonate (primary) increased PCO ₂ (compensatory) increased	
Signs of Compensation Increased respirations (rate and depth) Hyperkalemia Acid urine Increased ammonia in urine	Signs of Compensation Decreased respirations (rate and depth) with various degrees of hypoxia and respiratory acidosis	
Gastrointestinal Effects Anorexia Nausea and vomiting Abdominal pain		
Nervous System Effects Weakness Lethargy Confusion Stupor Coma Depression of vital functions	Nervous System Effects Hyperactive reflexes Tetany Confusion Seizures	
Cardiovascular Effects Peripheral vasodilation Decreased heart rate Cardiac dysrhythmias	Cardiovascular Effects Hypotension Cardiac dysrhythmias	
<mark>Skin</mark> Warm and flushed		
Skeletal System Effects Bone disease (chronic acidosis)		

bolic acidosis, regardless of cause. A person with metabolic acidosis often reports weakness, fatigue, general malaise, and a dull headache. Anorexia, nausea, vomiting, and abdominal pain also may be reported. The skin is warm and flushed, and there is a decrease in tissue turgor when fluid deficit accompanies acidosis. In persons with undiagnosed diabetes mellitus, the nausea, vomiting, and abdominal symptoms may be misinterpreted as being caused by gastrointestinal flu or other abdominal disease, such as appendicitis.

Neural activity becomes depressed as body pH declines. Acidosis directly depresses membrane excitability, and it decreases binding of calcium to plasma proteins so that more free calcium is available to decrease neural activity. As acidosis progresses, the level of consciousness declines, and stupor and coma develop. The skin becomes warm and flushed as the cutaneous blood vessels become less responsive to sympathetic stimulation and lose tone.

When the pH falls to 7.0, cardiac contractility and cardiac output decrease, the heart becomes less responsive to catecholamines (*i.e.*, epinephrine and norepinephrine), and dysrhythmias, including fatal ventricular dysrhythmias, can develop.

Metabolic acidosis also is accompanied by signs and symptoms related to the recruitment of compensatory mechanisms. In situations of acute metabolic acidosis, the respiratory system compensates for a decrease in pH by increasing ventilation to reduce PCO₂; this is accomplished through deep and rapid respirations. There may be complaints of dyspnea with exertion; with severe acidosis, dyspnea may be present even at rest.

When kidney function is normal, net acid excretion increases promptly in response to acidosis, and the urine becomes more acid. Most of the initial acid secretion into the urine is facilitated through use of the phosphate buffer system. Over several days, ammonia production by the kidney increases and becomes the most important mechanism for excreting excess H⁺ ions.

Chronic acidemia, as in renal failure, can lead to a variety of skeletal problems, some of which result from the release of calcium and phosphate during bone buffering of excess H⁺ ions.³⁷ Of particular importance is impaired growth in children. In infants and children, acidemia may be associated with a variety of nonspecific symptoms such as anorexia, weight loss, muscle weakness, and listlessness.^{33,37}

Treatment. The treatment of metabolic acidosis focuses on correcting the condition that caused the disorder and restoring the fluids and electrolytes that have been lost from the body. The treatment of diabetic ketoacidosis is discussed in Chapter 32.

Metabolic Alkalosis

Metabolic alkalosis involves a primary excess of base bicarbonate along with an increased plasma pH. It can be caused by a gain in HCO_3^- or loss of H⁺ ions. The body compensates for the increase in pH by decreasing the respiratory rate as a means of increasing PCO₂ and H₂CO₃ levels.

Causes. Metabolic alkalosis can occur because of (1) ingestion or administration of excess NaHCO₃ or other alkali, (2) excess H^+ , Cl^- , and K^+ loss, or (3) ECF volume contraction.

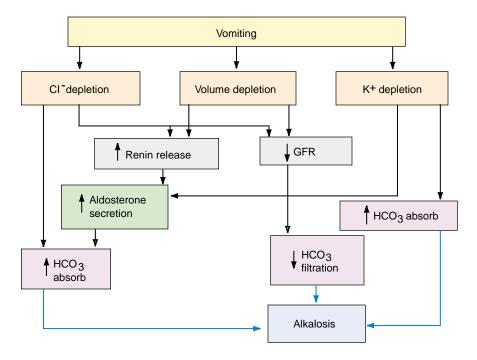
Excessive alkali ingestion, as in the use of bicarbonatecontaining antacids or sodium bicarbonate administration during cardiopulmonary resuscitation, can cause metabolic alkalosis. Other sources of alkali intake are acetate in hyperalimentation solutions and lactate in parenteral solutions such as Ringer's lactate. A condition called the milk-alkali syndrome may develop in persons who consume excessive amounts of milk (Ca⁺⁺ source) along with alkaline antacids. In this case metabolic alkalosis develops as a consequence of vomiting (volume depletion and K⁺ loss), hypercalcemia (increased HCO_3^- reabsorption), and reduced glomerular filtration rate (increased HCO_3^- reabsorption).^{34,38}

Vomiting, removal of gastric secretion through use of nasogastric suction, and low serum potassium levels resulting from hyperaldosteronism or diuretic therapy are the most common causes of metabolic alkalosis in hospitalized patients. The binge-purge syndrome, or self-induced vomiting, is also associated with metabolic alkalosis.³⁸ Gastric secretions contain high concentrations of HCl and lesser concentrations of K⁺. As Cl⁻ is taken from the blood and secreted into the stomach with H⁺, it is replaced by HCO₃⁻. Normally, the increase in serum HCO₃⁻ concentration is only transient because the entry of acid into the duodenum stimulates an equal secretion of pancreatic HCO₃⁻. However, the continued loss of H⁺ and Cl⁻ ions from the stomach because of vomiting or gastric suction stimulates continued production of gastric acid and thus the addition of more bicarbonate into the blood. Metabolic alkalosis can also result from the use of diuretics (*e.g.*, loop and thiazide diuretics) and other conditions that cause excessive loss of potassium in the urine. Hypokalemia produces a compensatory increase in K⁺ reabsorption and H⁺ secretion by the kidney, along with a simultaneous increase in HCO_3^- reabsorption.³⁴ The adrenocorticosteroid hormone, aldosterone, increases H⁺ ion secretion as it increases Na⁺ and HCO_3^- ion reabsorption. In hyperaldosteronism, the concurrent loss of K⁺ in the urine serves to perpetuate the alkalosis.

Chronic respiratory acidosis produces a compensatory loss of H⁺ and Cl⁻ ions in the urine along with HCO_3^- retention. When respiratory acidosis is corrected abruptly, as with mechanical ventilation, a "posthypercapnic" metabolic alkalosis may develop because of a rapid drop in PCO₂, while the HCO_3^- concentration, which requires renal elimination, remains elevated.

Vomiting results in the loss of water, Na⁺, Cl⁻, and K⁺. The resultant volume depletion, hypochloremia, and hypokalemia produce a metabolic alkalosis by increasing renal reabsorption of HCO_3^- (Fig. 6-20). Volume depletion and hypokalemia also increase the activity of the renin-angiotensin-aldosterone system, with increased Na⁺ reabsorption. Na⁺ reabsorption requires a concomitant anion reabsorption. Because of a Cl⁻ deficit, HCO_3^- is reabsorbed along with Na⁺, contributing to the development of metabolic alkalosis.

Manifestations. Metabolic alkalosis is characterized by a plasma pH greater than 7.45, plasma HCO_3^- level greater than 29 mEq/L, and base excess greater than 3.0 mEq/L (Table 6-9). Persons with metabolic alkalosis often have no symptoms or have signs related to volume depletion or hypokalemia. The neurologic signs (*e.g.*, hyperexcitability) occur less frequently with metabolic alkalosis than with other acid-base disorders because the HCO_3^- ion enters the cerebrospinal fluid more slowly than does CO_2 . When neurologic manifestations occur, as in acute and severe metabolic alkalosis, they include mental confusion, hyperactive reflexes, tetany, and carpopedal spasm. Metabolic alkalosis also leads to a compensatory hypoventila-



■ FIGURE 6-20 ■ The mechanisms by which vomiting maintains metabolic alkalosis. See text for explanation. (Adapted from Galla J.H., Luke R.G. [1987]. Pathophysiology of metabolic alkalosis. *Hospital Practice* 22 [10], 130)

tion, with the development of various degrees of hypoxemia and respiratory acidosis. Significant morbidity occurs with severe metabolic alkalosis (pH >7.55), including respiratory failure, dysrhythmias, seizures, and coma.

Treatment. The treatment of metabolic alkalosis usually is directed toward correcting the cause of the condition. A chloride deficit requires correction. Potassium chloride usually is the treatment of choice for metabolic alkalosis when there is an accompanying potassium deficit. When potassium chloride is used as a therapy, the chloride anion replaces the bicarbonate anion, and the administration of potassium corrects the potassium deficit and allows the kidneys to conserve H+ while eliminating the K⁺. Fluid replacement with normal saline or onehalf normal saline often is used in the treatment of patients with volume contraction alkalosis.

Respiratory Acidosis

Respiratory acidosis involves an increase in PCO₂ and H₂CO₃ along with a decrease in pH. Acute respiratory failure is associated with severe acidosis and only a small change in serum bicarbonate levels. Within a day or so, renal compensatory mechanisms begin to conserve and generate HCO₃⁻ ions. In chronic respiratory acidosis, there is a compensatory increase in bicarbonate levels.39

Causes. Respiratory acidosis occurs in conditions that impair alveolar ventilation and cause an increase in PCO2. It can occur as an acute or chronic disorder. Because renal compensatory mechanisms take time to exert their effects, blood pH tends to drop sharply in persons with acute respiratory acidosis.

Acute respiratory acidosis can be caused by impaired function of the respiratory center in the medulla (as in narcotic overdose), lung disease, chest injury, weakness of the respiratory muscles, or airway obstruction. Acute respiratory acidosis can also result from breathing air with a high CO₂ content. Almost all persons with acute respiratory acidosis are hypoxemic if they are breathing room air. In many cases, signs of hypoxemia develop before those of respiratory acidosis because CO₂ diffuses across the alveolar capillary membrane 20 times more rapidly than does oxygen.^{33,34}

Chronic respiratory acidosis is associated with chronic lung diseases such as chronic bronchitis and emphysema (see Chapter 21). In people with these disorders, the persistent elevation of PCO₂ stimulates renal H⁺ ion secretion and HCO₃⁻ reabsorption. The effectiveness of these compensatory mechanisms can often return the pH to near-normal values as long as oxygen levels are maintained within a range that does not unduly suppress the chemoreceptor-mediated control of ventilation.

An acute episode of respiratory acidosis can develop in persons with chronic lung disease who receive oxygen therapy at a flow-rate sufficient to raise the PO₂ to a level that produces a decrease in ventilation. In these persons, the respiratory center has become adapted to the elevated levels of CO₂ and no longer responds to increases in PCO₂. Instead, the oxygen content of their blood becomes the major stimulus for respiration. If oxygen is administered at a flow rate that is sufficient to suppress this stimulus, the rate and depth of respiration decrease, and the CO₂ content of the blood increases.

Manifestations. Respiratory acidosis is associated with a plasma pH less than 7.35 and an arterial PCO₂ greater than 50 mm Hg (Table 6-10). The signs and symptoms of respiratory acidosis depend on the rapidity of onset and whether the condition is acute or chronic. Because respiratory acidosis often

TABLE 6-10	Manifestations of Respiratory Acidosis and Alkalosis		
Respiratory Acidosis		Respiratory Alkalosis	
Laboratory Test pH decreased PCO ₂ (primary) in Bicarbonate (com		Laboratory Tests pH increased PCO ₂ (primary) decreased Bicarbonate (compensatory) decreased	
Signs of Compe Acid urine	nsation	Signs of Compensation Alkaline urine	
Nervous System Dilation of cerebr Headache Behavioral cha Confusion Depression Paranoia Hallucinati Weakness Tremors Paralysis Stupor and co	al vessels and decreased neuronal activity inges i	Nervous System Effects Constriction of cerebral vessels and increased neuronal activity Dizziness, panic, light-headedness Tetany Numbness and tingling of fingers and toes Seizures (severe)	
Skin Warm and flushed	ł	Cardiovascular Effects Cardiac dysrhythmias	

117

is accompanied by hypoxemia, the manifestations of respiratory acidosis often are intermixed with those of oxygen deficit.

Sudden elevations in arterial CO_2 can cause headache, blurred vision, an increase in heart rate and blood pressure, warm and flushed skin, irritability, muscle twitching, and psychological disturbances. Carbon dioxide readily crosses the blood-brain barrier, exerting its effects by changing the pH of brain fluids. Elevated levels of CO_2 produce vasodilation of cerebral blood vessels. If the condition is severe and prolonged, it can cause an increase in cerebrospinal fluid pressure and papilledema. Impaired consciousness, ranging from lethargy to coma, develops as the PCO₂ rises. Paralysis of extremities may occur, and there may be respiratory depression.

Treatment. The treatment of acute and chronic respiratory acidosis is directed toward improving ventilation (see Chapter 21). In severe cases, mechanical ventilation may be necessary.

Respiratory Alkalosis

Respiratory alkalosis involves a decrease in PCO_2 and a primary deficit in carbonic acid (H_2CO_3) along with an increase in pH. Because respiratory alkalosis can occur suddenly, a compensatory decrease in HCO_3^- may not occur before respiratory correction has been accomplished. The increase in pH is less in chronic compensated respiratory alkalosis, and the fall in HCO_3^- is greater.

Causes. Respiratory alkalosis is caused by hyperventilation or a respiratory rate in excess of that needed to maintain normal PCO_2 levels. One of the most common causes of respiratory alkalosis is the hyperventilation syndrome, which is characterized by recurring episodes of overbreathing often associated with anxiety. Persons experiencing panic attacks frequently present in the emergency room with acute respiratory alkalosis. Other causes of hyperventilation are fever, hypoxemia, early salicylate toxicity, and encephalitis. Hyperventilation can also occur during anesthesia or with use of mechanical ventilatory devices. Hypoxemia exerts its effect through stimulation of the peripheral chemoreceptors. Salicylate toxicity and encephalitis produce hyperventilation by directly stimulating the respiratory center in the brain stem.

Manifestations. Respiratory alkalosis is characterized by a pH greater than 7.45, arterial PCO₂ less than 35 mm Hg, and serum HCO₃⁻ levels less than 24 mEq/L (Table 6-10). The signs and symptoms of respiratory alkalosis are mainly associated with hyperexcitability of the nervous system and a decrease in cerebral blood flow. Alkalosis increases protein binding of serum calcium. This reduces ionized serum calcium levels, causing an increase in neuromuscular excitability. A decrease in the CO₂ content of the blood causes constriction of cerebral blood vessels. Because CO₂ crosses the blood-brain barrier rather quickly, the manifestations of acute respiratory alkalosis often are of sudden onset. The person often experiences light-headedness, dizziness, tingling, and numbness of the fingers and toes. These manifestations may be accompanied by sweating, palpitations, panic, air hunger, and dyspnea. Chvostek's and Trousseau's signs may be positive, and tetany and convulsions may occur. Because CO₂ provides the stimulus for short-term regulation of respiration, short periods of apnea may occur in persons with acute episodes of hyperventilation.

Treatment. The treatment of respiratory alkalosis focuses on measures to increase the PCO₂. Attention is directed toward correcting the disorder that caused the overbreathing. Rebreathing of small amounts of expired air (breathing into a paper bag) may prove useful in restoring PCO₂ levels in persons with anxiety-produced respiratory alkalosis.

In summary, normal body function depends on the precise regulation of acid-base balance to maintain a pH within the narrow physiologic range of 7.35 to 7.45. Metabolic processes produce volatile and nonvolatile metabolic acids that must be buffered and eliminated from the body. The volatile acid H₂CO₃ is in equilibrium with dissolved CO₂, which is eliminated through the lungs. The nonvolatile metabolic acids, most of which are excreted by the kidneys, are derived mainly from protein metabolism and incomplete carbohydrate and fat metabolism. It is the ratio of the bicarbonate ion concentration to dissolved CO₂ (carbonic acid concentration) that determines body pH. When this ratio is 20:1, the pH is 7.4. The ability of the body to maintain pH within the normal physiologic range depends on respiratory and renal mechanisms and on intracellular and extracellular buffers; the most important of these is the bicarbonate buffer system. The respiratory regulation of pH is rapid but does not return pH completely to normal. The kidney aids in regulation of pH by eliminating H⁺ ions or conserving HCO₃⁻ ions. In the process of eliminating H⁺ ions, it uses the phosphate and ammonia buffer systems. Body pH is also affected by the distribution of exchangeable cations (K^+ and H^+) and anions (Cl^- and HCO_3^-). Metabolic acidosis is defined as a decrease in HCO₃⁻, and metabolic alkalosis as an increase in HCO₃⁻. Metabolic acidosis is caused by an excessive production and accumulation of metabolic acids or an excessive loss of HCO₃⁻.

Metabolic alkalosis is caused by an increase in HCO_3^- or from a decrease in H^+ or CI^- , with a resultant increase in renal reabsorption of HCO_3^- . Respiratory acidosis reflects a decrease in pH that is caused by conditions that produce hypoventilation, with a resultant increase in PCO_2 levels. Respiratory alkalosis reflects an increase in pH that is caused by conditions that produce hyperventilation, with a resultant decrease in PCO_2 levels.

The signs and symptoms of acidosis and alkalosis reflect alterations in body function associated with the disorder causing the acid-base disturbance, the effect of the change of pH on body function, and the body's attempt to correct and maintain the pH within a normal physiologic range. In general, neuromuscular excitability is decreased in acidosis and increased in alkalosis.

REVIEW QUESTIONS

Describe factors that control fluid exchange between the vascular and interstitial fluid compartments and relate them to the development of edema and third spacing of extracellular fluids.

Compare and contrast the causes, manifestations, and treatment of isotonic fluid volume deficit, isotonic fluid volume excess, hyponatremia with water excess, and hypernatremia with water deficit.

Characterize the distribution of potassium in the body and explain how extracellular potassium levels are regulated in relation to body gains and losses.

Persons with severe hyperkalemia are at risk for cardiac arrest. Explain the effects of potassium on membrane potentials that are responsible for this risk.

A person may have hypocalcemia according to laboratory reports but have no clinical signs of the disorder. Explain.

Persons with renal failure experience both hyperphosphatemia and hypocalcemia. Explain.

Describe the three forms of carbon dioxide transport and their contribution to acid-base balance.

Use the Henderson-Hasselbalch equation to calculate pH and compare compensatory mechanisms for regulating pH.

Describe a clinical situation involving an acid-base disorder in which primary and compensatory mechanisms might be active.

Define metabolic acidosis, metabolic alkalosis, respiratory acidosis, and respiratory alkalosis.

Explain the use of the plasma anion gap in differentiating types of metabolic acidosis.

connection-

Visit the Connection site at connection.lww.com/go/porth for links to chapter-related resources on the Internet.

REFERENCES

- 1. Krieger J.N., Sherrad D.J. (1991). Practical fluid and electrolytes (pp. 104-105). Norwalk, CT: Appleton & Lange.
- 2. Guyton A., Hall J.E. (2000). Textbook of medical physiology (10th ed., pp. 157-171, 264-278, 322-345, 346-363, 820-826). Philadelphia: W.B. Saunders.
- 3. Cogan M.G. (1991). Fluid and electrolytes (pp. 1, 43, 80-84, 100-111, 112-123, 125-130, 242-245). Norwalk, CT: Appleton & Lange.
- 4. Stearns R.H., Spital A., Clark E.C. (1996). Disorders of water balance. In Kokko J., Tannen R.L. (Eds.), Fluids and electrolytes (3rd ed., pp. 65, 69, 95). Philadelphia: W.B. Saunders.
- 5. Metheney N.M. (2000). Fluid and electrolyte balance (4th ed., pp. 3, 16, 18, 43, 47, 56, 162, 256). Philadelphia: Lippincott Williams & Wilkins.
- 6. Porth C.J.M., Erickson M. (1992). Physiology of thirst and drinking: Implications for nursing practice. Heart and Lung 21, 273-284.
- 7. Ayus J.C., Arieff A.I. (1996). Abnormalities of water metabolism in the elderly. Seminars in Nephrology 16 (4), 277-288.
- 8. Kugler J.P., Hustead T. (2000). Hyponatremia and hypernatremia in the elderly. American Family Physician 61, 3623-3630.
- 9. Illowsky B.P., Kirch D.G. (1988). Polydipsia and hyponatremia in psychiatric patients. American Journal of Psychiatry 145, 675-683.
- 10. Vieweg W.V.R. (1994). Treatment strategies for polydipsiahyponatremia syndrome. Journal of Clinical Psychiatry 55 (4), 154–159.
- 11. Rose B.D., Post T.W. (2001). Clinical physiology of acid-base and electrolyte disorders (5th ed., pp. 168-178, 822, 835, 858, 909). New York: McGraw-Hill.
- 12. Berne R.M., Levy M. (2000). Principles of physiology (3rd ed., p. 438). St. Louis: Mosby.

- 13. Robertson G.L. (1983). Thirst and vasopressin function in normal and disordered states of water balance. Journal of Laboratory and Clinical Medicine 101, 351-371.
- 14. Robertson G.L. (1995). Diabetes insipidus. Endocrinology and Metabolic Clinics of North America 24, 549-571.
- 15. Bendz H., Aurell M. (1999). Drug-induced diabetes insipidus. Drug Safety 21, 449-456.
- 16. Batchell J. (1994). Syndrome of inappropriate antidiuretic hormone. Critical Care Clinics of North America 69, 687-691.
- 17. Fried L.F., Palevsky P.M. (1997). Hyponatremia and hypernatremia. Medical Clinics of North America 81, 585–606.
- 18. Androgue H.J., Madias N.E. (2000). Hyponatremia. New England Journal of Medicine 343, 1581-1589.
- 19. Oh M.S., Carroll H.J. (1992). Disorders of sodium metabolism: Hypernatremia and hyponatremia. Critical Care Medicine 20, 94-103.
- 20. Androgue H.J. (2000). Hypernatremia. New England Journal of Medicine 342, 1493-1499.
- 21. Behrman R.E., Kliegman R.M., Jenson H.B. (2000). Nelson textbook of pediatrics (16th ed., pp. 215-218). Philadelphia: W.B. Saunders.
- 22. Bräxmeyer D.L., Keyes J.L. (1996). The pathophysiology of potassium balance. Critical Care Nurse 16 (5), 59-71.
- 23. Gennari F.J. (1998). Hypokalemia. New England Journal of Medicine 339, 451-458.
- 24. Tannen R.L (1996). Potassium disorders. In Kokko J., Tannen R.L. (Eds.), Fluids and electrolytes (3rd ed., pp. 116-118). Philadelphia: W.B. Saunders.
- 25. Whang G., Whang G.G., Ryan M.P. (1992). Refractory potassium repletion: A consequence of magnesium deficiency. Archives of Internal Medicine 152 (1), 40-45.
- 26. Clark B.A., Brown R.S. (1995). Potassium homeostasis and hyperkalemic syndromes. Endocrinology and Metabolic Clinics of North America 24, 573-590.
- 27. Zaloga G.F. (1992). Hypocalcemia in critically ill patients. Critical Care Medicine 20, 251-262.
- 28. Yucha C.B., Toto K.H. (1994). Calcium and phosphorous derangements. Critical Care Clinics of North America 6, 747-765.
- 29. Barnett M.L. (1999). Hypercalcemia. Seminars in Oncology Nursing 15, 190-201.
- 30. Workman L. (1992). Magnesium and phosphorus: The neglected electrolytes. AACN Clinical Issues 3, 655-663.
- 31. Swain R., Kaplan-Machlis B. (1999). Magnesium for the next millennium. Southern Medical Journal 92, 1040-1046.
- 32. Toto K., Yucha C.B. (1994). Magnesium: Homeostasis, imbalances, and therapeutic uses. Critical Care Nursing Clinics of North America 6, 767-778.
- 33. Rose B.D., Post T.W. (2001). Clinical physiology of acid-base and electrolyte disorders (5th ed., pp. 325, 363, 578-646, 647, 669). New York: McGraw-Hill.
- 34. Abelow B. (1998). Understanding acid-base (pp. 43-49, 83-93, 139-169, 171-188, 189-198, 224-230). Baltimore: Williams & Wilkins.
- 35. Umpierrez G.E., DiGirolamo M., Tuvlin J.A., et. al. (2000). Differences in metabolic and hormonal milieu in diabetic and alcoholinduced ketoacidosis. Journal of Critical Care 15 (2), 52-59.
- 36. Powers F. (1999). The role of chloride in acid-base balance. Journal of Intravenous Nursing 22, 286-291.
- 37. Alpern R.J., Sakhaee K. (1997). The clinical spectrum of chronic metabolic acidosis: Homeostatic mechanisms produce significant morbidity. American Journal of Kidney Diseases 29, 291–302.
- 38. Galla J.H. (2000). Metabolic alkalosis. Journal of the American Society of Nephrology 11, 369-375.
- 39. Adrogue H.J., Madias N.E. (1998). Management of life-threatening acid-base disorders. New England Journal of Medicine 338, 107-111.

Dr Murtadha Alshareifi e-Library

TWO

Alterations in Body Defenses

Stress and Adaptation

Homeostasis

CHAPTER

Constancy of the Internal Environment Control Systems Feedback Systems Stress

The Stress Response Neuroendocrine-Immune Interactions Neuroendocrine Responses Immune Responses Coping and Adaptation to Stress Adaptation Factors Affecting the Ability to Adapt Disorders of the Stress Response Effects of Acute Stress Effects of Chronic Stress Post-traumatic Stress Disorder Treatment of Stress Disorders

Stress has become an increasingly discussed topic in today's world. The concept is discussed extensively in the health care fields, and it is found as well in economics, political science, business, and education. At the level of the popular press, the term is exploited with messages about how stress can be prevented, managed, and even eliminated.

Whether stress is more prevalent today than it was in centuries past is uncertain. Certainly, the pressures that existed were equally challenging, although of a different type. Social psychologists Richard Lazarus and Susan Folkman related that as early as the 14th century the term was used to indicate hardship, straits, adversity, or affliction.¹ In the 17th century, *stress* and related terms appeared in the context of physical sciences: *load* was defined as an external force, *stress* as the ratio of internal force created by the load to the area over which the force acted, and *strain* was the deformation or distortion of the object.¹ These concepts are still used in engineering today.

The concepts of stress and strain survived, and throughout the 19th and early 20th centuries, stress and strain were thought to be the cause of "ill health" and "mental disease."² By the 20th century, stress had drawn considerable attention both as a health concern and as a research focus. In 1910, when Sir William Osler delivered his Lumleian Lectures on "angina pectoris," he described the relationship of stress and strain to angina pectoris.³ Approximately 15 years later, Walter Cannon, well known for his work in physiology, began to use the word *stress* in relation to his laboratory experiments on the "fight-orflight" response. It seems possible that the term emerged from his work with the homeostatic features of living organisms and their tendency to "bound back" and "resist disruption" when acted on by an "external force."⁴ At about the same time, Hans Selye, who became known for his research and publications on stress, began using the term *stress* in a very special way to mean an orchestrated set of bodily responses to any form of noxious stimulus.⁵

The content in this chapter has been organized into three sections: homeostasis, the stress response and adaptation to stress, and the acute and chronic effects of stress.

HOMEOSTASIS

The concepts of stress and adaptation have their origin in the complexity of the human body and the interactions between the body's cells and its many organ systems. The body requires that a level of homeostasis or constancy be maintained during the many changes that occur in the internal and external environments. Stress and adaptation involve feedback control systems that regulate cellular function and integrate the function of the different body systems.

Constancy of the Internal Environment

The environment in which body cells live is not the external environment that surrounds the organism, but rather the local fluid environment that surrounds each cell. Claude Bernard, a 19th century physiologist, was the first to describe clearly the central importance of a stable internal environment, which he termed the milieu intèrieur. Bernard recognized that body fluids surrounding the cells and the various organ systems provide the means for exchange between the external and the internal environments. It is from this internal environment that body cells receive their nourishment, and it is into this fluid that they secrete their wastes. Even the contents of the gastrointestinal tract and lungs do not become part of the internal environment until they have been absorbed into the extracellular fluid. A multicellular organism is able to survive only as long as the composition of the internal environment is compatible with the survival needs of the individual cells. For example, even a small change in the pH of the body fluids can disrupt the metabolic processes of individual cells.

The concept of a stable internal environment was supported by Walter B. Cannon. He proposed that this kind of stability, which he called *homeostasis*, was achieved through a system of carefully coordinated physiologic processes that oppose change.⁶ Cannon pointed out that these processes were largely automatic and emphasized that homeostasis involves resistance to both internal and external disturbances.

In his book *The Wisdom of the Body*, published in 1939, Cannon presented four tentative propositions to describe the general features of homeostasis.⁶ With this set of propositions, Cannon emphasized that when a factor is known to shift homeostasis in one direction, it is reasonable to expect the existence of mechanisms that have the opposite effect. For example, in the homeostatic regulation of blood sugar, mechanisms that both raise and lower blood sugar levels would be expected to play a part. As long as the mechanism responding to the initiating disturbance can recover homeostasis, the integrity of the body and the status of normality are retained.

KEY CONCEPTS

HOMEOSTASIS

- Homeostasis is the purposeful maintenance of a stable internal environment maintained by coordinated physiologic processes that oppose change.
- The physiologic control systems that oppose change operate by negative feedback mechanisms that are composed of a sensor that detects a change, an integrator/comparator that sums and compares incoming data with a set point, and an effector system that returns the sensed function to within the range of the set point.

Control Systems

The ability of the body to function and maintain homeostasis under conditions of change in the internal and external environment depends on the thousands of physiologic *control systems* that regulate body function. A homeostatic control system is a collection of interconnected components that function to keep a physical or chemical parameter of the body relatively constant. The body's control systems regulate cellular function, control life processes, and integrate functions of the different organ systems.

Of recent interest have been the neuroendocrine control systems that influence behavior. Biochemical messengers that exist in our brain control nerve activity, information flow, and ultimately, behavior.⁷ These control systems function in producing the emotional reactions to stressors. In persons with mental health disorders, they can interact in the production of

Constancy of the Internal Environment

- Constancy in an open system, such as our bodies represent, requires mechanisms that act to maintain this constancy. Cannon based this proposition on insights into the ways by which steady states such as glucose concentrations, body temperature, and acid-base balance were regulated.
- Steady-state conditions require that any tendency toward change automatically meet with factors that resist change. An increase in blood sugar results in thirst as the body attempts to dilute the concentration of sugar in the extracellular fluid.
- 3. The regulating system that determines the homeostatic state consists of a number of cooperating mechanisms acting simultaneously or successively. Blood sugar is regulated by insulin, glucagon, and other hormones that control its release from the liver or its uptake by the tissues.
- Homeostasis does not occur by chance, but is the result of organized self-government.
- (Cannon W.B. [1932]. *The wisdom of the body* (pp. 299–300). New York: W.W. Norton)

symptoms associated with the disorder. The field of neuropharmacology has focused on the modulation of the endogenous messengers and signaling systems that control behavior in the treatment of mental disorders such as anxiety disorders, depression, and schizophrenia.

Feedback Systems

Most control systems in the body operate by *negative feedback mechanisms*, which function in a manner similar to the thermostat on a heating system. When the monitored function or value decreases below the set point of the system, the feedback mechanism causes the function or value to increase, and when the function or value is increased above the set point, the feedback mechanism causes it to decrease (Fig. 7-1). For example, in the negative feedback mechanism that controls blood glucose levels, an increase in blood glucose stimulates an increase in insulin, which enhances the removal of glucose from the blood. When sufficient glucose has left the bloodstream to cause blood glucose levels to fall, insulin secretion is inhibited and glucagon and other counter regulatory mechanisms stimulate the release of glucose from the liver, which causes the blood glucose level to return to normal.

The reason most physiologic control systems function under negative rather than *positive feedback mechanisms* is that a positive feedback mechanism interjects instability rather than stability into a system. It produces a cycle in which the initiating stimulus produces more of the same. For example, in a positive feedback system, exposure to an increase in environmental temperature would invoke compensatory mechanisms designed to increase, rather than decrease, body temperature.

In summary, physiologic and psychological adaptation involves the ability to maintain the constancy of the internal environment (homeostasis) and behavior in the face of a wide range of changes in the internal and external environments. It involves negative feedback control systems that regulate cellular function, control life's processes, regulate behavior, and integrate the function of the different body systems.

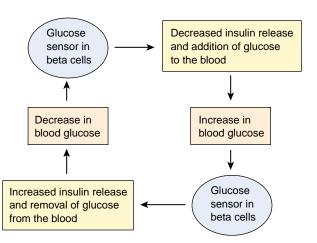


FIGURE 7-1 Illustration of negative feedback control mechanisms using blood glucose as an example.

STRESS

The increased focus on health promotion has heightened interest in the roles of stress and biobehavioral stress responses in the development of disease.⁸ Stress may contribute directly to the production or exacerbation of a disease, or it may contribute to the development of behaviors such as smoking, overeating, and drug abuse that increase the risk of disease.⁹

The Stress Response

In the early 1930s, the world-renowned endocrinologist Hans Selve was the first to describe a group of specific anatomic changes that occurred in rats that were exposed to a variety of different experimental stimuli. He came to an understanding that these changes were manifestations of the body's attempt to adapt to stimuli. Selve described stress as "a state manifested by a specific syndrome of the body developed in response to any stimuli that made an intense systemic demand on it."¹⁰ As a young medical student, Selve noticed that patients with diverse disease conditions had many signs and symptoms in common. He observed that "whether a man suffers from a loss of blood, an infectious disease, or advanced cancer, he loses his appetite, his muscular strength, and his ambition to accomplish anything; usually the patient also loses weight and even his facial expression betrays that he is ill."¹¹ Selve referred to this as the "syndrome of just being sick."

In his early career as an experimental scientist, Selye noted a triad of adrenal enlargement, thymic atrophy, and gastric ulcer appeared in rats he was using for his studies. These same three changes developed in response to many different or nonspecific experimental challenges. He assumed that the hypothalamic-pituitary-adrenal (HPA) axis played a pivotal role in the development of this response. To Selye, the response to stressors was a process that enabled the rats to resist the experimental challenge by using the function of the system best able to respond to it. He labeled the response the general adaptation syndrome (GAS): general because the effect was a general systemic reaction, adaptive because the response was in reaction to a stressor, and syndrome because the physical manifestations were coordinated and dependent on each other.¹⁰

According to Selve, the GAS involves three stages: the alarm stage, the stage of resistance, and the stage of exhaustion. The *alarm stage* is characterized by a generalized stimulation of the sympathetic nervous system and the HPA axis, resulting in the release of catecholamines and cortisol. During the resistance stage, the body selects the most effective and economic channels of defense. During this stage, the increased cortisol levels present during the first stage drop because they are no longer needed. If the stressor is prolonged or overwhelms the ability of the body to defend itself, the stage of exhaustion ensues, during which resources are depleted and signs of "wear and tear" or systemic damage appear.12 Selve contended that many ailments, such as various emotional disturbances, mildly annoying headaches, insomnia, upset stomach, gastric and duodenal ulcers, certain types of rheumatic disorders, and cardiovascular and kidney diseases appear to be initiated or encouraged by the "body itself because of its faulty adaptive reactions to potentially injurious agents."11

124 Unit Two: Alterations in Body Defenses

The events or environmental agents responsible for initiating the stress response were called stressors. According to Selye, stressors could be endogenous, arising from within the body, or exogenous, arising from outside the body.¹¹ In explaining the stress response, Selve proposed that two factors determine the nature of the stress response: the properties of the stressor and the conditioning of the person being stressed. Selve indicated that not all stress was detrimental; thus, he coined the terms eustress and distress.12 He suggested that mild, brief, and controllable periods of stress could be perceived as positive stimuli to emotional and intellectual growth and development. It is the severe, protracted, and uncontrolled situations of psychological and physical distress that are disruptive of health.¹¹ For example, the joy of becoming a new parent and the sorrow of losing a parent are completely different experiences, yet their stressor effect-the nonspecific demand for adjustment to a new situation-can be similar.

Stressors tend to produce different responses in different persons or in the same person at different times, indicating the influence of the adaptive capacity of the person, or what Selye called *conditioning factors*. These conditioning factors may be internal (*e.g.*, genetic predisposition, age, gender) or external (*e.g.*, exposure to environmental agents, life experiences, dietary factors, level of social support).¹¹ The relative risk for development of a stress-related pathologic process seems, at least in part, to depend on these factors.

Neuroendocrine-Immune Interactions

The manifestations of the stress response are strongly influenced by both the nervous and endocrine systems. The neuroendocrine systems integrate signals received along neurosensory pathways and from circulating mediators that are carried in the bloodstream. In addition, the immune system both affects and is affected by the stress response. The stress response is meant to protect the person against acute threats to homeostasis and is normally time limited. Therefore, under normal circumstances, the neural responses and the hormones that are released during the response are not around long enough to cause damage to vital tissues. However, in situations in which the stress response is hyperactive or becomes habituated, the physiologic and behavioral changes (*e.g.*, immunosuppression, sympathetic system activation) induced by the response can themselves become a threat to homeostasis. If the stress response is hypoactive, the person may be more susceptible to diseases associated with overactivity of the immune response.⁹

Neuroendocrine Responses

The integration of the stress responses, which occurs at the level of the central nervous system (CNS), is complex and not completely understood. It relies on communication along neuronal pathways of the cerebral cortex, the limbic system, the thalamus, the hypothalamus, the pituitary gland, and the reticular activating system (RAS) (Fig. 7-2). The cerebral cortex is involved with vigilance, cognition, and focused attention, and the limbic system with emotional components (e.g., fear, excitement, rage, anger) of the stress response. The thalamus functions as the relay center and is important in receiving, sorting out, and distributing sensory input. The hypothalamus coordinates the responses of the endocrine and autonomic nervous systems (ANS). The RAS modulates mental alertness, ANS activity, and skeletal muscle tone, using input from other neural structures. The musculoskeletal tension that occurs during the stress response reflects the increased activity of the RAS and its influence on the reflex circuits that control muscle tone.

Locus Ceruleus. Central to the neural component of the neuroendocrine response to the stress is an area in the brain stem

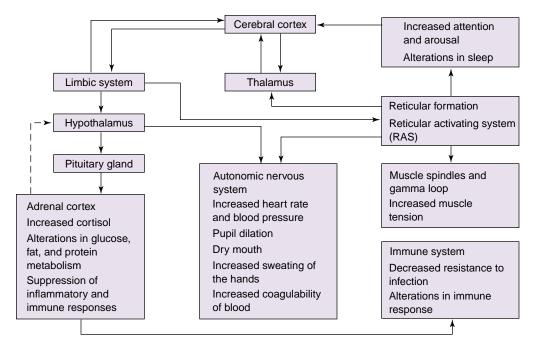


FIGURE 7-2 Stress pathways. The broken line represents negative feedback.

called the locus ceruleus (LC).¹³ The locus ceruleus is densely populated with neurons that produce norepinephrine (NE) and is thought to be the central integrating site for the ANS response to stressful stimuli (Fig. 7-3). The LC-NE system has afferent pathways to the hypothalamus, the limbic system, the hippocampus, and the cerebral cortex.

The LC-NE system confers an adaptive advantage during a stressful situation. The sympathetic nervous system manifestation of the stress reaction has been called the *fight-or-flight response*. This is the most rapid of the stress responses and represents the basic survival response of our primitive ancestors when confronted with the perils of the wilderness and its inhabitants. The increase in sympathetic activity in the brain increases attention and arousal and thus probably intensifies memory. The heart and respiratory rates increase, the hands and feet become moist, the pupils dilate, the mouth becomes dry, and the activity of the gastrointestinal tract decreases.

The Corticotropin-Releasing Factor. The corticotropinreleasing factor (CRF) is central to the endocrine component of the neuroendocrine response to stress (Fig. 7-3). CRF is a small peptide hormone found in both the hypothalamus and in extrahypothalamic structures, such as the limbic system and the brain stem. It is both an important endocrine regulator of pituitary and adrenal activity and a neurotransmitter involved in autonomic nervous system activity, metabolism, and behavior.^{9,14,15} Receptors for CRF are distributed throughout the brain as well as many peripheral sites. CRF from the hypothalamus induces the secretion of the adrenocorticotropic hormone (ACTH) from the anterior pituitary gland. ACTH, in turn, stimulates the adrenal gland to synthesize and secrete the glucocorticoid hormones (*e.g.*, cortisol).

The glucocorticoid hormones have a number of direct or indirect physiologic effects that mediate the stress response, enhance the action of other stress hormones, or suppress other components of the stress system. In this regard, cortisol acts both as a mediator of the stress response and an inhibitor of the stress response such that overactivation does not occur.¹⁶ Cortisol maintains blood glucose levels by antagonizing the effects of insulin and enhances the effect of catecholamines on the cardiovascular system. It also suppresses osteoblast activity, hematopoiesis, protein and collagen synthesis, and immune responses. All of these functions are meant to protect the organism against the effects of a stressor and to focus energy on regaining balance in the face of an acute challenge to homeostasis.

Other Hormones. A wide variety of other hormones, including growth hormones, thyroid hormone, and the reproductive hormones, also are responsive to stressful situations. Systems responsible for reproduction, growth, and immunity are directly linked to the stress system, and the hormonal effects of the stress response profoundly influence these systems.

Although growth hormone is initially elevated at the onset of stress, the prolonged presence of cortisol leads to suppression of growth hormone, somatomedin C, and other growth factors, exerting a chronically inhibitory effect on growth. In addition, CRF directly increases somatostatin, which in turn inhibits growth hormone secretion. Although the connection is speculative, the effects of stress on growth hormone may provide one of the vital links to understanding failure to thrive in children.

Stress-induced cortisol secretion also is associated with decreased levels of thyroid-stimulating hormone and inhibition of conversion of thyroxine to the more biologically active triiodothyronine in peripheral tissues. Both changes may serve as a means to conserve energy at times of stress.

Antidiuretic hormone (ADH) also is involved in the stress response, particularly in hypotensive stress or stress caused by



fluid volume loss. ADH, also known as vasopressin, increases water retention by the kidneys, produces vasoconstriction of blood vessels, and appears to synergize CRF's capacity to increase the release of ACTH.

The reproductive hormones are inhibited by CRF at the hypophyseal level and by cortisol at the pituitary, gonadal, and target tissue level.¹¹ Sepsis and severe trauma can induce anovulation and amenorrhea in women and decreased spermatogenesis and decreased levels of testosterone in men.

Immune Responses

The hallmark of the stress response, as first described by Selve, is the endocrine-immune interactions (i.e., increased corticosteroid production and atrophy of the thymus) that are known to suppress the immune response. In concert, these two components of the stress system, through endocrine and neurotransmitter pathways, produce the physical and behavioral changes designed to adapt to acute stress. Much of the literature regarding stress and the immune response focuses on the causal role of stress in immune-related diseases. It has also been suggested that the reverse may occur; emotional and psychological manifestations of the stress response may be a reflection of alterations in the CNS resulting from the immune response (Fig. 7-3). Immune cells such as monocytes and lymphocytes can penetrate the blood-brain barrier and take up residence in the brain, where they secrete cytokines and other inflammatory mediators that influence the stress response. In the case of cancer, this could mean that the subjective feelings of helplessness and hopelessness that have been repeatedly related to the onset and progression of cancers may arise secondary to the CNS effects of products released by immune cells during the early stage of the disease.¹⁷

The exact mechanism by which stress produces its effect on the immune response is unknown and probably varies from person to person, depending on genetic endowment and environmental factors. The most significant arguments for interactions between the neuroendocrine and immune systems derive from evidence that the immune and neuroendocrine systems share common signal pathways (i.e., messenger molecules and receptors), that hormones and neuropeptides can alter the function of immune cells, and that the immune system and its mediators can modulate neuroendocrine function.¹⁸ Receptors for a number of CNS-controlled hormones and neuromediators reportedly have been found on lymphocytes. Among these are receptors for glucocorticoids, insulin, testosterone, prolactin, catecholamines, estrogens, acetylcholine, and growth hormone, suggesting that these hormones influence lymphocyte function. For example, cortisol is known to suppress immune function, and pharmacologic doses of cortisol are used clinically to suppress the immune response. There is evidence that the immune system, in turn, influences neuroendocrine function.¹⁹ It has been observed that the HPA axis is activated by cytokines such as interleukin-1, interleukin-6, and tumor necrosis factor that are released from immune cells (see Chapter 8).

A second possible route for neuroendocrine regulation of immune function is through the sympathetic nervous system and the release of catecholamines. The lymph nodes, thymus, and spleen are supplied with ANS nerve fibers. Centrally acting CRF activates the ANS through multisynaptic descending pathways, and circulating epinephrine acts synergistically with CRF and cortisol to inhibit the function of the immune system.

Not only is the quantity of immune expression changed because of stress, but the quality of the response is changed. Stress hormones differentially stimulate the proliferation of subtypes of T lymphocyte helper cells. Because these T helper cell subtypes secrete different cytokines, they stimulate different aspects of the immune response. One subtype tends to stimulate the cellular-mediated immune response; whereas, a second type tends to activate B lymphocytes and humoral-mediated immune response.⁷

Coping and Adaptation to Stress

The ability to adapt to a wide range of environments and stressors is not peculiar to humans. According to René Dubos (a microbiologist noted for his study of human responses to the total environment), "adaptability is found throughout life and is perhaps the one attribute that distinguishes most clearly the world of life from the world of inanimate matter."20 Living organisms, no matter how primitive, do not submit passively to the impact of environmental forces. They attempt to respond adaptively, each in its own unique and most suitable manner. The higher the organism on the evolutionary scale, the larger its repertoire of adaptive mechanisms and its ability to select and limit aspects of the environment to which it responds. The most fully evolved mechanisms are the social responses through which individuals or groups modify their environments, their habits, or both to achieve a way of life that is best suited to their needs.

Adaptation

Human beings, because of their highly developed nervous system and intellect, usually have alternative mechanisms for adapting and have the ability to control many aspects of their environment. Air conditioning and central heating limit the need to adapt to extreme changes in environmental temperature. The availability of antiseptic agents, immunizations, and antibiotics eliminates the need to respond to common infectious agents. At the same time, modern technology creates new challenges for adaptation and provides new sources of stress, such as increased noise, air pollution, exposure to harmful chemicals, and changes in biologic rhythms imposed by shift work and transcontinental air travel.

Of particular interest are the differences in the body's response to events that threaten the integrity of the body's physiologic environment and those that threaten the integrity of the person's psychosocial environment. Many of the body's responses to physiologic disturbances are controlled on a moment-by-moment basis by feedback mechanisms that limit their application and duration of action. For example, the baroreflex-mediated rise in heart rate that occurs when a person moves from the recumbent to the standing position is almost instantaneous and subsides within seconds. Furthermore, the response to physiologic disturbances that threaten the integrity of the internal environment is specific to the threat; the body usually does not raise the body temperature when an increase in heart rate is needed. In contrast, the response to psychological disturbances is not regulated with the same degree of specificity and feedback control; instead, the effect may be inappropriate and sustained.

KEY CONCEPTS

STRESS AND ADAPTATION

- Stress is a state manifested by symptoms that arise from the coordinated activation of the neuroendocrine and immune systems, which Selye called the general adaptation syndrome.
- The hormones and neurotransmitters (catecholamines and cortisol) that are released during the stress response function to alert the individual to a threat or challenge to homeostasis, to enhance cardiovascular and metabolic activity in order to manage the stressor, and to focus the energy of the body by suppressing the activity of other systems that are not immediately needed.
- Adaptation is the ability to respond to challenges of physical or psychological homeostasis and to return to a balanced state.
- The ability to adapt is influenced by previous learning, physiologic reserve, time, genetic endowment, age, health status and nutrition, sleep-wake cycles, and psychosocial factors.

Factors Affecting the Ability to Adapt

Adaptation implies that an individual has successfully created a new balance between the stressor and the ability to deal with it. The means used to attain this balance are called *coping strategies* or *coping mechanisms*. Coping mechanisms are the emotional and behavioral responses used to manage threats to our physiologic and psychological homeostasis. According to Lazarus, how we cope with stressful events depends on how we perceive and interpret the event.²¹ Is the event perceived as a threat of harm or loss? And, is the event perceived as a challenge rather than a threat? Physiologic reserve, time, genetic endowment and age, health status, nutrition, sleep-wake cycles, hardiness, and psychosocial factors influence a person's appraisal of a stressor and the coping mechanisms used to adapt to the new situation (Fig. 7-4).

Physiologic and Anatomic Reserve. The trained athlete is able to increase cardiac output sixfold to sevenfold during exercise. The safety margin for adaptation of most body systems is considerably greater than that needed for normal activities. The red blood cells carry more oxygen than the tissues can use, the liver and fat cells store excess nutrients, and bone tissue stores calcium in excess of that needed for normal neuromuscular function. The ability of body systems to increase their function given the need to adapt is known as the *physiologic reserve*. Many of the body organs, such as the lungs, kidneys, and adrenals, are paired to provide anatomic reserve as well. Both organs are not needed to ensure the continued existence and maintenance of the internal environment. Many persons function normally with only one lung or one kidney. For example, in kidney disease signs of renal failure do not occur

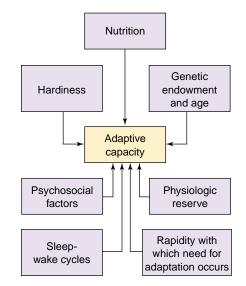


FIGURE 7-4 Factors affecting adaptation.

until approximately 90% of the functioning nephrons have been destroyed.

Time. Adaptation is most efficient when changes occur gradually, rather than suddenly. For instance, it is possible to lose a liter or more of blood through chronic gastrointestinal bleeding during the period of a week without manifesting signs of shock. However, a sudden hemorrhage that causes rapid loss of an equal amount of blood is likely to cause hypotension and shock.

Genetic Endowment. Adaptation is further affected by the availability of adaptive responses and flexibility in selecting the most appropriate and economical response. The greater the number of available responses, the more effective is the capacity to adapt.

Genetic endowment can ensure that the systems that are essential to adaptation function adequately. Even a gene that has deleterious effects may prove adaptive in some environments. In Africa, the gene for sickle cell anemia persists in some populations because it provides some resistance to infection with the parasite that causes malaria.

Age. The capacity to adapt is decreased at the extremes of age. The ability to adapt is impaired by the immaturity of an infant, much as it is by the decline in functional reserve that occurs with age. For example, the infant has difficulty concentrating urine because of immature renal structures and therefore is less able than an adult to cope with decreased water intake or exaggerated water losses. A similar situation, caused by age-related changes in renal function, exists in the elderly.

Health Status. Physical and mental health status determines physiologic and psychological reserves and is a strong determinant of the ability to adapt. For example, persons with heart disease are less able to adjust to stresses that require the recruitment of cardiovascular responses. Severe emotional stress often produces disruption of physiologic function and limits

the ability to make appropriate choices related to long-term adaptive needs. Those who have worked with acutely ill persons know that the will to live often has a profound influence on survival during life-threatening illnesses.

Nutrition. There are 50 to 60 essential nutrients, including minerals, lipids, certain fatty acids, vitamins, and specific amino acids. Deficiencies or excesses of any of these nutrients can alter a person's health status and impair the ability to adapt. The importance of nutrition to enzyme function, immune response, and wound healing is well known. On a worldwide basis, malnutrition may be one of the most common causes of immunodeficiency.

Among the problems associated with dietary excess are obesity and alcohol abuse. Obesity is a common problem. It predisposes to a number of health problems, including atherosclerosis and hypertension. Alcohol is commonly used in excess. It acutely affects brain function and, with long-term use, can seriously impair the function of the liver, brain, and other vital structures.

Sleep-Wake Cycles. Sleep is considered to be a restorative function in which energy is restored and tissues are regenerated.²² Sleep occurs in a cyclic manner, alternating with periods of wakefulness and increased energy use. Biologic rhythms play an important role in adaptation to stress, development of illness, and response to medical treatment. Many rhythms, such as rest and activity, work and leisure, and eating and drinking, oscillate with a frequency similar to that of the 24-hour light-dark solar day. The term *circadian*, from the Latin *circa* ("about") and *dies* ("day"), is used to describe these 24-hour diurnal rhythms.

Sleep disorders and alterations in the sleep-wake cycle have been shown to alter immune function, the normal circadian pattern of hormone secretion, and physical and psychological functioning.²³ The two most common manifestations of an alteration in the sleep-wake cycle are insomnia and sleep deprivation or increased somnolence. In some persons, stress may produce sleep disorders, and in others, sleep disorders may lead to stress. Acute stress and environmental disturbances, loss of a loved one, recovery from surgery, and pain are common causes of transient and short-term insomnia. Air travel and jet lag constitute additional causes of altered sleep-wake cycles, as does shift work. In persons with chronic insomnia, the bed often acquires many unpleasant secondary associations and becomes a place of stress and worry, rather than a place of rest.²⁴

Hardiness. Studies by social psychologists have focused on individuals' emotional reactions to stressful situations and their coping mechanisms to determine those characteristics that help some people remain healthy despite being challenged by high levels of stressors. For example, the concept of *hardiness* describes a personality characteristic that includes a sense of having control over the environment, a sense of having a purpose in life, and an ability to conceptualize stressors as a challenge, rather than a threat.²⁵ Many studies by nurses and social psychologists suggest that hardiness is correlated with positive health outcomes.²⁵

Psychosocial Factors. Several studies have related social factors and life events to illness. Scientific interest in the social en-

vironment as a cause of stress has gradually broadened to include the social environment as a resource that modulates the relation between stress and health. Presumably, persons who can mobilize strong supportive resources from within their social relationships are better able to withstand the negative effects of stress on their health. Studies suggest that social support has direct and indirect positive effects on health status and serves as a buffer or modifier of the physical and psychosocial effects of stress.²⁶

Social networks contribute in a number of ways to a person's psychosocial and physical integrity. The configuration of significant others that constitutes this network functions to mobilize the resources of the person; these friends, colleagues, and family members share the person's tasks and provide monetary support, materials and tools, and guidance in improving problem-solving capabilities.²⁶ Persons with ample social networks are not as likely to experience many types of stress, such as being homeless or being lonely.⁹ There is also evidence that persons who have social supports or social assets may live longer and have a lower incidence of somatic illness.²⁷

Social support has been viewed in terms of the number of relationships a person has and the person's perception of these relationships.²⁸ Close relationships with others can involve positive effects as well as the potential for conflict and may, in some situations, leave the person less able to cope with life stressors.

In summary, the stress response is an activation of several physiologic systems (sympathetic nervous system, the HPA axis, and the immune system) that work in a coordinated fashion to protect the body against damage from the intense demands made on it. Selye called this response the *general adaptation syndrome*. The stress response is divided into three stages: the *alarm stage*, with activation of the sympathetic nervous system and the HPA axis; the *resistance stage*, during which the body selects the most effective defenses; and the *stage of exhaustion*, during which physiologic resources are depleted and signs of systemic damage appear.

The activation and control of the stress response are mediated by the combined efforts of the nervous and endocrine systems. The neuroendocrine systems integrate signals received along neurosensory pathways and from circulating mediators that are carried in the bloodstream. In addition, the immune system both affects and is affected by the stress response.

Adaptation is affected by a number of factors, including experience and previous learning, the rapidity with which the need to adapt occurs, genetic endowment and age, health status, nutrition, sleep-wake cycles, hardiness, and psychosocial factors.

DISORDERS OF THE STRESS RESPONSE

For the most part, the stress response is meant to be acute and time limited. The time-limited nature of the process renders the accompanying catabolic and immunosuppressive effects advantageous. It is the chronicity of the response that is thought to be disruptive to physical and mental health. Stressors can assume a number of patterns in relation to time. They may be classified as acute time-limited, chronic intermittent, or chronic sustained. An acute time-limited stressor is one that occurs over a short time and does not recur; a chronic intermittent stressor is one to which a person is chronically exposed. The frequency or chronicity of circumstances to which the body is asked to respond often determines the availability and efficiency of the stress responses. For example, the response of the immune system is more rapid and efficient on second exposure to a pathogen than it is on first exposure, but chronic exposure to a stressor can fatigue the system and impair its effectiveness.

Effects of Acute Stress

The reactions to acute stress are those associated with the autonomic nervous system, the fight-or-flight response. The manifestations of the stress response—a pounding headache, cold moist skin, a stiff neck—are all part of the acute stress response. Centrally, there is facilitation of neural pathways mediating arousal, alertness, vigilance, cognition, and focused attention, as well as appropriate aggression. The acute stress response can result from either psychologically or physiologically threatening events. In situations of life-threatening trauma, these acute responses may be lifesaving in that they divert blood from less essential to more essential body functions. Increased alertness and cognitive functioning enables rapid processing of information and arrival at the most appropriate solution to the threatening situation.

However, for persons with limited coping abilities, either because of physical or mental health, the acute stress response may be detrimental. This is true of persons with pre-existing heart disease in whom the overwhelming sympathetic behaviors associated with the stress response can lead to dysrhythmias. For people with other chronic health problems, such as headache disorder, acute stress may precipitate a recurrence. In healthy individuals, the acute stress response can redirect attention from behaviors that promote health, such as attention to proper meals and getting adequate sleep. For those with health problems, it can interrupt compliance with medication regimens and exercise programs. In some situations, the acute arousal state actually can be life threatening, physically immobilizing the person when movement would avert catastrophe (*e.g.*, moving out of the way of a speeding car).

Effects of Chronic Stress

The stress response is designed to be an acute self-limited response in which activation of the ANS and the HPA axis is controlled in a negative feedback manner. As with all negative feedback systems, including the stress response system, pathophysiologic changes can occur. Function can be altered in several ways, including when a component of the system fails; when the neural and hormonal connections among the components of the system are dysfunctional; and when the original stimulus for the activation of the system is prolonged or of such magnitude that it overwhelms the ability of the system to respond appropriately. In these cases, the system may become overactive or underactive.

Chronicity and excessive activation of the stress response can result from chronic illnesses as well as contribute to the development of long-term health problems. Chronic activation of the stress response is an important public health issue from both a health and a cost perspective. The National Institute for Occupational Safety and Health declared stress a hazard of the workplace.²⁹ It is linked to a myriad of health disorders, such as diseases of the cardiovascular, gastrointestinal, immune, and neurologic systems, as well as depression, chronic alcoholism and drug abuse, eating disorders, accidents, and suicide.

Occurrence of the oral disease acute necrotizing gingivitis, in which the normal bacterial flora of the mouth become invasive, is known by dentists to be associated with acute stress, such as final examinations.³⁰ Similarly, herpes simplex type 1 infection (*i.e.*, cold sores) often develops during periods of inadequate rest, fever, ultraviolet radiation, and emotional upset. The resident herpesvirus is kept in check by body defenses, probably T lymphocytes, until a stressful event occurs that causes suppression of the immune system. Psychological stress is associated in a dose-response manner with an increased risk for development of the common cold, and this risk is attributable to increased rates of infection, rather than frequency of symptoms after infection.³¹

In a study in which participants were infected with the influenza virus, persons who reported the greatest amount of premorbid stress reported the most intense influenza symptoms and had a statistically greater production of interleukin-6, a cytokine that acts as a chemotactic agent for immune cells.³² Elderly caregivers of a spouse with dementia had a significantly higher score for emotional distress and higher salivary cortisol than did matched control subjects. The higher stress was correlated with a decreased immune response to the influenza vaccine.³⁴ The experience of stress also has been associated with delays in wound healing.

Post-traumatic Stress Disorder

The post-traumatic stress disorder (PTSD) is an example of chronic activation of the stress response as a result of experiencing a potentially life-threatening event. It was formerly called *battle fatigue* or *shell shock* because it was first characterized in men and women returning from combat. Although war is still a significant cause of PTSD, other major catastrophic events, such as major weather-related disasters, airplane crashes, terrorist bombings, and rape or child abuse, also may result in the development of the disorder. The terrorist attacks on the World Trade Center and Pentagon on September 11, 2001, represented an amalgam of interpersonal violence, loss, and threat to tens of thousands of people.35 These events have influenced and will continue to influence the development of PTSD in a substantial number of people. On the basis of data obtained after the 1995 bombing of the Murrah Federal Building in Oklahoma City, it is predicted that PTSD will develop in approximately 35% of those who were directly exposed to the September 11 attacks.36

PTSD is characterized by a constellation of symptoms that are experienced as states of intrusion, avoidance, and hyperarousal. *Intrusion* refers to the occurrence of "flashbacks" during waking hours or nightmares in which the past traumatic event is relived, often in vivid and frightening detail. *Avoidance* refers to the emotional numbing that accompanies this disorder and disrupts important personal relationships. Because a

130 Unit Two: Alterations in Body Defenses

person with PTSD has not been able to resolve the painful feelings associated with the trauma, depression is commonly a part of the clinical picture. Survivor guilt also may be a product of traumatic situations in which the person survived the disaster but loved ones did not. *Hyperarousal* refers to the presence of increased irritability and exaggerated startle reflex. In addition, memory problems, sleep disturbances, and anxiety are commonly experienced by persons with PTSD.

Although the pathophysiology of PTSD is not completely understood, the elucidation of physiologic changes related to the disorder has shed light on why some people recover from the disorder but others do not. It has been hypothesized that the intrusive symptoms of PTSD may arise from exaggerated sympathetic nervous system activation in response to the traumatic event. Persons with chronic PTSD have been shown to have increased levels of norepinephrine and increased activity of α_2 -adrenergic receptors.³⁵ The increase in catecholamines, in tandem with increased thyroid levels in persons with PTSD, is thought to explain some of the intrusive and somatic symptoms of the disorder.^{35,37}

Recent neuroanatomic studies have identified alterations in two brain structures—the amygdala and hippocampus. Positron-emission tomography and functional magnetic resonance imaging (MRI) have shown increased reactivity of the amygdala and hippocampus and decreased reactivity of the anterior cingulate and orbitofrontal areas. These areas of the brain are involved in fear responses. The hippocampus also functions in memory processes. Differences in hippocampal function and memory processes suggest a neuroanatomic basis for the intrusive recollections and other cognitive problems that characterize PTSD.³⁵

Despite the observed neuroanatomical changes, persons with PTSD do not uniformly undergo the biological responses associated with other types of stress.³⁸ For example, persons with PTSD demonstrate decreased cortisol levels, increased sensitivity of cortisol receptors, and an enhanced negative feedback inhibition of cortisol release with the dexamethasone suppression test. Dexamethasone is a synthetic glucocorticoid that mimics the effects of cortisol and directly inhibits the action of CRF and ACTH. This is in contrast to patients with major depression, who have a decreased sensitivity of glucocorticoid receptors, a high plasma level of cortisol, and a decreased dexamethasone suppression.³⁸ The hypersuppression of cortisol observed with the dexamethasone test suggests that persons with PTSD do not exhibit a classic stress response as described by Selye. Because this hypersuppression has not been described in other psychiatric disorders, it may serve as a relatively specific marker for PTSD.

Little is known about the risk factors that predispose people to the development of PTSD. It is important to note that less than half of all people who are exposed to a traumatic event experience PTSD. For example, only 15% to 30% of soldiers exposed to combat experience the disorder.³⁹ It also has been found that children exposed to violent events but who have strong family relationships rarely experience PTSD.⁴⁰ Statistics indicate there is a need for studies to determine risk factors for PTSD as a means of targeting individuals who may need intensive therapeutic measures after a life-threatening event. Research also is needed to determine the mechanisms by which the disorder develops so that it can be prevented, or if that is not possible, so that treatment methods can be created to decrease the devastating effects that this disorder has on affected individuals and their families.³⁷

Health care professionals need to be aware that clients who present with symptoms of depression, anxiety, and alcohol or drug abuse may in fact be experiencing PTSD. The client history should include questions concerning the occurrence of violence, major loss, or traumatic events in the person's life. Debriefing, or talking about the traumatic event at the time it happens, often is an effective therapeutic tool. Crisis teams are among the first people to attend to the emotional needs of those caught in catastrophic events. Some people may need continued individual or group therapy. Often concurrent pharmacotherapy, such as antidepressants and antianxiety agents, is useful and helps the individual participate more fully in therapy.

Most important, the person with PTSD must not be made to feel responsible for the disorder or that it is evidence of a character flaw. It is not uncommon for persons with this disorder to be told to "get over it" or "just get on with it, because others have." There is ample evidence to suggest that there is a biologic basis for the individual differences in responses to traumatic events, and these differences need to be taken into account.

Treatment of Stress Disorders

The treatment of stress should be directed toward helping people avoid coping behaviors that impose a risk to their health and providing them with alternative stress-reducing strategies. Purposeful priority setting and problem solving can be used by persons who are overwhelmed by the number of life stresses to which they have been exposed. Other nonpharmacologic methods used for stress reduction are relaxation techniques, guided imagery, music therapy, massage, and biofeedback.

Relaxation

Practices for evoking the relaxation response are numerous. They are found in virtually every culture and are credited with producing a generalized decrease in sympathetic system activity and musculoskeletal tension. According to Herbert Benson, a physician who worked in developing the technique, four elements are integral to the various relaxation techniques: a repetitive mental device, a passive attitude, decreased mental tonus, and a quiet environment.⁴¹ Benson developed a noncultural method that is commonly used for achieving relaxation (see accompanying box).

Progressive muscle relaxation, originally developed by Edmund Jacobson, who did extensive research on the muscle correlates of anxiety and tension, is another method of relieving tension. He observed that tension can be defined physiologically as the inappropriate contraction of muscle fibers. His procedure, which has been modified by a number of therapists, consists of systematic contraction and relaxation of major muscle groups.⁴² As the person learns to relax, the various muscle groups are combined. Eventually, the person learns to relax individual muscle groups without first contracting them.

Imagery

Guided imagery is another technique that can be used to achieve relaxation. One method is scene visualization, in which the person is asked to sit back, close the eyes, and con-

- Sit quietly in a comfortable position.
- Deeply relax all your muscles, beginning at your feet and progressing up to your face.
- Breathe through your nose. Become aware of your breathing. As you breathe out, say the word "one" silently to yourself. Continue for 20 minutes. When you have finished, sit quietly for several minutes, first with your eyes closed and then with them open.
- Do not worry about whether you are successful in achieving a deep level of relaxation. Maintain a positive attitude and permit the relaxation to occur at its own rate. Expect distracting thoughts, ignore them, and continue repeating "one" as you breathe out.
- (Modified from Benson H. [1977]. Systemic hypertension and the relaxation response. New England Journal of Medicine 296, 1152)

centrate on a scene narrated by the therapist. Whenever possible, all five senses are involved: the person attempts to see, feel, hear, and taste aspects of the visual experience. Other types of imagery involve imagining the appearance of each of the major muscle groups and how they feel during tension and relaxation.

Music Therapy

Music therapy is used for both its physiologic and psychological effects. It involves listening to selected pieces of music as a means of ameliorating anxiety or stress, reducing pain, decreasing feelings of loneliness and isolation, buffering noise, and facilitating expression of emotion. Music is defined as having three components: rhythm, melody, and harmony.43,44 Rhythm is the order in the movement of the music. Rhythm is the most dynamic aspect of music, and particular pieces of music often are selected because they harmonize with body rhythms, such as heart rhythm, respiratory rhythm, or gait. The melody is created by the musical pitch and distance (or interval) between the musical tone. The melody contributes to the listener's emotional response to the music. The harmony results from the way pitches are blended together, with the combination of sounds described as consonant or dissonant by the listener. Music usually is selected based on a person's musical preference and past experiences with music. Depending on the setting, headphones may be used to screen distracting noises. Radio and television music is inappropriate for music therapy because of the inability to control the selection of pieces that are played, the interruptions that occur (e.g., commercials and announcements), and the quality of the reception.

Massage Therapies

Massage is the manipulation of the soft tissues of the body to promote relaxation and relief of muscle tension. The technique that is used may involve a gentle stroking along the length of a muscle (effleurage), application of pressure across the width of a muscle (pétrissage), deep massage movements applied by a circular motion of the thumbs or fingertips (friction), squeezing across the width of a muscle (kneading), or use of light slaps or chopping actions (hacking).⁴⁵ Massage may be administered by practitioners who have received special training in its use or by less prepared persons, such as parents of small children^{46,47} or caregivers of confused elders.⁴⁸ It often is used as a means of physiologic relaxation and stress relief in critically ill patients.49

Biofeedback

Biofeedback is a technique in which an individual learns to control physiologic functioning. It involves electronic monitoring of one or more physiologic responses to stress with immediate feedback of the specific response to the person undergoing treatment. Several types of responses are used: electromyographic (EMG), electrothermal, and electrodermal (EDR).⁵⁰ The EMG response involves the measurement of electrical potentials from muscles, usually the forearm extensor or frontalis. This is used to gain control over the contraction of skeletal muscles that occurs with anxiety and tension. The electrodermal sensors monitor skin temperature in the fingers or toes. The sympathetic nervous system exerts significant control over blood flow in the distal parts of the body such as the digits of the hands and feet. Consequently, anxiety often is manifested by a decrease in skin temperature in the fingers and toes. EDR sensors measure conductivity of skin (usually the hands) in response to anxiety. Fearful and anxious people often have cold and clammy hands, which leads to a decrease in conductivity.

In summary, stress is neither negative nor deleterious to health. The stress response is designed to be time limited and protective, but in situations of prolonged activation of the response because of overwhelming or chronic stressors, it could be damaging to health. PTSD is an example of chronic activation of the stress response as a result of experiencing a severe trauma. In this disorder, memory of the traumatic event seems to be enhanced. Flashbacks of the event are accompanied by intense activation of the neuroendocrine system.

Treatment of stress should be aimed at helping people avoid coping behaviors that can adversely affect their health and providing them with other ways to reduce stress. Nonpharmacologic methods used in the treatment of stress include relaxation techniques, guided imagery, music therapy, massage techniques, and biofeedback.

REVIEW QUESTIONS

Describe the concept of homeostasis as it relates to a specific body function.

State Selve's definition of stress.

Relate Selve's description of the general adaptation syndrome to the signs and symptoms experienced by a person with a condition such as hemorrhagic shock.

Describe the contributions of the autonomic nervous system, endocrine system, and the immune system to the stress response.

List at least six factors that influence a person's adaptive apacity.

132 Unit Two: Alterations in Body Defenses

Describe the physiologic and psychological effects of a chronic stress response.

Describe the three states characteristic of post-traumatic stress disorder.

Describe at least five nonpharmacologic methods of treating stress.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- 1. Lazarus R.S., Folkman S. (1984). Stress, appraisal, and coping. New York: Springer.
- Hinkle L.E. (1977). The concept of 'stress' in the biological and social sciences. In Lipowskin Z.J., Lipsitt D.R., Whybrow P.C. (Eds.), *Psychosomatic medicine* (pp. 27–49). New York: Oxford University Press.
- 3. Osler W. (1910). The Lumleian lectures in angina pectoris. *Lancet* 1, 696–700, 839–844, 974–977.
- Cannon W.B. (1935). Stresses and strains of homeostasis. American Journal of Medical Science 189, 1–5.
- Selye H. (1946). The general adaptation syndrome and diseases of adaptation. *Journal of Clinical Endocrinology* 6, 117–124.
- Cannon W.B. (1939). The wisdom of the body (pp. 299–300). New York: W.W. Norton.
- Wilcox R.E., Gonzales R.A. (1995). Introduction to neurotransmitters, receptors, signal transduction, and second messengers. In Schatzberg A.F., Nemeroff C.B. (Eds.), *Textbook of psychopharmacology* (pp. 3–29). Washington, DC: American Psychiatric Press.
- Elenkov I.J., Webster E.L., Torpy D.J., et al. (1999). Stress, corticotrophin-releasing hormone, glucocorticoids, and the immune/ inflammatory response: Acute and chronic effects. Annals of the New York Academy of Sciences 876, 1–11.
- Chrousos G.P. (1998). Stressors, stress, and neuroendocrine integration of the adaptive response. Annals of the New York Academy of Sciences 851, 311–335.
- 10. Selye H. (1976). The stress of life (rev. ed.). New York: McGraw-Hill.
- 11. Selye H. (1973). The evolution of the stress concept. *American Scientist* 61, 692–699.
- 12. Selye H. (1974). *Stress without distress* (p. 6). New York: New American Library.
- Lopez J.F., Akil H., Watson S.J. (1999). Neural circuits mediating stress. *Biological Psychiatry* 46, 1461–1471.
- 14. Koob G.F. (1999) Corticotropin-releasing factor, norepinephrine, and stress. *Biological Psychiatry* 46, 1167–1180.
- Lehnert H., Schulz C., Dieterich K. (1998). Physiological and neurochemical aspects of corticotrophin-releasing factor actions in the brain: The role of the locus ceruleus. *Neurochemical Research* 23, 1039–1052.
- Sapolsky R.M., Romero L.M., Munck A.U. (2000). How do glucocorticoids influence stress responses? Integrating permissive, suppressive, stimulatory, and preparative actions. *Endocrine Reviews* 21, 55–89.
- Dantzer R., Kelley K.W. (1989). Stress and immunity: An integrated view of relationships between the brain and immune system. *Life Sciences* 44, 1995–2008.

- 18. Falaschi P., Martocchia A., Proietti A., et al. (1994). Immune system and the hypothalamus-pituitary-adrenal axis. *Annals of the New York Academy of Sciences* 741, 223–231.
- Woiciechowsky C., Schoning F., Lanksch W.R., et al. (1999). Mechanisms of brain mediated systemic anti-inflammatory syndrome causing immunodepression. *Journal of Molecular Medicine* 77, 769–780.
- 20. Dubos R. (1965). *Man adapting* (pp. 256, 258, 261, 264). New Haven: Yale University Press.
- Lazarus R. (2000). Evolution of a model of stress, coping, and discrete emotions. In Rice V.H. (Ed.), *Handbook of stress, coping, and health* (pp. 195–222). Thousand Oaks, CA: Sage Publications.
- 22. Adams K., Oswold I. (1983). Protein synthesis, bodily renewal and sleep-wake cycle. *Clinical Science* 65, 561–567.
- 23. Gillin J.C., Byerley W.F. (1990). The diagnosis and management of insomnia. *New England Journal of Medicine* 322, 239–248.
- Moldofsky H., Lue F.A., Davidson J.R., et al. (1989). Effects of sleep deprivation on human immune functions. *FASEB Journal* 3, 1972–1977.
- 25. Ford-Gilboe M., Cohen J.A. (2000). Hardiness: A model of commitment, challenge, and control. In Rice V.H. (Ed.), *Handbook of stress*, *coping, and health* (pp. 425–436). Thousand Oaks, CA: Sage Publications.
- Broadhead W.E., Kaplan B.H., James S.A., et al. (1983). The epidemiologic evidence for a relationship between social support and health. *American Journal of Epidemiology* 117, 521–537.
- Greenblatt M., Becerra R.M., Serafetinides E.A. (1982). Social networks and mental health: An overview. *American Journal of Psychiatry* 139, 977–984.
- Tilden V.P., Weinert C. (1987). Social support and the chronically ill individual. Nursing Clinics of North America 33, 613–620.
- National Institute for Occupational Safety and Health. (1999). *Stress at work* (pp. 1–26). Publication no. 99-101, HE 20.7102:ST 8/4. Bethesda, MD: U.S. Department of Health and Human Services.
- Dworkin S.F. (1969). Psychosomatic concepts and dentistry: Some perspectives. Journal of Periodontology 40, 647.
- Cohen S., Tyrrell D.A.J., Smith A.P. (1991). Psychological stress and susceptibility to the common cold. *New England Journal of Medicine* 325, 606–612.
- Cohen S., Doyle W.J., Skoner D.P. (1999). Psychological stress, cytokine production, and severity of upper respiratory illness. *Psychosomatic Medicine* 61 (2), 175–180.
- Vedhara K., Wilcock G.K., Lightman S.L., et al. (1999). Chronic stress in elderly carers of dementia patients and antibody response to influenza vaccination. *Lancet* 353, 627–631.
- Rozlog L.A., Kiecolt-Glaser J.K., Marucha P.T., et al. (1999). Stress and immunity: Implication for viral disease and wound healing. *Journal of Periodontology* 70, 786–792.
- 35. Yehuda R. (2002). Post-traumatic stress disorder. *New England Journal of Medicine* 346, 108–114.
- North C.S., Nixon S.J., Shariat S., et al. (1999). Psychiatric disorders among survivors of the Oklahoma City bombing. *Journal of the American Medical Association* 282, 755–762.
- Yehuda R. (2000). Biology of posttraumatic stress disorder. *Journal of Clinical Psychiatry* 61 (Suppl. 7), 14–21.
- Yehuda R. (1998). Psychoneuroendocrinology of posttraumatic stress disorder. Psychiatric Clinics of North America 21, 359–379.
- Sapolsky R. (1999). Stress and your shrinking brain (posttraumatic stress disorder's effect on the brain). *Discover* 20 (3), 116.
- McCloskey L.A. (2000). Posttraumatic stress in children exposed to family violence and single event trauma. *Journal of the American Academy of Child and Adolescent Psychiatry* 39, 108–115.
- Benson H. (1977). Systemic hypertension and the relaxation response. New England Journal of Medicine 296, 1152–1154.
- 42. Jacobson E. (1958). *Progressive relaxation*. Chicago: University of Chicago Press.

- 43. Chlan L., Tracy M.F. (1999). Music therapy in critical care: Indications and guidelines for intervention. *Critical Care Nurse* 19 (3), 35–41.
- 44. White J.M. (1999). Effects of relaxing music on cardiac autonomic balance and anxiety after acute myocardial infarction. *American Journal of Critical Care* 8, 220–230.
- 45. Vickers A., Zollman C. (1999). ABC of complementary therapies: Massage therapies. *British Medical Journal* 319, 1254–1257.
- Rusy L.M., Weisman S.J. (2000). Complementary therapies for acute pediatric pain management. *Pediatric Clinics of North America* 47, 589–599.
- Huhtala V., Lehtonen L., Heinonen R., et al. (2000). Infant massage compared with crib vibrator in treatment of colicky infants. *Pediatrics* 105 (6), E84.
- Rowe M., Alfred D. (1999). The effectiveness of slow-stroke massage in diffusing agitated behaviors in individuals with Alzheimer's disease. *Journal of Gerontological Nursing* 25 (6), 22–34.
- Richards K.C. (1998). Effect of back massage and relaxation intervention on sleep in critically ill patients. *American Journal of Critical Care* 7, 288–299.
- 51. Fischer-Williams M., Nigl A.J., Sovine D.L. (1986). A textbook of biological feedback. New York: Human Sciences Press.

The Immune Response

Immune System

CHAPTER

Properties of the Immune System Nonspecific Immunity Specific Immunity Specificity, Diversity, Memory, Self-limitation, and Self-nonself Recognition Lymphoid Organs Thymus Lymph Nodes **Spleen** Other Secondary Lymphoid Tissues Antigens **Immune Cells** Clusters of Differentiation Major Histocompatibility Complex Molecules Monocytes, Macrophages, and Dendritic Cells **B** Lymphocytes T Lymphocytes Natural Killer Cells Cytokines and the Immune Response Interleukin 1 and 2 Interferons **Tumor Necrosis Factor** Hematopoietic Colony-Stimulating Factors Mechanisms of the Immune Response Active versus Passive Immunity Humoral Immunity **Cell-Mediated Immunity** The Complement System **Regulation of the Immune Response Developmental Aspects of the Immune System** Transfer of Immunity From Mother to Infant Immune Response in the Elderly

he immune system is clearly essential for survival. It constantly defends the body against bacteria, viruses, and other foreign substances it encounters. It also detects and responds to abnormal cells and molecules that periodically develop in the body so that diseases such as cancers do not occur. An essential aspect of the immune response is the ability to recognize almost limitless numbers of foreign cells and nonself substances, distinguishing them from self molecules that are native to the body.

The major focus of this chapter is to present an overview of the immune cells, molecules, and tissues and to describe the normal mechanisms used to protect the body against foreign invaders.

THE IMMUNE SYSTEM

The *immune system* consists of the central and peripheral lymphoid tissues and the immune cells that protect the body against a myriad of microbes and foreign substances. The individual components of the substance that the immune system recognizes as foreign are called *antigens*. The interaction of the collective and coordinated components of the immune system and the antigens of a foreign agent is called the *immune response*.

Fundamental to the appropriate functioning of the immune system is the ability to regulate the recognition, amplification, and the response of the immune cells to a foreign agent. The immune system must recognize and differentiate one foreign pathogen from another, while simultaneously distinguishing these foreign molecules from normal cells and proteins in the body.

Properties of the Immune System

The body protects again bacteria, viruses, and other foreign substances by many different mechanisms. These include physical barriers, phagocytic cells in the blood and tissues, a class of lymphocytes called natural killer cells, and various bloodborne molecules that protect individuals from a potentially harmful environment. These mechanisms can be divided into two cooperative defense systems: the nonspecific or innate defense system and the specific or acquired immune system.

As a first line of defense system, the nonspecific immune system distinguishes self from non-self but does not distinguish one type of pathogen from another. Nonspecific resistance to microbe invasion results from two general lines of defense. Microorganisms encounter the first line of resistance on exposure to the epithelial layers of our skin and mucous membranes that line our respiratory, gastrointestinal, and urogenital tracts. The second line of nonspecific defense involves chemical signals, antimicrobial substances, phagocytic and natural killer cells, and fever associated with the inflammatory response. These two lines of nonspecific defense mechanisms are important for excluding pathogens from our body and removing them if they enter. They also aid in proper signaling of the second defense system—specific immunity.

Specific Immunity

Specific or acquired immunity develops during an individual's lifetime, distinguishes self from nonself, and responds specifically to different pathogens and foreign molecules. White blood cells called *lymphocytes* are key players in the specific or acquired immune response. These cells include the T lymphocytes (also called T cells), which participate in cell-mediated immunity, and the B lymphocytes (also called B cells), which participate in humoral immunity. Cell-mediated immunity involves the production of cytotoxic T cells, which have the ability to destroy antigen-bearing cells. Humoral immunity is characterized by the transformation of B cells into plasma cells, which secrete immunoglobulins (antibodies) that have specific activity against the inciting antigen.

Specificity, Diversity, Memory, Self-limitation, and Self-nonself Recognition

A cardinal feature of the specific immunity provided by the T and B lymphocytes is that of specificity, diversity, memory, selflimitation, and self-nonself recognition. These cells can exactly recognize a particular microorganism or foreign molecule. Each lymphocyte targets a specific antigen and distinguishes subtle differences between distinct antigens. The approximately 10¹² lymphocytes in the body have tremendous diversity. They can respond to the millions of different kinds of antigens encountered daily. This diversity occurs because an enormous variety of lymphocyte populations have been programmed during development, each to respond to a particular antigen.

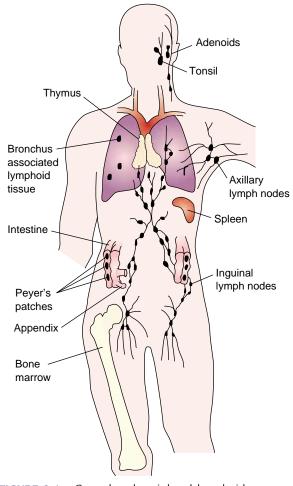
An evolutionary adaptation that is unique to the immune system is a memory response—the ability to recall and quickly produce a heightened immune response on subsequent exposure to the same foreign agent. After lymphocytes are stimulated by an antigen, they acquire a memory response. The memory T and B lymphocytes that are generated remain in the body for a long time and can respond more rapidly on repeat exposure than can naive cells. Because of this heightened state of immune reactivity, the immune system usually can respond to commonly encountered microorganisms so quickly and efficiently that one is unaware of the response.

Self-limitation refers to the ability of the stimulated lymphocytes to perform their functions for a brief period of time, sufficient to destroy the invading pathogen, and then die or differentiate into functionally quiescent memory cells. Selflimitation allows the immune system to return to a state of rest after it eliminates each antigen, thus enabling it to respond optimally to other antigens that an individual encounters.

Discrimination of self from nonself is one of the most important properties of the immune system. Immunologic unresponsiveness to self antigens, or self-tolerance, is essential for preventing reactions against one's own cells and tissues while maintaining a diverse repertoire of lymphocytes specific for foreign antigens.

Lymphoid Organs

The lymphoid organs are at the center of the immune response. These organs and tissues are widely distributed in the body and provide different, but often overlapping, functions (Fig. 8-1). The central lymphoid organs, the bone marrow and the thymus, provide the environment for immune cell production and maturation. The peripheral lymphoid organs function to trap and process antigen and promote its interaction with mature immune cells. Lymph nodes, spleen, tonsils, appendix, Peyer's patches in the intestine, and mucosa-associated lymphoid tissues in the respiratory, gastrointestinal, and reproductive systems comprise the peripheral lymphoid organs. Networks of lymph channels, blood vessels, and capillaries connect the lymphoid organs. The immune cells continuously circulate through



■ FIGURE 8-1 ■ Central and peripheral lymphoid organs and tissues.

KEY CONCEPTS

COMPONENTS OF THE IMMUNE SYSTEM

- The immune system consists of immune cells; the central immune structures (the bone marrow and thymus), where immune cells are produced and mature; and the peripheral immune structures (lymph nodes, spleen, and other accessory structures), where the immune cells interact with antigen.
- The immune cells consist of the lymphocytes (T and B lymphocytes), which are the primary cells of the immune system, and the accessory cells such as the macrophages, which aid in processing and presentation of antigens to the lymphocytes.
- Cytokines are molecules that form a communication link between immune cells and other tissues and organs of the body.
- Recognition of self from nonself by the immune cells depends on a system of MHC membrane molecules that differentiate viral-infected and abnormal cells from normal cells (MHC I) and identify immune cells from other types of cells (MHC II).

the various tissues and organs to seek out and destroy foreign material.

Thymus

The thymus is an elongated, bilobed structure that is located in the neck region above the heart. The function of the thymus is central to the development of the immune system because it generates mature immunocompetent T lymphocytes. The thymus is a fully developed organ at birth, weighing approximately 15 to 20 g. At puberty, when the immune cells are well established in peripheral lymphoid tissues, the thymus begins regressing and is replaced by adipose tissue. Nevertheless, some thymus tissue persists into old age. Precursor T (pre-T) cells enter the thymus as functionally and phenotypically immature T cells. They progressively differentiate into mature T cells under the influence of the thymic hormones and cytokines. As the T cells multiple and mature, they acquire T cell receptors, surface markers that distinguish among the different types of T cells, and antigens that distinguish self from nonself. More than 95% of the thymocytes die in the thymus because they do not produce the appropriate type of self-antigens. Only those T cells able to recognize foreign antigen and not react to selfantigens are allowed to mature. This process is called *thymic* selection. Mature immunocompetent T cells leave the thymus in 2 to 3 days and enter the peripheral lymphoid tissues through the bloodstream.

Lymph Nodes

Lymph nodes are small aggregates of lymphoid tissue located along lymphatic vessels throughout the body. Each lymph node processes lymph from a discrete, adjacent anatomic site. Many lymph nodes are in the axillae, groin, and along the great vessels of the neck, thorax, and abdomen. These tissues are located along the lymph ducts, which lead from the tissues to the thoracic duct. Lymph nodes have two functions: removal of foreign material from lymph before it enters the bloodstream and serving as centers for proliferation of immune cells.

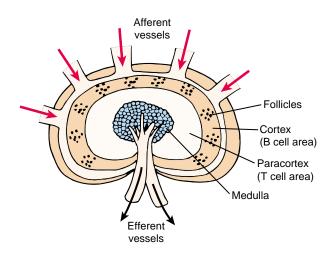
A lymph node is a bean-shaped tissue surrounded by a connective tissue capsule. Lymph enters the node through afferent channels that penetrate the capsule, and the lymph leaves through the efferent lymph vessels located in the deep indentation of the hilus (Fig. 8-2). Lymphocytes and macrophages flow slowly through the node, which allows trapping and interaction of antigen and immune cells. The reticular meshwork serves as a surface on which macrophages can more easily phagocytize antigens. Dendritic cells, which also permeate the lymph node, aid antigen presentation.

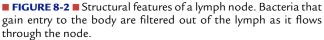
Spleen

The spleen is a large, ovoid organ located high in the left abdominal cavity. The spleen filters antigens from the blood and is important in response to systemic infections. The spleen is composed of red and white pulp. The red pulp is well supplied with arteries and is the area where senescent and injured red blood cells are destroyed. The white pulp contains concentrated areas of B and T lymphocytes permeated by macrophages and dendritic cells.

Other Secondary Lymphoid Tissues

Other secondary lymphoid tissues include the *mucosa-associated lymphoid tissues*. These nonencapsulated clusters of lymphoid tissues are located around membranes lining the respiratory, digestive, and urogenital tract. These gateways into the body contain the immune cells needed to respond to a large and diverse population of microorganisms. In some tissues, the lymphocytes are organized in loose clusters, but in other tissues such as the tonsils, Peyer's patches in the intestine, and the appendix, organized structures are evident. These tissues contain all the necessary cell components (*i.e.*, T cells, B cells, macrophages, and dendritic cells) for an immune response.





Immunity at the mucosal layers helps to protect the vulnerable internal organs.

Antigens

Before discussing the cells and responses inherent to immunity, it is important to understand the substances that elicit a response from the host. Antigens or immunogens are substances foreign to the host that can stimulate an immune response. These foreign molecules are recognized by receptors on immune cells and by proteins, called antibodies or immunoglobulins, that are generated in response to the antigen. Antigens include bacteria, fungi, viruses, protozoans, and parasitic worms. Antigens also can include substances such as pollen, poison ivy plant resin, insect venom, and transplanted organs. Most antigens are macromolecules such as proteins and polysaccharides, although lipids and nucleic acids occasionally can serve as antigens. Chemically complex molecules tend to be good stimulators of immunity.

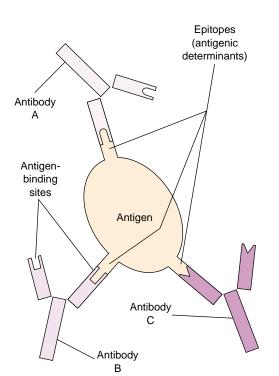
Antigens, which in general are large and complex, are biologically degraded into smaller chemical units or peptides. These discrete, immunologically active sites on antigens are called antigenic determinants or epitopes (Fig. 8-3). It is the unique molecular shape of an epitope that is recognized by a specific receptor found on the surface of the lymphocyte or by the antigen-binding site of an antibody. A single antigen may contain several antigenic determinants; each can stimulate a distinct clone of lymphocytes to respond. For example, different proteins that comprise a virus may function as unique antigens, each of which contains several antigenic determinants. Hundreds of antigenic determinants are found on complex structures such as the bacterial cell wall.

Smaller substances (molecular masses <10,000 daltons) usually are unable to stimulate an adequate immune response by themselves. When these low-molecular-weight compounds, known as haptens, combine with larger protein molecules, they function as antigens. The proteins act as carrier molecules for the haptens to form antigenic hapten-carrier complexes. An allergic response to the antibiotic penicillin is an example of a hapten-carrier complex that has medical importance. Penicillin (molecular mass of approximately 350 daltons) is incapable of causing an immune response by itself. However, penicillin can chemically combine with body proteins to form larger complexes that can then generate in some individuals an immune response to the penicillin epitope.

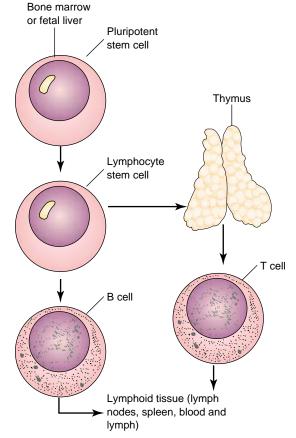
Immune Cells

The primary cells of the specific immune system are the lymphocytes. However, the recognition and activation of the specific immune responses depend on non-lymphoid cells, called accessory cells, which are not specific for different antigens. The accessory cells include the mononuclear phagocytes, dendritic cells, and other specialized antigen-presenting cells (APCs).

Lymphocytes represent 25% to 35% of blood leukocytes. Like other blood cells, lymphocytes are generated from stem cells in the bone marrow (Fig. 8-4). These undifferentiated cells congregate in the central lymphoid tissues, where they mature into distinct types of lymphocytes. One class of lymphocyte, the B lymphocytes (B cells), matures in the bone marrow and is essential for humoral or antibody-mediated immunity. The other class of lymphocytes, the T lymphocytes



■ FIGURE 8-3 ■ Multiple epitopes on a complex antigen being recognized by their respective (A, B, C) antibodies.



■ FIGURE 8-4 ■ Pathway for T- and B-cell differentiation.

138 Unit Two: Alterations in Body Defenses

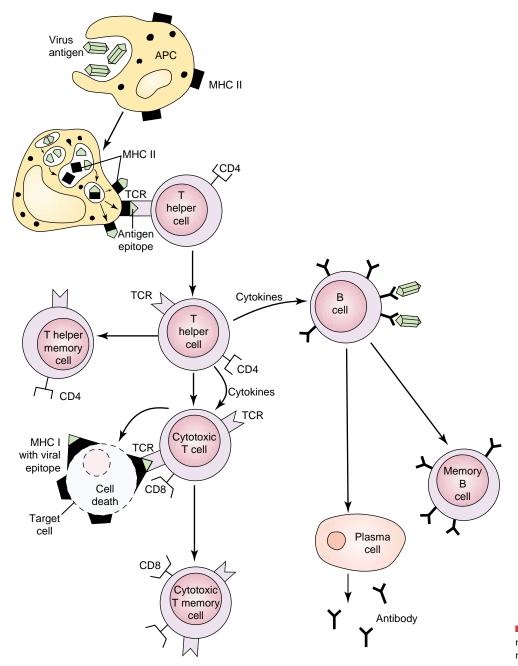
(T cells), complete their maturation in the thymus and are essential for cell-mediated immunity, as well as aiding with antibody production. Approximately 60% to 70% of blood lymphocytes are T cells, and 10% to 20% are B cells. The various types of lymphocytes are distinguished by their function and response to antigen, their cell membrane molecules and receptors, their types of secreted proteins, and their tissue location.

Activation of the lymphocytes is dependent upon the appropriate processing and presentation of antigen to the T lymphocytes by APCs such as macrophages (Fig. 8-5). On recognition of antigen and after additional stimulation by various secreted signaling molecules called *cytokines*, the T and B lymphocytes divide several times to form populations or clones of

cells that continue to differentiate into effector cells that destroy the antigen and memory cells that retain the information needed for future encounters with the antigen.

Clusters of Differentiation

Mature T and B cells display surface molecules called *clusters of differentiation* (CD). These molecules serve to define functionally distinct T-cell subsets such as CD4⁺ T helper cells and CD8⁺ T cytotoxic cells. The many cell surface CD molecules detected on immune cells have allowed scientists to study the normal and abnormal processes displayed by these cells. In cell-mediated immunity, regulatory CD4⁺ helper T cells enhance the response of other T and B cells, and effector cytotoxic T cells



■ **FIGURE 8-5** ■ Pathway for immune cell participation in an immune response.

(CD8⁺) kill virus-infected cells and tumor cells. The human immunodeficiency virus (HIV) that causes acquired immunodeficiency syndrome (AIDS) infects and destroys the CD4⁺ helper T cell (see Chapter 10).

Major Histocompatibility Complex Molecules

An essential feature of specific immunity is the ability to discriminate between the body's own molecules and foreign antigens. Failure to distinguish self from nonself can lead to conditions such as autoimmune disease where the immune system destroys the body's own cells. Key recognition molecules essential for distinguishing self from nonself are the cell surface MHC antigens. These molecules, which in humans are coded by closely linked genes on chromosome 6, were first identified because of their role in organ and tissue transplantation. When cells are transplanted between individuals who are not identical for their MHC molecules, the immune system produces a vigorous immune response, leading to rejection of the transferred cells or organs. MHC molecules did not evolve to reject transplanted tissues, a situation not encountered in nature. Rather, these molecules are essential for correct cell-to-cell interactions among immune and body cells.

The MHC molecules involved in self-recognition and cell-tocell communication fall into two classes, class I and class II (Fig. 8-6). Class I MHC molecules are cell surface glycoproteins that interact with antigen receptors and the CD8 molecule on T cytotoxic lymphocytes. They are found on nearly all nucleated cells in the body and thus are capable of alerting the immune system of any cell changes caused by viruses or products of mutated genes in cancer cells. In viral infected cells, viral protein antigens become associated with class I MHC molecules. As the virus multiplies, small peptides from degraded viral proteins

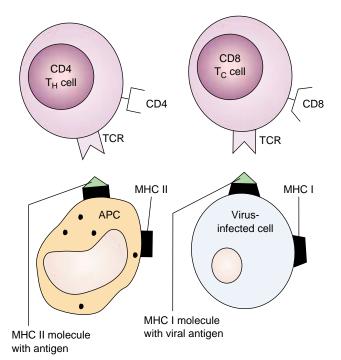


FIGURE 8-6 Interaction of a T-cell receptor (TCR) on a CD4 helper T (T_H) cell with class II MHC molecule on an antigenpresenting (APC) cell and CD8 cytotoxic (T_c) T cell with class I MHC molecule on a virus-infected cell.

complex with class I MHC molecules and are then transported to the infected cell membrane. This antigen-MHC I complex communicates to the T cytotoxic cell that the cell must be destroved for the overall survival of the host. Class II MHC molecules, which are found primarily on APCs such as macrophages, dendritic cells, and B lymphocytes, communicate with an antigen receptor and a CD4 molecule on T helper lymphocytes. Class II MHC molecules bind to a fragment of antigen from pathogens that have been engulfed and digested during the process of phagocytosis. The engulfed pathogen is degraded in cytoplasmic vesicles and its peptide components complexed with class II MHC molecules. Thelper cells recognize these complexes on the surface of APCs and become activated. These activated T cells multiply quickly and direct the immune response to the invading pathogen.

Each individual has a unique collection of several MHC proteins, and a variety of MHC molecules can exist in a population. Because of the number of MHC genes and the possibility of several alleles for each gene, it is almost impossible for any two individuals to be identical, unless they are identical twins. The uniqueness of these genes is essential for the immune system to distinguish self from nonself. In contrast to the receptors on T and B lymphocytes that bind a unique antigen molecule, each MHC protein can bind a broad spectrum of antigen peptides. Incorporation of the MHC molecules into the antigen presentation process allow for self/nonself recognition in establishing an appropriate immune response.

Human MHC proteins are called human leukocyte antigens (HLA) because they were first detected on white blood cells. Because these molecules play a role in transplant rejection and are detected by immunologic tests, they are commonly called antigens. The human class I MHC molecules are divided into types called HLA-A, HLA-B, and HLA-C, and the class II MHC molecules are identified as HLA-DR, HLA-DP, and HLA-DQ (Table 8-1). Additional, less well-defined class I and II MHC genes also have been described. Each of the gene loci that describes an HLA molecule can be occupied by multiple alleles or alternate genes. For example, there are more than 120 possible genes for the A locus and 250 genes for the B locus. Each of the gene products or antigens is designated by a number, such as HLA-B27.

Because the class I and II MHC genes are closely linked on one chromosome, the combination of HLA genes usually is inherited as a unit, called a haplotype. Each person inherits a chromosome from each parent and therefore has two HLA haplotypes. The identification or typing of HLA molecules is important in tissue or organ transplantation, forensics, and paternity evaluations. In organ or tissue transplantation, the closer the matching of HLA types, the greater is the probability of identical antigens and the lower the chance of rejection.

Monocytes, Macrophages, and Dendritic Cells

Monocytes and tissue macrophages are a part of the mononuclear phagocyte system, which in turn is part of the reticuloendothelial system. All of the cells of the mononuclear phagocytic system arise from common precursors in the bone marrow that produce the blood monocytes (see Chapter 11). The monocytes migrate to various tissues where they mature into macrophages. Macrophages are characterized as large cells with extensive cytoplasm and numerous vacuoles. The tissue macrophages are scattered in connective tissue or clustered in

TABLE 8-1	Properties of MHC Class I and MHC Class II Molecules			
Properties	MHC Class I	MHC Class II		
HLA antigens Distribution	HLA-A, HLA-B, HLA-C Virtually all nucleated cells	HLA-DR, HLA-DP, HLA-DQ Restricted to immune cells, antigen-presenting cells, B cells, and macrophages		
Functions	Present processed antigen to cytotoxic CD8+ T cells; restrict cytolysis to virus-infected cells, tumor cells, transplanted cells	Present processed antigenic fragments to CD4+ T cells; necessary for effective interaction among immune cells		

HLA, human leukocyte antigen; MHC, major histocompatibility complex.

organs such as the lung (*i.e.*, alveolar macrophages), liver (*i.e.*, Kupffer's cells), spleen, lymph nodes, peritoneum, central nervous system (*i.e.*, microglial cells), and other areas.

Macrophages have important functions in both innate and antigen-specific immune responses. As phagocytic cells with antigen nonspecific activity, they help to contain infectious agents until specific immunity can be marshaled. In addition, early in the host response, the macrophage functions as an accessory cell to ensure amplification of the inflammatory response and initiation of specific immunity. Macrophages are activated by the presence of antigen to engulf and digest foreign particles (Fig. 8-7). Activated macrophages act as APCs that break down complex antigens into peptide fragments that can associate with class II MHC molecules. Macrophages can then

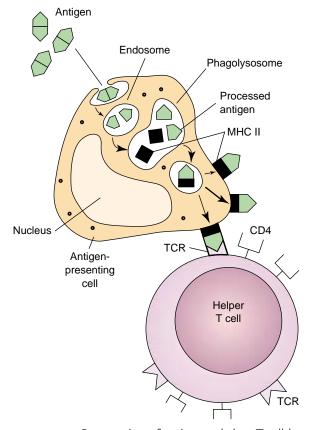


FIGURE 8-7 Presentation of antigen to helper T cell by an antigen-presenting cell (APC).

present these complexes to the helper T cell so that nonself-self recognition and activation of the immune response can occur. Macrophages also secrete cytokines that produce fever and prime T and B lymphocytes that have recognized antigen.

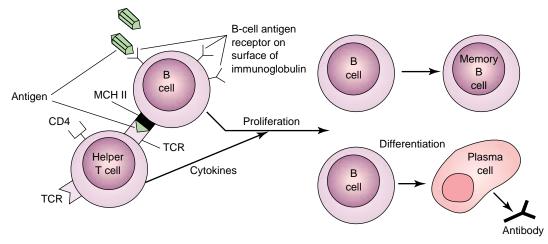
As the general scavenger cell of the body, the macrophage can be fixed in a tissue or can be free to migrate from an organ to lymphoid tissues. Macrophages also can serve as phagocytic effector cells in humoral and cell-mediated immune responses. They can remove antigen–antibody aggregates or, under the influence of T-cell cytokines, they can destroy virus-infected cells or tumor cells.

Dendritic cells share with the macrophage the important task of presenting antigen to T lymphocytes. These distinctive, star-shaped cells with long extensions of their cytoplasmic membrane provide an extensive surface rich in class II MHC molecules, which is essential for initiation of an acquired immune response. Dendritic cells are found in lymphoid tissues and other body areas where antigen enters the body. In these different environments, dendritic cells can acquire specialized functions and appearances, as do macrophages. Langerhans' cells are specialized dendritic cells in the skin, whereas follicular dendritic cells are found in the lymph nodes. Langerhans' cells are constantly surveying the skin for antigen and can transport foreign material to a nearby lymph node. Skin dendritic cells and macrophages also are involved in cell-mediated immune reactions of the skin, such as delayed allergic contact hypersensitivity.

B Lymphocytes

B lymphocytes can be identified by the presence of surface immunoglobulin that functions as the antigen receptor, class II MHC proteins, complement receptors, and specific CD molecules. During the maturation of B cells, which occurs in the bone marrow, stem cells change into immature precursor (pre-B) cells (Fig. 8-8). This B cell then acquires a unique surface receptor and a specific type of effector antibody (*e.g.*, immunoglobulin M [IgM] or IgD). This stage of maturation is programmed into the B cells and does not require antigen for its stimulation. CD molecules also change as the B cell matures. The CD surface markers are useful for defining immature and undifferentiated cells in B-cell malignancies. The mature B cell leaves the bone marrow, enters the circulation, and migrates to the various peripheral lymphoid tissues, where it is stimulated to respond to a specific antigen.

The commitment of a B-cell line to a specific antigen is evident by the expression of the surface immunoglobulin receptors. B cells that encounter antigen complementary to their sur-



■ FIGURE 8-8 ■ Pathway for B-cell differentiation.

face immunoglobulin receptor and receive T-cell help undergo a series of changes that transform them into antibody-secreting plasma cells or into memory B cells (Fig. 8-8). The activated B cell that divides and maturates into a plasma cell can produce thousands of antibody molecules per second. The antibodies are released into the blood and lymph, where they bind and remove their unique antigen with the help of other immune cells and molecules. Longer-lived memory B cells are generated and distributed into the peripheral tissues in preparation for subsequent antigen exposure.

Immunoglobulins. Antibodies comprise a class of proteins called immunoglobulins. The immunoglobulins have been divided into five classes: IgG, IgA, IgM, IgD, and IgE (Table 8-2), each with a different role in the immune defense strategy. Immunoglobulins have a characteristic four-polypeptide structure consisting of at least two identical antigen-binding sites (Fig. 8-9). Each immunoglobulin is composed of two identical light (L) chains and two identical heavy (H) chains to form a Y-shaped molecule. The two forked ends of the immunoglobulin molecule bind antigen and are called Fab (i.e., antigenbinding) fragments, and the tail of the molecule, which is called the *Fc* fragment, determines the biologic properties that are characteristic of a particular class of immunoglobulins. The amino acid sequence of the heavy and light chains shows constant (C) regions and variable (V) regions. The constant regions have sequences of amino acids that vary little among the antibodies of a particular class of immunoglobulin. The constant region allows for separation of immunoglobulins into classes (e.g., IgM, IgG) and it allows each class of antibody to interact with certain effector cells and molecules. For example, IgG can tag an antigen for recognition and destruction by phagocytes. The variable regions contain the antigen-binding sites of the molecule. The wide variation in the amino acid sequence of the variable regions seen from antibody to antibody allows this region to serve as the antigen-binding site. A unique amino acid sequence in this region determines a distinctive three-dimensional pocket that is complementary to the antigen, allowing recognition and binding of the antigen. Each B-cell clone produces antibody with one specific antigen-binding variable region or domain. During the course of the immune response, class

switching (*e.g.*, from IgM to IgG) can occur, causing the B-cell clone to produce any of the following antibody types.

IgG (gamma globulin) is the most abundant of the circulating immunoglobulins. It is present in body fluids and readily enters the tissues. IgG is the only immunoglobulin that crosses the placenta and can transfer immunity from the mother to the fetus. This class of immunoglobulin protects against bacteria, toxins, and viruses in body fluids and activates the complement system. There are four subsets of IgG (*i.e.*, IgG1, IgG2, IgG3, and IgG4) that have some restrictions in their response to certain types of antigens. For example, IgG2 appears to be responsive to bacteria that are encapsulated with a polysaccharide

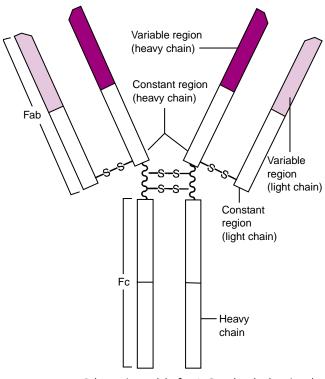


FIGURE 8-9 Schematic model of an IgG molecule showing the constant and variable regions of the light and dark chains.

142 Unit Two: Alterations in Body Defenses

TABLE 8-2	Classes and Characteristics of Immunoglobulins			
Figure	Class	Percentage of Total	Characteristics	
	lgG	75.0	Displays antiviral, antitoxin, and antibacterial properties; only lg that crosses the placenta; responsible for protection of new- born; activates complement and binds to macrophages	
	IgA	15.0	Predominant Ig in body secretions, such as saliva, nasal and respiratory secretions, and breast milk; protects mucous membranes	
	IgM	10.0	Forms the natural antibodies such as those for ABO blood antigens; prominent in early immune responses; activates complement	
\ <u></u> //	, IgD	0.2	Found on B lymphocytes; needed for maturation of B cells	
	IgE	0.004	Binds to mast cells and basophils; involved in parasitic infections, allergic and hypersensitivity reactions	

covering, such as *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Neisseria meningitidis*.

IgA is the second most abundant of the immunoglobulins. It is a secretory immunoglobulin found in saliva, tears, colostrum (*i.e.*, first milk of a nursing mother), and in bronchial, gastrointestinal, prostatic, and vaginal secretions. IgA prevents the attachment of viruses and bacteria to epithelial cells and is considered a primary defense against local infections in mucosal tissues.

IgM is a macromolecule that forms a polymer of five basic immunoglobulin units. It cannot cross the placenta and does not transfer maternal immunity. It is the first circulating immunoglobulin to appear in response to an antigen and is the first antibody type made by a newborn. This is diagnostically useful because the presence of IgM suggests a current infection by a specific pathogen. The identification of newborn IgM, rather than maternally transferred IgG, to a specific pathogen is indicative of an in utero or newborn infection. *IgD* is found primarily on the cell membranes of B lymphocytes. It serves as an antigen receptor for initiating the differentiation of B cells.

IgE is involved in inflammation, allergic responses, and combating parasitic infections. It binds to mast cells and basophils. The binding of antigen to mast cell- or basophil-bound IgE triggers these cells to release histamine and other mediators important in inflammation and allergies.

T Lymphocytes

T lymphocytes function in the activation of other T cells and B cells, in the control of viral infections, in the rejection of foreign tissue grafts, and in delayed hypersensitivity reactions (see Chapter 10). Collectively, these immune responses are called *cell-mediated* or *cellular immunity*. Besides the ability to respond to cell-associated antigens, the T cell is integral to immunity because it regulates self-recognition and amplifies the response of B and T lymphocytes.

T lymphocytes arise from bone marrow stem cells, but unlike B cells, pre-T cells migrate to the thymus for their maturation. There, the immature T lymphocyte acquires a T-cell receptor (TCR). The TCR for antigen is composed of membrane proteins expressed only on the T lymphocytes and binds specifically to antigen-peptide-MHC complexes on the surface of APCs or target cells. As with the variable proteins on the immunoglobulin molecule, the TCR proteins differ among T cells with different antigen specificity.

Maturation of subpopulations of T cells (*i.e.*, CD4⁺ and CD8⁺) also occurs in the thymus. Mature T cells migrate to the peripheral lymphoid tissues and, on encountering antigen, multiply and differentiate into memory T cells and various effector T cells.

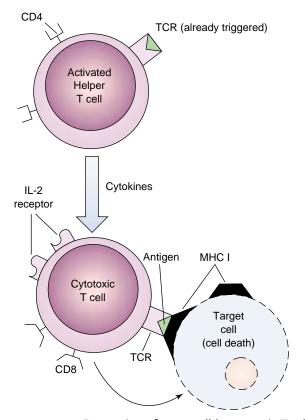
The CD4⁺ helper T cell (T_H) serves as a master switch for the immune system. Activation of helper T cells depends on the recognition of antigen in association with class II MHC molecules. Activated helper T cells secrete cytokines that influence the function of nearly all other cells of the immune system. These cytokines activate and regulate B cells, cytotoxic T cells, natural killer (NK) cells, macrophages, and other immune cells. Distinct subpopulations of helper T cells (*i.e.*, T_H1 and T_H2) have been identified and shown to secrete different patterns of cytokines. The pattern of cytokine production determines whether an antibody- or cell-mediated immune response develops. This differential expression of cytokines can influence expressions of some diseases (*i.e.*, lepromatous and tuberculoid leprosy).

The cytotoxic CD8⁺ T cells become activated after recognition of class I MHC-antigen complexes on target cell surfaces such as body cells infected by viruses or transformed by cancer (Fig. 8-10). The recognition of class I MHC-antigen complexes on infected target cells ensures that neighboring uninfected host cells, which express class I MHC molecules alone or with self-peptide, are not indiscriminately destroyed. The CD8⁺ T cells destroy target cells by releasing cytolytic enzymes, toxic cytokines, and pore-forming molecules (*i.e.*, perforins) or by triggering programmed cell death (apoptosis) in the target cell. The perforin proteins produce pores in the target cell membrane, allowing entry of toxic molecules and loss of cell constituents. The CD8⁺ T cells are especially important in controlling replicating viruses and intracellular bacteria because antibody cannot penetrate living cells.

Natural Killer Cells

Natural killer cells are lymphocytes that are functionally, genotypically, and phenotypically distinct from T cells, B cells, and monocyte-macrophages. The NK cell is a nonspecific effector cell that can kill tumor cells and virus-infected cells. They are called *natural killer cells* because, unlike T cytotoxic cells, they do not need to recognize a specific antigen before being activated. Both NK cells and T cytotoxic cells kill after contact with a target cell. The NK cell is programmed automatically to kill foreign cells, in contrast with the CD8⁺ T cells, which need to be activated to become cytotoxic. However, programmed killing is inhibited if the NK cell membrane molecules contact MHC self-molecules on normal host cells.

The mechanism of NK cytotoxicity is similar to T-cell cytotoxicity in that it depends on production of pore-forming proteins (*i.e.*, NK perforins), enzymes, and toxic cytokines. NK cell activity can be enhanced in vitro on exposure to



■ **FIGURE 8-10** Destruction of target cell by cytotoxic T cell. Cytokines released from the activated helper T cell enhance final destruction of the target cell by the cytotoxic T cell.

interleukin-2 (IL-2), a phenomenon called *lymphokine-activated killer* activity. These activated NK cells are used in the treatment of cancer. NK cells also participate in *antibody-dependent cellular cytotoxicity*, a mechanism by which a cytotoxic effector cell can kill an antibody-coated target cell. The role of NK cells probably is one of immune surveillance for cancerous or virally infected cells.

Cytokines and the Immune Response

Cytokines are low-molecular-weight regulatory proteins that are produced during all phases of an immune response. Cytokines are made primarily by and act predominantly on immune cells. These intercellular signal molecules are very potent, act at very low concentrations, and usually regulate neighboring cells. Cytokines modulate reactions of the host to foreign antigens or injurious agents by regulating the movement, proliferation, and differentiation of leukocytes and other cells (Table 8-3). Cytokines are synthesized by many cell types but are made primarily by activated T helper cells and macrophages.

These regulator molecules can be named for the general cell type that produces them (*e.g.*, lymphokines, monokines). More specifically, they are named by an international nomenclature (*i.e.*, interleukins 1 through 18) for the biologic property that was first ascribed to them. For example, *interferons* (IFNs) were named because they interfered with virus multiplication. Cytokines commonly affect more than one cell type

Characteristic Biologic Properties of Human Cytokin

TARI F 8-3

TABLE 6-5 Characteristic Biologic Properties of Human Cytokines				
Cytokine	Biologic Activity			
Interleukin-1 (alpha and beta)	Activates resting T cells; is cofactor for hematopoietic growth factor; induces fever, sleep, adreno- corticotropic hormone release, neutropenia, and other systemic acute-phase responses; stimu- lates synthesis of cytokines, collagen, and collegenases; activates endothelial and macrophagic cells; mediates inflammation, catabolic processes, and nonspecific resistance to infection			
Interleukin-2	Growth factor for activated T cells; induces synthesis of other cytokines, activates cytotoxic lymphocytes			
Interleukin-3	Support growth of pluripotent (multilineage) bone marrow stem cells; is growth factor for mast cells			
Interleukin-4	Growth factor for activated B cells, resting T cells, and mast cells; induces MHC class I antigen ex- pression on B cells; enhances cytotoxic T cells; activates macrophages			
Interleukin-5	B-cell differentiating and growth factor; promotes differentiation of eosinophils; promotes antibody production (IgA)			
Interleukin-6	Acts as cofactor for immunoglobulin production by B cells; stimulates hepatocytes to produce acute-phase proteins			
Interleukin-7	Stimulates pre-B cells and thymocytes; stimulates myeloid precursors and megakaryocytes			
Interleukin-8	Chemoattracts neutrophils and T lymphocytes; regulates lymphocyte homing and neutrophil infiltration			
Interleukin-10	Suppresses cytokine production by T helper cells; inhibits antigen presentation			
Interleukin-12	Enhances activation of cytotoxic T, NK, and macrophages; acts opposite to IL-10			
Interferon-gamma (γ)	Induces MHC class I, class II, and other surface antigens on a variety of cells; activates			
	macrophages and endothelial cells; augments or inhibits other cytokine activities; augments NK cell activity; exerts antiviral activity			
Interferon (alpha and beta) (α and β)	Exerts antiviral activity; induces class I antigen expression; augments NK cell activity; has fever- inducing and antiproliferative properties			
Tumor necrosis factor (alpha) (α)	Direct cytotoxin for some tumor cells; induces fever, sleep, and other acute-phase responses; stim- ulates the synthesis of other cytokines, collagen, and collagenases; activates endothelial and macrophagic cells; mediates inflammation, catabolic processes, and septic shock			
Colony-stimulating factor (CSF) Granulocyte-macrophage CSF	Promotes neutrophilic, eosinophilic, and macrophagic bone marrow colonies; activates mature granulocytes			
Granulocyte CSF	Promotes neutrophilic colonies			
Macrophage CSF	Promotes macrophagic colonies			

MHC, major histocompatibility complex; NK, natural killer.

and have more than one biologic effect. For example, IFN- γ inhibits virus replication and is a potent activator of macrophages and NK cells. Specific cytokines can have biologic activities that overlap. Maximization of the immune response and protection against detrimental mutations in a single cytokine are possible benefits of redundancy.

The production of cytokines often occurs in a cascade, in which one cytokine affects the production of subsequent cytokines or cytokine receptors. Some cytokines function as antagonists to inhibit the biologic effects of earlier cytokines. This pattern of expression and feedback ensures appropriate control of cytokine synthesis and subsequently of the immune response. Excessive cytokine production can have serious adverse effects, including those associated with septic shock, food poisoning, and types of cancer.

Cytokines generate their responses by binding to specific receptors on their target cells. Many cytokine receptors share a common structural shape and a cytoplasmic tail that interacts with a family of cytoplasmic signaling proteins responsible for the induction of the genes for cell responses. Most cytokines are released at cell-to-cell interfaces, where they bind to receptors on nearby cells. The short half-life of cytokines ensures that excessive immune responses and systemic activation do not occur.

The biologic properties of cytokines fall into several major groups. One group of cytokines (e.g., IL-1, IL-6, TNF) mediates

inflammation by producing fever and the acute-phase response and by attracting and activating phagocytes (*e.g.*, IL-8, IFN- γ). Other cytokines are maturation factors for the hematopoiesis of white or red blood cells (*e.g.*, IL-3, granulocyte-macrophage colony-stimulating factor [GM-CSF]). Recombinant CSF molecules are being used to increase the success rates of bone marrow transplantations. Most of the interleukin cytokines function as cell communication molecules among T cells, B cells, macrophages, and other immune cells. The availability of recombinant cytokines offers the possibility of several clinical therapies for which stimulation or inhibition of the immune response is desirable.

Interleukin 1 and 2

The major function of IL-1 is as a mediator of the inflammatory response. In concert with IL-6 and TNF- α , IL-1 can stimulate the production of an acute-phase response, mobilize neutrophils, produce a fever, and activate the vascular epithelium. IL-1 also can serve as a priming signal in the activation of CD4⁺ T cells and the growth and differentiation of B cells. The major source of IL-1 is the macrophage, although it also is produced by keratinocytes, Langerhans' cells, normal B cells, cultured T cells, fibroblasts, neutrophils, and smooth muscle cells.

The presence of IL-2, formerly known as *T-cell growth factor*, is necessary for the proliferation and function of helper T, cyto-

toxic T, B, and NK cells. IL-2 interacts with T lymphocytes by binding to specific membrane receptors that are present on activated T cells but not on resting T cells. The expression of highaffinity IL-2 receptors can be triggered by specific antigen and other stimulatory signals. Sustained T-cell proliferation relies on the presence of IL-2 and IL-2 receptors; if either is missing, cell proliferation ceases, and the cell dies. This cytokine ensures maximum amplification of immune responses if antigen is present. The drugs cyclosporine and tacrolimus, which are used to prevent rejection of heart, kidney, and liver transplants, function primarily by inhibiting the synthesis of IL-2.

Interferons

The IFNs are a family of cytokines that protect neighboring cells from invasion by intracellular parasites, including viruses, rickettsiae, malarial parasites, and other organisms. Bacterial toxins, complex polysaccharides, and several other chemical substances can induce IFN production. Not all the substances that induce IFN are antigenic.

There are three types of IFN: IFN- α , produced by leukocytes; IFN- β , produced by fibroblasts; and IFN- γ , produced by T and NK cells. IFN- α and IFN- β are grouped as type I IFNs to distinguish them from IFN- γ (type II). Type I-secreted IFNs interact with receptors on neighboring cells to stimulate the translation of an antiviral protein that affects viral synthesis and its spread to uninfected cells. The actions of IFNs are not pathogen specific; they are effective against different types of viruses and intracellular parasites. However, they are species specific. Animal IFNs do not provide protection in humans. The IFN produced during immune reactions is primarily IFN- γ . IFN- γ functions to activate macrophages, generate cytotoxic lymphocytes, and enhance NK cell activity.

Tumor Necrosis Factor

Like IL-1, TNF-α is a cytokine with multiple immunologic and inflammatory effects. It was first described as an activity in serum that induced hemorrhagic necrosis in certain tumors, and can function as a circulating mediator of wasting disease. TNF is produced by activated macrophages and other activated cells, such as T cells. Besides functioning as a major chemical mediator in the inflammatory response and indirectly affecting the fever response, TNF may function as a costimulator of T cells. This cytokine is an especially potent stimulator of IL-1, IL-6, and IL-8. In bacterial sepsis, high serum levels of TNF may mediate endotoxic shock. TNF is primarily responsible for the tissue wasting seen in cases of chronic inflammation.

Hematopoietic Colony-Stimulating Factors

Colony-stimulating factors are cytokines that stimulate bone marrow pluripotent stem and progenitor or precursor cells to produce large numbers of platelets, erythrocytes, neutrophils, monocytes, eosinophils, and basophils (see Chapter 11). The CSFs were named according to the type of target cell on which they act (see Table 8-3). GM-CSF acts on the granulocyte– monocyte progenitor cells to produce monocytes, neutrophils, and dendritic cells; G-CSF more specifically induces neutrophil proliferation; and M-CSF specifically directs the mononuclear phagocyte progenitor. Other cytokines, including IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, and IL-11, also may influence hematopoiesis.

Mechanisms of the Immune Response

The immune response is a complex series of interactions among the components of the immune system and the antigens of infectious agents and other pathogens. It consists of active or passive immunity and involves humoral and cellmediated immune mechanisms. The complement system links the humoral immune response with the inflammatory response and the lysis and phagocytosis of pathogens.

Active Versus Passive Immunity

Active or acquired immunity can be achieved through exposure to a specific antigen or through transfer of protective antibodies to an antigen. It is acquired through immunization or actually having a disease. Active immunity, although long lasting once established, requires a few days to weeks after a first exposure to become sufficiently developed to contribute to the destruction of the pathogen. However, on subsequent exposure to the same agent, the immune system usually is able to react within hours because of the presence of memory B and T lymphocytes.

Passive immunity is immunity transferred from one source to another source. An infant receives passive immunity naturally from the transfer of antibodies from its mother in utero and through a mother's breast milk. Normally, an infant has few infectious diseases during the first 3 to 6 months because of the protection provided by the mother's antibodies. Passive immunity also can be artificially provided by the transfer of antibodies produced by other people or animals. Some protection against infectious disease can be provided by the injection of hyperimmune serum, which contains high concentrations of antibodies for a specific disease, or immune serum or gamma globulin, which contains a pool of antibodies for many infectious agents. Passive immunity produces only short-term protection that lasts weeks to months.

KEY CONCEPTS

THE IMMUNE RESPONSE

- The immune response involves a complex series of interactions between components of the immune system and the antigens of a foreign pathogen.
- Passive immunity represents a temporary type of immunity that is transferred from another source (in utero transfer of antibodies from mother to infant).
- Active immunity depends on a response by the person's immune system and is acquired through immunization or actually having a disease.
- Humoral immunity consists of protection provided by the B lymphocyte-derived plasma cells, which produce antibodies that travel in the blood and interact with circulating and cell surface antigens.
- Cell-mediated immunity consists of protection provided by cytotoxic T lymphocytes, which protect against virus-infected or cancer cells.

Humoral Immunity

Humoral, or antibody-mediated immunity, relies on the presence of antibodies in the blood or body fluids. The combination of antigen with antibody can result in several effector responses, such as precipitation of antigen–antibody complexes, agglutination or clumping of cells, neutralization of bacterial toxins and viruses, lysis and destruction of pathogens or cells, adherence of antigen to immune cells, complement activation, and facilitation of phagocytosis. Phagocytic cells can more effectively bind, engulf, and digest antigen–antibody aggregates or immune complexes than they can antigen alone. Antibody can also neutralize a virus by blocking the sites on the virus that it uses to bind to the host cell, thereby negating its ability to infect the cell.

Two types of responses occur in the development of humoral immunity: a primary and a secondary response (Fig. 8-11). A primary immune response occurs when the antigen is first introduced into the body. During this primary response, there is a latent period or lag before the antibody can be detected in the serum. During this latent period, B cells are activated to proliferate and differentiate into antibody-secreting plasma cells. Recovery from many infectious diseases occurs at the time during the primary response when the antibody concentration is reaching its peak. The secondary or memory response occurs on second or subsequent exposures to the antigen. During the secondary response, the rise in antibody occurs sooner and reaches a higher level because of the available memory cells. The booster immunization given for some infectious diseases, such as tetanus, makes use of the secondary or memory response. For a person who has been previously immunized, administration of a booster shot causes an almost immediate rise in antibody to a level sufficient to prevent development of the disease.

Cell-mediated Immunity

Cell-mediated immunity provides protection against viruses, intracellular bacteria, and cancer cells. In cell-mediated immunity, the actions of T lymphocytes and macrophages predominate. The most aggressive phagocyte, the macrophage, becomes activated after exposure to T-cell cytokines, espe-

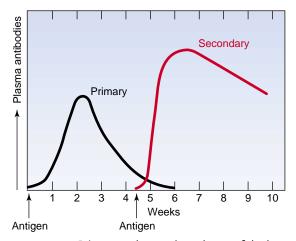


FIGURE 8-11 Primary and secondary phases of the humoral immune response to the same antigen.

cially IFN- γ . As in humoral immunity, the initial stages of cellmediated immunity are directed by an APC displaying the antigen peptide–class II MHC complex to the helper T cell. Helper T cells become activated after recognition by the TCR of the antigen–MHC complex and by priming with IL-1. The activated helper T cell then synthesizes IL-2 and the IL-2 receptor. These molecules drive the multiplication of clones of helper T cells, which amplify the response. Further differentiation of the helper T cells leads to production of additional cytokines (*e.g.*, IFN- γ , TNF, IL-12), which enhance the activity of cytotoxic T cells and effector macrophages. A cell-mediated immune response usually occurs through the cytotoxic activity of cytotoxic T cells and the enhanced engulfment and killing by macrophages.

The Complement System

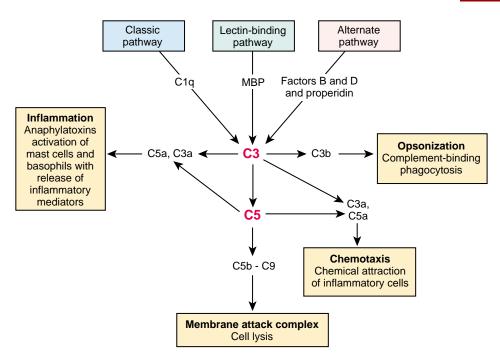
The complement system is a primary mediator of the humoral immune response that enables the body to produce an inflammatory response, lyse foreign cells, and increase phagocytosis. The complement system, like the blood coagulation system, consists of a group of proteins that normally are present in the circulation as functionally inactive precursors. These proteins make up 10% to 15% of the plasma protein fraction. For a complement reaction to occur, the complement components must be activated in the proper sequence. Uncontrolled activation of the complement system is prevented by inhibitor proteins and the instability of the activated complement proteins at each step of the process. There are three parallel but independent mechanisms for recognizing microorganisms that result in the activation of the complement system: the classic, the alternate, and the lectin-mediated pathways. All three pathways of activation generate a series of enzymatic reactions that proteolytically cleave successive complement proteins in the pathway. The consequence is the deposition of some complement protein fragments on the pathogen surface, thereby producing tags for better recognition by the phagocytic cells. Other complement fragments that are released into the tissue fluids further stimulate the inflammatory response.

The classic pathway of complement activation is initiated by antibody bound to antigens on the surface of microbes or through soluble immune complexes (Fig. 8-12). The alternate and the lectin pathways do not use antibodies and are part of the innate immune defenses. The alternate pathway of complement activation is initiated by the interaction with certain polysaccharide molecules characteristic of bacterial surfaces. The lectin-mediated pathway is initiated following the binding of a mannose-binding protein to mannose-containing molecules commonly present on the surface of bacteria and yeast.

The activation of the three pathways produces similar effects on C3 and subsequent complement proteins. The classic pathway of complement activation was the first discovered and is the best studied. The major proteins of the classic system are designated by a numbering system from C1 to C9. The classic pathway is triggered when complement-fixing antibodies, such as IgG or IgM, bind to antigens. The immune complexes trigger a series of enzyme reactions that act in a cascade fashion. Modified or split complement proteins (*e.g.*, C3b, C3a, C5a) released during activation function in the next step of the pathway or are released into the tissue fluid to produce biologic effects important in inflammation. C3 has a central role in the

Chapter 8: The Immune Response

147



■ FIGURE 8-12 ■ Classic, lectin, and alternative complement pathways.

complement pathway because it is integral to all three pathways. The triggering of C3 initiates several mechanisms for microbial destruction. One result of activation of C3 is the formation of the membrane attack complex formed by C5 to C9. Several structurally modulated complement proteins bind to form pores in the membrane of foreign cells that lead to eventual cell lysis.

The alternate and lectin pathways are activated by microbial surface molecules and substitute other molecules for the proteins in the first two steps of the classic complement pathway. The *alternate pathway* uses protein factors B, D, and properdin for activation, whereas the *lectin pathway* uses mannosebinding protein (MBP) and accessory proteins. Both pathways require the presence of C3b and subsequent complement proteins to generate biologic effects similar to those of the classic complement pathway. Whatever the mechanism of activation of the complement system, the effects produced range from lysis of a variety of different cells to direct mediation of the inflammatory process.

The activation of complement can result in formation of products that produce opsonization, chemotaxis, anaphylaxis, and cell membrane attack (Fig. 8-12). A major biologic function of complement activation is opsonization—the coating of antigen–antibody complexes such that antigens are engulfed and cleared more efficiently by macrophages. Chemotactic complement products (C3a and C5a) can trigger an influx of leukocytes. These white blood cells remain fixed in the area of complement activation through attachment to specific sites on C3b and C4b molecules. Anaphylatoxins (C3a and C5a) lead to activation of basophils and mast cells and release of inflammatory mediators that produce smooth muscle contract and increased vascular permeability. The late phase of the complement cascade involves the mediation of the membrane attack complex (MAC) that leads to the lytic destruction of many kinds of cells, including red blood cells, platelets, bacteria, and lymphocytes.

Regulation of the Immune Response

Self-regulation is an essential property of the immune system. An inadequate immune response may lead to immunodeficiency, but an inappropriate or excessive response may lead to conditions varying from allergic reactions to autoimmune diseases. This regulation is not well understood and involves all aspects of the immune response—antigen, antibody, cytokines, regulatory T cells, and the neuroendocrine system.

With each exposure to antigen, the immune system must determine the branch of the immune system to be activated and the extent and duration of the immune response. After exposure to an antigen, the immune response to that antigen develops after a brief lag, reaches a peak, and then recedes. Normal immune responses are self-limited because the response eliminates the antigen, and the products of the response, such as cytokines and antibodies, have a short or limited life span and are secreted only for brief periods after antigen recognition. Evidence suggests that cytokine feedback from the helper T cell controls several aspects of the immune response.

Another facet of immune self-regulation is inhibition of immune responses by tolerance. The term *tolerance* is used to define the ability of the immune system to be nonreactive to self-antigens while producing immunity to foreign agents. Tolerance to self-antigens protects an individual from harmful autoimmune reactions. Exposure of an individual to foreign antigens may lead to tolerance and the inability to respond to potential pathogens that cause infection. Tolerance exists not only to self-tissues, but to maternal-fetal tissues. Special regulation of the immune system also is evident in defined privileged sites, such as the brain, testes, ovaries, and eyes. Immune damage in these areas could result in serious consequences to the individual and the species.

In summary, immunity is the resistance to a disease that is provided by the immune system. It can be acquired actively through immunization or by having a disease, or passively by receiving antibodies or immune cells from another source. Antigens have antigenic determinant sites or epitopes, which the immune system recognizes with specific receptors that distinguish the antigens as nonself and as unique foreign molecules. Immune mechanisms can be classified into two types: specific or acquired and nonspecific or innate immunity. Specific or acquired immunity involves humoral and cellular mechanisms whereby the immune cells differentiate self from nonself and recognize and respond to a unique antigen. The humoral immune response involves antibodies produced by activated B lymphocytes. Cell-mediated immunity depends on T-cell responses to cellular antigens. Nonspecific immune mechanisms can distinguish between self and nonself but cannot differentiate among antigens. They include the complement system, cytokines, and the phagocytic activities of neutrophils and macrophages. The cytokines, produced largely by T cells, function as intercellular signals that regulate immune and inflammatory responses.

DEVELOPMENTAL ASPECTS OF THE IMMUNE SYSTEM

Embryologically, the immune system develops in several stages, beginning at 5 to 6 weeks as the fetal liver becomes active in hematopoiesis. Development of the primary lymphoid organs (*i.e.*, thymus and bone marrow) begins during the middle of the first trimester and proceeds rapidly. Secondary lymphoid organs (*i.e.*, spleen, lymph nodes, and tonsils) develop soon after. These secondary lymphoid organs are rather small but well developed at birth and mature rapidly during the postnatal period. The thymus at birth is the largest lymphoid tissue relative to body size and normally is approximately two thirds its mature weight, which it achieves during the first year of life.

Transfer of Immunity From Mother to Infant

Protection of a newborn against antigens occurs through transfer of maternal antibodies. Maternal IgG antibodies cross the placenta during fetal development and remain functional in the newborn for the first months of life (Fig. 8-13). IgG is the only class of immunoglobulins to cross the placenta. Levels of maternal IgG decrease significantly during the first 3 to 6 months of life while infant synthesis of immunoglobulins increases. Maternally transmitted IgG is effective against most microorganisms and viruses. The largest amount of IgG crosses the placenta during the last weeks of pregnancy and is stored in fetal tissues, and infants born prematurely have deficient amounts. Because of the transfer of IgG antibodies to the fetus, an infant born to a mother infected with HIV has a positive HIV antibody test result, although he or she may not be infected with the virus. Cord blood does not normally contain IgM or IgA. If present, these antibodies are of fetal origin and represent exposure to intrauterine infection. The infant begins producing IgM antibodies within a few months after birth in response to the immense antigenic stimulation of his or her new environment. Premature infants appear to be able to produce IgM as well as do term infants. At approximately 6 days of age, the neonate's IgM rises sharply, and this rise continues until approximately 1 year of age, when the adult level is achieved.

Serum IgA normally is first detected at approximately 13 days after birth. The level increases during early childhood until adult levels are reached between the sixth and seventh year. Maternal IgA also is transferred to the infant in colostrum or milk by breast-feeding. These antibodies provide local immunity for the intestinal system and have been shown to decrease diarrheal infections in underdeveloped countries. These evolutionary adaptations of the immune system have increased the survival of our species and optimized the development of other important organs in the early months of life.

Immune Response in the Elderly

Aging is characterized by a declining ability to adapt to environmental stresses. One of the factors thought to contribute to this problem is a decline in immune responsiveness. This includes changes in cell-mediated and antibody-mediated immune responses. Elderly persons tend to be more susceptible to infections, have more evidence of autoimmune and immune complex disorders than do younger persons, and have a higher incidence of cancer. Experimental evidence suggests that vaccination is less successful in inducing immunization in older persons than in younger adults. However, the effect of altered immune function on the health of elderly persons is clouded by the fact that age-related changes or disease may affect the immune response.

The alterations in immune function that occur with advanced age are not fully understood. There is a decrease in the size of the thymus gland, which is thought to affect T-cell function. The size of the gland begins to decline shortly after sexual maturity, and by 50 years of age, it usually has diminished to 15% or less of its maximum size. A common finding is a slight decrease in the proportion of T cells to other lymphocytes and a decrease in CD4⁺ and CD8⁺ cells.

More evident are altered responses of the immune cells to antigen stimulation, increasing the proportion of lymphocytes that become unresponsive, while the remainder continue to function relatively normally. T and B cells show deficiencies in activation. In the T-cell types, the CD4⁺ subset is most severely affected. Evidence indicates that aged T cells have a decreased rate of synthesis of the cytokines that drive the proliferation of lymphocytes and a diminished expression of the receptors that interact with those cytokines. For example, it has been shown that IL-2 synthesis decreases markedly with aging. Although Bcell function is compromised with age, the range of antigens that can be recognized is not diminished. If anything, the repertoire is increased to the extent that B cells begin to recognize some self-antigens as foreign antigens. This may be the basis for the increased incidence of autoimmune disease in the elderly.

In summary, a newborn is protected against antigens in early life by passive transfer of maternal antibodies through the placenta (IgG) and in colostrum (IgA) through breastfeeding. Some changes are seen with aging, including an increase in autoimmune diseases. The impact of alterations in immune function that occur with aging is not fully understood.

REVIEW QUESTIONS

Compare the properties of innate or nonspecific and specific or acquired immunity.

Differentiate between the central and peripheral lymphoid structures in terms of generation of immune cells and interaction with antigens.

Characterize the significance and function of major histocompatibility complex molecules in terms of recognizing self from nonself and distinguishing between body cells and immune cells.

Describe the phagocytic and antigen presenting functions of the macrophage as they relate to the immune response.

Explain why destruction of the CD4⁺ helper T cells by the human immunodeficiency virus has such a devastating effect on the immune system.

Explain the body's rationale for using natural killer (NK) cells, rather than B lymphocytes, to destroy mutant cells with the potential for development into cancer cells.

State the function of the five classes of immunoglobulins.

Describe the properties of cytokines and their functions in terms of the immune response and production of symptoms such as fever and malaise that accompany an acute infectious process.

Characterize the role of the complement system in mediation of the immune response.

Explain the transfer of passive immunity from mother to fetus and explain why infants who have high circulating levels of IgM at birth are thought to have contracted the infection in utero.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

BIBLIOGRAPHY

- Abbas A.K., Litchman A.H., Pober J.H., et al. (2000). *Cellular and molecular immunology* (4th ed.). Philadelphia: W.B. Saunders.
- Ahmed R., Gray D. (1996). Immunological memory and protective immunity: Understanding their relation. *Science* 272, 54–60.
- Benjamini E., Sunshine G., Leskowitz S. (2000). *Immunology: A short course* (4th ed.). New York: John Wiley & Sons.
- Cotran R.S., Kumar V., Tucker T. (1999). Pathologic basis of disease (6th ed., pp. 188–196). Philadelphia: W.B. Saunders.
- Delves P.J., Roitt I.M. (2000). The immune system: Parts I and II. New England Journal of Medicine 343, 37–49, 108–117.
- Goldsby R.A., Kindt T.J., Osborne B.A. (2000). Kuby immunology (4th ed.). San Francisco: W.H. Freeman.
- Janeway C.A., Jr., Travers P. (1999). Immunobiology: The immune system in health and disease (4th ed.). New York: Garland Publishing.
- Miller R.A. (1996). The aging immune system: Primer and prospectus. *Science* 273, 70–73.
- Moretta L. (1996). Receptors for HLA class-I molecules in human natural killer cells. *Annual Review of Immunology* 14, 619–648.
- Parham P. (2000). The immune system. New York: Garland Publishing.

Parslow T.G., Stites D., Abba I., et al.(Eds.). (2001). *Medical immunology* (10th ed.). East Norwalk, CT: Appleton & Lange.

Roitt I., Brostoff J., Male D. (1998). Immunology (5th ed.). St. Louis: Mosby.

Inflammation, Tissue Repair, and Fever

The Inflammatory Response **Acute Inflammation Cardinal Signs** The Vascular Response The Cellular Stage Inflammatory Mediators **Chronic Inflammation** Nonspecific Chronic Inflammation Granulomatous Inflammation Local Manifestations of Inflammation Systemic Manifestations of Inflammation **Acute-Phase Response** White Blood Cell Response (Leukocytosis and Leukopenia) Lymphadenitis **Tissue Repair and Wound Healing** Regeneration **Connective Tissue Repair Inflammatory Phase Proliferative Phase Remodeling Phase** Factors That Affect Wound Healing Malnutrition Blood Flow and Oxygen Delivery Impaired Inflammatory and Immune Responses Infection, Wound Separation, and Foreign Bodies **Temperature Regulation and Fever Body Temperature Regulation** Mechanisms of Heat Production Mechanisms of Heat Loss Fever Patterns **Manifestations Diagnosis and Treatment** Fever in Children Fever in the Elderly

CHAPTER

The ability of the body to sustain injury, resist attack by microbial agents, and repair damaged tissue is dependent upon the inflammatory reaction, the immune system response, and tissue repair and wound healing. Although the effects of inflammation are often viewed as undesirable because they are unpleasant and cause discomfort, the process is essentially a beneficial one that allows a person to live with the effects of everyday stress. Without the inflammatory response, wounds would not heal, and minor infections would become overwhelming. Inflammation also produces undesirable effects. For example, the crippling effects of rheumatoid arthritis result from chronic inflammation.

This chapter focuses on the manifestations of acute and chronic inflammation, tissue repair and wound healing, and temperature regulation and fever. The immune response is discussed in Chapter 8.

THE INFLAMMATORY RESPONSE

Inflammation is the reaction of vascularized tissue to local injury. The causes of inflammation are many and varied. Inflammation commonly results because of an immune response to infectious microorganisms. Other causes of inflammation are trauma, surgery, caustic chemicals, extremes of heat and cold, and ischemic damage to body tissues.

Inflammatory conditions are named by adding the suffix *-itis* to the affected organ or system. For example, *appendicitis* refers to inflammation of the appendix, *pericarditis* to inflammation of the pericardium, and *neuritis* to inflammation of a nerve. More descriptive expressions of the inflammatory process might indicate whether the process was acute or chronic and what type of exudate was formed (*e.g.*, acute fibrinous pericarditis).

Acute Inflammation

Acute inflammation is the early (almost immediate) response to injury. It is nonspecific and may be evoked by any injury short of one that is immediately fatal. It is usually of short duration and typically occurs before the immune response becomes established and is aimed primarily at removing the injurious agent and limiting the extent of tissue damage.

Cardinal Signs

The classic description of acute inflammation has been handed down through the ages. In the first century AD, the Roman physician Celsus described the local reaction of injury in terms that have come to be known as the *cardinal signs* of inflammation. These signs are *rubor* (redness), *tumor* (swelling), *calor* (heat), and *dolor* (pain). In the second century AD, the Greek physician Galen added a fifth cardinal sign, *functio laesa*, or loss of function. These signs and symptoms, which are apparent when inflammation occurs on the surface of body, may not be present when internal organs are involved. For example, inflammation of the lung does not usually cause pain unless the pleura, where pain receptors are located, is affected. In addition, an increase in heat is uncommon in inflammation involving internal organs, where tissues are normally maintained at core temperature.

In addition to the cardinal signs that appear at the site of injury, systemic manifestations (*e.g.*, fever) may occur as chemical mediators produced at the site of inflammation gain entrance to the circulatory system. The constellation of systemic manifestations that may occur during an acute inflammatory response is known as the *acute-phase response* (to be discussed).

The manifestation of acute inflammation can be divided into two categories: vascular and cellular responses.^{1,2,3} At the biochemical level, many of the responses that occur during acute inflammation are associated with the release of chemical mediators.

The Vascular Response

The vascular, or hemodynamic, changes that occur with inflammation begin almost immediately after injury and are initiated by a momentary constriction of small blood vessels in the area. This vasoconstriction is followed rapidly by vasodilation of the arterioles and venules that supply the area (Fig. 9-1). As a result, the area becomes congested, causing the

KEY CONCEPTS

THE INFLAMMATORY RESPONSE

- Inflammation represents the response of body tissue to immune reactions, injury, or ischemic damage.
- The classic response to inflammation includes redness, swelling, heat, pain or discomfort, and loss of function.
- The manifestations of an acute inflammatory response can be attributed to the immediate vascular changes that occur (vasodilation and increased capillary permeability), the influx of inflammatory cells such as neutrophils, and, in some cases, the widespread effects of inflammatory mediators, which produce fever and other systemic signs and symptoms.
- The manifestations of chronic inflammation are due to infiltration with macrophages, lymphocytes, and fibroblasts, leading to persistent inflammation, fibroblast proliferation, and scar formation.

redness (erythema) and warmth associated with acute inflammation. Accompanying this hyperemic vascular response is an increase in capillary permeability, which causes fluid to move into the tissues and cause swelling, pain, and impaired function. The exudation or movement of the fluid out of the capillaries and into the tissue spaces dilutes the offending agent. As fluid moves out of the capillaries, stagnation of flow and clotting of blood in the small capillaries occurs at the site of injury. This aids in localizing the spread of infectious microorganisms.

Depending on the severity of injury, the vascular changes that occur with inflammation follow one of three patterns of response.³ The first is an immediate transient response, which occurs with minor injury. The second is an immediate sustained response, which occurs with more serious injury and continues for several days and damages the vessels in the area. The third type of response is a delayed hemodynamic response, which involves an increase in capillary permeability that occurs 4 to 24 hours after injury. A delayed response often accompanies radiation types of injuries, such as sunburn.

The Cellular Stage

The cellular stage of acute inflammation is marked by movement of phagocytic white blood cells (leukocytes) into the area of injury. Two types of leukocytes participate in the acute inflammatory response—the granulocytes and monocytes.

Granulocytes. Granulocytes are identifiable because of their characteristic cytoplasmic granules. These white blood cells have distinctive multilobed nuclei. The granulocytes are divided into three types (*i.e.*, neutrophils, eosinophils, and basophils) according to the staining properties of the granules (Fig. 9-2).

The neutrophil is the primary phagocyte that arrives early at the site of inflammation, usually within 90 minutes of injury. The neutrophils' cytoplasmic granules contain enzymes and other antibacterial substances that are used in destroying and degrading the engulfed particles. They also have oxygendependent metabolic pathways that generate toxic oxygen (e.g., hydrogen peroxide) and nitrogen (e.g., nitric oxide) products. Because these white blood cells have nuclei that are divided into three to five lobes, they often are called *polymor*phonuclear neutrophils (PMNs) or segmented neutrophils (segs). The neutrophil count in the blood often increases greatly during the inflammatory process, especially with bacterial infections. After being released from the bone marrow, circulating neutrophils have a life span of only approximately 10 hours and therefore must be constantly replaced if their numbers are to remain adequate. This requires an increase in circulating white blood cells, a condition called leukocytosis. With excessive demand for phagocytes, immature forms of neutrophils are released from the bone marrow. These immature cells often are called *bands* because of the horseshoe shape of their nuclei.

The cytoplasmic granules of the *eosinophils* stain red with the acid dye eosin. These granulocytes increase in the blood during allergic reactions and parasitic infections. The granules of eosinophils contain a protein that is highly toxic to large parasitic worms that cannot be phagocytized. They also regulate inflammation and allergic reactions by controlling the release of specific chemical mediators during these processes.

The granules of the *basophils* stain blue with a basic dye. The granules of these granulocytes contain histamine and other

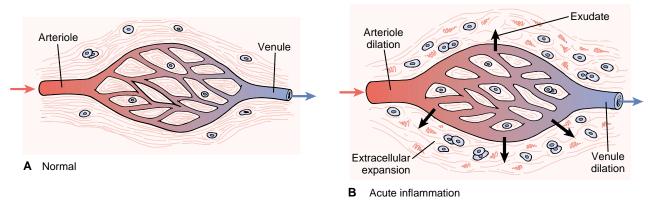
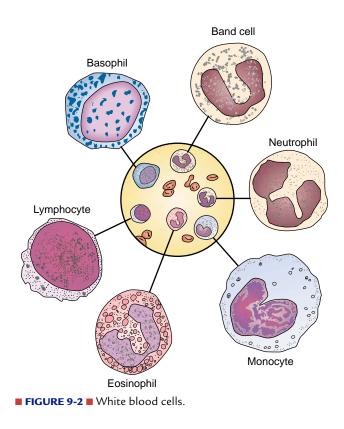


FIGURE 9-1 Vascular phase of acute inflammation. (A) Normal capillary bed. (B) Acute inflammation with vascular dilation causing increased redness (erythema) and heat (calor), movement of fluid into the interstitial spaces (swelling), extravasation of plasma proteins into the extracellular spaces (exudate), and emigration of leukocytes.

bioactive mediators of inflammation. The basophils are involved in producing the symptoms associated with inflammation and allergic reactions. The mast cell, which is widely distributed in connective tissues throughout the body, is very similar in many of its properties to the basophil. It can participate in acute and chronic inflammatory responses.² Sensitized mast cells, which are "armed" with IgE, play a central role in allergic and hypersensitivity responses (see Chapter 10). They also may play a role in parasitic infections. Mast cells can also elaborate tumor necrosis factor (TNF), thereby participating in chronic inflammatory responses.

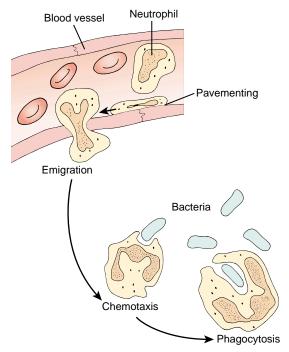


Mononuclear Phagocytes. The monocytes are the largest of the white blood cells and constitute 3% to 8% of the total blood leukocytes. The circulating life span of the monocyte is three to four times longer than that of the granulocytes, and these cells survive for a longer time in the tissues. These longer-lived phagocytes help to destroy the causative agent, aid in the signaling processes of specific immunity, and serve to resolve the inflammatory process.

The monocytes, which migrate in increased numbers into the tissues in response to inflammatory stimuli, mature into macrophages. Within 24 hours, mononuclear cells arrive at the inflammatory site, and by 48 hours, monocytes and macrophages are the predominant cell types. The macrophages engulf larger and greater quantities of foreign material than do the neutrophils. They also migrate to the local lymph nodes to prime specific immunity. These leukocytes play an important role in chronic inflammation, where they can surround and wall off foreign material that cannot be digested.

Cellular Response. The sequence of events in the cellular response to inflammation includes: (1) pavementing, (2) emigration, (3) chemotaxis, and (4) phagocytosis (Fig. 9-3). During the early stages of the inflammatory response, fluid leaves the capillaries, causing blood viscosity to increase. The release of chemical mediators (*i.e.*, histamine, leukotrienes, and kinins) and cytokines affects the endothelial cells of the capillaries and causes the leukocytes to increase their expression of adhesion molecules. As this occurs, the leukocytes slow their migration and begin to marginate, or move to and along the periphery of the blood vessels.

Emigration is a mechanism by which the leukocytes extend pseudopodia, pass through the capillary walls by ameboid movement, and migrate into the tissue spaces. The emigration of leukocytes also may be accompanied by an escape of red blood cells. Once they have exited the capillary, the leukocytes wander through the tissue guided by secreted cytokines (chemokines; interleukin [IL]-8), bacterial and cellular debris, and complement fragments (C3a, C5a). The process by which leukocytes migrate in response to a chemical signal is called *chemotaxis*. The positive movement up the



■ FIGURE 9-3 ■ Cellular phase of acute inflammation. Neutrophil margination, pavementing, chemotaxis, and phagocytosis.

concentration gradient of chemical mediators to the site of injury increases the probability of a sufficiently localized cellular response.

During the next and final stage of the cellular response, the neutrophils and macrophages engulf and degrade the bacteria and cellular debris in a process called *phagocytosis*. Phagocytosis involves three distinct steps: (1) adherence plus opsonization, (2) engulfment, and (3) intracellular killing (see Fig. 9-4). Contact of the bacteria or antigen with the phagocyte cell membrane is essential for trapping the agent and triggering the final steps of phagocytosis. If the antigen is coated with antibody or

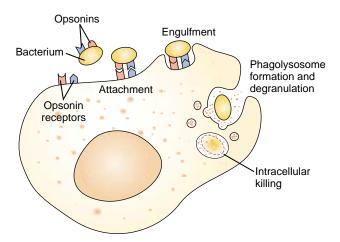


FIGURE 9-4 Phagocytosis of a particle (*e.g.*, bacterium): opsonization, attachment, engulfment, and intracellular killing.

Chapter 9: Inflammation, Tissue Repair, and Fever

complement, its adherence is increased because of binding to complement. This process of enhanced binding of an antigen caused by antibody or complement is called *opsonization* (Chapter 8). Engulfment follows the recognition of the agent as foreign. Cytoplasmic extensions (pseudopods) surround and enclose the particle in a membrane-bounded phagocytic vesicle or phagosome. In the cell cytoplasm, the phagosome merges with a lysosome containing antibacterial molecules and enzymes that can digest the microbe.

Intracellular killing of pathogens is accomplished through several mechanisms, including enzymes, defensins, and toxic oxygen and nitrogen products produced by oxygen-dependent metabolic pathways. The metabolic burst pathways that generate toxic oxygen and nitrogen products (*i.e.*, nitric oxide, peroxyonitrites, hydrogen peroxide, and hypochlorous acid) require oxygen and metabolic enzymes such as myeloperoxidase, NADPH oxidase, and nitric oxide synthetase. Individuals born with genetic defects in some of these enzymes have immunodeficiency conditions that make them susceptible to repeated bacterial infection.

Inflammatory Mediators

Although inflammation is precipitated by injury, its signs and symptoms are produced by chemical mediators. Mediators can be classified by function: those with vasoactive and smooth muscle-constricting properties such as histamine, prostaglandins, leukotrienes, and platelet-activating factor; chemotactic factors such as complement fragments and cytokines; plasma proteases that can activate complement and components of the clotting system; and reactive molecules and cytokines liberated from leukocytes, which when released into the extracellular environment can damage the surrounding tissue. Table 9-1 describes some chemical mediators and their major impact on inflammation.

Histamine. Histamine is widely distributed throughout the body. It is found in high concentration in platelets, basophils, and mast cells. Histamine causes dilation and increased permeability of capillaries. It is one of the first mediators of an inflammatory response. Antihistamine drugs inhibit this immediate, transient response.

Plasma Proteases. The plasma proteases consist of the kinins, activated complement proteins, and clotting factors. One kinin, bradykinin, causes increased capillary permeability and pain. The clotting system (see Chapter 12) contributes to the vascular phase of inflammation, mainly through fibrinopeptides that are formed during the final steps of the clotting process.

Prostaglandins. The prostaglandins are ubiquitous, lipidsoluble molecules derived from arachidonic acid, a fatty acid liberated from cell membrane phospholipids. Several prostaglandins are synthesized from arachidonic acid through the cyclooxygenase metabolic pathway (Fig. 9-5). Prostaglandins contribute to vasodilation, capillary permeability, and the pain and fever that accompany inflammation. The stable prostaglandins (PGE₁ and PGE₂) induce inflammation and potentiate the effects of histamine and other inflammatory mediators. The prostaglandin thromboxane A₂ promotes platelet aggregation and vasoconstriction. Aspirin and the nonsteroidal

154 Unit Two: Alterations in Body Defenses

TABLE 9-1 Signs of Inflammation and Corresponding Chemical Mediator				
Inflammatory Response	Chemical Mediator			
Swelling, redness, and tissue warmth (vasodilation and increased capillary permeability)	Histamine, prostaglandins, leukotrienes, bradykinin, platelet-activating factor			
Tissue damage	Lysosomal enzymes and products released from neutrophils, macrophages, and other inflammatory cells			
Chemotaxis	Complement fragments			
Pain	Prostaglandins			
	Bradykinin			
Fever	Interleukin-1 and interleukin-6			
Leukocytosis	Tumor necrosis factor and interleukin-8			

anti-inflammatory drugs (NSAIDs) reduce inflammation by inactivating the first enzyme in the cyclooxygenase pathway for prostaglandin synthesis.

Leukotrienes. Like the prostaglandins, the leukotrienes are formed from arachidonic acid, but through the lipoxygenase pathway (see Fig. 9-5). Histamine and leukotrienes are complementary in action in that they have similar functions. Histamine is produced rapidly and transiently while the more potent leukotrienes are being synthesized. Leukotrienes C4 and D4 are recognized as the primary components of the *slowreacting substance of anaphylaxis* (SRS-A) that causes slow and sustained constriction of the bronchioles and is an important inflammatory mediator in bronchial asthma and anaphylaxis. The leukotrienes also have been reported to affect the permeability of the postcapillary venules, the adhesion properties of endothelial cells, and the chemotaxis and extravascularization of neutrophils, eosinophils, and monocytes.

Platelet-Activating Factor. Platelet-activating factor (PAF), which is generated from a complex lipid stored in cell membranes, affects a variety of cell types and induces platelet aggregation. It activates neutrophils and is a potent eosinophil chemoattractant. When injected into the skin, PAF causes a

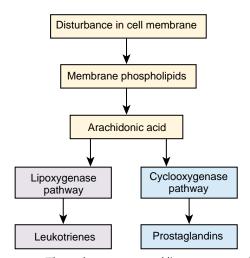


FIGURE 9-5 The cyclooxygenase and lipoxygenase pathways.

wheal-and-flare reaction and the leukocyte infiltrate characteristic of immediate hypersensitivity reactions. When inhaled, PAF causes bronchospasm, eosinophil infiltration, and nonspecific bronchial hyperreactivity.

Chronic Inflammation

Acute infections usually are self-limiting and rapidly controlled by the host defenses. In contrast, chronic inflammation is selfperpetuating and may last for weeks, months, or even years. It may develop during a recurrent or progressive acute inflammatory process or from low-grade, smoldering responses that fail to evoke an acute response.

Characteristic of chronic inflammation is an infiltration by mononuclear cells (macrophages) and lymphocytes, instead of the influx of neutrophils commonly seen in acute inflammation. Chronic inflammation also involves the proliferation of fibroblasts instead of exudates. As a result, the risk of scarring and deformity usually is considered greater than in acute inflammation. Agents that evoke chronic inflammation typically are low-grade, persistent irritants that are unable to penetrate deeply or spread rapidly. Among the causes of chronic inflammation are foreign bodies such as talc, silica, asbestos, and surgical suture materials. Many viruses provoke chronic inflammatory responses, as do certain bacteria, fungi, and larger parasites of moderate to low virulence. Examples are the tubercle bacillus, the treponema of syphilis, and the actinomyces. The presence of injured tissue such as that surrounding a healing fracture also may incite chronic inflammation. Immunologic mechanisms are thought to play an important role in chronic inflammation. The two patterns of chronic inflammation are a nonspecific chronic inflammation and granulomatous inflammation.

Nonspecific Chronic Inflammation

Nonspecific chronic inflammation involves a diffuse accumulation of macrophages and lymphocytes at the site of injury. Ongoing chemotaxis causes macrophages to infiltrate the inflamed site, where they accumulate because of prolonged survival and immobilization. These mechanisms lead to fibroblast proliferation, with subsequent scar formation that in many cases replaces the normal connective tissue or the functional parenchymal tissues of the involved structures. For example, scar tissue resulting from chronic inflammation of the bowel causes narrowing of the bowel lumen.

Granulomatous Lesions

A granulomatous lesion results from chronic inflammation. A granuloma typically is a small, 1- to 2-mm lesion in which there is a massing of macrophages surrounded by lymphocytes. These modified macrophages resemble epithelial cells and sometimes are called epithelioid cells. Like other macrophages, these epithelioid cells are derived originally from blood monocytes. Granulomatous inflammation is associated with foreign bodies such as splinters, sutures, silica, and asbestos and with microorganisms that cause tuberculosis, syphilis, sarcoidosis, deep fungal infections, and brucellosis. These types of agents have one thing in common: they are poorly digested and usually are not easily controlled by other inflammatory mechanisms. The epithelioid cells in granulomatous inflammation may clump in a mass (granuloma) or coalesce, forming a large, multinucleated giant cell that attempts to surround the foreign agent (Fig. 9-6). A dense membrane of connective tissue eventually encapsulates the lesion and isolates it.

A *tubercle* is a granulomatous inflammatory response to *Mycobacterium tuberculosis* infection. Peculiar to the tuberculosis granuloma is the presence of a caseous (cheesy) necrotic center.

Local Manifestations of Inflammation

The local manifestations of acute and chronic inflammation are dependent upon its cause and the particular tissue involved. These manifestations can range from swelling and the formation of exudates to abscess formation or ulceration.

Characteristically, the acute inflammatory response involves production of exudates. These exudates vary in terms of fluid, plasma proteins, and cell count. Acute inflammation can produce serous, hemorrhagic, fibrinous, membranous, or purulent exudates. Inflammatory exudates often are composed of a combination of these types. *Serous exudates* are watery fluids low in protein content that result from plasma entering the inflammatory site. *Hemorrhagic exudates* occur when there is severe tissue injury that causes damage to blood vessels or when there is significant leakage of red cells from the capillaries. *Fibrinous exudates* contain large amounts of fibrinogen and form a thick and sticky meshwork, much like the fibers of a

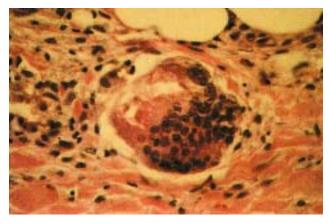
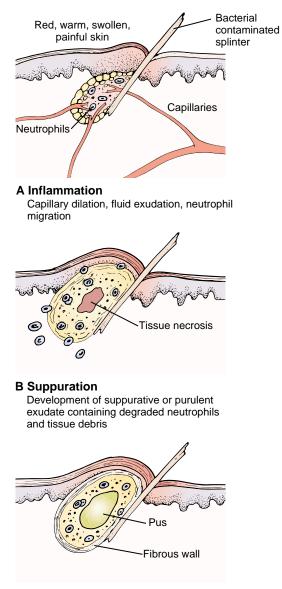


FIGURE 9-6 Foreign body giant cell. The numerous nuclei are randomly arranged in the cytoplasm. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p 40]. Philadelphia: Lippincott-Raven)

blood clot. *Membranous* or *pseudomembranous exudates* develop on mucous membrane surfaces and are composed of necrotic cells enmeshed in a fibropurulent exudate.

A purulent or suppurative exudate contains pus, which is composed of degraded white blood cells, proteins, and tissue debris (Fig. 9-7). An abscess is a localized area of inflammation containing a purulent exudate. Certain microorganisms (*e.g.*, staphylococcus) are more likely to induce localized suppurative inflammation and are referred to as *pyogenic*. Abscesses typically have a central necrotic core containing purulent



C Abscess formation Walling off of the area of purulent (pus) exudate to form an abscess

FIGURE 9-7 Abscess formation. (**A**) Bacterial invasion and development of inflammation. (**B**) Continued bacterial growth, neutrophil migration, liquefaction tissue necrosis, and development of a purulent exudate. (**C**) Walling off of the inflamed area with its purulent exudate to form an abscess.

exudates surrounded by a layer of neutrophils.² Fibroblasts may eventually enter the area and wall of the abscess. Because antimicrobial agents cannot penetrate the abscess wall, surgical incision and drainage may be required to effect a cure.

An *ulceration* refers to a site of inflammation where an epithelial surface (*e.g.*, skin or gastrointestinal epithelium) has become necrotic and eroded, often with associated subepithelial inflammation. Ulceration may occur as the result of traumatic injury to the epithelial surface (*e.g.*, peptic ulcer) or because of vascular compromise (foot ulcers associated with diabetes). In chronic lesions where there is repeated insult, the area surrounding the ulcer develops fibroblastic proliferation, scarring, and accumulation of chronic inflammatory cells.²

Systemic Manifestations of Inflammation

Under optimal conditions, the inflammatory response remains confined to a localized area. However, in some cases local injury can result in prominent systemic manifestations as inflammatory mediators are released into the circulation. The most prominent systemic manifestations of inflammation are the acute phase response, alterations in white blood cell count (leukocytosis or leukopenia), and fever. Sepsis and septic shock, also called the *systemic inflammatory response*, represent the severe systemic manifestations of inflammation (see Chapter 18).

Acute-Phase Response

Along with the cellular responses that occur during the inflammatory response, a constellation of systemic effects called the acute-phase response occurs. The acute-phase response, which usually begins within hours or days of the onset of inflammation or infection, includes changes in the concentrations of plasma proteins, increased erythrocyte sedimentation rate (ESR), fever, increased numbers of leukocytes, skeletal muscle catabolism, and negative nitrogen balance. These responses are generated after the release of the cytokines, IL-1, IL-6, and TNF- α . These cytokines affect the thermoregulatory center in the hypothalamus to produce fever, the most obvious sign of the acute-phase response. IL-1 and other cytokines induce an increase in the number and immaturity of circulating neutrophils by stimulating their production in the bone marrow. Lethargy, a common feature of the acute-phase response, results from the effects of IL-1 and TNF- α on the central nervous system.

During the acute-phase response, the liver dramatically increases the synthesis of acute-phase proteins such as fibrinogen and C-reactive protein that serve several different nonspecific host defense functions. The change in the types of plasma proteins contributes to the increased ESR. The metabolic changes, including skeletal muscle catabolism, provide amino acids that can be used in the immune response and for tissue repair. The total systemic process coordinates various activities in the body to enable an optimum host response.

White Blood Cell Response (Leukocytosis and Leukopenia)

Leukocytosis is common feature of the inflammatory response, especially those caused by bacterial infection. The white blood cell count usually increases to 15,000 to 20,000 cells/ μ L (normal 4000 to 10,000 cells/ μ L). After being released from the

bone marrow, circulating neutrophils have a life span of only about 10 hours and therefore must be constantly replaced if their numbers are to be adequate. With excessive demand for phagocytes, immature forms of neutrophils (bands) are released from the bone marrow. The phase, which is referred to as a "shift to the left" in a white blood cell differential count, refers to the increase in immature neutrophils seen in severe infections.

Bacterial infections produce a relatively selective increase in neutrophils (neutrophilia), whereas parasitic and allergic responses induce eosinophilia. Viral infections tend to produce neutropenia (decreased numbers of neutrophils) and lymphocytosis.³ Leukopenia is also encountered in infections that overwhelm persons with other debilitating diseases such as cancer.

Lymphadenitis

Localized acute and chronic inflammation may lead to a reaction in the lymph nodes that drain the affected area. This response represents a nonspecific response to mediators released from the injured tissue or an immunologic response to a specific antigen. Painful palpable nodes are more commonly associated with inflammatory processes, whereas nonpainful lymph nodes are more characteristic of neoplasms.¹

In summary, inflammation describes a local response to tissue injury and can present as an acute or chronic condition. The classic signs of inflammation are redness, swelling, local heat, pain, and loss of function. It involves a hemodynamic phase during which blood flow and capillary permeability are increased, and a cellular phase during which phagocytic white blood cells move into the area to engulf and degrade the inciting agent. The inflammatory response is orchestrated by chemical mediators such as histamine, prostaglandins, PAF, complement fragments, and reactive molecules that are liberated by leukocytes.

In contrast to acute inflammation, which is self-limiting, chronic inflammation is prolonged and usually is caused by persistent irritants, most of which are insoluble and resistant to phagocytosis and other inflammatory mechanisms. Chronic inflammation involves the presence of mononuclear cells (lymphocytes and macrophages), rather than granulocytes.

The local manifestations of acute and chronic inflammation depend upon the agent and extent of injury. Acute inflammation may involve the production of exudates containing serous fluid (serous exudate), red blood cells (hemorrhagic exudate), fibrinogen (fibrinous exudate), or tissue debris and white blood cell breakdown products (purulent exudate). *Ulceration* occurs at the site of inflammation where an epithelial surface (skin or mucous membranes) has become necrotic and eroded. In chronic lesions where there is repeated insult, the area surrounding the ulcer develops fibroblastic proliferation, scarring, and the accumulation of chronic inflammatory cells.

The systemic manifestations of inflammation include an increased ESR, fever, and leukocytosis (or in some cases, leukopenia). These responses are mediated by release of the cytokines, IL-1, TNF- α , and IL-6. Localized acute and chronic inflammation may lead to a reaction in the lymph nodes and enlargement of the lymph nodes that drain the affected area.

Body organs and structures contain two types of tissues: parenchymal and stromal. The parenchymal tissues contain the functioning cells of an organ or body part (*e.g.*, hepatocytes, renal tubular cells). The stromal tissues consist of the supporting connective tissues, blood vessels, and nerve fibers.

Injured tissues are repaired by regeneration of parenchymal cells or by connective tissue repair in which scar tissue is substituted for the parenchymal cells of the injured tissue. The primary objective of the healing process is to fill the gap created by tissue destruction and to restore the structural continuity of the injured part. When regeneration cannot occur, healing by replacement with connective scar tissue provides the means for maintaining this continuity. Although scar tissue fills the gap created by tissue death, it does not repair the structure with functioning parenchymal cells. Because the regenerative capabilities of most tissues are limited, wound healing usually involves some connective tissue repair.

Considerable research has contributed to the understanding of chemical mediators and growth factors that orchestrate the healing process.^{4,5} These chemical mediators and growth factors are released in an orderly manner from many of the cells that participate in the healing process. Some growth factors act as chemoattractants, enhancing the migration of white blood cells and fibroblasts to the wound site, and others act as mitogens, causing increased proliferation of cells that participate in the healing process.⁶ For example, platelet-derived growth factor, which is released from activated platelets, attracts white blood cells and acts as a growth factor for blood vessels and fibroblasts. Many of the cytokines discussed in Chapter 8 function as growth factors that are involved in wound healing.

Regeneration

Regeneration involves replacement of the injured tissue with cells of the same parenchymal type, leaving little or no evidence of the previous injury. The ability to regenerate varies with the tissue and cell type. Body cells are divided into three types according to their ability to undergo regeneration: labile, stable, or permanent cells.⁴

Labile cells are those that continue to divide and replicate throughout life, replacing cells that are continually being destroyed. Labile cells can be found in tissues that have a daily turnover of cells. They include the surface epithelial cells of the skin, the oral cavity, vagina, and cervix; the columnar epithelium of the gastrointestinal tract, uterus, and fallopian tubes; the transitional epithelium of the urinary tract; and bone marrow cells.

Stable cells are those that normally stop dividing when growth ceases. However, these cells are capable of undergoing regeneration when confronted with an appropriate stimulus. For stable cells to regenerate and restore tissues to their original state, the supporting stromal framework must be present. When this framework has been destroyed, the replacement of tissues is haphazard. The hepatocytes of the liver are one form of stable cell, and the importance of the supporting framework to regeneration is evidenced by two forms of liver disease. For example, in some types of viral hepatitis there is selective destruction of the parenchymal liver cells, while the cells of the

supporting tissue remain unharmed. After the disease has subsided, the injured cells regenerate, and liver function returns to normal. In cirrhosis of the liver, fibrous bands of tissue form and replace the normal supporting tissues of the liver, causing disordered replacement of liver cells and disturbance of hepatic blood flow and liver function.

Permanent or fixed cells cannot undergo mitotic division. The fixed cells include nerve cells, skeletal muscle cells, and cardiac muscle cells. These cells cannot regenerate; once destroyed, they are replaced with fibrous scar tissue that lacks the functional characteristics of the destroyed tissue. For example, the scar tissue that develops in the heart after a heart attack (Fig. 9-8) cannot conduct impulses or contract to pump blood.

Connective Tissue Repair

Connective tissue replacement is an important process in the repair of tissue. It allows replacement of nonregenerated parenchymal cells by a connective tissue scar. Depending on the extent of tissue loss, wound closure and healing occur by *primary* or *secondary* intention. A sutured surgical incision is an example of healing by primary intention. Larger wounds (*e.g.*, burns and large surface wounds) that have a greater loss of tissue and contamination, heal by secondary intention. Healing by secondary intention is slower than healing by primary intention of larger amounts of scar tissue. A wound that might otherwise have healed by primary intention.

Wound healing is commonly divided into three phases: the inflammatory phase, the proliferative phase, and the maturational or remodeling phase (Fig. 9-9).^{6,7} In wounds healing by primary intention, the duration of the phases is fairly predictable. In wounds healing by secondary intention, the process depends on the extent of injury and the healing environment.

Inflammatory Phase

The inflammatory phase of wound healing begins at the time of injury and is a critical period because it prepares the wound environment for healing. It includes hemostasis and the

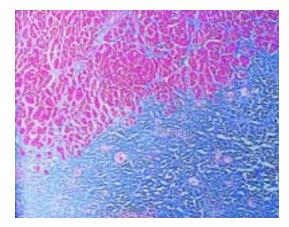
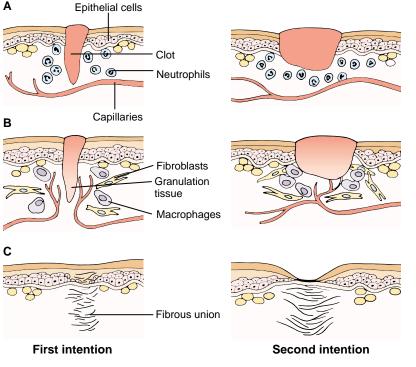


FIGURE 9-8 Healed myocardial infarct. Tissues with permanent cells are replaced with scar tissue only. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 102]. Philadelphia: Lippincott Williams & Wilkins)





■ FIGURE 9-9 ■ Healing of a skin wound by primary and secondary intention. (A) The inflammatory phase with formation of a blood clot and migration of neutrophils. (B) The proliferative phase with migration of macrophages and fibroblasts, proliferation of vascular endothelial cells, and development of granulation tissue. (C) Remodeling stage with development of the fibrous scar, disappearance of increased vascularity, and exit of inflammatory cells.

vascular and cellular phases of inflammation. Hemostatic processes are activated immediately at the time of injury. There is constriction of injured blood vessels and initiation of blood clotting by way of platelet activation and aggregation (see Chapter 12). After a brief period of constriction, these same vessels dilate and capillaries increase their permeability, allowing plasma and blood components to leak into the injured area. In small surface wounds, the clot loses fluid and becomes a hard, desiccated scab that protects the area.

The cellular phase of inflammation follows and is evidenced by the migration of phagocytic white blood cells that digest and remove invading organisms, fibrin, extracellular debris, and other foreign matter. The neutrophils or PMNs are the first cells to arrive and are usually gone by day 3 or 4. They ingest bacteria and cellular debris. Approximately 24 hours after the arrival of the PMNs, a larger and less specific phagocytic cell, the macrophage, which is an essential cell in the healing process, enters the wound area and remains for an extended period. Its functions include phagocytosis and release of growth factors that stimulate epithelial cell growth, angiogenesis (i.e., growth of new blood vessels), and attraction of fibroblasts. When a large defect occurs in deeper tissues, PMNs and macrophages are required to remove the debris and facilitate wound closure. Although a wound may heal in the absence of PMNs, it cannot heal in the absence of macrophages.

Proliferative Phase

The proliferative phase of healing usually begins within 2 to 3 days of injury and may last as long as 3 weeks in wounds healing by primary intention. The primary processes during this time focus on the building of new tissue to fill the wound space. The key cell during this phase is the *fibroblast*. The fibroblast is a connective tissue cell that synthesizes and secretes

collagen and other intercellular elements needed for wound healing. Fibroblasts also produce a family of growth factors that induce the growth of blood vessels (angiogenesis) and endothelial cell proliferation and migration.

As early as 24 to 48 hours after injury, fibroblasts and vascular endothelial cells begin proliferating to form a specialized type of soft, pink granular tissue, called *granulation tissue*, that

KEY CONCEPTS

TISSUE REPAIR AND WOUND HEALING

- Injured tissues can be repaired by regeneration of the injured tissue cells with cells of the same tissue or parenchymal type, or by connective repair processes in which scar tissue is used to effect healing.
- Regeneration is limited to tissues with cells that are able to undergo mitosis.
- Connective tissue repair occurs by primary or secondary intention and involves the inflammatory phase, the proliferative phase, and remodeling phases of the wound healing process.
- Wound healing is impaired by conditions that diminish blood flow and oxygen delivery, restrict nutrients and other materials needed for healing, and depress the inflammatory and immune responses; and by infection, wound separation, and the presence of foreign bodies.

serves as the foundation for scar tissue development. This tissue is fragile and bleeds easily because of the numerous, newly developed capillary buds. Wounds that heal by secondary intention have more necrotic debris and exudate that must be removed, and they involve larger amounts of granulation tissue. The newly formed blood vessels are leaky and allow plasma proteins and white blood cells to leak into the tissues. At approximately the same time, epithelial cells at the margin of the wound begin to regenerate and move toward the center of the wound, forming a new surface layer that is similar to that destroyed by the injury. In wounds that heal by primary intention, these epidermal cells proliferate and seal the wound within 24 to 48 hours.8 When a scab has formed on the wound, the epithelial cells migrate between it and the underlying viable tissue; when a significant portion of the wound has been covered with epithelial tissue, the scab can be easily removed. At times, excessive granulation tissue, sometimes referred to as proud flesh, may form and extend above the edges of wound, preventing re-epithelialization. Surgical removal or chemical cauterization of the defect allows healing to proceed.

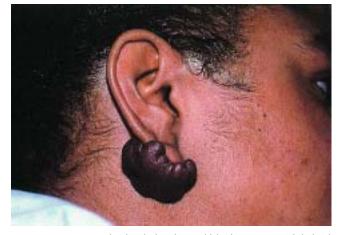
As the proliferative phase progresses, there is continued accumulation of collagen and proliferation of fibroblasts. Collagen synthesis reaches a peak within 5 to 7 days and continues for several weeks, depending on wound size. By the second week, the white blood cells have largely left the area, the edema has diminished, and the wound begins to blanch as the small blood vessels become thrombosed and degenerate.

Remodeling Phase

The third or remodeling phase of wound healing begins approximately 3 weeks after injury and can continue for 6 months or longer, depending on the extent of the wound. As the term implies, there is continued remodeling of scar tissue by simultaneous synthesis of collagen by fibroblasts and lysis by collagenase enzymes. As a result of these two processes, the architecture of the scar becomes reoriented to increase the tensile strength of the wound.

Most wounds do not regain the full tensile strength of unwounded skin after healing is completed. Carefully sutured wounds immediately after surgery have approximately 70% of the strength of unwounded skin, largely because of the placement of the sutures. This allows persons to move about freely after surgery without fear of wound separation. When the sutures are removed, usually at the end of the first week, wound strength is approximately 10%. It increases rapidly during the next 4 weeks and then slows, reaching a plateau of approximately 70% to 80% of the tensile strength of unwounded skin at the end of 3 months.⁵ An injury that heals by secondary intention undergoes wound contraction during the proliferative and remodeling phases. As a result, the scar that forms is considerably smaller than the original wound. Cosmetically, this may be desirable because it reduces the size of the visible defect. However, contraction of scar tissue over joints and other body structures tends to limit movement and cause deformities. As a result of loss of elasticity, scar tissue that is stretched fails to return to its original length.

An abnormality in healing by scar tissue repair is *keloid* formation (Fig. 9-10). Keloids are tumorlike masses caused by excess production of scar tissue. The tendency toward development of keloids is more common in African Americans and seems to have a genetic basis.



■ FIGURE 9-10 ■ Keloid. A light-skinned black woman with keloid that developed following ear piercing. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 90]. Philadelphia: Lippincott Williams & Wilkins)

Factors That Affect Wound Healing

Many local and systemic factors influence wound healing. Although there are many factors that impair healing, science has found a few ways to hasten the normal process of wound repair. Among the causes of impaired wound healing are malnutrition; impaired blood flow and oxygen delivery; impaired inflammatory and immune responses; infection, wound separation, and foreign bodies; and age effects.

Malnutrition

Successful wound healing depends in part on adequate stores of proteins, carbohydrates, fats, vitamins, and minerals. It is well recognized that malnutrition slows the healing process, causing wounds to heal inadequately or incompletely.⁹ Protein deficiencies prolong the inflammatory phase of healing and impair fibroblast proliferation, collagen and protein matrix synthesis, angiogenesis, and wound remodeling. Carbohydrates are needed as an energy source for white blood cells. Carbohydrates also have a protein-sparing effect and help to prevent the use of amino acids for fuel when they are needed for the healing process. Fats are essential constituents of cell membranes and are needed for the synthesis of new cells.

Although most vitamins are essential cofactors for the daily functions of the body, only vitamins A and C have been shown to play an essential role in the healing process. Vitamin C is needed for collagen synthesis. In vitamin C deficiency, improper sequencing of amino acids occurs, proper linking of amino acids does not take place, the by-products of collagen synthesis are not removed from the cell, and new wounds do not heal properly. Vitamin A functions in stimulating and supporting epithelialization, capillary formation, and collagen synthesis. The B vitamins are important cofactors in enzymatic reactions that contribute to the wound-healing process. All are water soluble, and with the exception of vitamin B_{12} , which is stored in the liver, almost all must be replaced daily. Vitamin K plays an indirect role in wound healing by preventing bleeding disorders that contribute to hematoma formation and subsequent infection.

The role of minerals in wound healing is less clearly defined. The macrominerals, including sodium, potassium, calcium, and phosphorus, as well as the microminerals, such as copper and zinc, must be present for normal cell function. Zinc is a cofactor in a variety of enzyme systems responsible for cell proliferation. In animal studies, zinc has been found to aid in re-epithelialization.

Blood Flow and Oxygen Delivery

For healing to occur, wounds must have adequate blood flow to supply the necessary nutrients and to remove the resulting waste, local toxins, bacteria, and other debris. Impaired wound healing caused by poor blood flow may occur as a result of wound conditions (*e.g.*, swelling) or pre-existing health problems. Arterial disease and venous pathology are well-documented causes of impaired wound healing. In situations of trauma, a decrease in blood volume may cause a reduction in blood flow to injured tissues.

Molecular oxygen is required for collagen synthesis. It has been shown that even a temporary lack of oxygen can result in the formation of less stable collagen.¹⁰ Wounds in ischemic tissue become infected more frequently than do wounds in well-vascularized tissue. PMNs and macrophages require oxygen for destruction of microorganisms that have invaded the area. Although these cells can accomplish phagocytosis in a relatively anoxic environment, they cannot digest bacteria.

Impaired Inflammatory and Immune Responses

Inflammatory and immune mechanisms function in wound healing. Inflammation is essential to the first phase of wound healing, and immune mechanisms prevent infections that impair wound healing. Among the conditions that impair inflammation and immune function are disorders of phagocytic function, diabetes mellitus, and therapeutic administration of corticosteroid drugs.

Phagocytic disorders may be divided into extrinsic and intrinsic defects. Extrinsic disorders are those that impair attraction of phagocytic cells to the wound site, prevent engulfment of bacteria and foreign agents by the phagocytic cells (*i.e.*, opsonization), or cause suppression of the total number of phagocytic cells (*e.g.*, immunosuppressive agents). Intrinsic phagocytic disorders are the result of enzymatic deficiencies in the metabolic pathway for destroying the ingested bacteria by the phagocytic cell (see Chapter 10).

Wound healing is a problem in persons with diabetes mellitus, particularly those who have poorly controlled blood glucose levels. Studies have shown delayed wound healing, poor collagen formation, and poor tensile strength in diabetic animals. Of particular importance is the effect of hyperglycemia on the phagocytic function of white blood cells. For example, neutrophils have diminished chemotactic and phagocytic function, including engulfment and intracellular killing of bacteria, when exposed to elevated glucose levels. Small blood vessel disease is also common among persons with diabetes, impairing the delivery of inflammatory cells, oxygen, and nutrients to the wound site.¹¹

The therapeutic administration of corticosteroid drugs decreases the inflammatory process and may delay the healing process. These hormones decrease capillary permeability during the early stages of inflammation, impair the phagocytic property of the leukocytes, and inhibit fibroblast proliferation and function.

Infection, Wound Separation, and Foreign Bodies

Wound contamination, wound separation, and foreign bodies delay wound healing. Infection impairs all dimensions of wound healing. It prolongs the inflammatory phase, impairs the formation of granulation tissue, and inhibits proliferation of fibroblasts and deposition of collagen fibers. All wounds are contaminated at the time of injury. Although body defenses can handle the invasion of microorganisms at the time of wounding, badly contaminated wounds can overwhelm host defenses. Trauma and existing impairment of host defenses also can contribute to the development of wound infections.

Approximation of the wound edges (*i.e.*, suturing of an incision type of wound) greatly enhances healing and prevents infection. Epithelialization of a wound with closely approximated edges occurs within 1 to 2 days. Large, gaping wounds tend to heal more slowly because it is often impossible to effect wound closure with this type of wound. Foreign bodies tend to invite bacterial contamination and delay healing. Fragments of wood, steel, glass, and other compounds may have entered the wound at the site of injury and can be difficult to locate when the wound is treated. Sutures are also foreign bodies, and although needed for the closure of surgical wounds, they are an impediment to healing. This is why sutures are removed as soon as possible after surgery. Wound infections are of special concern in persons with implantation of foreign bodies such as orthopedic devices (e.g., pins, stabilization devices), cardiac pacemakers, and shunt catheters. These infections are difficult to treat and may require removal of the device.

In summary, the ability of tissues to repair damage caused by injury depends on the body's ability to replace the parenchymal cells and to organize them as they were originally. Regeneration describes the process by which tissue is replaced with cells of a similar type and function. Healing by regeneration is limited to tissue with cells that are able to divide and replace the injured cells. Body cells are divided into types according to their ability to regenerate: labile cells, such as the epithelial cells of the skin and gastrointestinal tract, which continue to regenerate throughout life; stable cells, such as those in the liver, which normally do not divide but are capable of regeneration when confronted with an appropriate stimulus; and permanent or fixed cells, such as nerve cells, which are unable to regenerate. Scar tissue repair involves the substitution of fibrous connective tissue for injured tissue that cannot be repaired by regeneration.

Wound healing occurs by primary and secondary intention and is commonly divided into three phases: the inflammatory phase, the proliferative phase, and the maturational or remodeling phase. In wounds healing by primary intention, the duration of the phases is fairly predictable. In wounds healing by secondary intention, the process depends on the extent of injury and the healing environment. Wound healing can be impaired or complicated by factors such as malnutrition; restricted blood flow and oxygen delivery; diminished inflammatory and immune responses; and infection, wound separation, and the presence of foreign bodies.

TEMPERATURE REGULATION AND FEVER

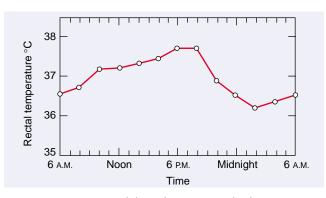
Fever is a clinical hallmark of infection and inflammation. This section of the chapter focuses on regulation of body temperature and fever caused by infectious and noninfectious conditions.

Body Temperature Regulation

The temperature within the deep tissues of the body (core temperature) is normally maintained within a range of 36.0°C to 37.5°C (97.0°F to 99.5°F).¹² Within this range, there are individual differences and diurnal variations; internal core temperatures reach their highest point in late afternoon and evening and their lowest point in the early morning hours (Fig. 9-11). Virtually all biochemical processes in the body are affected by changes in temperature. Metabolic processes speed up or slow down, depending on whether body temperature is rising or falling.

Body temperature reflects the difference between heat production and heat loss. Body heat is generated in the tissues of the body, transferred to the skin surface by the blood, and then released into the environment surrounding the body. The thermoregulatory center in the hypothalamus functions to modify heat production and heat losses as a means of regulating body temperature.

It is the core body temperature, rather than the surface temperature, that is regulated by the thermoregulatory center in the hypothalamus. This center integrates input from cold and warm thermal receptors located throughout the body and generates output responses that conserve body heat or increase its dissipation. The thermostatic set point of the thermoregulatory center is set so that the core temperature is regulated within the normal range. When body temperature begins to rise above the normal range, heat-dissipating behaviors are initiated; when the temperature falls below the normal range, heat production is increased. A core temperature greater than 41°C (105.8°F) or less than 34°C (93.2°F) usually indicates that the body's ability to thermoregulate is impaired (Fig. 9-12). Body responses that produce, conserve, and dissipate heat are described in Table 9-2. Spinal cord injuries that transect the cord at T6 or above can seriously impair temperature regulation because



■ FIGURE 9-11 ■ Normal diurnal variations in body temperature.

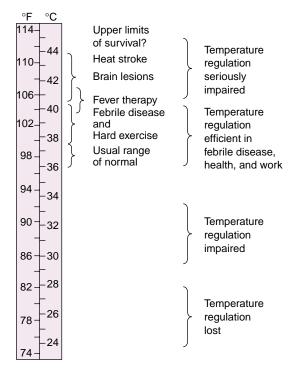
KEY CONCEPTS

FEVER

- Fever represents an increase in body temperature due to resetting of the hypothalamic thermoregulatory set point as the result of endogenous pyrogens released from host macrophages or endothelial cells.
- In response to the increase in set point, the hypothalamus initiates physiologic responses to increase core temperature to match the new set point.
- Fever is an adaptive response to bacterial and viral infections or to tissue injury. The growth rate of microorganisms is inhibited, and immune function is enhanced.

the hypothalamus can no longer control skin blood flow or sweating.

In addition to physiological thermoregulatory mechanisms, humans engage in voluntary behaviors to help regulate body temperature. These behaviors include the selection of proper clothing and regulation of environmental temperature through heating systems and air conditioning. Body positions that hold the extremities close to the body (*e.g.*, huddling or holding the extremities close to the body) prevent heat loss and are commonly assumed in cold weather.



■ **FIGURE 9-12** ■ Body temperatures under different conditions. (Dubois, E.F. [1948]. *Fever and the regulation of body temperature*. Courtesy of Charles C. Thomas, Publisher, Ltd., Springfield, IL)

TABLE 9-2

Heat Gain and Heat Loss Responses Used in Regulation

of Body Temperatur

Hea	at Gain	Heat Loss		
Body Response	Mechanism of Action	Body Response	Mechanism of Action	
Vasoconstriction of the superficial blood vessels	Confines blood flow to the inner core of the body, with the skin and sub- cutaneous tissues acting as insula- tion to prevent loss of core heat	Dilatation of the super- ficial blood vessels	Delivers blood containing core heat to the periphery where it is dissipated through radiation, conduction, and convection	
Contraction of the pilomotor mus- cles that surround the hairs on the skin	Reduces the heat loss surface of the skin	Sweating	Increases heat loss through evaporation	
Assumption of the huddle position with the extremities held close to the body	Reduces the area for heat loss			
Shivering	Increases heat production by the muscles			
Increased production of epinephrine	Increases the heat production associ- ated with metabolism			
Increased production of thyroid hormone	Is a long-term mechanism that in- creases metabolism and heat production			

Mechanisms of Heat Production

Metabolism is the body's main source of heat production. The sympathetic neurotransmitters, epinephrine and norepinephrine, which are released when an increase in body temperature is needed, act at the cellular level to shift metabolism so energy production is reduced and heat production is increased. This may be one of the reasons fever tends to produce feelings of weakness and fatigue. Thyroid hormone increases cellular metabolism, but this response usually requires several weeks to reach maximal effectiveness.

Fine involuntary actions such as shivering and chattering of the teeth can produce a threefold to fivefold increase in body temperature. *Shivering* is initiated by impulses from the hypothalamus. The first muscle change that occurs with shivering is a general increase in muscle tone, followed by an oscillating rhythmic tremor involving the spinal-level reflex that controls muscle tone. Because no external work is performed, all of the energy liberated by the metabolic processes from shivering is in the form of heat.

Physical exertion increases body temperature. With strenuous exercise, more than three quarters of the increased metabolism resulting from muscle activity appears as heat within the body, and the remainder appears as external work.

Mechanisms of Heat Loss

Most of the body's heat is produced by the deeper core tissues (*i.e.*, muscles and viscera), which are insulated from the environment and protected against heat loss by the subcutaneous tissues. Adipose tissue is a particularly good insulator, conducting heat only one third as effectively as other tissues.

Heat is lost from the body through radiation and conduction from the skin surface; through the evaporation of sweat and insensible perspiration; through the exhalation of air that has been warmed and humidified; and through heat lost in urine and feces. Contraction of the *pilomotor muscles* of the skin, which raises the skin hair and produces goose bumps, reduces the surface area available for heat loss. Of these mechanisms, only heat losses that occur at the skin surface are directly under hypothalamic control.

Most of the body's heat losses occur at the skin surface as heat from the blood moves to the skin and from there into the surrounding environment. There are numerous *arteriovenous* (*AV*) *shunts* under the skin surface that allow blood to move directly from the arterial to the venous system (Fig. 9-13). These AV shunts are much like the radiators in a heating system. When the shunts are open, body heat is freely dissipated to the skin and surrounding environment; when the shunts are closed, heat is retained in the body. The blood flow in the AV shunts is controlled almost exclusively by the sympathetic ner-

vous system in response to changes in core temperature and environmental temperature. The transfer of heat from the skin to the environment occurs by means of radiation, conduction, convection, and evaporation.

Radiation. Radiation involves the transfer of heat through the air or a vacuum. Heat from the sun is carried by radiation. The human body radiates heat in all directions. The ability to dissipate body heat by radiation depends on the temperature of the environment. Environmental temperature must be less than that of the body for heat loss to occur.

Conduction. Conduction involves the direct transfer of heat from one molecule to another. Blood carries, or conducts, heat from the inner core of the body to the skin surface. Normally, only a small amount of body heat is lost through conduction to a cooler surface. However, loss of heat by conduction to air represents a sizable proportion of the body's heat loss.

The conduction of heat to the body's surface is influenced by blood volume. In hot weather, the body compensates by increasing blood volume as a means of dissipating heat. Exposure to cold produces a cold diuresis and a reduction in blood volume as a means of controlling the transfer of heat to the body's surface.

Convection. Convection refers to heat transfer through the circulation of air currents. Normally, a layer of warm air tends to remain near the body's surface; convection causes continual removal of the warm layer and replacement with air from the surrounding environment. The wind-chill factor that often is included in the weather report combines the effect of convection caused by wind with the still-air temperature.

Evaporation. Evaporation involves the use of body heat to convert water on the skin to water vapor. Water that diffuses through the skin independent of sweating is called *insensible perspiration*. Insensible perspiration losses are greatest in a dry

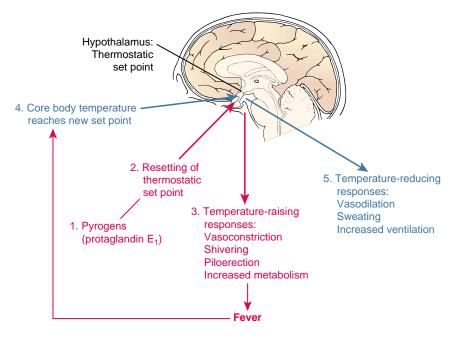
environment. Sweating occurs through the sweat glands and is controlled by the sympathetic nervous system. In contrast to other sympathetically mediated functions, sweating relies on acetylcholine, rather that the catecholamines, as a neurotransmitter. This means that anticholinergic drugs, such as atropine, can interfere with heat loss by interrupting sweating.

Evaporative heat losses involve insensible perspiration and sweating, with 0.58 calories being lost for each gram of water that is evaporated.¹² As long as body temperature is greater than the atmospheric temperature, heat is lost through radiation. However, when the temperature of the surrounding environment becomes greater than skin temperature, evaporation is the only way the body can rid itself of heat. Any condition that prevents evaporative heat losses causes the body temperature to rise.

Fever

Fever, or *pyrexia*, describes an elevation in body temperature that is caused by a cytokine-induced upward displacement of the set point of the hypothalamic thermoregulatory center. Fever is resolved or "broken" when the factor that caused the increase in the set point is removed. Fevers that are regulated by the hypothalamus usually do not rise above 41°C (105.8°F), suggesting a built-in thermostatic safety mechanism. Temperatures above that level are usually the result of superimposed activity, such as convulsions, hyperthermic states, or direct impairment of the temperature control center.

Fever can be caused by a number of microorganisms and substances that are collectively called pyrogens (Fig. 9-14). Many proteins, breakdown products of proteins, and certain other substances, including lipopolysaccharide toxins released from bacterial cell membranes, can cause the set point of the hypothalamic thermostat to increase. Some pyrogens can act directly and immediately on the hypothalamic thermoregulatory center to increase its set point. Other pyrogens, sometimes called *exogenous pyrogens*, act indirectly and may require several hours to produce their effect.¹²



■ FIGURE 9-14 ■ Mechanisms of fever. (1) Release of endogenous pyrogen from inflammatory cells, (2) resetting of hypothalamus thermostatic set point to a higher level (prodrome), (3) generation of hypothalamicmediated responses that raise body temperature (chill), (4) development of fever with elevation of body to new thermostatic set point, and (5) production of temperaturelowering responses (flush and defervescence) and return of body temperature to a lower level.

164 Unit Two: Alterations in Body Defenses

Exogenous pyrogens induce host cells, such as blood leukocytes and tissue macrophages, to produce fever-producing mediators called *endogenous pyrogens* (e.g., interleukin-1). For example, the phagocytosis of bacteria and breakdown products of bacteria that are present in the blood lead to the release of endogenous pyrogens into the circulation. The endogenous pyrogens are thought to increase the set point of the hypothalamic thermoregulatory center through the action of prostaglandin E2.12 In response to the sudden increase in set point, the hypothalamus initiates heat production behaviors (shivering and vasoconstriction) that increase the core body temperature to the new set point, and fever is established. In addition to their fever-producing actions, the endogenous pyrogens mediate a number of other responses. For example, interleukin-1 is an inflammatory mediator that produces other signs of inflammation, such as leukocytosis, anorexia, and malaise.

Many noninfectious disorders, such as myocardial infarction, pulmonary emboli, and neoplasms, produce fever. In these conditions, the injured or abnormal cells incite the production of pyrogen. For example, trauma and surgery can be associated with several days of fever. Some malignant cells, such as those of leukemia and Hodgkin's disease, secrete pyrogen.

A fever that has its origin in the central nervous system is sometimes referred to as a *neurogenic fever*.¹³ It usually is caused by damage to the hypothalamus caused by central nervous system trauma, intracerebral bleeding, or an increase in intracranial pressure. Neurogenic fevers are characterized by a high temperature that is resistant to antipyretic therapy and is not associated with sweating.

The purpose of fever is not completely understood. However, from a purely practical standpoint, fever is a valuable index to health status. For many, fever signals the presence of an infection and may legitimize the need for medical treatment. In ancient times, fever was thought to "cook" the poisons that caused the illness. With the availability of antipyretic drugs in the late 19th century, the belief that fever was useful began to wane, probably because most antipyretic drugs also had analgesic effects.

There is little research to support the belief that fever is harmful unless the temperature is greater than 40°C (104°F). Animal studies have demonstrated a clear survival advantage in infected members with fever compared with animals that were unable to produce a fever.¹⁴ It has been shown that small elevations in temperature, such as those that occur with fever, enhance immune function. There is increased motility and activity of the white blood cells, stimulation of interferon production, and activation of T cells.^{15,16} Many of the microbial agents that cause infection grow best at normal body temperatures, and their growth is inhibited by temperatures in the fever range. For example, the rhinoviruses responsible for the common cold are cultured best at 33°C (91.4°F), which is close to the temperature in the nasopharynx. Temperaturesensitive mutants of the virus that cannot grow at temperatures greater than 37.5°C (99.5°F), produce fewer signs and symptoms.17

Patterns

The patterns of temperature change in persons with fever vary and may provide information about the nature of the causative agent.¹⁸ These patterns can be described as intermittent, remittent, sustained, or relapsing. An *intermittent fever* is one in which temperature returns to normal at least once every 24 hours. Intermittent fevers are commonly associated with conditions such as gram-negative/positive sepsis, abscesses, and acute bacterial endocarditis. In a *remittent fever*, the temperature does not return to normal and varies a few degrees in either direction. It is associated with viral upper respiratory tract, legionella, and mycoplasma infections. In a *sustained* or *continuous* fever, the temperature remains above normal with minimal variations (usually less than 0.55°C or 1°F). Sustained fevers are seen in persons with drug fever. A *recurrent* or *relapsing fever* is one in which there is one or more episodes of fever, each as long as several days, with one or more days of normal temperature between episodes. Relapsing fevers may be caused by a variety of infectious diseases, including tuberculosis, fungal infections, Lyme disease, and malaria.

Critical to the analysis of a fever pattern is the relation of heart rate to the level of temperature elevation. Normally, a 1°C rise in temperature produces a 15-bpm (beats per minute) increase in heart rate (1°F, 10 bpm).¹⁹ Most persons respond to an increase in temperature with an appropriate increase in heart rate. The observation that a rise in temperature is not accompanied by the anticipated change in heart rate can provide useful information about the cause of the fever. For example, a heart rate that is slower than would be anticipated can occur with Legionnaires' disease and drug fever, and a heart rate that is more rapid than anticipated can be symptomatic of hyperthyroidism and pulmonary emboli.

Manifestations

The physiologic behaviors that occur during the development of fever can be divided into four successive stages: a prodrome; a chill, during which the temperature rises; a flush; and defervescence. During the first or prodromal period, there are nonspecific complaints, such as mild headache and fatigue, general malaise, and fleeting aches and pains. During the second stage or chill, there is the uncomfortable sensation of being chilled and the onset of generalized shaking, although the temperature is rising. Vasoconstriction and piloerection usually precede the onset of shivering. At this point the skin is pale and covered with goose flesh. There is a feeling of being cold and an urge to put on more clothing or covering and to curl up in a position that conserves body heat. When the shivering has caused the body temperature to reach the new set point of the temperature control center, the shivering ceases, and a sensation of warmth develops. At this point, the *third stage* or *flush* begins, during which cutaneous vasodilation occurs and the skin becomes warm and flushed. The fourth, or defervescence, stage of the febrile response is marked by the initiation of sweating. Not all persons proceed through the four stages of fever development. Sweating may be absent, and fever may develop gradually, with no indication of a chill or shivering.

Common manifestations of fever are anorexia, myalgia, arthralgia, and fatigue. These discomforts are worse when the temperature rises rapidly or exceeds 39.5°C (103.1°F). Respiration is increased, and the heart rate usually is elevated. Dehydration occurs because of sweating and the increased vapor losses caused by the rapid respiratory rate. The occurrence of chills commonly coincides with the introduction of pyrogen into the circulation. This is one of the reasons that blood cultures to identify the organism causing the fever are usually drawn during the first signs of a chill.

Chapter 9: Inflammation, Tissue Repair, and Fever

Many of the manifestations of fever are related to the increases in the metabolic rate, increased need for oxygen, and use of body proteins as an energy source. During fever, the body switches from using glucose (an excellent medium for bacterial growth) to metabolism based on protein and fat breakdown. With prolonged fever, there is increased breakdown of endogenous fat stores. If fat breakdown is rapid, metabolic acidosis may result (see Chapter 6).

Headache is a common accompaniment of fever and is thought to result from the vasodilation of cerebral vessels occurring with fever. Delirium is possible when the temperature exceeds 40°C (104°F). In the elderly, confusion and delirium may follow moderate elevations in temperature. Because of the increasingly poor oxygen uptake by the aging lung, pulmonary function may prove to be a limiting factor in the hypermetabolism that accompanies fever in older persons. Confusion, incoordination, and agitation commonly reflect cerebral hypoxemia. Febrile seizures can occur in some children.²⁰ They usually occur with rapidly rising temperatures or at a threshold temperature that differs with each child.

The herpetic lesions, or fever blisters, that develop in some persons during fever are caused by a separate infection by the type 1 herpes simplex virus that established latency in the regional ganglia and is reactivated by a rise in body temperature.

Diagnosis and Treatment

Fever usually is a manifestation of a disease state, and as such, determining the cause of a fever is an important aspect of its treatment. For example, fevers caused by infectious diseases usually are treated with antibiotics, whereas other fevers, such as those resulting from a noninfectious inflammatory condition, may be treated symptomatically.

Sometimes it is difficult to establish the cause of a fever. A prolonged fever for which the cause is difficult to ascertain is often referred to as fever of unknown origin (FUO). FUO is defined as a temperature elevation of 38.3°C (101°F) or higher that is present for 3 weeks or longer.²¹ Among the causes of FUO are malignancies (i.e., lymphomas, metastases to the liver and central nervous system); infections such as human immunodeficiency virus or tuberculosis, or abscessed infections; and drug fever. Malignancies, particularly non-Hodgkin's lymphoma, are important causes of FUO in the elderly. Cirrhosis of the liver is another cause of FUO.

The methods of fever treatment focus on modifications of the external environment intended to increase heat transfer from the internal to the external environment, support of the hypermetabolic state that accompanies fever, protection of vulnerable body organs and systems, and treatment of the infection or condition causing the fever. Because fever is a disease symptom, its manifestation suggests the need for treatment of the primary cause.

Modification of the environment ensures that the environmental temperature facilitates heat transfer away from the body. Sponge baths with cool water or an alcohol solution can be used to increase evaporative heat losses. More profound cooling can be accomplished through the use of a cooling mattress, which facilitates the conduction of heat from the body into the coolant solution that circulates through the mattress. Care must be taken so that cooling methods do not produce vasoconstriction and shivering that decrease heat loss and increase heat production.

Adequate fluids and sufficient amounts of simple carbohydrates are needed to support the hypermetabolic state and prevent the tissue breakdown that is characteristic of fever. Additional fluids are needed for sweating and to balance the insensible water losses from the lungs that accompany an increase in respiratory rate. Fluids also are needed to maintain an adequate vascular volume for heat transport to the skin surface.

Antipyretic drugs, such as aspirin and acetaminophen, often are used to alleviate the discomforts of fever and protect vulnerable organs, such as the brain, from extreme elevations in body temperature. These drugs act by resetting the hypothalamic temperature control center to a lower level, presumably by blocking the activity of cyclooxygenase, an enzyme that is required for the conversion of arachidonic acid to prostaglandin E2.22

Fever in Children

The mechanisms for controlling temperature are not well developed in the infant. In infants younger than 3 months, a mild elevation in temperature (i.e., rectal temperature of 38°C [100.4°F]) can indicate serious infection that requires immediate medical attention.^{23,24} Fever without a source occurs frequently in infants and children and is a common reason for visits to the clinic or emergency department.

Both minor and life-threatening infections are common in the infant to 3-year age group.^{23,24} The most common causes of fever in children are minor or more serious infections of the respiratory system, urinary system, gastrointestinal tract, or central nervous system. Occult bacteremia and meningitis also occur in this age group and should be excluded as diagnoses. The Agency for Health Care Policy and Research Expert Panel has developed clinical guidelines for use in the treatment of infants and children 0 to 36 months of age with fever without a source.²⁵ The guidelines define fever in this age group as a rectal temperature of at least 38°C (100.4°F). The guidelines also point out that fever may result from overbundling or a vaccine reaction. When overbundling is suspected, it is suggested that the infant be unbundled and the temperature retaken after 15 to 30 minutes.

Fever in infants and children can be classified as low risk or high risk, depending on the probability of the infection progressing to bacteremia or meningitis. Infants usually are considered at low risk if they were delivered at term and sent home with their mother without complications and have been healthy with no previous hospitalizations or previous antimicrobial therapy. A white blood cell count and urinalysis are recommended as a means of confirming low-risk status. Signs of toxicity (and high risk) include lethargy, poor feeding, hypoventilation, poor tissue oxygenation, and cyanosis. Blood and urine cultures, chest radiographs, and lumbar puncture usually are done in high-risk infants and children to determine the cause of fever.

Infants with fever who are considered to be at low risk usually are managed on an outpatient basis providing the parents or caregivers are deemed reliable. Older children with fever without source also may be treated on an outpatient basis. Parents or caregivers require full instructions, preferably in writing, regarding assessment of the febrile child. They should be instructed to contact their health care provider should their child show signs suggesting sepsis. Infants younger than 3 months are evaluated carefully. Infants and children with signs of toxicity and/or petechiae (a sign of meningitis) usually are hospitalized for evaluation and treatment.²⁶ Parenteral antimicrobial therapy usually is initiated after samples for blood, urine, and spinal fluid cultures have been taken.

Fever in the Elderly

In the elderly, even slight elevations in temperature may indicate serious infection or disease. This is because the elderly often have a lower baseline temperature, and although they increase their temperature during an infection, it may fail to reach a level that is equated with significant fever.^{27,28}

Normal body temperature and the circadian pattern of temperature variation often are altered in the elderly. Elderly persons are reported to have a lower basal temperature (36.4°C [97.6°F] in one study) than do younger persons.²⁹ It has been recommended that the definition of fever in the elderly be expanded to include an elevation of temperature of at least 1.1°C (2°F) above baseline values.²⁸

It has been suggested that 20% to 30% of elders with serious infections present with an absent or blunted febrile response.²⁸ When fever is present in the elderly, it usually indicates the presence of serious infection, most often caused by bacteria. The absence of fever may delay diagnosis and initiation of antimicrobial treatment. Unexplained changes in functional capacity, worsening of mental status, weakness and fatigue, and weight loss are signs of infection in the elderly. They should be viewed as possible signs of infection and sepsis when fever is absent. The probable mechanisms for the blunted fever response include a disturbance in sensing of temperature by the thermoregulatory center in the hypothalamus, alterations in release of endogenous pyrogens, and the failure to elicit responses such as vasoconstriction of skin vessels, increased heat production, and shivering that increase body temperature during a febrile response.

Another factor that may delay recognition of fever in the elderly is the method of temperature measurement. Oral temperature remains the most commonly used method for measuring temperature in the elderly. It has been suggested that rectal and tympanic membrane methods are more effective in detecting fever in the elderly. This is because conditions such as mouth breathing, tongue tremors, and agitation often make it difficult to obtain accurate oral temperatures in the elderly.

In summary, body temperature is normally maintained within a range of 36.0°C to 37.4°C. Body heat is produced by metabolic processes that occur within deeper core structures of the body and is lost at the body's surface when core heat is transported to the skin by the circulating blood. The transfer of heat from the skin to the environment occurs through radiation, conduction, convection, and evaporation. The thermoregulatory center in the hypothalamus functions to modify heat production and heat losses as a means of regulating body temperature.

Fever represents an increase in body temperature outside the normal range. Fever can be caused by a number of factors, including microorganisms, trauma, and drugs or chemicals, all of which incite the release of endogenous pyrogens and subsequent resetting of the hypothalamic thermoregulatory center. The reactions that occur during fever consist of four stages: a prodrome, a chill, a flush, and defervescence. Many of the manifestations of fever are related to the increases in the metabolic rate, increased need for oxygen, and use of body proteins as an energy source.

Fever in infants and children can be classified as low risk or high risk, depending upon the probability of the infection progressing to bacteremia or meningitis. Infants younger than 28 days and those at high risk usually are hospitalized for evaluation of their fever and treatment. In the elderly, even slight elevations in temperature may indicate serious infection or disease. The elderly often have a lower baseline temperature, so serious infections may go unrecognized because of the perceived lack of a significant fever.

REVIEW QUESTIONS

State the five cardinal signs of acute inflammation and describe the physiologic mechanisms and mediators involved in production of these signs.

Describe the systemic manifestations associated with an acute inflammatory response.

Compare the etiology and pathogenesis of acute and chronic inflammation.

- Trace the wound-healing process through the inflammatory, proliferative, and remodeling phases.
- Explain the effect of malnutrition; ischemia and oxygen deprivation; impaired immune and inflammatory responses; and infection, wound separation, and foreign bodies on wound healing.
- Apply the physiologic mechanisms involved in body temperature regulation to describe the four stages of fever.
- Describe the criteria used when determining the seriousness of fever without source in children younger than 36 months.
- State the definition for fever in the elderly and cite possible mechanisms for altered febrile responses in the elderly.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- Fantone J.C., Ward P.A. (1999). Inflammation. In Rubin E., Farber J.L. (Eds.), *Pathology* (3rd ed., pp. 37–75). Philadelphia: Lippincott Williams & Wilkins.
- Mitchell R.N., Cotran R.S. (2003). Acute and chronic inflammation. In Kumar V., Cotran R.S., Robbins S. (Eds.), *Basic pathology* (7th ed., pp. 33–59). Philadelphia: W.B. Saunders.
- Chandrasoma P., Taylor C.R. (1998). Concise pathology (3rd ed., pp. 31–92). Stamford, CT: Appleton & Lange.

Chapter 9: Inflammation, Tissue Repair, and Fever

- Martinez-Hernandez A. (1999). Repair, regeneration, and fibrosis. In Rubin E., Farber J.L. (Eds.), *Pathology* (3rd ed., pp. 77–103). Philadelphia: Lippincott Williams & Wilkins.
- Mitchell R.N., Cotran R.S. (2003). Tissue repair: Cell regeneration and fibrosis. In Kumar V., Cotran R.S., Robbins S. (Eds.), *Basic pathology* (7th ed., pp. 61–78). Philadelphia: W.B. Saunders.
- Waldorf H., Fewkes J. (1995). Wound healing. Advances in Dermatology 10, 77–95.
- 7. Flynn M.B. (1996). Wound healing and critical illness. *Critical Care Clinics of North America* 8, 115–124.
- Orgill D., Deming H.R. (1988). Current concepts and approaches to healing. *Critical Care Medicine* 16, 899–908.
- 9. Albina J.E. (1995). Nutrition and wound healing. *Journal of Parenteral and Enteral Nutrition* 18, 367–376.
- Whitney J.D. (1990). The influence of tissue oxygenation and perfusion on wound healing. *Clinical Issues in Critical Care Nursing* 1, 578–584.
- King L. (2000). Impaired wound healing in patients with diabetes. Nursing Standards 15(38), 39–45.
- Guyton A.C., Hall J.E. (2000). *Textbook of medical physiology* (10th ed., pp. 822–833). Philadelphia: W.B. Saunders.
- Saper C.B., Breder C.D. (1994). The neurologic basis of fever. New England Journal of Medicine 330, 1880–1886.
- Roberts N.J. (1979). Temperature and host defenses. *Microbiological Reviews* 43(2), 241–259.
- Mackowiak P.A. (1998). Concepts of fever. Archives of Internal Medicine 158, 1870–1881.
- Blatteis C.M. (1998). Fever. In Blatteis C.M. (Ed.), Physiology and pathophysiology of temperature regulation (pp. 178–192). River Edge, NJ: World Scientific Publishing.
- Rodbard D. (1981). The role of regional temperature in the pathogenesis of disease. *New England Journal of Medicine* 305, 808–814.

- Cunha B.A. (1996). The clinical significance of fever patterns. Infectious Disease Clinics of North America 10, 33–43.
- 19. McGee Z.A., Gorby G.L. (1987). The diagnostic value of fever patterns. *Hospital Practice* 22(10), 103–110.
- Champi C., Gaffney-Yocum P.A. (1999). Managing febrile seizures in children. Nurse Practitioner 24 (10), 28–30, 34–35.
- Cunha B.A. (1996). Fever without source. Infectious Disease Clinics of North America 10, 111–127.
- Plaisance K.I., Mackowiak P.A. (2000). Antipyretic therapy: Physiologic rationale, diagnostic implications, and clinical consequences. *Archives of Internal Medicine* 160, 449–456.
- 23. Baker M.D. (1999). Evaluation and management of infants with fever. *Pediatric Clinics of North America* 46, 1061–1072.
- 24. Park J.W. (2000). Fever without source in children. *Postgraduate Medicine* 107, 259–266.
- Baraff L.J., Bass J.W., Fleisher G.R., et al. (1993). Practice guidelines for the management of infants and children 0 to 36 months of age with fever without source. Agency for Health Care Policy and Research. (Erratum appears in Ann Emerg Med. [1993]. 22(9), 1490.) Annals of Emergency Medicine 22(7), 1198–1210.
- Powell K.R. (2000). Fever without focus. In Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.), *Nelson's textbook of pediatrics* (16th ed., pp. 742–747). Philadelphia: W.B. Saunders.
- Yoshikawa T.T., Norman D.C. (1996). Approach to fever and infections in the nursing home. *Journal of the American Geriatric Society* 44, 74–82.
- Yoshikawa T.T., Norman, D.C. (1998). Fever in the elderly. *Infectious Medicine* 15, 704–706, 708.
- Castle S.C., Yeh M., Toledo S., et al. (1993). Lowering the temperature criterion improves detection of infections in nursing home residents. *Aging Immunology and Infectious Disease* 4, 67–76.

CHAPTER 10

Alterations in the Immune Response

Allergic and Hypersensitivity Disorders Type I IgE-Mediated Disorders Systemic Anaphylactic Reactions Localized Atopic Disorders Type II, Antibody-Mediated Cytotoxic Disorders Type III, Immune Complex-Mediated Disorders Type IV, Cell-Mediated Hypersensitivity Disorders Allergic Contact Dermatitis Hypersensitivity Pneumonitis Latex Allergy **Transplantation Immunopathology** Host-Versus-Graft Disease Graft-Versus-Host Disease **Autoimmune Disorders** Immunologic Tolerance Mechanisms of Autoimmune Disease Heredity and Gender Failure of Self-tolerance **Diagnosis and Treatment Immunodeficiency Disorders Primary Immunodeficiency Disorders** Humoral (B-Cell) Immunodeficiencies Cellular (T-Cell) Immunodeficiencies Combined T-Cell and B-Cell Immunodeficiencies Acquired Immunodeficiency Syndrome Transmission of HIV Infection Pathophysiology of AIDS **Diagnosis and Classification Clinical Course Opportunistic Infections** Cancers and Malignancies Wasting Syndrome Metabolic Disorders Treatment Infection in Pregnancy and in Infants and Children The human immune network is a multifaceted defense system that has evolved to protect against invading microorganisms, prevent the proliferation of cancer cells, and mediate the healing of damaged tissue. Under normal conditions, the immune response deters or prevents disease. However, occasionally the inadequate, inappropriate, or misdirected activation of the immune system can lead to debilitating or life-threatening illnesses, typified by allergic or hypersensitivity reactions, transplantation immunopathology, autoimmune disorders, and immunodeficiency states.

ALLERGIC AND HYPERSENSITIVITY DISORDERS

Hypersensitivity is defined as an exaggerated immune response to a foreign agent resulting in injury to the host.^{1,2} Allergic or hypersensitivity disorders are caused by immune responses to environmental antigens that produce inflammation and cause tissue injury. In the context of an allergic response, these antigens usually are referred to as allergens. Allergens are any foreign substances capable of inducing an immune response. Many different chemicals of natural and synthetic origin are known allergens. Complex natural organic chemicals, especially proteins, are more likely to cause an immediate hypersensitivity response, whereas simple organic compounds, inorganic chemicals, and metals more commonly cause delayed hypersensitivity reactions. Exposure to the allergen can be through inhalation, ingestion, injection, or skin contact. Sensitization of a specific individual to a specific allergen is the result of a particular interplay between the chemical or physical properties of the allergen, the mode and quantity of exposure, and the unique genetic makeup of the person.

The manifestations of allergic responses reflect the effect of an immunologically induced inflammatory response in the organ or tissue involved. These manifestations usually are independent of the agent involved. For example, the symptoms of hay fever are the same whether the allergy is caused by ragweed pollen or mold spores. The diversity of allergic responses derives from the different immunologic effector pathways that are involved (*e.g.*, hay fever vs. allergic dermatitis).

Historically, hypersensitivity disorders have been categorized as four types: type I, IgE-mediated disorders; type II,

ALLERGIC AND HYPERSENSITIVITY DISORDERS

- Allergic and hypersensitivity disorders result from immune responses to exogenous and endogenous antigens that produce inflammation and cause tissue damage.
- Type I hypersensitivity is an IgE-mediated immune response that leads to the release of inflammatory mediators for sensitized mast cells.
- Type II disorders involve humoral antibodies that participate directly in injuring cells by predisposing them to phagocytosis or lysis.
- Type III disorders result in generation of immune complexes in which humoral antibodies bind antigen and activate complement. The fractions of complement attract inflammatory cells that release tissue-damaging products.
- Type IV disorders involve tissue damage in which cell-mediated immune responses with sensitized T lymphocytes cause cell and tissue injury.

antibody-mediated (cytotoxic) disorders; type III, immune complex-mediated disorders; and type IV, cell-mediated hypersensitivity reactions (Table 10-1). Latex allergy is a newly emerging disorder that can result from an IgE-mediated or T-cell-mediated hypersensitivity response.

Type I, IgE-Mediated Disorders

Type I reactions are immediate-type hypersensitivity reactions that are triggered by binding of an allergen to a specific IgE that is found on the surface of mast cells or basophils. In addition to its role in allergic responses, IgE is involved in acquired immunity to parasitic infections.

The mast cells, which are tissue cells, and basophils, which are blood cells, are derived from hematopoietic (blood) precursor cells. Mast cells normally are distributed throughout connective tissue, especially in areas beneath the skin and mucous membranes of the respiratory, gastrointestinal, and genitourinary tracts, and adjacent to blood and lymph vessels.³ This location places mast cells near surfaces that are exposed to environmental antigens and parasites. Mast cells and basophils have granules that contain potent mediators of allergic reactions. These mediators are preformed in the cell or activated through enzymatic processing. During the sensitization or priming stage, the allergen-specific IgE antibodies attach to receptors on the surface of mast cells and basophils. With subsequent exposure, the sensitizing allergen binds to the cellassociated IgE and triggers a series of events that ultimately lead to degranulation of the sensitized mast cells or basophils, causing release of their allergy-producing mediators (Fig. 10-1).

The primary (preformed) mediators of allergic reactions include histamine, acetylcholine, adenosine, chemotactic mediators, and neutral proteases that are released from mast cells. Histamine, the most important of these preformed mediators, is a potent vasodilator that increases the permeability of capillaries and venules, and causes bronchoconstriction and increased secretion of mucus. Acetylcholine produces bronchial smooth muscle contraction and dilation of small blood vessels. The proteases generate kinins and cleave complement components to produce additional chemotactic and inflammatory mediators. Secondary mediators include leukotrienes and prostaglandins that are generated from arachidonic acid in the mast cell membrane (see Chapter 9). The leukotrienes and prostaglandins produce responses similar to those of histamine and acetylcholine, although their effects are delayed and prolonged by comparison. Platelet-activating factor is another secondary mediator, resulting in platelet aggregation, histamine release, and bronchospasm. It also acts as a chemotactic factor for neutrophils and eosinophils. Mast cells also produce a number of cytokines that play a role in type I hypersensitivity responses through their ability to recruit and activate a variety of inflammatory cells.

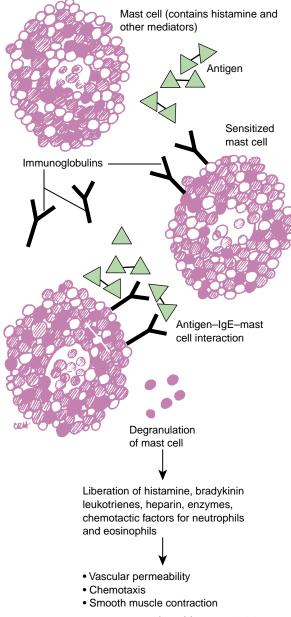
Type I hypersensitivity reactions may present as a systemic disorder (anaphylaxis) or a localized reaction (atopy).²

Systemic Anaphylactic Reactions

Systemic anaphylactic disorders often result from injected allergens (*e.g.*, penicillin, radiographic contrast dyes, bee or wasp stings). More rarely, they may result from ingested allergens (seafood, nuts, legumes). In sensitized individuals, only a small amount of the allergen may be required to produce a

TABLE 10-1	Classification of Hypersensitivity Responses			
Туре		Mechanism	Examples	
I — Anaphylactic hypersensitiv	. ,	IgE-mediated—mast cell degranulation	Hay fever, asthma, anaphylaxis	
II — Cytotoxic	,	Formation of antibodies (IgG, IgM) against cell sur- face antigens. Complement usually is involved.	Autoimmune hemolytic anemia, hemolytic disease of the newborn, Goodpasture's disease	
III— Immune con	nplex disease	Formation of antibodies (IgG, IgM, IgA) that interact with exogenous or endogenous anti- gens to form antigen- antibody complexes.	Arthus reaction, autoimmune diseases (syste- mic lupus erythematosus, rheumatoid arthri- tis), certain forms of acute glomerulosclerosis	
IV— Cell-mediate hypersensitiv		Sensitized T lymphocytes release cytokines and produce T-cell-mediated cytotoxicity.	Tuberculosis, contact dermatitis, transplant rejection	

170 Unit Two: Alterations in Body Defenses



■ FIGURE 10-1 ■ Type I, IgE-mediated hypersensitivity reaction. Exposure to the antigen causes sensitization of the mast cell; subsequent binding of the antigen to the sensitized degranulation of mast cell with release of potent inflammatory mediators, such as histamine, that are responsible for the hypersensitivity reactions.

reaction. Anaphylaxis has a rapid onset, often within minutes, of itching, urticaria (hives), gastrointestinal cramps, and difficulty breathing caused by bronchospasm. Angioedema (swelling of face and throat) may develop, causing upper airway obstruction. Massive vasodilation may lead to peripheral pooling of blood, a profound drop in blood pressure, and life-threatening circulatory shock (see Chapter 18).

Localized Atopic Disorders

Localized reactions generally occur when the antigen is confined to a particular site, usually related to the route of exposure. Localized type I hypersensitivity reactions appear to be genetically determined and the term *atopy* is often used to imply a hereditary predisposition to such reactions. Persons with atopic disorders commonly are allergic to more than one, and often many, environmental allergens. They tend to have high serum levels of IgE and increased numbers of basophils and mast cells. Although the IgE-triggered response is likely to be a key factor in the pathophysiology of the disorders, it is not the only factor. It is possible that persons with atopic disorders are exquisitely responsive to the chemical mediators of allergic reactions, rather than having a hyperactive IgE immune response.

Atopic disorders include food allergies, allergic rhinitis (hay fever), allergic dermatitis, and certain forms of bronchial asthma. The discussion in this section focuses on allergic rhinitis and food allergy. Allergic asthma is discussed in Chapter 21 and atopic dermatitis in Chapter 44.

Allergic Rhinitis. Allergic rhinitis (i.e., allergic rhinoconjunctivitis) is characterized by symptoms of sneezing, itching, and watery discharge from the eyes and nose. Allergic rhinitis not only produces nasal symptoms but also frequently is associated with other chronic airway disorders, such as sinusitis and bronchial asthma.⁴ Severe attacks may be accompanied by systemic malaise, fatigue, and muscle soreness from sneezing. Fever is absent. Sinus obstruction may cause headache. Typical allergens include pollens from ragweed, grasses, trees, and weeds; fungal spores; house dust mites; animal dander; and feathers. Allergic rhinitis can be divided into perennial and seasonal allergic rhinitis, depending on the chronology of symptoms. Persons with the perennial type of allergic rhinitis experience symptoms throughout the year, but those with seasonal allergic rhinitis (i.e., hay fever) are plagued with intense symptoms in conjunction with periods of high allergen (e.g., pollens, fungal spores) exposure. Symptoms that become worse at night suggest a household allergen, and symptoms that disappear on weekends suggest occupational exposure.

Diagnosis depends on a careful history and physical examination, microscopic identification of nasal eosinophilia, and skin testing to identify the offending allergens.

Treatment is symptomatic in most cases and includes the use of oral antihistamines and decongestants.⁴ Intranasal corticosteroids often are effective when used appropriately. Intranasal cromolyn, a drug that stabilizes mast cells and prevents their degranulation, may be useful, especially when administered before expected contact with an offending allergen. The anticholinergic agent ipratropium, which is available as a nasal spray, also may be used. When possible, avoidance of the offending allergen is recommended.

Food Allergies. Virtually any food can produce atopic allergies. The primary target of food allergy may be the skin, the gastrointestinal tract, or the respiratory system. The foods most commonly causing these reactions in children are milk, eggs, peanuts, soy, tree nuts, fish, and shellfish foods (*i.e.*, crustaceans and mollusks).⁵ In adults, such foods are peanuts, shellfish, and fish. The allergenicity of a food may be changed by heating or cooking. A person may be allergic to drinking milk but may not have symptoms when milk is included in cooked foods. Both acute reactions (*e.g.*, anaphylaxis) and chronic reactions (*e.g.*, asthma, atopic dermatitis, and gastrointestinal disorders) can occur. Anaphylaxis is a serious and sometimes fatal systemic reaction. The foods most responsible for anaphylaxis are peanuts, tree nuts (*e.g.*, walnuts, almonds, pecans, cashews, hazelnuts), and shellfish.

Although food allergies can occur at any age, they tend to manifest during childhood. The allergic response is thought to occur after contact between specific food allergens and IgEsensitized mast cells found in the intestinal mucosa causes local and systemic release of histamine and other mediators of the allergic response. In this disorder, allergens usually are food proteins and partially digested food products. Carbohydrates, lipids, or food additives, such as preservatives, colorings, or flavorings, also are potential allergens. Closely related food groups can contain common cross-reacting allergens. For example, some persons are allergic to all legumes (*i.e.*, beans, peas, and peanuts).

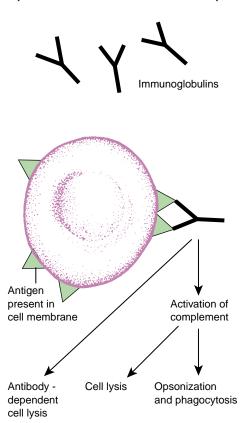
Diagnosis of food allergies usually is based on careful food history and provocative diet testing. Provocative testing involves careful elimination of a suspected allergen from the diet for a period of time to see if the symptoms disappear and reintroducing the food to see if the symptoms reappear. Only one food should be tested at a time. Treatment focuses on avoidance of the food or foods responsible for the allergy. However, avoidance may be difficult for persons who are exquisitely sensitive to a particular food protein because foods may be contaminated with the protein during processing or handling of the food. For example, contamination may occur when chocolate candies without peanuts are processed with the same equipment used for making candies with peanuts. In predisposed persons, even using the same spatula to serve cookies with and without peanuts can cause enough contamination to produce a severe anaphylactic reaction.

Type II, Antibody-Mediated Cytotoxic Disorders

Type II (cytotoxic) hypersensitivity reactions are the end result of direct interaction between IgG and IgM class antibodies and tissue or cell surface antigens, with subsequent activation of complement- or antibody-dependent cell-mediated cytotoxicity (Fig. 10-2). Examples of type II reactions include mismatched blood transfusions, hemolytic disease of the newborn caused by ABO or Rh incompatibility, and certain drug reactions. In the latter, the binding of certain drugs to the surface of red or white blood cells elicits an antibody and complement response that lyses the drug-coated cell. Lytic drug reactions can produce transient anemia, leukopenia, or thrombocytopenia, which are corrected by the removal of the offending drug.

Type III, Immune-Complex Disorders

Immune complex disorders are mediated by the formation of insoluble antigen-antibody complexes that activate complement (Fig. 10-3). Activation of complement by the immune complex generates chemotactic and vasoactive mediators that cause tissue damage by a variety of mechanisms, including alterations in blood flow, increased vascular permeability, and the destructive action of inflammatory cells. The reaction occurs when the antigen combines with antibody, whether in the

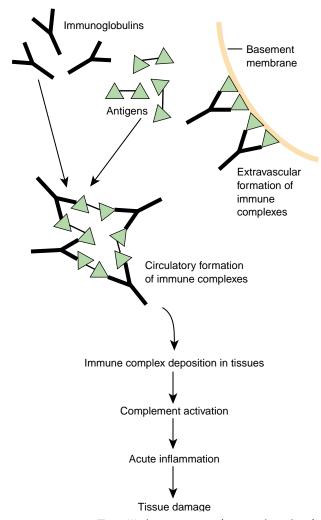


■ FIGURE 10-2 ■ Type II, cytotoxic hypersensitivity reactions involve formation of immunoglobulins (IgG and IgM) against cell surface antigens. The antigen-antibody response leads to (1) complement-mediated mechanisms of cell injury or to (2) antibody cytotoxicity that does not require the complement system.

circulation (circulating immune complexes) or at extravascular sites where antigen may have been deposited. Immune complexes formed in the circulation produce damage when they come in contact with the vessel lining or are deposited in tissues, including the renal glomerulus, skin venules, the lung, and joint synovium. Once deposited, the immune complexes elicit an inflammatory response by activating complement, thereby leading to chemotactic recruitment of neutrophils and other inflammatory cells.

There are two general types of antigens that cause immunecomplex mediated injury: (1) exogenous antigens such as viral and bacterial proteins and (2) endogenous antigens such as self-antigens associated with autoimmune disorders. Type III reactions are responsible for the acute glomerulonephritis that follows a streptococcal infection and the manifestations of autoimmune disorders such as systemic lupus erythematosus (SLE). Unlike type II reactions, in which the damage is caused by binding of antibody to body cells, the harmful effects of type III reactions are indirect (*i.e.*, secondary to the inflammatory response induced by activated complement).

Acute serum sickness is the prototype of a systemic immune complex disease. The term *serum sickness* was originally coined to describe a syndrome consisting of rash, lymphadenopathy, arthralgias, and occasionally neurologic disorders that appeared 7 or more days after injections of horse antisera for prevention of tetanus. Although this therapy is not used



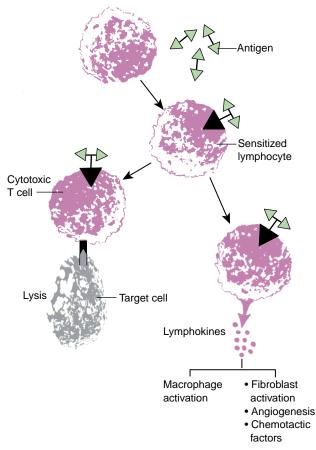
■ FIGURE 10-3 ■ Type III, immune complex reactions involve complement-activating IgG and IgM immunoglobulins with formation of blood-borne immune complexes that are eventually deposited in tissues. Complement activation at the site of immune complex deposition leads to recruitment of leukocytes, which are eventually responsible for tissue injury.

today, the name remains. The most common contemporary causes of this allergic disorder include antibiotics (especially penicillin), various foods, drugs, and insect venoms. Serum sickness is triggered by the deposition of insoluble antigenantibody (IgM and IgG) complexes in blood vessels, joints, heart, and kidney tissue. The deposited complexes activate complement, increase vascular permeability, and recruit phagocytic cells, all of which can promote focal tissue damage and edema. The signs and symptoms include urticaria, patchy or generalized rash, extensive edema (usually of the face, neck, and joints), and fever. In most cases, the damage is temporary, and symptoms resolve within a few days. However, a prolonged and continuous exposure to the sensitizing antigen can lead to irreversible damage. In previously sensitized persons, severe and even fatal forms of serum sickness may occur immediately or within several days after the sensitizing drug or serum is administered.

Type IV, Cell-Mediated Hypersensitivity Disorders

Unlike other hypersensitivity reactions, type IV, delayed hypersensitivity, is mediated by cells, not antibodies. Type IV hypersensitivity reactions usually occur 24 to 72 hours after exposure of a sensitized individual to the offending antigen. They are mediated by T lymphocytes that are directly cytotoxic (CD8⁺ T cells) or that secrete inflammatory mediators (CD4⁺ T cells) that cause tissue changes (Fig. 10-4). The reaction is initiated by antigen-specific CD4⁺ helper T cells, which release numerous immunoregulatory and proinflammatory cytokines into the surrounding tissue. These substances attract antigenspecific and antigen-nonspecific T or B lymphocytes as well as monocytes, neutrophils, eosinophils, and basophils. Some of the cytokines promote differentiation and activation of macrophages that function as phagocytic and antigen-presenting cells (APCs). Activation of the coagulation cascade leads to formation and deposition of fibrin.

The best-known type of delayed hypersensitivity response is the reaction to the tuberculin test, in which inactivated tuberculin or purified protein derivative is injected under the skin. In a previously sensitized person, redness and induration of the area develop within 8 to 12 hours, reaching a peak in



■ FIGURE 10-4 ■ Type IV, cell-mediated or delayed-type hypersensitivity reactions involve sensitization of T lymphocytes with the subsequent formation of cytotoxic T cells that lyse target cells or T cells that release cell-damaging lymphokines.

24 to 72 hours. A positive tuberculin test indicates that a person has had sufficient exposure to the *M. tuberculosis* organism to incite a hypersensitivity reaction; it does not mean that the person has tuberculosis. Certain types of antigens induce cellmediated immunity with an especially pronounced macrophage response. This type of delayed hypersensitivity commonly develops in response to particulate antigens that are large, insoluble, and difficult to eliminate. The accumulated macrophages are often transformed into so-called *epithelioid cells* because they resemble epithelium. A microscopic aggregation of epithelioid cells, which usually are surrounded by a layer of lymphocytes, is called a *granuloma*. Inflammation that is characterized by this type of type IV hypersensitivity is called *granulomatous inflammation* (see Chapter 9).

Direct T-cell-mediated cytotoxicity, which causes necrosis of antigen-bearing cells, is believed to be important in the eradication of virus-infected cells, autoimmune diseases such as Hashimoto's thyroiditis, and host-versus-graft or graft-versushost transplant rejection. Allergic contact dermatitis and hypersensitivity pneumonitis are presented as examples of cellmediated hypersensitivity reactions.

Allergic Contact Dermatitis

Allergic contact dermatitis denotes an inflammatory response confined to the skin that is initiated by re-exposure to an allergen to which a person had previously become sensitized (*e.g.*, cosmetics, hair dyes, metals, topical drugs). Contact dermatitis usually consists of erythematous macules, papules, and vesicles (*i.e.*, blisters). The affected area often becomes swollen and warm, with exudation, crusting, and development of a secondary infection. The location of the lesions often provides a clue about the nature of the antigen causing the disorder. The most common form of this condition is the dermatitis that follows an intimate encounter with poison ivy or poison oak antigens, although many other substances can trigger a reaction.

The mechanism of events leading to prior sensitization to an antigen is not completely understood. It is likely that sensitization follows transdermal transport of an antigen, with subsequent presentation to T lymphocytes. Subpopulations of sensitized lymphocytes are distributed throughout the body so that subsequent cutaneous exposure to the offending antigen promotes a localized reaction regardless of the initial site of contact. The severity of the reaction associated with contact dermatitis ranges from mild to intense, depending on the person and the allergen. Because this condition follows the time course of a delayed hypersensitivity response, the reaction does not become apparent for at least 12 hours and usually more than 24 hours after exposure. Depending on the antigen and the duration of exposure, the reaction may last from days to weeks and is typified by erythematous, vesicular, or papular lesions associated with intense pruritus and weeping.

Diagnosis of contact dermatitis is made by observing the distribution of lesions on the skin surface and associating a particular pattern with exposure to possible allergens. If a particular allergen is suspected, a patch test can be used to confirm the suspicion. Treatment usually is limited to removal of the irritant and application of topical preparations (*e.g.*, ointments, corticosteroid creams) to relieve symptomatic skin lesions and prevent secondary bacterial infections. Severe reactions may require systemic corticosteroid therapy.

Hypersensitivity Pneumonitis

Hypersensitivity pneumonitis or allergic alveolitis is associated with exposure to inhaled organic dusts or related occupational antigens. The disorder is thought to involve a susceptible host and activation of pulmonary T cells, followed by the release of cytokine mediators of inflammation.6 The inflammatory response that ensues (usually several hours after exposure) produces labored breathing, dry cough, chills and fever, headache, and malaise. The symptoms usually subside within hours after the sensitizing antigens are removed. A primary example of hypersensitivity pneumonitis is "farmer's lung," a condition resulting from exposure to moldy hay. Other sensitizing antigens include tree bark, sawdust, animal danders, and Actinomycetes bacteria that are occasionally found in humidifiers, hot tubs, and swimming pools. Exposure to small amounts of antigen for a long period may lead to chronic lung disease with minimal reversibility. This can happen to persons exposed to avian or animal antigens or a contaminated home air humidifier.6

The most important element in the diagnosis of hypersensitivity pneumonitis is to obtain a good history (occupational and otherwise) of exposure to possible antigens. Skin tests, when available, and serum tests for precipitating antibody can be done. Occasionally, direct observation of the person's work and other environments may help to establish a diagnosis. Treatment consists of identifying and avoiding the offending antigens. Severe forms of the disorder may be treated with systemic corticosteroid therapy.

Latex Allergy

With the advent of the human immunodeficiency virus (HIV) and other blood-borne pathogens, the use of natural latex gloves has spiraled. Along with the expanded use of latex gloves have come increased reports of latex allergy among health care workers. Other persons at high risk of sensitization are those with prolonged exposure to latex, including persons who have undergone repeated surgeries. Exposure to latex may occur by cutaneous, mucous membrane, inhalation, internal tissue, or intravascular routes. Most severe reactions have resulted from latex proteins coming in contact with the mucous membranes of the mouth, vagina, urethra, or rectum. Children with meningomyelocele (spina bifida) who undergo frequent examinations and treatments involving the mucosal surface of the bladder or rectum are at particular risk for the development of latex allergy.^{7,8} A large number of latex products are used in dentistry, and oral mucosal contact is common during dental procedures. Anaphylactic reactions have been caused by exposure of the internal organs to the surgeon's gloves during surgery.

Allergic reactions to latex products can be triggered by the latex proteins or by the additives used in the manufacturing process. Natural rubber latex is derived from the milky sap of the *Hevea brasiliensis* plant or rubber tree. Various accelerants, curing agents, antioxidants, and stabilizers are added to the liquid latex during the manufacturing process. Cornstarch powder is applied to the gloves during the manufacturing process to prevent stickiness and give the gloves a smooth feel. The cornstarch glove powder has an important role in the allergic response. Latex proteins are readily absorbed by glove powder and become airborne during removal of the gloves. Operating rooms and other high exposure areas where powdered gloves are used contain sufficiently high levels of aerosolized latex to produce symptoms in sensitized persons.

Latex allergy can involve a type I, IgE-mediated hypersensitivity reaction, or a type IV, cell-mediated response. The distinction between the type I and type IV reactions to latex products is not always clear. Affected individuals may experience both types of reactions. Persons with latex allergy commonly show cross-sensitivity to bananas, avocado, kiwi, tomatoes, and chestnuts, probably because latex proteins are similar to the proteins in these foods.^{7,8} These foods have been responsible for anaphylactic reactions in latex-sensitive persons.

The most common type of allergic reaction to latex gloves is a contact dermatitis caused by a type IV, delayed hypersensitivity reaction to rubber additives. It usually develops 48 to 96 hours after direct contact with latex additives. It often affects the dorsum of the hands and is characterized by a vesicular rash. When latex contact is continued, the area becomes crusted and thickened. The type I, IgE-mediated hypersensitivity reactions that occur in response to the latex proteins are less common but far more serious. They may manifest as urticaria, rhinoconjunctivitis, asthma, or anaphylaxis.

Diagnosis of latex allergy often is based on careful history and evidence of skin reactions caused by latex exposure. Symptoms after use of a rubber condom or diaphragm should raise suspicion of latex allergy. Because many of the reported reactions to latex gloves have been the result of a nonimmunologic dermatitis, it is important to differentiate between nonallergic and allergic types of dermatitis.

Treatment of latex allergy consists of avoiding latex exposure. Use of powder-free gloves can reduce the amount of airborne latex particles. Health care workers with severe and lifethreatening allergy may be forced to change employment. Patients at high risk for latex allergy (*e.g.*, children with spina bifida, health care workers with atopic disorders) should be offered clinical testing for latex allergy before undergoing procedures that expose them to natural rubber latex. All surgical or other procedures on persons with latex allergy should be done in a latex-free environment.

In summary, hypersensitivity and allergic disorders are responses to environmental, food, or drug antigens that would not affect most of the population. There are four basic categories of hypersensitivity responses: type I responses such as allergic rhinitis, bronchial asthma, and food allergies, which are mediated by IgE-class immunoglobulins; type II cytotoxic reactions such as blood transfusion reactions, which result from immunoglobulin (IgG and IgM) activation of complement; type III reactions, which result from the formation of insoluble antigen-antibody complexes that become deposited in blood vessels or in the kidney and cause localized tissue injury; and type IV, cell-mediated responses which cause conditions such as allergic dermatitis and hypersensitivity pneumonitis. Latex allergy, which is becoming more common with the spiraling use of latex gloves to prevent transmission of blood-borne diseases, can involve a type I, IgE-mediated anaphylactic reaction or a type IV, T-cell-mediated contact dermatitis.

TRANSPLANTATION IMMUNOPATHOLOGY

Not long ago, transplantation of solid organs (*e.g.*, liver, kidney, heart) and bone marrow was considered experimental and reserved for persons for whom alternative methods of therapy were exhausted and survival was unlikely. However, with a greater understanding of humoral and cellular immune regulation, the development of immunosuppressive drugs such as cyclosporine, and an appreciation of the role of the major histocompatibility complex (MHC) antigens, transplantation has become nearly routine, and the subsequent success rate has been greatly enhanced.

The cell surface antigens that determine whether transplanted tissue is recognized as foreign are the MHC or human leukocyte antigen (HLA) (see Chapter 8). Transplanted tissue can be categorized as *allogeneic* if the donor and recipient are related or unrelated but share similar HLA types, syngeneic if the donor and recipient are identical twins, and autologous if donor and recipient are the same person. Donors of solid organ transplants can be living or dead (cadaver) and related or nonrelated (heterologous). When tissues bearing foreign MHC antigens are transplanted, the recipient's immune system attempts to eliminate the donor cells, a process referred to as host-versusgraft disease (HVGD). Conversely, the cellular immune system of the transplanted tissues can attack unrelated recipient tissue, causing a graft-versus-host disease (GVHD). The likelihood of rejection varies indirectly with the degree of HLA or MCH relatedness between donor and recipient.

Host-Versus-Graft Disease

In HVGD, the immune cells of the transplant recipient attack the donor cells of the transplanted organ. HVGD usually is limited to allogeneic organ transplants, although even HLAidentical siblings may differ in some minor HLA loci, which can evoke slow rejection. Rejection caused by HVGD is a complex process that involves cell-mediated and circulating antibodies. Although many cells may participate in the process of acute transplant rejection, only the T lymphocytes seem to be absolutely required.² The activation of CD8⁺ cytotoxic T cells and CD4⁺ helper T cells is triggered in response to the donor's HLA antigens. Activation of CD4⁺ helper cells leads to proliferation of B-cell-mediated antibody production and a delayed-type hypersensitivity reaction. The initial target of the recipient antibodies is graft vasculature. The antibodies can produce injury to the transplanted organ by complementmediated cytotoxicity, generation of antigen-antibody complexes, or through antibody-mediated cytolysis.²

There are three basic patterns of transplant rejection: hyperacute, acute, and chronic.² A hyperacute reaction occurs almost immediately after transplantation; in kidney transplants, it can often be seen at the time of surgery. As soon as blood flow from the recipient to the donor organ begins, it takes on a cyanotic, mottled appearance. Sometimes, the reaction takes hours to days to develop. The hyperacute response is produced by existing recipient antibodies to graft antigens that initiate a type III immune-complex reaction in the blood vessels of the graft. These antibodies usually have developed in response to pre-

Chapter 10: Alterations in the Immune Response

vious blood transfusions, pregnancies in which the mother makes antibodies to fetal antigens, or infections with bacteria or viruses possessing antigens that mimic MHC antigens.

Acute rejection usually occurs within the first few months after transplantation. It also may occur suddenly months or even years later, after immunosuppression has been used and terminated. In the person with an organ transplant, acute rejection is evidenced by signs of organ failure. Acute rejection often involves both humoral and cell-mediated immune responses. Histologically, humoral rejection is associated with vasculitis, whereas a cell-mediated rejection response is marked by interstitial infiltration by mononuclear cells. Acute rejection vasculitis is mediated primarily by antidonor antibodies and is characterized by lesions that lead to arterial narrowing or obliteration.

Chronic HVGD occurs over a prolonged period. It is manifest by dense fibrosis of the intimal layer of blood vessels in the transplanted organ. In renal transplantation, it is characterized by a gradual rise in serum creatinine during a period of 4 to 6 months. The actual mechanism of this type of response is unclear but may include release of inflammatory mediators such as interleukin-1 and platelet-derived growth factor.

Graft-Versus-Host Disease

GVHD occurs mainly in patients who undergo bone marrow transplant and in severely immunocompromised patients who have received blood products containing HLA-incompatible lymphocytes.1 Although GVHD occurs mainly in patients who undergo bone marrow transplant, it may also occur after transplantation of solid organs rich in lymphoid cells (e.g., the liver) or transfusion of nonirradiated blood.² Three basic requirements are necessary for GVHD to develop: (1) the donor bone marrow must have a functional cellular immune component; (2) the recipient's tissue must bear antigens foreign to the donor tissue; and (3) the recipient's immunity must be compromised to the point that it cannot destroy the transplanted cells. The primary agents of GVHD are the donor T cells, and the antigens they recognize and attack are the host HLA. The greater the difference in tissue antigens between the donor and recipient, the greater is the likelihood of GVHD. Recipients of bone marrow transplants are usually immunodeficient, either because of the primary disease or prior treatment with immunosuppressant drugs or irradiation. When such recipients receive normal bone marrow cells from allogenic donors, the immunocompetent T cells derived from the donor recognize the recipient's HLA antigens as foreign and react against them.

If GVHD occurs, the primary targets of the acute illness are the skin, liver, intestine, and cells of the immune system. Acute GVHD is characterized by a pruritic, maculopapular rash, which begins on the palms and soles and frequently extends over the entire body, with subsequent desquamation. The epithelial layer is the primary site of injury. When the intestine is involved, symptoms include nausea, bloody diarrhea, and abdominal pain. GVHD of the liver can lead to bleeding disorders and coma. GVHD is considered chronic if symptoms persist or begin 100 days or more after transplantation. Chronic GVHD is characterized by abnormal humoral and cellular immunity, severe skin disorders, and liver disease. Another type of GVHD has been recognized after the transplantation of genetically identical tissue (*i.e.*, syngeneic or autologous [from self]). This variety of GVHD stems from the pretreatment conditioning regimen (*e.g.*, total-body irradiation) or treatment with cytotoxic drugs. The conditioning therapy disrupts the normal immune surveillance system and allows "rogue" autoreactive T cells to proliferate and attack native tissue. This type of GVHD usually is self-limited and not severe.

GVHD can often be prevented by selectively blocking the mechanisms responsible for the rejection process. For example, donor T cells can be selectively removed from the transplanted tissue or destroyed using various treatments such as monoclonal antibodies with attached toxins, equivalent to heat-seeking missiles. Alternatively, immunosuppressive or anti-inflammatory drugs such as cyclosporine and tacrolimus or glucocorticoids can be used to block T-cell activation and the action of cytokines.

In summary, organ and bone marrow transplantation has been enhanced by a greater understanding of humoral and cellular immune regulation, the development of immunosuppressive drugs such as cyclosporine, and an appreciation of the role of the MHC antigens in transplant rejection. The likelihood of rejection varies with the degree of HLA (or MHC) relatedness between donor and recipient. A rejection can involve an attempt by the recipient's immune system to eliminate the donor cells, as in HVGD, or an attack by the cellular immunity of the transplanted tissue on the unrelated recipient tissue, as in GVHD.

AUTOIMMUNE DISORDERS

Autoimmune diseases represent a group of disorders that are caused by a breakdown in the ability of the immune system to differentiate between self- and nonself antigens. To function properly, the immune system must be able to differentiate foreign antigens from self-antigens. Normally, there is a high degree of immunologic tolerance to self-antigens, which prevents the immune system from destroying the host.

Autoimmune diseases can affect almost any cell or tissue in the body. Some autoimmune disorders, such as Hashimoto's thyroiditis, are tissue specific; others, such as SLE, affect multiple organs and systems. Chart 10-1 lists some of the probable autoimmune diseases. Many of these disorders are discussed elsewhere in this book.

Immunologic Tolerance

The ability of the immune system to differentiate self from nonself is called *self-tolerance*. It is the HLA antigens encoded by MHC genes that serve as recognition markers of self and nonself for the immune system (see Chapter 8). To elicit an immune response, an antigen must first be processed by an antigen-presenting cell (APC), such as a macrophage, which then presents the antigenic determinants along with an MHC II molecule to a CD4⁺ helper T cell for binding to its T-cell receptor (TCR). It is the dual recognition of the MHC-antigen

CHART 10-1 Probable Autoimmune Disease*

Systemic

Mixed connective tissue disease Polymyositis-dermatomyositis Rheumatoid arthritis Scleroderma Sjögren's syndrome Systemic lupus erythematosus

Blood

Autoimmune hemolytic anemia Autoimmune neutropenia and lymphopenia Idiopathic thrombocytopenic purpura

Other Organs

Acute idiopathic polyneuritis Atrophic gastritis and pernicious anemia Autoimmune adrenalitis Goodpasture's syndrome Hashimoto's thyroiditis Insulin-dependent diabetes mellitus Myasthenia gravis Premature gonadal (ovarian) failure Primary biliary cirrhosis Sympathetic ophthalmia Temporal arteritis Thyrotoxicosis (Graves' disease) Ulcerative colitis

*Examples are not inclusive.

complex by the TCR that acts as a security check that affects all T cells, including CD4⁺ helper T cells, which orchestrate T- and B-cell immune responses, and cytotoxic T cells, which act directly to destroy target cells. A number of chemical messengers (*e.g.*, interleukins) and costimulatory signals are essential to the activation of immune responses and the preservation of self-tolerance.

Several mechanisms have been postulated to explain the tolerant state, including central tolerance and peripheral tolerance.⁹ Central tolerance refers to the elimination of self-reactive T cells in the thymus and B cells in the bone marrow. Peripheral tolerance refers to the deletion or inactivation of autoreactive T cells or B cells that escaped elimination in the central lymphoid organs. Autoreactive B cells are deleted in the spleen and lymph nodes. Autoreactive T cells may undergo activation-induced cell death, be rendered inactive to the extent that they cannot recognize self antigens, or their activity may be suppressed by other regulatory T cells.²

Mechanisms of Autoimmune Disease

There are multiple explanations for the loss of self-tolerance and formation of autoantibodies or failure to recognize host antigens as self. Among the possible mechanisms responsible for development of autoimmune disease are aberrations in immune cell function or antigen structure. Heredity and gender may play a role in the development of autoimmunity. Because

KEY CONCEPTS

IMMUNOLOGIC TOLERANCE AND AUTOIMMUNE DISEASE

- Immunologic tolerance is the ability of the immune system to differentiate self from nonself.
- Central tolerance involves the elimination of selfreactive T and B cells in the central lymphoid organs. Self-reactive T cells are deleted in the thymus and self-reactive B cells in the bone marrow.
- Peripheral tolerance derives from the deletion or inactivation of self-reactive T and B cells that escaped deletion in the central lymphoid organs.
- Autoimmune disorders result from the breakdown in the integrity of immune tolerance such that a humoral or cellular immune response can be mounted against host tissue or antigens, leading to localized or systemic injury.

of the complexity of the immune system, it seems unlikely that autoimmune disorders arise from a single defect.

Heredity and Gender

Genetic factors can increase the incidence and severity of autoimmune diseases, ¹⁰ as shown by the familial clustering of several autoimmune diseases and the observation that certain inherited HLA types occur more frequently in persons with a variety of immunologic and lymphoproliferative disorders. For example, 90% of persons with ankylosing spondylitis carry the HLA-B27 antigen. Other HLA-associated diseases are Reiter's syndrome and HLA-B27, rheumatoid arthritis and HLA-DR4, and systemic lupus erythematosus (SLE) and HLA-DR3 (see Chapter 44). The molecular basis for these associations is unknown. Because autoimmunity does not develop in all persons with genetic predisposition, it appears that other factors, such as a "trigger event," interact to precipitate the altered immune state. The event or events that trigger the development of an autoimmune response are unknown. It has been suggested that the trigger may be a virus or other microorganism, a chemical substance, or a self-antigen from a body tissue that has been hidden from the immune system during development.

A number of autoimmune disorders such as SLE occur more commonly in women than men, suggesting that estrogens may play a role in the development of autoimmune disease. Evidence suggests that estrogens stimulate and androgens suppress the immune response.¹¹ For example, estrogen stimulates a DNA sequence that promotes the production of interferon- γ , which is thought to assist in the induction of an autoimmune response.

Failure of Self-tolerance

Autoimmune disorders can result from one or more mechanisms of self-tolerance. Immunologic cells are undoubtedly involved in the tissue injury that results, but the precise mechanisms involved in initiating the response are largely unknown. More than one defect might be present in each disease, and each mechanism may be involved in more than one disease. Among the proposed mechanisms involved in loss of selftolerance are: failure of T-cell-mediated suppression, breakdown of T-cell anergy, disorders of MHC-antigen receptor/ complex interactions, release of sequestered antigens, molecular mimicry, and superantigens.

Failure of T-cell–Mediated Suppression. Disorders of immune regulatory or surveillance function can result from failure to delete autoreactive immune cells or suppress the immune response.² Because T cells regulate the immune response, an increasing ratio of helper T to suppressor T cells may lead to the development of autoimmune disorders.

Breakdown in T-cell Anergy. Anergy involves the prolonged or irreversible inactivation of T cells under certain conditions. Activation of antigen-specific CD4+ T cells requires two signals: recognition of the antigen in association with class II MHC molecules on the surface of the APCs and a set of costimulatory signals provided by the APCs. If the second costimulatory signal is not delivered, the T cell becomes anergic. Most normal tissues do not express the costimulatory molecules and thus are protected from autoreactive T cells. This protection can be broken if the normal cells that do not normally express the costimulatory molecules are induced to do so. Some inductions can occur after an infection, or in situations where there is tissue necrosis and local inflammation. For example, up-regulation of the costimulator molecule B7-1 has been observed in the central nervous system of persons with multiple sclerosis, in the synovium of persons with rheumatoid arthritis, and in the skin of persons with psoriasis.2

Disorders in MHC-Antigen Complex/Receptor Interactions.

The immune system recognizes antigen in the context of MHC-antigen complex and TCR interactions. Aberrations in any of these three stages of the immune response—antigen structure, TCR recognition of antigen, or MHC antigen presentation—have the potential for initiating an autoimmune response.

There are many ways in which chemical or microbial antigens can be modified to evoke an altered immune response, leading to an autoimmune disorder. Autoantigenic drugs and viruses can be complexed to a carrier that is recognized by nontolerant CD4⁺ helper T cells as foreign. Virus-encoded antigens expressed on the cell surface can serve as carriers for selfantigens. In this case, the self-antigen would appear as a hapten for which an immune response could be induced.

Partial degradation of self-antigens also may occur. For example, partially degraded collagen or enzymatically altered thyroglobulin or gamma globulin may be sufficiently foreign to promote an autoimmune response.

Release of Sequestered Antigens. Normally the body does not produce antibodies against self-antigens. Thus, any self-antigen that was completely sequestered during development and then reintroduced to the immune system is likely to be regarded as foreign. Among the sequestered tissues that could be regarded as foreign are spermatozoa and ocular antigens such as those

found in uveal tissue. Post-traumatic uveitis and orchiditis after vasectomy may fall into this category.

Molecular Mimicry. It is possible that certain autoimmune disorders are caused by molecular mimicry, in which a foreign antigen so closely resembles a self-antigen that antibodies produced against the former react with the latter.^{12,13} A humoral or cellular response can be mounted against antigenically altered or injured tissue, creating an immune process. For example, in rheumatic fever and acute glomerulonephritis, a protein in the cell wall of group A β-hemolytic streptococci has considerable similarity with antigens in heart and kidney tissue, respectively. After infection, antibodies directed against the microorganism cause a classic case of mistaken identity, which leads to inflammation of the heart or kidney. Certain drugs, when bound to host proteins or glycoproteins, form a complex to which a humoral response is directed with substantial cross-reactivity to the original self-protein. The antihypertensive agent methyldopa can bind to surface antigens on red cells to induce an antibody-mediated hemolytic anemia.

Not everyone exposed to group A β -hemolytic streptococci has an autoimmune reaction. The reason that only certain persons are targeted for autoimmune reactions to a particular selfmimicry molecule may be determined by differences in HLA types. The HLA type determines exactly which fragments of a pathogen are displayed on the cell surface for presentation to T cells. One individual's HLA may bind self-mimicry molecules for presentation to T cells, and another's HLA type may not. In the spondyloarthropathies, particularly Reiter's syndrome and reactive arthritis, there is a clear relationship between arthritis and a prior bacterial infection, combined with the inherited HLA-B27 antigens.¹³

Superantigens. Superantigens are a family of related substances, including staphylococcal and streptococcal exotoxins, that can short-circuit the normal sequence of events in an immune response, leading to inappropriate activation of CD4⁺ helper T cells. Superantigens do not require processing and presentation of antigen by APCs to induce a T-cell response.¹⁴ Instead, they are able to interact with a TCR outside the normal antigenbinding site. Normally, only a small percentage (0.01%) of the T-cell population is stimulated by the presence of processed antigens on the surface of macrophages; however, superantigens can interact with 5% to 30% of T cells.14 Superantigens directly link the MHC II complex molecules of APCs such as macrophages to TCRs, causing a massive release of T-cell inflammatory cytokines, primarily interleukin-2 and tumor necrosis factor, and an uncontrolled proliferation of T cells. At least one disease in adults, toxic shock syndrome, is mediated by superantigens (see Chapter 18). Kawasaki's disease in children (see Chapter 17) probably has a similar cause.

Diagnosis and Treatment

Suggested criteria for determining that a disorder is an autoimmune disorder are evidence of an autoimmune reaction, determination that the immunologic findings are not secondary to another condition, and the lack of other identifiable causes for the disorder. Currently, the diagnosis of autoimmune disease is based primarily on clinical findings and serologic testing. The basis for most serologic assays is the demonstration of antibodies directed against tissue antigens or cellular components. For example, a serological assay for antinuclear antibodies is used in the diagnosis of SLE.

Treatment of autoimmune disease is based on the tissue or organ that is involved, the effector mechanism involved, and the magnitude and chronicity of the effector processes. Ideally, treatment should focus on the mechanism underlying the autoimmune disorder.

In summary, autoimmune diseases represent a disruption in self-tolerance that results in damage to body tissues by the immune system. Autoimmune diseases can affect almost any cell or tissue of the body. The ability of the immune system to differentiate self from nonself is called *self-tolerance*. Normally, self-tolerance is maintained through central and peripheral mechanisms that delete autoreactive B or T cells or otherwise suppress or inactivate immune responses that would be destructive to host tissues. Defects in any of these mechanisms could impair self-tolerance and predispose to development of autoimmune disease.

The ability of the immune system to differentiate foreign from self-antigens is the responsibility of HLA encoded by MHC genes. Antigen is presented to receptors of T cells in combination with MHC molecules. Among the possible mechanisms responsible for the development of autoimmune disease are failure of T-cell-mediated immune suppression; aberrations in MHC-antigen-TCR interactions; molecular mimicry; and superantigens.

Suggested criteria for determining that a disorder results from an autoimmune disorder are evidence of an autoimmune reaction, determination that the immunologic findings are not secondary to another condition, and the lack of other identifiable causes for the disorder.

IMMUNODEFICIENCY DISORDERS

Immunodeficiency can be defined as an abnormality in one or more branches of the immune system that renders a person susceptible to diseases normally prevented by an intact immune system. Two major categories of immune mechanisms defend the body against infectious or neoplastic disease: humoral or antibody-mediated immunity (*i.e.*, B lymphocytes) and cell-mediated immunity (*i.e.*, T lymphocytes and lymphokines).

Abnormalities of the immune system can be classified as primary (*i.e.*, congenital or inherited) or secondary if the immunodeficiency is acquired later in life. Secondary immunodeficiencies are more common than primary disorders of genetic origin. Secondary deficiencies in humoral immunity can develop as a consequence of selective loss of immunoglobulins through the gastrointestinal or genitourinary tracts. Secondary deficiencies of T-cell function have been described in conjunction with acute viral infections (*e.g.*, measles virus, cytomegalovirus) and with certain malignancies (*e.g.*, Hodgkin's disease and other lymphomas). HIV/AIDS (to be discussed) is the most devastating example of a secondary immunodeficiency. Regardless of the cause, primary and secondary deficiencies can produce the same spectrum of disease. The severity and symptomatology of the various immunodeficiencies depend on the disorder and extent of immune system involvement.

Primary Immunodeficiency Disorders

Until recently, little was known about the causes of primary immunodeficiency diseases. As a result of recent advances in mapping the human genome, the genetic origin of many of the defects has been identified.¹⁵ In addition, previous classifications of the disorders were based on specific clinical manifestations and alterations in immune function. Advances in molecular genetics now allow many of these disorders to be grouped according to the types of genetically altered molecules that are involved. Although genes essential to immune function are located throughout the genome, a large number are located on the X chromosome. Thus, there is a clear dominance of X-linked immunodeficiencies in males who have only one X chromosome and a single copy of these genes.¹⁵ In addition, spontaneous mutations in these X-linked genes are relatively common.

Humoral (B-Cell) Immunodeficiencies

Humoral immunodeficiency can range from a transient decrease in immunoglobulin levels during early infancy to inherited disorders that interrupt the production of one or all of the immunoglobulins. During the first few months of life, infants are protected from infection by immunoglobulin G (IgG) class

KEY CONCEPTS

PRIMARY IMMUNODEFICIENCY DISORDERS

- Primary immunodeficiency disorders are congenital or inherited abnormalities of immune function that render a person susceptible to diseases normally prevented by an intact immune system.
- Disorders of B-cell function impair the ability to produce antibodies and defend against microorganisms and toxins that circulate in body fluids (IgM and IgG) or enter the body through the mucosal surface of the respiratory or gastrointestinal tract (IgA). Persons with primary B-cell immunodeficiency are particularly prone to infections due to encapsulated organisms.
- Disorders of T-cell function impair the ability to orchestrate the immune response (CD4+ helper T cells) and to protect against fungal, protozoan, viral, and intracellular bacterial infections (CD8+ cytotoxic T cells).
- Combined T-cell and B-cell immunodeficiency states affect all aspects of immune function. Severe combined immunodeficiency represents a lifethreatening absence of immune function that requires bone marrow transplantation for survival.

Chapter 10: Alterations in the Immune Response 179

antibodies that have been transferred from the maternal circulation during fetal life. IgA, IgM, IgD, and IgE do not normally cross the placenta (see Chapter 9). An infant's level of maternal IgG gradually declines during a period of approximately 6 months. Concomitant with the loss of maternal antibody, the infant's immature humoral immune system begins to function, and between the ages of 1 and 2 years, the child's antibody production reaches that of adult levels.

Any abnormality that blocks or prevents the maturation of B-lymphocyte stem cells can produce a state of immunodeficiency. For example, certain infants experience a delay in the maturation process of B cells that leads to a prolonged deficiency in IgG levels (IgM and IgA levels are normal) beyond 6 months of age. The total number and antigenic response of circulating B cells is normal, but the chemical communication between B and T cells that leads to clonal proliferation of antibody-producing plasma cells seems to be reduced.¹⁶ This condition is referred to as *transient hypogammaglobulinemia of infancy*. The result of this condition usually is limited to repeated bouts of upper respiratory and middle ear infections. This condition usually resolves by the time the child is 2 to 4 years of age.

Primary B-cell immunodeficiencies are genetic disorders of B lymphocyte maturation. They account for 70% of primary immunodeficiencies and are manifested by decreased IgG production.¹⁷ Antibody production depends on the differentiation of B-lymphocyte stem cells in the bone marrow to mature, immunoglobulin-producing plasma cells. This maturation cycle initially involves the production of surface IgM, migration from the marrow to the peripheral lymphoid tissue, and switching to the specialized production of IgG, IgA, IgD, IgE, or IgM antibodies after antigenic stimulation (Fig. 10-5). Defects in B-cell function increase the risk of recurrent pyogenic infections, including those caused by *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Staphylococcus aureus*, and by gram-negative organisms such as *Pseudomonas* species. Humoral immunity usually is not as important in defending against intracellular bacteria (mycobacteria), fungi, and protozoa. Viruses usually are handled normally, except for the enteroviruses that cause gastrointestinal infections.

X-linked Agammaglobulinemia. *X-linked* or *Bruton's agammaglobulinemia* is a recessive trait that affects only males.^{15–18} As the name implies, persons with this disorder have essentially undetectable levels of all serum immunoglobulins. Therefore, they are susceptible to meningitis and recurrent otitis media and to sinus and pulmonary infections with encapsulated organisms such as *S. pneumoniae*, *H. influenzae* type b, *S. aureus*, and *Neisseria meningitidis*.¹⁵ Many boys with this disorder have severe tooth decay.

The central defect in this syndrome is a genetic mutation that blocks the differentiation of pre-B cells, creating an absence of mature circulating B cells and plasma cells. However, T lymphocytes are normal in number and function. Symptoms of the disorder usually coincide with the loss of maternal antibodies at about 6 months of age. A clue to the presence of the disorder is failure of an infection to respond completely and promptly to antibiotic therapy. Diagnosis is based on demonstration of low or absent serum immunoglobulins. Therapy consists of prophylaxis with intravenous immunoglobulin and prompt antimicrobial therapy for suspected infections. The prognosis of this condition depends on the prompt recognition and treatment of infections. Chronic pulmonary disease is an ever-present danger.

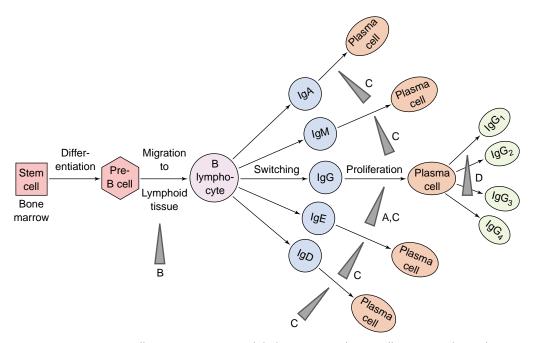


FIGURE 10-5 Stem cells to mature immunoglobulin-secreting plasma cells. *Arrows* indicate the stage of the maturation process that is interrupted in (**A**) transient hypoglobulinemia, (**B**) X-linked hypogammaglobulinemia, (**C**) common variable immunodeficiency, and (**D**) IgG subclass deficiency.

180 Unit Two: Alterations in Body Defenses

Common Variable Immunodeficiency. Another disorder of B-cell maturation, which is similar to X-linked agammaglobulinemia, is a condition called *common variable immunodeficiency*. In this syndrome, the terminal differentiation of mature B cells to plasma cells is blocked. The result is markedly reduced serum immunoglobulin levels, normal numbers of circulating B lymphocytes, and a complete absence of germinal centers and plasma cells in lymph nodes and the spleen.

The symptomatology of common variable immunodeficiency is similar to that of X-linked agammaglobulinemia (*i.e.*, recurrent otitis media and sinus and pulmonary infections with encapsulated organisms), but the onset of symptoms occurs much later, usually between the ages of 15 and 35 years, and distribution of disease between the sexes is equal. Persons with late-onset hypogammaglobulinemia also have an increased tendency toward development of chronic lung disease, autoimmune disorders, hepatitis, gastric carcinoma, and chronic diarrhea with associated intestinal malabsorption. Approximately one half of persons with the disorder have evidence of abnormal T-cell immunity, suggesting that this syndrome is a complex immunodeficiency. Treatment methods for late-onset hypogammaglobulinemia are similar to those used for X-linked hypogammaglobulinemia.

Selective Immunoglobulin A Deficiency. Selective IgA deficiency is the most common type of immunoglobulin deficiency, affecting 1 in 400 to 1 in 1000 persons.¹⁵ The syndrome is characterized by moderate to marked reduction in levels of serum and secretory IgA. It is likely that the cause of this deficiency is a block in the pathway that promotes terminal differentiation of mature B cells to IgA-secreting plasma cells.

Approximately two thirds of persons with selective IgA deficiency have no overt symptoms, presumably because IgG and IgM levels are normal and compensate for the defect. At least 50% of affected children overcome the deficiency by the age of 14 years. Persons with markedly reduced levels of IgA often experience repeated upper respiratory and gastrointestinal infections and have increased incidence of allergies such as asthma and autoimmune disorders. It has been estimated that as many as 50% of persons with selective IgA deficiency have some form of allergy.¹⁹ It has been suggested that the lack of IgA allows inhaled and ingested antigens to cross the mucosal epithelium and elicit antibody responses in the gastrointestinal and bronchial lymphoid tissues. Persons with IgA deficiency also can develop antibodies against IgA, which can lead to an anaphylactic response when blood components containing IgA are given.18

There is no treatment available for selective IgA deficiency unless there is a concomitant reduction in IgG levels. Administration of IgA is of little benefit because it has a short halflife and is not secreted across the mucosa. There also is the risk associated with IgA antibodies.

Immunoglobulin G Subclass Deficiency. An IgG subclass deficiency can affect one or more of IgG subtypes, despite normal levels or elevated serum concentrations of IgG. As discussed in Chapter 9, IgG immunoglobulins can be divided into four subclasses (IgG1 through IgG4) based on structure and function. Most circulating IgG belongs to the IgG1 (70%) and IgG2 (20%) subclasses. In general, antibodies directed against protein antigens belong to the IgG1 and IgG3 sub-

classes, and antibodies directed against carbohydrate and polysaccharide antigens are primarily IgG2 subclass. As a result, persons who are deficient in IgG2 subclass antibodies can be at greater risk for development of sinusitis, otitis media, and pneumonia caused by polysaccharide-encapsulated microorganisms such as *S. pneumoniae*, *H. influenzae* type b, and *N. meningitidis*. Children with mild forms of the deficiency can be treated with prophylactic antibiotics to prevent repeated infections. Intravenous immune globulin can be given to children with severe manifestations of this deficiency. The use of polysaccharide vaccines conjugated to protein carriers can provide protection against some of these infections because protein conjugated to protein carriers stimulates an IgG1 response.

Cellular (T-Cell) Immunodeficiencies

There are few primary forms of T-cell immunodeficiency, probably because persons with defects in this branch of the immune response rarely survive beyond infancy or childhood. However, exceptions are being recognized as newer T-cell defects, such as X-linked immunodeficiency with hyper IgM and CD2, are being identified. Other primary T-cell immunodeficiency disorders result from defective expression of the T-cell receptor complex, defective cytokine production, and defects in T-cell activation.

Unlike the B-cell lineage, in which a well-defined series of differentiation steps ultimately leads to the production of immunoglobulins, mature T lymphocytes are composed of distinct subpopulations whose immunologic assignments are diverse. T cells can be functionally divided into helper and cytotoxic subtypes and a population of T cells that promote delayed hypersensitivity reactions. Collectively, T lymphocytes protect against fungal, protozoan, viral, and intracellular bacterial infections; control malignant cell proliferation; and are responsible for coordinating the overall immune response.

DiGeorge Syndrome. DiGeorge syndrome stems from a developmental defect that occurs during the time (*i.e.*, before the 12th week of gestation) when the thymus gland, parathyroid gland, and parts of the head, neck, and heart are developing. The disorder affects both sexes. Formerly thought to be caused by a variety of factors, including extrinsic teratogens, this defect has been traced to a gene on chromosome 22 (22q11).¹⁸⁻²⁰

Infants born with this defect have partial or complete failure of development of the thymus and parathyroid glands and have congenital defects of the head, neck, and heart. The extent of immune and parathyroid abnormalities is highly variable, as are the other defects. Occasionally, a child has no heart defect. In some children, the thymus is not absent but is in an abnormal location and is extremely small. These infants can have partial DiGeorge syndrome, in which hypertrophy of the thymus occurs with development of normal immune function. The facial disorders can include hypertelorism (i.e., increased distance between the eyes); micrognathia (i.e., fish mouth); low-set, posteriorly angulated ears; split uvula; and high-arched palate. Urinary tract abnormalities also are common. The most common presenting sign is hypocalcemia and tetany that develops during the first 24 hours of life. It is caused by the absence of the parathyroid gland and is resistant to standard therapy.

Children who survive the immediate neonatal period may have recurrent or chronic infections because of impaired T-cell immunity. Children also may have an absence of immunoglobulin production, caused by a lack of helper T-cell function. For children who do require treatment, thymus transplantation can be performed to reconstitute T-cell immunity. Bone marrow transplantation also has been successfully used to restore normal T-cell populations. If blood transfusions are needed, as during corrective heart surgery, special processing is required to prevent graft-versus-host disease.

X-Linked Immunodeficiency With Hyper-IgM. The X-linked immunodeficiency of hyper-IgM, also known as the *hyper-IgM syndrome*, is characterized by low IgG and IgA levels with normal or, more frequently, high IgM concentrations. Being X-linked, the disorder is confined to males. Formerly classified as a B-cell defect, it now has been traced to a T-cell defect. The disorder results from the inability of T cells to signal B cells to undergo isotype switching to IgG and IgA; thus, they produce only IgM.¹⁸

Like boys with X-linked agammaglobulinemia, affected boys become symptomatic during the first and second years of life. They have recurrent pyogenic infections, including otitis media, sinusitis, tonsillitis, and pneumonia. They are also more susceptible to *Pneumocystis carinii* infection. Thymicdependent lymphoid tissues and T-cell function usually are normal, as are B-cell counts. Hemolytic anemia and thrombocytopenia may occur, and transient, persistent, or cyclic neutropenia is a common feature. The occurrence of concomitant autoimmune disorders is higher than with other immunoglobulin deficiency disorders.¹⁸

Combined T-Cell and B-Cell Immunodeficiencies

Disorders of the immune response that have elements of T-cell and B-cell dysfunction fall under the broad classification of combined immunodeficiency syndrome (CIDS) and include a spectrum of inherited (autosomal recessive and X-linked) conditions. A single mutation in any one of the many genes that influence lymphocyte development or response, including lymphocyte receptors, cytokines, or major histocompatibility antigens, could lead to combined immunodeficiency. Regardless of the affected gene, the net result is a disruption in the normal communication system of T and B lymphocytes and deregulation of the immune response. The spectrum of disease resulting from CIDS ranges from mild to severe to ultimately fatal forms.

Severe Combined Immunodeficiency. The most severe form of T- and B-cell deficiency often is referred to as *severe combined immunodeficiency syndrome* (SCIDS). SCIDS is caused by diverse genetic mutations that lead to absence of all immune function.^{18,19} A family history of similarly affected relatives occurs in approximately 50% of cases.¹⁹ Both autosomal recessive and X-linked inheritance are involved. Infants with SCIDS have a disease course that resembles acquired immunodeficiency syndrome (AIDS), with failure to thrive, chronic diarrhea, and opportunistic infections that usually lead to death by the age of 2 years. If recognized at birth or within the first 3 months of life, 95% of infants can be successfully treated with human leukocyte antigen (HLA)-identical or T-cell-depleted bone marrow stem cell transplantation.¹⁸

Approximately 50% of persons with the autosomal recessive form of SCIDS have an associated deficiency in the enzyme adenosine deaminase (ADA).¹⁶ Absence of this enzyme leads to accumulation of toxic metabolites that kill dividing and resting T cells. Bone marrow and stem cell transplantation has been successful in treating children with ADA-negative SCIDS.^{18,21} Enzyme replacement therapy also may be used in the care of persons with this form of SCIDS.²¹

Acquired Immunodeficiency Syndrome

AIDS is an infectious disease of the immune system caused by the HIV retrovirus. First described in June 1981, the disease is prevalent worldwide and is one of the leading causes of death among young adults in the United States. At the end of 2000, nearly 36.1 million people worldwide were living with HIV/AIDS, 22 million had died of the infection, and 5.3 million people had become newly infected during the year.²² Most of the new infections are in people younger than 25 years who live in developing countries. Sub-Saharan Africa has been hardest hit by HIV, with close to 70% of the world's infections, although it is home to only 10% of the world's population.²²

The virus responsible for most HIV infection worldwide is called *HIV type 1*. A second type of human immunodeficiency, *HIV type 2* (HIV-2) is endemic in many countries in West Africa but generally much more rare in other parts of the world.²³ HIV-2 appears to be transmitted in the same manner as HIV-1. HIV-2 can also cause immunodeficiency evidenced by a reduction in the number of CD4⁺ T cells and the development of AIDS. Although the spectrum of disease for HIV-2 is similar to that of HIV-1, it spreads more slowly and causes disease more slowly than HIV-1.²³ Long-term consequences of HIV-2 will depend on its spread in the population.

KEY CONCEPTS

ACQUIRED IMMUNODEFICIENCY SYNDROME

- AIDS is a secondary immunodeficiency disorder that results from HIV infection, which is transmitted from one person to another through blood, semen, or vaginal fluids.
- The main effect of HIV infection is the destruction of CD4⁺ T cells, which constitutes an attack on the entire immune system, because this subset of T cells exerts critical regulatory and effector functions involving both cellular and humoral immunity.
- The three phases of HIV are primary HIV, latency, and overt AIDS. The classification of HIV/AIDS is based on laboratory counts of CD4⁺ T cells and the manifestations of the immunodeficiency state (development of opportunistic infections, neoplasms, and other related problems).

Transmission of HIV Infection

Human immunodeficiency virus is transmitted from one person to another through sexual contact, blood, or perinatally. Transmission can occur when infected blood, semen, or vaginal secretions from one person are deposited onto a mucous membrane or into the bloodstream of another person.

HIV is transmitted most frequently via sexual contact. Worldwide, 75% to 85% of HIV infections are transmitted through unprotected sex.24 HIV is present in semen and vaginal fluids. There is risk of transmitting HIV when these fluids come in contact with a part of the body that lets them enter the bloodstream. This can include the vaginal mucosa, anal mucosa, and wounds or a sore on the skin.²⁴ Contact with semen occurs during vaginal and anal sexual intercourse, oral sex (*i.e.*, fellatio), and donor insemination. Exposure to vaginal or cervical secretions occurs during vaginal intercourse and oral sex (i.e., cunnilingus). Condoms are highly effective in preventing transmission of HIV. Evidence increasingly shows that people with other sexually transmitted diseases (STDs) are at increased risk for HIV infection. The risk of HIV transmission is further increased in the presence of STDs with genital ulcerations (i.e., syphilis, herpes simplex virus infection, and chancroid) as well as nonulcerative STDs (i.e., gonorrhea, chlamydial infection, and trichomoniasis).

Because HIV is found in blood, the use of needles, syringes, and other drug injection paraphernalia is a direct route for transmission. HIV-infected injecting drug users can pass the virus to their needle-sharing and sex partners and, in the case of pregnant women, to their offspring. Although alcohol, cocaine, and other noninjected drugs do not directly transmit infection, their use alters perception of risk and reduces inhibitions about engaging in behaviors that pose a high risk of transmitting HIV infection.

Transfusions of whole blood, plasma, platelets, or blood cells before 1985 resulted in the transmission of HIV. Since 1985, all blood donations in the United States have been screened for HIV so the risk of transmission has virtually been eliminated. The clotting factor used by persons with hemophilia is derived from the pooled plasma of hundreds of donors. Before HIV testing of plasma donors was implemented in 1985, the virus was transmitted to persons with hemophilia through infusions of these clotting factors.

HIV may be transmitted from infected women to their offspring in utero, during labor and delivery, or through breastfeeding. Transmission from mother to infant is the most common way that children become infected with HIV. Ninety percent of infected children acquired the virus from their mother. The risk of transmission of HIV from mother to infant is approximately 25%, with estimates ranging from 15% to 45%, depending on the country in which they reside.²⁵

Occupational HIV infection among health care workers is uncommon.²⁶ Universal Blood and Body Fluid Precautions should be used in encounters with all patients in the health care setting because HIV status is not always known. The occupational risk of infection for health care workers most often is associated with percutaneous inoculation (*i.e.*, needle stick) of blood from a patient with HIV.

The HIV-infected person can transmit the virus even when no symptoms are present and the antibody test is negative. The point at which an infected person converts from being negative for the presence of HIV antibodies in the blood to being positive is called *seroconversion*. Seroconversion typically occurs within 1 to 3 months after exposure to HIV but can take up to 6 months.²⁷ The time after infection and before seroconversion is known as the *window period*. During the window period, an HIV-infected person could transmit the virus through the blood. Blood collection centers have implemented more stringent processes to avoid HIV transmission from a donor who has not seroconverted. Potential donors are now screened through interviews designed to identify risk behaviors for HIV infection, and blood is tested for the HIV antibody as well as viral nucleic acid.

Pathophysiology of AIDS

HIV belongs to a class of viruses called retroviruses, which carry their genetic information in ribonucleic acid (RNA), rather than deoxyribonucleic acid (DNA). HIV infects a limited number of cell types in the body, including a subset of lymphocytes called *CD4*⁺ T helper cells and macrophages. The CD4⁺ T cells are necessary for normal immune function (see Chapter 8). Among other functions, the CD4⁺ T cell recognizes foreign antigens and infected cells and helps activate antibody-producing B lymphocytes. The CD4⁺ T cells also orchestrate cell-mediated immunity, in which cytotoxic CD8⁺ T cells and natural killer cells directly destroy viral infected cells and foreign antigens. Phagocytic monocytes and macrophages are also activated by CD4⁺ T cells.

Viral Replication. Replication of HIV has been divided into eight steps (Fig. 10-6). Each of these steps provides insights into the development of methods for preventing and treating HIV infection. The first step involves the binding of the virus to the CD4⁺ T cell. Once the HIV has entered the bloodstream, it attaches to the surface of a CD4⁺ T cell by binding to the CD4 molecule and a chemokine coreceptor.² This is known as attachment. The second step allows for the internalization of the virus. Binding to the coreceptor allows the virus to be internalized into the cell, where it releases its coat along with two single strands of viral RNA that carry the instructions for producing more HIV. This is called *uncoating*. The *third step* consists of DNA synthesis. In order for the HIV to reproduce, it must change its RNA into DNA. It does this by using an enzyme called *reverse transcriptase*. Reverse transcriptase makes a copy of the viral RNA, and then in reverse order makes another mirror image copy. The result is double-stranded DNA. The fourth step is called integration. It involves the entry of the double stranded viral DNA into the nucleus of the CD4⁺ T cell and, with the help of the enzyme *integrase*, insertion of the HIV DNA into the cell's original DNA. The fifth step involves transcription of the double stranded viral DNA to form a single stranded messenger RNA (mRNA) with the instructions for building new viruses. Transcription involves activation of the T cell and induction of host cell transcription factors. The sixth step includes the translation of mRNA. During translation, the ribosomal RNA (rRNA) uses the instructions in viral mRNA to create a chain of proteins and enzymes called a *polyprotein*. These polyproteins are the components for the new viruses that are formed. The seventh step is called cleavage. During the cleavage stage, protease, one of the enzymes in the polypeptide chain, cuts the chain into individual proteins, which will make up the

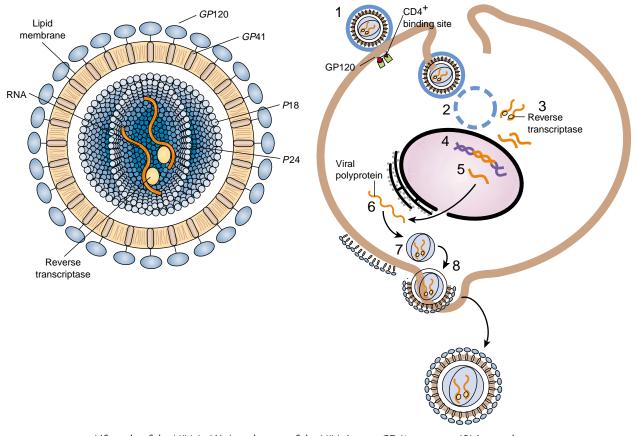


FIGURE 10-6 Life cycle of the HIV-1: (1) Attachment of the HIV virus to CD4⁺ receptor; (2) internalization and uncoating of the virus with viral RNA and reverse transcriptase; (3) reverse transcription, which produces a mirror image of the viral RNA and double-stranded DNA molecule; (4) integration of viral DNA into host DNA using the integrase enzyme; (5) transcription of the inserted viral DNA to produce viral messenger RNA; (6) translation of viral messenger RNA to create viral polyprotein; (7) cleavage of viral polyprotein into individual viral proteins that make up the new virus; and (8) assembly and release of the new virus from the host cell.

new HIV viruses. Finally, during the *eighth step*, the proteins and the new RNA are assembled into new HIV viruses and released from the cell.

Viral Latency and Activation. In some CD4⁺ T cells, the infection enters a latent phase that serves as a reservoir from which the virus can continue to be released for several years. In other CD4⁺ T cells, the virus replicates, killing the cell and releasing copies of HIV into the bloodstream. These viral particles, or virions, invade other CD4+ T cells, allowing the infection to progress. Initially, the infected cells are replaced and the viral particles destroyed. However, with time the CD4+ T-cell count gradually decreases and the viral count detected in the blood increases. The initiation of viral replication in latent HIV infection is critically dependent on host proteins and transcription factors that are present during T-cell activation. These factors may be activated by proteins produced by other viruses known to affect persons with AIDS, such as herpes viruses, Epstein-Barr virus, adenovirus, and cytomegalovirus.² Thus, activation of the immune system by a variety of infectious agents may promote HIV replication.

Until the CD4⁺ T-cell count falls to a very low level, infected persons can remain symptom free, although there is active viral replication, and serologic tests can identify antibodies to HIV. Unfortunately, these antibodies do not convey protection against the virus. Although symptoms are not evident, the infection proceeds on a microbiologic level, including the invasion and selective destruction of CD4⁺ T cells. The continual decline of CD4⁺ T cells, which are pivotal cells in the immune response, strips the person with AIDS of protection against common organisms and cancerous cells.

Diagnosis and Classification

Diagnosis. The most accurate and inexpensive method for identifying HIV is the HIV antibody test. The first commercial assays for HIV were introduced in 1985 to screen donated blood. Since then, the use of antibody detection tests has been expanded to include evaluating persons at increased risk for HIV infection. The HIV antibody test procedure consists of screening with an *enzyme immunoassay* (*EIA*), also known as *enzyme-linked immunosorbent assay* (*ELISA*), followed by a confirmatory test, the *Western blot* assay, which is performed if the

EIA is positive.²⁸ The EIA is based on the reaction of antibodies to HIV in the blood sample with viral proteins in the test material. The Western blot is a more sensitive assay that looks for the presence of antibodies to specific viral antigens.

Polymerase chain reaction (PCR) is a technique for detecting HIV DNA. PCR detects the presence of the virus, rather than the antibody to the virus, which the EIA and Western blot tests detect.¹⁷ PCR is useful in diagnosing HIV infection in infants born to infected mothers because these infants have their mothers' HIV antibody, regardless of whether the children are infected.

Classification. Effective January 1, 1993, the United States Centers for Disease Control implemented a new classification system for HIV infection identifying two categories: one based on laboratory tests and the other on clinical manifestations²⁹ (Fig. 10-7). The classification system defines three laboratory test categories that correspond to CD4⁺ cell counts per micro-liter (μ L) of blood: *category 1:* >500 cells/ μ L, *category 2:* 200 to 499 cells/ μ L, and *category 3:* <200 cells/ μ L.

The clinical manifestations are also divided into three categories. *Clinical category A* includes persons who have no symptoms or have persistent generalized lymphadenopathy or symptoms of primary HIV infection (*i.e.*, acute seroconversion illness). *Clinical category B* includes persons with symptoms of immune deficiency not serious enough to be AIDS defining. *Clinical category C* includes AIDS-defining illnesses that are listed in the AIDS surveillance case definition shown in Chart 10-2. Each HIV-infected person has a CD4⁺ T-cell category and a clinical category. The combination of laboratory and clinical categorizations can guide clinical and therapeutic actions in the management of HIV infection. According to the 1993 case definition, persons in laboratory category 3 or clinical category C are considered to have AIDS.

Clinical Course

The typical course of HIV is defined by three phases, which usually occur during a period of 8 to 12 years. The three stages are the primary infection or the acute clinical syndrome,

AIDS-defining clinical category	Category A No AIDS- defining symptoms			
	Category B Symptoms not severe enough to be AIDS defining			
	Category C AIDS-defining illnesses present			
		Category 1 >500 cells u/L	Category 2 200–499 cells u/L	Category 3 <200 cells u/L

CD4⁺ count category

■ FIGURE 10-7 ■ CD4⁺ category and clinical category.

CHART 10-2 Conditions Included in the 1993 AIDS Surveillance Case Definition

Candidiasis of bronchi, trachea, or lungs
Candidiasis, esophageal
Cervical cancer, invasive*
Coccidioidomycosis, disseminated or extrapulmonary
Cryptococcosis, extrapulmonary
Cryptosporidiosis, chronic intestinal (>1 month's duration)
Cytomegalovirus disease (other than liver, spleen, or nodes)
Cytomegalovirus retinitis (with loss of vision)
Encephalopathy, HIV-related
Herpes simplex: chronic ulcer(s) (>1 month's duration) or
bronchitis, pneumonitis, or esophagitis
Histoplasmosis, disseminated or extrapulmonary
Isosporiasis, chronic intestinal (>1 month's duration)
Kaposi's sarcoma
Lymphoma, Burkitt's (or equivalent term)
Lymphoma, immunoblastic (or equivalent term)
Lymphoma, primary, of brain
Mycobacterium avium-intracellulare complex or M. kansasii,
disseminated or extrapulmonary
Mycobacterium tuberculosis, any site (pulmonary* or
extrapulmonary)
Mycobacterium, other species or unidentified species,
disseminated or extrapulmonary
Pneumocystis carinii pneumonia
Pneumonia, recurrent*
Progressive multifocal leukoencephalopathy
Salmonella septicemia, recurrent
Toxoplasmosis of brain
Wasting syndrome due to HIV
*Added to the 1993 expansion of the AIDS surveillance case definition.
(Centers for Disease Control and Prevention. [1992]. 1993 Revised classi- fication system for HIV infection and expanded surveillance case defini- tion for AIDS among adolescents and adults. <i>Morbidity and Mortality</i>
and the second and the second se

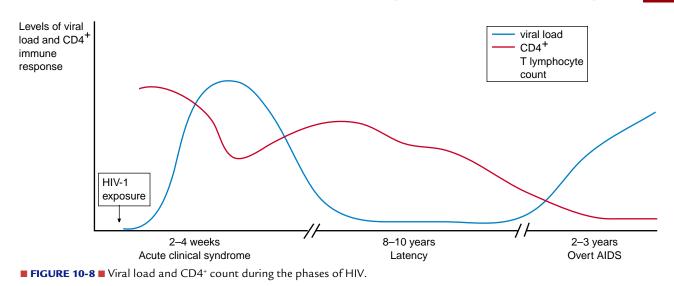
shronic asymptomatic phase or latency and evert AIDS

Weekly Report 41 [RR-17], 19)

chronic asymptomatic phase or latency, and overt AIDS³⁰ (Fig. 10-8).

Many persons, when they are initially infected with HIV, have an acute mononucleosis-like syndrome known as primary infection. This acute phase may include fever, fatigue, myalgias, sore throat, night sweats, gastrointestinal problems, lymphadenopathy, maculopapular rash, and headache³⁰ (Chart 10-3). During this time, there is an increase in viral replication, which leads to very high viral loads, sometime greater than 1,000,000 copies/mL, and a decrease in the CD4⁺ count. The signs and symptoms of primary HIV infection usually appear 2 to 4 weeks after exposure to HIV and last for a few days to 2 weeks.³⁰ After several weeks, the immune system acts to control viral replication and reduces it to a lower level, where it remains for several years.

The primary phase is followed by a latent period during which the person has no signs or symptoms of illness. The median time of the latent period is 10 years. During this time, the CD4⁺ count falls gradually from the normal range (800 to 1000 cells/ μ L) to 200 cells/ μ L or lower.³⁰ Lymphadenopathy



develops in some persons with HIV infection during this phase. Persistent generalized lymphadenopathy usually is defined as lymph nodes that are chronically swollen for more than 3 months in at least two locations, not including the groin. The lymph nodes may be sore or visible externally.

The third phase, overt AIDS, occurs when a person has a CD4⁺ count of less than 200 cells/ μ L or an AIDS-defining illness. Without antiretroviral therapy, this phase leads to death within 2 to 3 years. The risk of death and opportunistic infection increases significantly when the CD4⁺ count reaches this level.³⁰

The clinical course of HIV varies from person to person. Most (60% to 70%) of those infected with HIV acquire AIDS 10 to 11 years after infection. These people are the *typical progressors*.³⁰ Another 10% to 20% of those infected experience more rapid progression. They acquire AIDS in less than 5 years and are called *rapid progressors*. The final 5% to 15% are *slow progressors*, who do not experience progression to AIDS for more than 15 years. There is a subset of slow progressors, called *longterm nonprogressors*, who account for 1% of all HIV infections. These people have been infected for at least 8 years, are anti-

CHART 10-3 Signs and Symptoms of Acute HIV Infection

- Fever
- Fatigue
- Rash
- Headache
- Lymphadenopathy
- Pharyngitis
- Arthralgia
- Myalgia
- Night sweats
- Gastrointestinal problems
- Aseptic meningitis
- Oral or genital ulcers

retroviral naive, have high CD4⁺ counts, and usually have very low viral loads. $^{\rm 30}$

Opportunistic Infections

When the immune system becomes severely compromised, an opportunistic infection or malignancy may occur. The number of CD4⁺ T cells directly correlates with the risk of developing opportunistic infections. The risk of opportunistic infections increases greatly once the CD4⁺ count drops to less than 200 cells/ μ L.³¹ Opportunistic infections involve common organisms that normally do not produce infection unless there is impaired immune function. Although a person with AIDS may live for many years after the first serious illness, as the immune system fails, these opportunistic illnesses become progressively more severe and difficult to treat.

In the United States, the most common opportunistic infections are *Pneumocystis carinii* pneumonia (PCP), oropharyngeal or esophageal candidiasis (thrush), cytomegalovirus (CMV), and respiratory infections caused by *Mycobacterium avium-intracellulare complex* (MAC).³¹

Respiratory Infections. The most common causes of respiratory disease in persons with HIV infection are PCP and pulmonary tuberculosis (TB). Other organisms that cause opportunistic pulmonary infections in persons with AIDS include CMV, MAC, *Toxoplasma gondii*, and *Cryptococcus neoformans*. Pneumonia also may occur because of more common pulmonary pathogens, including *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Legionella pneumophila*. Some persons may be infected with multiple organisms. Kaposi's sarcoma (KS) also can occur in the lungs.

P. carinii pneumonia was the most common presenting manifestation of AIDS during the first decade of the epidemic. Since highly active antiretroviral therapy (HAART) and prophylaxis for PCP were instituted, the incidence has decreased.³² PCP still is common in people who do not know their HIV status, those who choose not to treat their HIV, and in those with poor access to health care. The best predictor of PCP is a CD4⁺ T-cell count of less than 200 cells/µL, and it is at this point that prophylaxis with trimethoprim-sulfamethoxazole

186 Unit Two: Alterations in Body Defenses

is started. PCP is caused by *P. carinii*, an organism that is common in soil, houses, and many other places in the environment. Although *P. carinii* does not cause infection in persons with healthy immune systems, it can multiply quickly in the lungs of persons with AIDS and cause pneumonia. The symptoms of PCP may be acute or gradually progressive. The person may present with mild cough, fever, shortness of breath, and weight loss. Physical examination may demonstrate only fever and tachypnea, and breath sounds may be normal. As the disease progresses, the alveoli become filled with foamy protein-rich fluid, causing impairment of gas exchange (Fig. 10-9). Diagnosis of PCP is made on recognition of the organism in pulmonary secretions. This can be done through examination of induced sputum, bronchoalveolar lavage, occasionally bronchoscopy, and rarely, lung biopsy.

Tuberculosis is the leading cause of death for people with HIV worldwide. There are more than 80,000 people coinfected with HIV and TB in North America and another 5 million in the rest of the world.³³

Although the lungs are the most common site of *M. tuber-culosis* infection, extrapulmonary infections of the kidney, bone marrow, and other organs also occur in people with HIV. Whether a person has pulmonary or extrapulmonary TB, most patients present with fever, night sweats, cough, and weight loss.³³ Persons infected with *M. tuberculosis* (*i.e.*, those with positive tuberculin skin tests) are more likely to have reactivated TB develop if they become infected with HIV; if they are coinfected, they are more likely to have a rapidly progressive form of TB.³³ Equally important, HIV-infected persons with TB coinfection usually have an increase in viral load, which decreases the success of TB therapy. They also have an increased number of other opportunistic infections and an increased mortality rate.³³

Since the late 1960s, most persons with TB have experienced good response to therapy. However, in 1991, there were outbreaks of multidrug-resistant (MDR) TB. Many cases of drug-resistant TB occur in HIV-infected persons.

Gastrointestinal Infections. Infections of the gastrointestinal tract are some of the most common complications of HIV and

AIDS. Esophageal candidiasis (thrush), CMV infection, and herpes simplex virus infection are common opportunistic infections that cause esophagitis in people with HIV. Persons experiencing these infections usually report painful swallowing or retrosternal pain. The clinical presentation can range from no symptoms to a complete inability to swallow and dehydration.

Diarrhea or gastroenteritis is common in persons with HIV. The most common protozoal infection that causes diarrhea is *Cryptosporidium parvum*. The clinical features of cryptosporidiosis can range from mild diarrhea to severe, watery diarrhea with a loss of as much as several liters of water per day. The most severe form usually occurs in persons with a CD4⁺ count of less than 50 cells/µL, and also can include malabsorption, electrolyte disturbances, dehydration, and weight loss.³⁴ Other organisms that cause gastroenteritis and diarrhea are *Salmonella*, CMV, *Clostridium difficile*, *Escherichia coli*, *Shigella*, *Giardia*, and Microsporida.

Nervous System Manifestations. HIV infection, particularly in its late stages of severe immunocompromise, leaves the nervous system vulnerable to an array of neurologic disorders, including AIDS dementia complex (ADC), toxoplasmosis, and progressive multifocal leukoencephalopathy (PML). These disorders can affect the peripheral (PNS) or central nervous system (CNS) and contribute to the morbidity and mortality of persons with HIV.³⁵

AIDS dementia complex is a syndrome of cognitive and motor dysfunction. ADC is caused by HIV itself, rather than an opportunistic infection, and usually is a late complication of HIV. The clinical features of ADC are impairment of attention and concentration, slowing of mental speed and agility, slowing of motor speed, and apathetic behavior. The diagnosis of ADC can be based on these clinical findings.

Toxoplasmosis is a common opportunistic infection in persons with AIDS. The organism responsible, *T. gondii*, is a parasite that most often affects the CNS. Toxoplasmosis usually is a reactivation of a latent *T. gondii* infection that has been dormant in the CNS.³⁶ The typical presentation includes fever, headaches, and neurologic dysfunction, including confusion

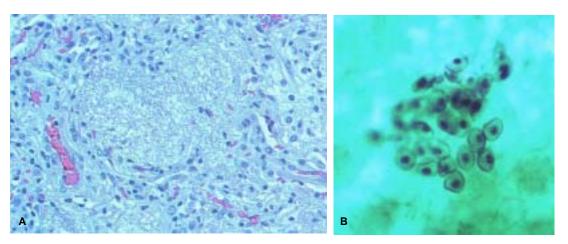


FIGURE 10-9 *Pneumocystis carinii* pneumonia. (**A**) The alveoli are filled with a foamy exudate, and the interstitium is thickened and contains a chronic inflammatory infiltrate. (**B**) A centrifuged bronchoalveolar lavage specimen impregnated with silver shows a cluster of *Pneumocystis carinii* cysts. (From Rubin E., Farber J.L. [1994]. *Pathology* [2nd ed.]. Philadelphia: J.B. Lippincott)

and lethargy, visual disturbances, and seizures. Computed tomography scans or magnetic resonance imaging (MRI) should be performed immediately to detect the presence of neurologic lesions. Prophylactic treatment with trimethoprimsulfamethoxazole is effective against *T. gondii* when the CD4⁺ T-cell count decreases to less than 200 cells/µL.

Progressive multifocal leukoencephalopathy is a demyelinating disease of the white matter of the brain caused by the JC virus, a DNA papovavirus that attacks the oligodendrocytes. PML advances slowly, and it can be weeks to months before the person seeks medical care.³⁷ It is characterized by progressive limb weakness, sensory loss, difficulty controlling the digits, visual disturbances, subtle alterations in mental status, hemiparesis, ataxia, diplopia, and seizures. Diagnosis is based on clinical findings and an MRI, and confirmed by the presence of the JC virus. The mortality rate is high, and the average survival time is 2 to 4 months.³⁷

Cancers and Malignancies

Persons with AIDS have a high incidence of certain malignancies, especially Kaposi's sarcoma (KS), non-Hodgkin's lymphoma, and noninvasive cervical carcinoma. The increased incidence of malignancies probably is a function of impaired cell-mediated immunity. Non-Hodgkin's lymphoma develops in 3% to 4% of people with HIV infection (see Chapter 11). Women with HIV infection experience a higher incidence of *cervical dysplasia* than do women without HIV infection.³⁸ Cervical dysplasia, which usually results from infection with a human papillomavirus, is a slowly developing precursor to cervical carcinoma (see Chapter 34). In women with HIV infection this progression is much more rapid.³⁸

Kaposi's Sarcoma. Kaposi's sarcoma is a malignancy of endothelial cells that line small blood vessels. An opportunistic cancer, KS occurs in immunosuppressed persons (e.g., transplant recipients or persons with AIDS). KS was one of the first opportunistic cancers associated with AIDS and still is the malignancy most frequently related to HIV.³⁹

The lesions of KS can be found on the skin and in the oral cavity, gastrointestinal tract, and lungs. More than 50% of people with skin lesions also have gastrointestinal lesions. The disease usually begins as one or more macules, papules, or violet skin lesions that enlarge and become darker (Fig. 10-10). They may enlarge to form raised plaques or tumors. These irregularly shaped tumors can be from one eighth of an inch to silver dollar size. Tumor nodules frequently are located on the head, neck, and trunk. They usually are painless in the early stages, but discomfort may develop as the tumor ages. Invasion to the internal organs, including the lungs, gastrointestinal tract, and lymphatic system, commonly occurs. Gastrointestinal tract KS often is asymptomatic but can cause pain, bleeding, or obstruction. Pulmonary KS usually is a late development of the disease. Pulmonary KS causes dyspnea, cough, and hemoptysis. The progression of KS may be slow or rapid.

A presumptive diagnosis of KS usually is made based on visual identification of red or violet skin or oral lesions. Biopsy of at least one lesion must be done to establish the diagnosis and to distinguish the KS from other skin lesions that may resemble it. Effective HAART, local therapy with liquid nitrogen or vinblastine, chemotherapy, radiation, and interferon injections are the most common therapies. These therapies are largely palliative and are not a cure.



■ FIGURE 10-10 ■ Disseminated Kaposi's sarcoma. Multiple red to brown papules distributed along the skin lines in a man with AIDS. (Hall J.C. [2000] *Sauer's manual of skin* [p.197]. Philadelphia: Lippincott Williams & Wilkins.)

There is recent evidence linking KS to a herpesvirus (herpesvirus 8, also called KS-associated herpesvirus [KSHV]).⁴⁰ More than 95% of KS lesions, regardless of the source or clinical subtype, have reportedly been found to be infected with KSHV. The virus is readily transmitted through homosexual and heterosexual activities. Maternal–infant transmission also occurs. The virus has been detected in saliva from infected persons, and other modes of transmission are suspected.

Wasting Syndrome

In 1997, wasting became an AIDS-defining illness. The syndrome is common in persons with HIV infection or AIDS. Wasting is characterized by involuntary weight loss of at least 10% of baseline body weight in the presence of diarrhea, more than two stools per day, or chronic weakness and a fever. This diagnosis is made when no other opportunistic infections or neoplasms can be identified as causing these symptoms. Factors that contribute to wasting are anorexia, metabolic abnormalities, endocrine dysfunction, malabsorption, and cytokine dysregulation. Treatment for wasting includes nutritional interventions such as oral supplements, or enteral or parenteral nutrition. There also are numerous pharmacologic agents used to treat wasting, including appetite stimulants, cannabinoids, and megestrol acetate.⁴¹

Metabolic Disorders

A wide range of metabolic disorders is associated with HIV infection, including lipodystrophy and mitochondrial disorders.

Lipodystrophy. A metabolic disorder called *lipodystrophy* is one of the newest group of problems for those infected with HIV. The symptoms of HIV-associated lipodystrophy fall into two categories: changes in body appearance and metabolic changes. The alterations in body appearance include an increase in abdominal girth, abnormal distribution of fat in the supraclavicular area (*i.e., buffalo hump*), wasting of fat from the face and

extremities, and breast enlargement in men and women. The metabolic changes include elevated serum cholesterol and triglyceride levels and insulin resistance. Originally attributed to the use of protease inhibitors, the pathogenesis of lipodystrophy still is not understood. It may be caused by protease inhibitor therapy or nucleoside reverse transcriptase inhibitor therapy, or may arise simply because people are living longer with HIV.

The diagnosis of lipodystrophy is based on appearance changes, elevated serum levels of triglycerides and cholesterol, and observed changes in body shape (measured changes in waist and hip girth).⁴² The management of lipodystrophy is a matter of controversy because the etiology is unknown. Some authorities recommend switching to a nonprotease inhibitor-based HAART regimen. The problem with this approach is that, although serum levels of triglycerides and cholesterol decrease and there may be some resolution of the fat redistribution, the viral load often increases and becomes detectable.⁴³

Mitochondrial Disorders. Mitochondrial disorders are metabolic disorders caused by antiretroviral therapy, in particular, the nucleoside reverse transcriptase inhibitors.⁴² The mitochondria control many of the oxidative chemical reactions that release energy from glucose and other organic molecules. In the absence of normal mitochondrial function, cells revert to anaerobic metabolism with generation of lactic acid. Patients often present with nonspecific gastrointestinal symptoms, including nausea, vomiting, and abdominal pain. On examination, they can have hepatomegaly with normal liver function tests. The only laboratory abnormality may be lactic acidosis.⁴² Mitochondrial dysfunction is the most feared complication of antiretroviral therapy. This fear is caused by the condition's unpredictability, its fatality in half the presenting patients, the nonspecific presenting symptoms, and the prevalence of elevated lactate levels in 8% to 22% of patients who have no symptoms.⁴²

Treatment

There is no cure for AIDS. Although many companies are working on a vaccine to prevent HIV infection, none has been approved. There currently are three different types of HIV antiretroviral medications: nucleoside reverse transcriptase inhibitors, non-nucleoside reverse transcriptase inhibitors, and protease inhibitors.⁴⁴ Each type of agent attempts to interrupt viral replication at a different point.

Reverse transcriptase inhibitors inhibit HIV replication by acting on the enzyme reverse transcriptase. There are two types of HIV medications that work on this enzyme: nucleoside reverse transcriptase inhibitors and non-nucleoside reverse transcriptase inhibitors. *Nucleoside reverse transcriptase inhibitors* act by blocking the elongation of the DNA chain by stopping more nucleosides from being added. *Non-nucleoside reverse transcriptase inhibitors* work by binding to the reverse transcriptase enzyme so it cannot copy the virus's RNA into DNA. *Protease inhibitors* bind to the protease enzyme and inhibit its action. This inhibition prevents the cleavage of the polyprotein chain into individual proteins, which would be used to construct the new virus. Because the information inside the nucleus is not put together properly, the new viruses that are released into the body are immature and noninfectious.

The treatment of HIV is one of the most rapidly evolving fields in medicine. Optimal treatment of HIV includes a com-

bination of drugs because different drugs act on different stages of the replication cycle. The goal of HAART, using a combination of antiviral drugs, is a sustained suppression of HIV replication, resulting in an undetectable viral load and an increasing CD4⁺ count. In general, antiviral therapies are prescribed to improve the overall survival time of persons with HIV infection and to slow the progression to AIDS.

Drugs and vaccines commonly are used for the prevention and treatment of opportunistic infections and conditions, including PCP, toxoplasmosis, MAC, candidiasis, CMV infection, influenza, hepatitis B, and *S. pneumoniae* infections. Prophylactic medications are used once an individual's CD4⁺ count has dropped below a certain level that indicates his or her immune system is no longer able to fight off the opportunistic infections.

Persons with HIV should be advised to avoid infections as much as possible and seek evaluation promptly when they occur. Immunization is important because persons infected with HIV are at risk for contracting many infectious diseases. Some of these diseases can be avoided by vaccination while the immune system's responsiveness is relatively intact.

Infection in Pregnancy and in Infants and Children

Early in the epidemic, children who contracted HIV could have become infected through blood products or perinatally. Now, almost all of the children who become infected with HIV at a young age in the United States get HIV perinatally. Infected women may transmit the virus to their offspring in utero, during labor and delivery, or through breast milk.⁴⁵ The risk of transmission is increased if the mother has advanced HIV disease as evidenced by low CD4⁺ counts, high levels of HIV in the blood (high viral load); if the time from the rupture of membranes to delivery is prolonged; if there is increased exposure of the fetus to maternal blood; or if the mother breast-feeds the child.²⁵

Diagnosis of HIV infection in children born to HIVinfected mothers is complicated by the presence of maternal HIV IgG antibody, which crosses the placenta to the fetus. Consequently, infants born to HIV-infected women can be HIV antibody positive by ELISA for as long as 18 months, even though they are not HIV infected.²⁵ PCR testing for HIV DNA is used most often to diagnose HIV in infants younger than 18 months. Two positive PCR tests for HIV DNA are needed to diagnose HIV in a child. Children born to mothers with HIV infection are considered uninfected if they become HIV antibody negative after 6 months of age, have no other laboratory evidence of HIV infection, and have not met the surveillance case definition criteria for AIDS in children.²⁵

Perinatal transmission can be lowered by approximately two thirds by administering zidovudine to the mother during pregnancy and labor and delivery and to the infant when it is born.⁴⁵ Thus, the U.S. Public Health Service recommends that HIV counseling and testing should be offered to all pregnant women and women of childbearing age in the United States.⁴⁵ The recommendations also stress that women who test positive for HIV antibodies should be informed of the perinatal prevention benefits of zidovudine therapy and offered treatment that includes zidovudine alone, or HAART therapy. The benefits of voluntary testing for mothers and newborns include reduced morbidity because of intensive treatment and supportive health care, the opportunity for early antiviral therapy for mother and child, and information regarding the risk of transmission from breast milk.45

Because pregnant women in less developed countries do not always have access to zidovudine, studies are being conducted in Africa to determine if any other simple and less expensive antiretroviral regimen can be used to decrease the transmission from mother to infant.

Children have a very different pattern of HIV infection than do adults. Failure to thrive, CNS abnormalities, and developmental delays are the most prominent primary manifestations of HIV infection in children.²⁸ Children born infected with HIV usually weigh less and are shorter than noninfected infants. A major cause of early mortality for HIV-infected children is PCP. As opposed to adults, in whom PCP occurs in the late stages, children experience PCP early, with the peak age of onset at 3 to 6 months. For this reason, prophylaxis with trimethoprimsulfamethoxazole is started by 4 to 6 weeks for all infants born to HIV-infected mothers, regardless of their CD4+ count or infection status.46

In summary, an immunodeficiency is defined as an absolute or partial loss of the normal immune response, which places a person in a state of compromise and increases the risk for development of infections or malignant complications. Any components of the immune response, including antibody or humoral (B-cell) immunity or cellular or T-cell immunity, may contribute to immunodeficiencies. The variety of defects known to involve the immune response can be classified as primary (i.e., endogenous or inherited) or secondary (*i.e.*, caused by exogenous factors such as drugs or infection).

Most primary immunodeficiency states are inherited and are either present at birth or become apparent shortly after birth. Primary immunodeficiencies can be categorized into three types: humoral (B-cell) immunodeficiencies, cellular (T-cell) immunodeficiencies, and combined (B-cell and T-cell) immunodeficiencies. B-cell immunodeficiencies can selectively affect a single type of immunoglobulin (e.g., IgA immunodeficiency) or all of the immunoglobulins. Defects in B-cell function increase the risk of recurrent pyogenic infections, including those caused by Streptococcus pneumoniae, Haemophilus influenzae, and Staphylococcus aureus, and by gram-negative organisms such as Pseudomonas species. There are few primary forms of T-cell immunodeficiency, probably because persons with defects in this branch of the immune response rarely survive beyond infancy or childhood. Combined immunodeficiency disorders involve both B-cell and T-cell dysfunction and include a spectrum of inherited (autosomal recessive and X-linked) conditions. The spectrum of disease resulting from combined immunodeficiency ranges from mild to severe to ultimately fatal forms.

AIDS is the most common type of secondary immunodeficiency. AIDS is an infectious disease of the immune system caused by the retrovirus HIV. The disease is prevalent worldwide and is one of the leading causes of death among young adults in the United States. HIV is transmitted from one person to another through sexual contact, through blood exchange, or perinatally. HIV is a retrovirus that infects the

body's CD4⁺ T cells and macrophages. The destruction of CD4⁺ T cells by HIV constitutes an attack on the entire immune system because this subset of lymphocytes exerts critical regulatory and effector functions that involve both humoral and cellular immunity.

Manifestations of infection, such as acute mononucleosislike symptoms, may occur shortly after infection, and this is followed by a latent phase that may last for many years. The end of the latent period is characterized by the marked decrease in CD4⁺ T cells and the development of opportunistic infections, cancers, and other disorders as the person moves toward an AIDS diagnosis. The complications of these infections, manifested throughout the respiratory, gastrointestinal, and nervous systems, include pneumonia, esophagitis, diarrhea, gastroenteritis, tumors, wasting syndrome, altered mental status, seizures, motor deficits, and metabolic disorders. There is no cure for AIDS. Treatment largely involves the use of drugs that interrupt the replication of the HIV virus and prevention or treatment of complications such as opportunistic infections.

Infected women may transmit the virus to their offspring in utero, during labor and delivery, or through breast milk. Diagnosis of HIV infection in children born to HIV-infected mothers is complicated by the presence of maternal HIV antibody, which crosses the placenta to the fetus. This antibody usually disappears within 18 months in uninfected children. The administration of zidovudine to the mother during pregnancy and labor and delivery and to the infant after birth can decrease perinatal transmission.

REVIEW QUESTIONS

Describe the immune mechanisms involved in a type I, type II, type III, and type IV hypersensitivity reaction and use these mechanisms to describe the pathogenesis of allergic rhinitis, food allergy, serum sickness, contact dermatitis, and hypersensitivity pneumonitis.

Compare the immune mechanisms involved in host-versusgraft and graft-versus-host transplant rejection.

Relate the mechanisms of self-tolerance to the possible explanations for development of autoimmune disease.

Describe three or more postulated mechanisms underlying autoimmune disease.

Compare and contrast immunodeficiency disorders caused by B-cell, T-cell, and combined B- and T-cell disorders.

Explain why the attack by the HIV virus on the CD4⁺ cell is so devastating in terms of the function of the immune system.

Explain why it is not possible to state with certainty that a person with a positive ELISA antibody test for HIV does in fact have the disease and needs to be retested using the Western blot test.

Relate the altered immune function in persons with HIV infection and AIDS to the development of opportunistic infections.

Discuss the vertical transmission of HIV from mother to child and explain why the HIV test might be positive even the infant does not have the virus.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- Johnson K.J., Chensue S.W., Ward P.A. (1999). In Rubin E., Farber J.L. (Eds.), *Pathology* (3rd ed., pp. 114–127, 134). Philadelphia: Lippincott Williams & Wilkins.
- Kumar V., Cotran R.S., Robbins S.L. (2003). Robbins basic pathology (7th ed., pp. 103–164, 1252). Philadelphia: W.B. Saunders.
- Galli S.J. (1993). New concepts about the mast cell. New England Journal of Medicine 328, 257–265.
- Rachelefsky G.S. (1999). National guidelines need to manage rhinitis and prevent complications. *Annals of Allergy, Asthma, and Immunology* 82, 296–305.
- Sicherer S.H. (1999). Manifestations of food allergy: Evaluation and management. *American Family Physician* 57, 93–102.
- Salvaggio J.E. (1995). The identification of hypersensitivity pneumonitis. *Hospital Practice* 30 (5), 57–66.
- Sussman G.L. (1995). Allergy to latex rubber. Annals of Internal Medicine 122, 43–46.
- 8. Poley G.E., Slater J.E. (2000). Latex allergy. *Journal of Allergy and Clinical Immunology* 105, 1054–1062.
- Kamradt T., Mitchison N.A. (2001). Advances in immunology: Tolerance and autoimmunity. New England Journal of Medicine 344, 655–664.
- Theofilopoulos A.N. (1995). The basis of autoimmunity: Part II: Genetic predisposition. *Immunology Today* 16, 150–158.
- 11. Cutolo M., Sulli A., Seriolo S., et al. (1995). Estrogens, the immune response and autoimmunity. *Clinical and Experimental Rheumatology* 13, 217–226.
- Rose N.R. (1997). Autoimmune disease: Tracing the shared threads. Hospital Practice 32(4), 147–154.
- 13. Albert L.J., Inman R.D. (1999). Molecular mimicry and autoimmunity. *New England Journal of Medicine* 341, 2068–2074.
- 14. Kotzin B.L. (1994). Superantigens and their role in disease. *Hospital Practice* 29 (11), 59–70.
- Buckley R.H. (2000). Primary immunodeficiency diseases due to defects in lymphocytes. *New England Journal of Medicine* 343, 1313–1324.
- Sorensen R.U., Moore C. (2000). Antibody deficiency syndromes. Pediatric Clinics of North America 47, 1225–1252.
- Ten R.M. (1998). Primary immunodeficiencies. Mayo Clinic Proceedings 73, 865–872.
- Buckley R. (2000). T-, B-, and NK-cell systems. In Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.), *Nelson textbook of pediatrics* (16th ed., pp. 590–606). Philadelphia: W.B. Saunders.
- Johnson K.B., Oski F.A. (1997). Oski's essential pediatrics (pp. 532– 539). Philadelphia: Lippincott-Raven.
- Elder M.E. (2000). T-cell immunodeficiencies. Pediatric Clinics of North America 47, 1253–1274.
- Candotti F. (2000). The potential for therapy of immune disorders with gene therapy. *Pediatric Clinics of North America* 47, 1389– 1405.
- 22. American Association for World Health. (2000). *AIDS: All men make a difference*. Washington, DC: Author.
- O'Brien T.R., George J.R., Holmberg S.D. (1992). Human immunodeficiency virus type-2 infection in the United States: Epidemiology, diagnosis, and public health implications. *Journal of the American Medical Association* 267, 2775–2779.
- 24. Colpin H. (1999). Prevention of HIV transmission through behavioral changes and sexual means. In Armstrong D., Cohen J.

(Eds.), Infectious diseases (Section 5, Chapter 2, pp. 1-4). London: Harcourt.

- Havens P.L. (1999). Pediatric AIDS. In Armstrong D., Cohen J. (Eds.), Infectious diseases (Section 5, Chapter 20). London: Harcourt.
- Henderson D.K. (1999). Preventing occupational infections with HIV in health care settings. In Armstrong D., Cohen J. (Eds.), *Infectious dis*eases (Section 5, Chapter 3, pp. 1–10). London: Harcourt.
- Hirschel B. (1999). Primary HIV infection. In Armstrong D., Cohen J. (Eds.), *Infectious diseases* (Section 5, Chapter 8, pp. 1–4). London: Harcourt.
- Brun-Vezinet F., Simon F. (1999). Diagnostic tests for HIV infection. In Armstrong D., Cohen J. (Eds.), *Infectious diseases* (Section 5, Chapter 23, pp. 1–10). London: Harcourt.
- Centers for Disease Control and Prevention. (1992). 1993 Revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. *Morbidity and Mortality Weekly Report* 41 (RR-17), 1–23.
- Rizzardi G.P., Pantaleo G. (1999). The immunopathogenesis of HIV-1 infection. In Armstrong D., Cohen J. (Eds.), *Infectious diseases* (Section 5, Chapter 6, pp. 1–12). London: Harcourt.
- Clumeck N., Dewit S. (1999). Prevention of opportunistic infections in the presence of HIV infection. In Armstrong D., Cohen J. (Eds.), *Infectious diseases* (Section 5, Chapter 9). London: Harcourt.
- Girard P.M. (1999). Pneumocystis carinii pneumonia. In Armstrong D., Cohen J. (Eds.), Infectious diseases (Section 5, Chapter 10, pp. 1–4). London: Harcourt.
- Gordin F. (1999). Mycobacterium tuberculosis. In Dolin R., Masur H., Saag M.S. (Eds.), AIDS therapy (pp. 359–374). Philadelphia: Churchill Livingstone.
- Wilcox C.M., Monkemuller K.E. (1999). Gastrointestinal disease. In Dolin R., Masur H., Saag M.S. (Eds.), *AIDS therapy* (pp. 752–765). Philadelphia: Churchill Livingstone.
- Price R.W. (1999). Neurologic disease. In Dolin R., Masur H., Saag M.S. (Eds.), AIDS therapy (pp. 620–638). Philadelphia: Churchill Livingstone.
- Katlama C. (1999). Parasitic infections. In Armstrong D., Cohen J. (Eds.), *Infectious diseases* (Section 5, Chapter 13, pp. 1–4). London: Harcourt.
- Hall C.D. (1999). JC virus neurologic infection. In Dolin R., Masur H., Saag M.S. (Eds.), *AIDS therapy* (pp. 565–572). Philadelphia: Churchill Livingstone.
- Krown S.E. (1999). Kaposi sarcoma. In Dolin R., Masur H., Saag M.S. (Eds.), AIDS therapy (pp. 580–591). Philadelphia: Churchill Livingstone.
- Anteman K., Chang Y. (2000). Kaposi's sarcoma. New England Journal of Medicine 342, 1027–1038.
- 40. Centers for Disease Control. (1990). Risk of cervical disease in HIV infected women. *Morbidity and Mortality Weekly Report* 39, 846–849.
- Von Ruenn J.H., Mulligan K. (1999). Wasting syndrome. In Dolin R., Masur H., Saag M.S. (Eds.), *AIDS therapy* (pp. 607–619). Philadelphia: Churchill Livingstone.
- Chaisson R.E., Triesman G.J. (2000). Antiretroviral therapy in perspective: Managing drug side effects to improve patient outcomes (Vol. I). Connecticut: Scientific Exchange.
- Lyon D., Truban E. (2000). HIV-related lipodystrophy: A clinical syndrome with implications for nursing practice. *Journal of the Association* of Nurses in AIDS Care 11 (2), 36–42.
- Vella S., Floridia M. (1999). Antiretroviral therapy. In Armstrong D., Cohen J. (Eds.), *Infectious diseases* (Section 5, Chapter 26, pp. 1–10). London: Harcourt.
- 45. U.S. Public Health Service. (2000). Revised public health service recommendations for human immunodeficiency virus screening of pregnant women. Washington, DC: Author.
- 46. Connor E.M., Sperling R.S., Gelber R., et al. (1994). Reduction of maternal-infant transmission of human immunodeficiency virus type 1 with zidovudine treatment. *New England Journal of Medicine* 331, 1173–1180.

Three

CHAPTER

Alterations in the Hematologic System

White Blood Cells

Alterations in

Hematopoietic and Lymphoid Tissue Leukocytes (White blood cells) Granulocytes Lymphocytes Monocytes and Macrophages The Bone Marrow and Hematopoiesis Lymphoid Tissues Non-neoplastic Disorders of White Blood Cells Neutropenia (Agranulocytosis) Acquired Neutropenia **Congenital Neutropenia Clinical Course** Infectious Mononucleosis **Pathogenesis Clinical Course** Neoplastic Disorders of Hematopoietic and Lymphoid Origin Malignant Lymphomas Hodgkin's Disease Non-Hodgkin's Lymphomas

Leukemias Classification Acute Leukemias Chronic Leukemias Multiple Myeloma

he white blood cells and lymphoid tissues where these cells originate and mature function to protect the body against invasion by foreign agents. Disorders of the white blood cells include a deficiency of leukocytes (leukopenia) or increased numbers as occurs with proliferative disorders. The proliferative disorders may be reactive, such as occurs with infection, or neoplastic, such as occurs with leukemias and lymphomas.

HEMATOPOIETIC AND LYMPHOID TISSUE

Blood consists of blood cells (*i.e.*, white blood cells, thrombocytes or platelets [see Chapter 12], and red blood cells [see Chapter 13]) and the plasma in which the cells are suspended. The generation of blood cells takes place in the *hematopoietic* (from the Greek *haima* for "blood" and *poiesis* for "making") system.¹ The hematopoietic system encompasses all of the blood cells and their precursors, the bone marrow where blood cells have their origin, and the lymphoid tissues where some blood cells circulate as they develop and mature.

Leukocytes (White Blood Cells)

The leukocytes, or white blood cells, constitute only 1% of the total blood volume. They originate in the bone marrow and circulate throughout the lymphoid tissues of the body. There they function in the inflammatory and immune processes. They include the granulocytes, the lymphocytes, and the monocytes (Fig. 11-1).

Granulocytes

192

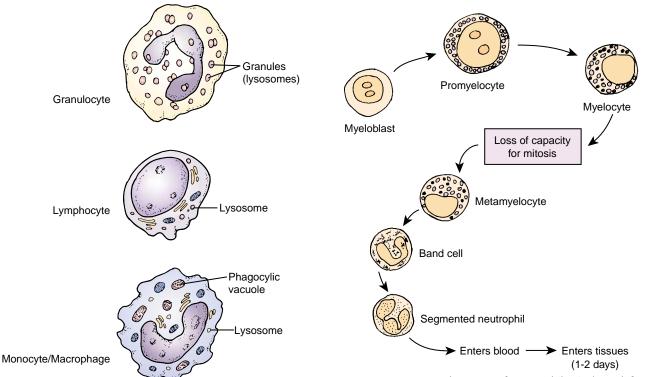
The granulocytes are all phagocytic cells and are identifiable because of their cytoplasmic granules. These white blood cells are spherical and have distinctive multilobar nuclei. The granulocytes are divided into three types (neutrophils, eosinophils, and basophils) according to the staining properties of the granules.

Neutrophils. The neutrophils, which constitute 55% to 65% of the total number of white blood cells, have granules that are neutral and thus do not stain with an acidic or a basic dye. Because their nuclei are divided into three to five lobes, they are often called polymorphonuclear leukocytes (PMNs).

The neutrophils are primarily responsible for maintaining normal host defenses against invading bacteria, fungi, products of cell destruction, and a variety of foreign substances. The cytoplasm of mature neutrophils contains fine granules. These granules contain degrading enzymes that are used in destroying foreign substances and correspond to lysosomes found in other cells. Enzymes and oxidizing agents associated with these granules are capable of degrading a variety of natural and synthetic substances, including complex polysaccharides, proteins, and lipids. The degradative functions of the neutrophil are important in maintaining normal host defenses and in mediating the inflammatory response (see Chapter 9).

The neutrophils have their origin in the myeloblasts that are found in the bone marrow (Fig. 11-2). The myeloblasts are the committed precursors of the granulocyte pathway and do not normally appear in the peripheral circulation. When they are present, it suggests a disorder of blood cell proliferation and differentiation. The myeloblasts differentiate into promyelocytes and then myelocytes. Generally, a cell is not called a myelocyte until it has at least 12 granules. The myelocytes mature to become metamyelocytes (Greek meta for "beyond"), at which point they lose their capacity for mitosis. Subsequent development of the neutrophil involves reduction in size, with transformation from an indented to an oval to a horseshoeshaped nucleus (i.e., band cell) and then to a mature cell with a segmented nucleus. Mature neutrophils are often referred to as segs because of their segmented nucleus. The development from stem cell to mature neutrophil takes about 2 weeks. It is at this point that the neutrophil enters the bloodstream.

After release from the marrow, the neutrophils spend only about 4 to 8 hours in the circulation before moving into the tissues.² Their survival in the tissues lasts about 4 to 5 days. They die in the tissues while discharging their phagocytic function or die of senescence. The pool of circulating neutrophils (*i.e.*, those that appear in the blood count) is in rapid equilibrium with a similar-sized pool of cells marginating along the



■ **FIGURE 11-1** White blood cells—granulocyte, lymphocyte, and monocyte.

■ FIGURE 11-2 ■ Development of neutrophils. (Adapted from Cormack D.H. [1993]. *Ham's histology* [9th ed.]. Philadelphia: J.B. Lippincott)

walls of small blood vessels. These are the neutrophils that respond to chemotactic factors and migrate into the tissues toward the offending agent during an inflammatory response. Epinephrine, exercise, stress, and corticosteroid drug therapy can cause rapid increases in the circulating neutrophil count by shifting cells from the marginating to the circulating pool. Endotoxins or microbes may have the opposite effect, by attracting neutrophils to move out of the circulation and into the tissues.

Eosinophils. The cytoplasmic granules of the eosinophils stain red with the acidic dye eosin. These leukocytes constitute 1% to 3% of the total number of white blood cells and increase in number during allergic reactions. They are thought to release enzymes or chemical mediators that detoxify agents associated with allergic reactions. Eosinophils are also involved in parasitic infections. Although most parasites are too large to be phagocytized by eosinophils, the eosinophils attach themselves to the parasite by special surface molecules and release hydrolytic enzymes and other substances from their granules that kill the parasite.

Basophils. The granules of the basophils stain blue with a basic dye. These cells constitute only about 0.3% to 0.5% of the white blood cells. The basophils in the circulating blood are similar to the large mast cells located immediately outside the capillaries in body tissues. Both the basophils and mast cells release heparin, an anticoagulant, into the blood. The mast cells and basophils also release histamine, a vasodilator, and other inflammatory mediators. The mast cells and basophils play an exceedingly important role in allergic reactions (see Chapter 10).

Lymphocytes

The lymphocytes have their origin in the lymphoid stem cells that are found in the bone marrow. The lymphocytes constitute 20% to 30% of the white blood cell count. They have no identifiable granules in the cytoplasm and are sometimes referred to as *agranulocytes*. They move between blood and lymphoid tissues, where they may be stored for hours or years. Their function in the lymph nodes or spleen is to defend against foreign microbes in the immune response (see Chapter 8). There are two types of lymphocytes: B lymphocytes and T lymphocytes. The lymphocytes play an important role in the immune response. The B lymphocytes differentiate to form antibodyproducing plasma cells and are involved in humoral-mediated immunity. The T lymphocytes are responsible for orchestrating the immune response (CD4⁺ T cells) and effecting cellmediated immunity (CD8⁺ T cells).

Monocytes and Macrophages

Monocytes are the largest of the white blood cells and constitute about 3% to 8% of the total leukocyte count. They have abundant cytoplasm and a darkly stained nucleus, which has a distinctive U or kidney shape. The circulating life span of the monocyte is about 1 to 3 days, three to four times longer than that of the granulocytes. These cells survive for months to years in the tissues. The monocytes, which are phagocytic cells, are often referred to as *macrophages* when they enter the tissues. The monocytes engulf larger and greater quantities of foreign material than do the neutrophils. These leukocytes play an important role in chronic inflammation and are also involved in the immune response by activating lymphocytes and by presenting antigen to T cells. When the monocyte leaves the vascular system and enters the tissues, it functions as a macrophage with specific activity. The macrophages are known as histiocytes in loose connective tissue, *microglial cells* in the brain, and *Kupffer's cells* in the liver.

The Bone Marrow and Hematopoiesis

The blood-forming population of bone marrow is made up of three types of cells: self-renewing stem cells, differentiated progenitor (parent) cells, and functional mature blood cells. All of the blood cell precursors of the erythrocyte (*i.e.*, red cell), myelocyte (*i.e.*, granulocyte or monocyte), lymphocyte (*i.e.*, platelet) series are derived from a small population of primitive cells called the pluripotent stem cells (Fig. 11-3). Their lifelong potential for proliferation and self-renewal makes them an indispensable and lifesaving source of reserve cells for the entire hematopoietic system.

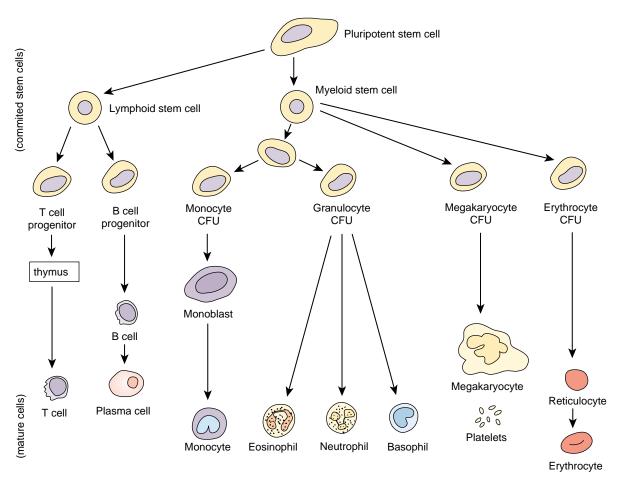
Several levels of differentiation lead to the development of committed stem cells, which are the progenitor for each of the blood cell types. A committed stem cell that forms a specific type of blood cell is called a *colony-forming unit* (CFU). Under normal conditions, the numbers and total mass for each type of circulating blood cell remain relatively constant. The blood cells are produced in different numbers according to needs and regulatory factors. This regulation of blood cells is controlled by a group of short-acting soluble mediators, called *cytokines*, that stimulate the proliferation, differentiation, and functional activation of the various blood cell precursors in bone marrow.³ The cytokines that stimulate hematopoiesis are called colony-stimulating factors (CSFs), based on their ability to promote the growth of the hematopoietic cell colonies from bone marrow precursors. Lineage-specific CSFs that act on committed progenitor cells include: erythropoietin (EPO), granulocyte-monocyte colony-stimulating factor (GM-CSF), and thrombopoietin (TPO). The major sources of the CSFs are lymphocytes and stromal cells of the bone marrow. Other cytokines, such as the interleukins, support the development of lymphocytes and act synergistically to aid the functions of the CSFs (see Chapter 8).

Lymphoid Tissues

The lymphoid tissues represent the structures where lymphocytes originate, mature, and interact with antigens. Lymphoid tissues can be classified into two groups: the central or generative organs and peripheral lymphoid organs (see Chapter 8). The central lymphoid structures consist of the bone marrow, where all lymphocytes arise, and the thymus, where T cells mature and reach a stage of functional competence. The thymus is also the site where self-reactive T cells are eliminated.

The peripheral lymphoid organs are the sites where mature lymphocytes respond to foreign antigens. They include the lymph nodes, the spleen, mucosa-associated lymphoid tissues, and the cutaneous immune system. In addition, poorly defined aggregates of lymphocytes are found in connective tissues and virtually all organs of the body.

194 Unit Three: Alterations in the Hematologic System



■ FIGURE 11-3 ■ Major maturational stages of blood cells. CFU, colony-forming units.

KEY CONCEPTS

HEMATOPOIESIS

- White blood cells are formed partially in the bone marrow (granulocytes, monocytes, and some lymphocytes) and partially in the lymph system (lymphocytes and plasma cells).
- They are formed from hematopoietic stem cells that differentiate into committed progenitor cells that in turn develop into the myelogenous and lymphocytic lineages needed for the formation of the different types of white blood cell.
- The growth and reproduction of the different stem cells is controlled by CSFs and other cytokines and chemical mediators.
- The life span of white blood cells is relatively short so that constant renewal is necessary to maintain normal blood levels. Any conditions that decrease the availability of stem cells or hematopoietic growth factors produce a decrease in white blood cells.

In summary, the hematopoietic system consists of a number of cells derived from the pluripotent stem cells originating in the bone marrow. These cells differentiate into committed cell lines that mature in red blood cells, platelets, and a variety of white blood cells. The development of the different types of blood cells is supported by chemical messengers called colony-stimulating factors. The lymphoid tissues are found in the central lymphoid structures (bone marrow and thymus) where lymphocytes arise, mature, and where self-reactive lymphocytes are eliminated and the peripheral lymphoid structures (lymph nodes, spleen, mucosalassociated lymphoid tissue, and the cutaneous immune system) where lymphocytes respond to foreign antigens.

NON-NEOPLASTIC DISORDERS OF WHITE BLOOD CELLS

The number of leukocytes, or white blood cells, in the peripheral circulation normally ranges from 5000 to $10,000/\mu$ L of blood. The term *leukopenia* describes an absolute decrease in white blood cell numbers. The disorder may affect any of the

specific types of white blood cells, but most often it affects the neutrophils, which are the predominant type of granulocyte.

Neutropenia (Agranulocytosis)

Neutropenia refers specifically to a decrease in neutrophils. It commonly is defined as a circulating neutrophil count of less than 1500 cells/µL. Agranulocytosis, which denotes a severe neutropenia, is characterized by a circulating neutrophil count of less than 200 cells/µL.4

The reduction in granulocytes can occur because there is reduced or ineffective production of neutrophils or because there is excessive removal of neutrophils from the blood. The causes of neutropenia are summarized in Table 11-1.

Acquired Neutropenia

Granulopoiesis may be impaired by a variety of bone marrow disorders, such as aplastic anemia or bone marrow depression caused by cancer chemotherapy and irradiation, that interfere with the formation of all blood cells. Overgrowth of neoplastic cells in cases of nonmyelogenous leukemia and lymphoma also may suppress the function of neutrophil precursors. Infections by viruses or bacteria may drain neutrophils from the blood faster than they can be replaced, thereby depleting the neutrophil storage pool in the bone marrow.⁴ Because of the neutrophil's short life span of less than 1 day in the peripheral blood, neutropenia occurs rapidly when granulopoiesis is impaired. Under these conditions, neutropenia usually is accompanied by thrombocytopenia (*i.e.*, platelet deficiency).

In aplastic anemia, all of the myeloid stem cells are affected, resulting in anemia, thrombocytopenia, and agranulocytosis. Autoimmune disorders or idiosyncratic drug reactions may cause increased and premature destruction of neutrophils. In splenomegaly, neutrophils may be trapped in the spleen along with other blood cells. In Felty's syndrome, a variant of rheumatoid arthritis, there is increased destruction of neutrophils in the spleen.

Most cases of neutropenia are drug related. Chemotherapeutic drugs used in the treatment of cancer (e.g., alkylating agents, antimetabolites) cause predictable dose-dependent suppression of bone marrow function. The term idiosyncratic is used to describe drug reactions that are different from the effects obtained in most persons and that cannot be explained in terms of allergy. A number of drugs, such as chloramphenicol (an antibiotic), phenothiazine tranquilizers, sulfonamides, propylthiouracil (used in the treatment of hyperthyroidism), and phenylbutazone (used in the treatment of arthritis), may cause idiosyncratic depression of bone marrow function. Some drugs, such as hydantoin derivatives and primidone (used in the treatment of seizure disorders), can cause intramedullary destruction of granulocytes and thereby impair production. Many idiosyncratic cases of drug-induced neutropenia are thought to be caused by immunologic mechanisms, with the drug or its metabolites acting as antigens (i.e., haptens) to incite the production of antibodies reactive against the neutrophils. Neutrophils possess human leukocyte antigens (HLA) and other antigens specific to a given leukocyte line. Antibodies to these specific antigens have been identified in some cases of drug-induced neutropenia.4

8 **Congenital Neutropenia**

A decreased production of granulocytes is a feature of a group of hereditary hematologic disorders, including cyclic neutropenia and Kostmann's syndrome. Periodic or cyclic neutropenia is an autosomal dominant disorder with variable expression that begins in infancy and persists for decades. It is characterized by periodic neutropenia that develops every 21 to 30 days and lasts approximately 3 to 6 days. Although the cause is undetermined, it is thought to result from impaired feedback regulation of granulocyte production and release. Kostmann's syndrome, which occurs sporadically or as an autosomal recessive disorder, causes severe neutropenia while preserving the erythroid and megakaryocyte cell lineages that result in red blood cell and platelet production. The total white blood cell count may be within normal limits, but the neutrophil count is less than 200/µL. Monocyte and eosinophil levels may be elevated.

A transient neutropenia may occur in neonates whose mothers have hypertension. It usually lasts from 1 to 60 hours

TABLE 11-1 Causes of Neutropenia	
Cause	Mechanism
Accelerated removal (<i>e.g.,</i> inflammation and infection) Drug-induced granulocytopenia Defective production	Removal of neutrophils from the circulation exceeds production
Cytotoxic drugs used in cancer therapy Phenothiazine, thiouracil, chloramphenicol, phenylbutazone, and others	Predictable damage to precursor cells, usually dose dependent Idiosyncratic depression of bone marrow function
Hydantoinates, primidone, and others Immune destruction Aminopyrine and others	Intramedullary destruction of granulocytes Immunologic mechanisms with cytolysis or leukoagglutination
Periodic or cyclic neutropenia (occurs during infancy and later)	Unknown
Neoplasms involving bone marrow (e.g., leukemias and lymphomas)	Overgrowth of neoplastic cells, which crowd out granulopoietic precursors
Idiopathic neutropenia that occurs in the absence of other disease or provoking influence	Autoimmune reaction
Felty's syndrome	Intrasplenic destruction of neutrophils

but can persist for 3 to 30 days. This type of neutropenia, which is associated with increased risk of nosocomial infection, is thought to result from transiently reduced neutrophil production.⁴

Clinical Course

The clinical features of neutropenia usually depend on the severity of neutropenia and the cause of the disorder. Because the neutrophil is essential to the cellular phase of inflammation, infections are common in persons with neutropenia, and extreme caution is needed to protect them from exposure to infectious organisms. Infections that may go unnoticed in a person with a normal neutrophil count could prove fatal in a person with neutropenia. These infections commonly are caused by organisms that normally colonize the skin, vagina, and the gastrointestinal tract.

The signs and symptoms of neutropenia initially are those of bacterial or fungal infections. They include malaise, chills, and fever, followed by extreme weakness and fatigue. The white blood cell count often is reduced to $1000/\mu$ L and, in certain cases, may fall to 200 to $300/\mu$ L. The most frequent site of serious infection is the respiratory tract, a result of bacteria, fungi, and protozoa that frequently colonize the airways. Ulcerative necrotizing lesions of the mouth are common in neutropenia. Ulcerations of the skin, vagina, and gastrointestinal tract also may occur.⁴

Antibiotics are used to treat infections in those situations in which neutrophil destruction can be controlled or the neutropoietic function of the bone marrow can be recovered. Hematopoietic growth factors such as GM-CSF are being used more commonly to stimulate the maturation and differentiation of the polymorphonuclear cell lineage. Treatment with these biologic response modifiers has reduced the period of neutropenia and the risk for development of potentially fatal septicemia.^{4,5}

Infectious Mononucleosis

Infectious mononucleosis is a self-limiting lymphoproliferative disorder caused by the Epstein-Barr virus (EBV), a member of the herpesvirus family.^{6,7} EBV is commonly present in all human populations. Infectious mononucleosis is most prevalent in adolescents and young adults in the upper socioeconomic classes in developed countries. This is probably because the disease, which is relatively asymptomatic when it occurs during childhood, confers complete immunity to the virus. In families from upper socioeconomic classes, exposure to the virus may be delayed until late adolescence or early adulthood. In such persons, the mode of infection, size of the viral pool, and physiologic and immunologic condition of the host may determine whether the infection occurs.

Pathogenesis

Infectious mononucleosis is largely transmitted through oral contact with EBV-contaminated saliva. The virus initially penetrates the nasopharyngeal, oropharyngeal, and salivary epithelial cells. It then spreads to the underlying oropharyngeal lymphoid tissue, and more specifically, to B lymphocytes, all of which have receptors for EBV. Infection of the B cells may take one of two forms—it may kill the infected B cell or it may become incorporated into its genome. A small number of infected B cells are killed and in the process release the virions. However, in most cells the virus associates with the B-cell genome. The B cells that harbor the EBV genome proliferate in the circulation and produce the well known *heterophil* antibodies that are used for the diagnosis of infectious mono-nucleosis.⁷ A heterophil antibody is an immunoglobulin that reacts with antigens from another species—in this case, sheep red blood cells.

The normal immune response is important in controlling the proliferation of the EBV-infected B cells and cell-free virus. Most important in controlling the proliferation of EBV-infected B cells are the cytotoxic CD8⁺ T cells and natural killer (NK) cells. The virus-specific T cells appear as large atypical lymphocytes that are characteristic of the infection.⁷ In otherwise healthy persons, the humoral and cellular immune responses control viral shedding by limiting the number of infected B cells, rather than eliminating them.

Although infectious B cells and free virions disappear from the blood after recovery from the disease, the virus remains in a few transformed B cells in the oropharyngeal region and is shed in the saliva. Once infected with the virus, persons have asymptomatic infection for life, and a few intermittently shed EBV. Immunosuppressed persons shed the virus more frequently. Asymptomatic shedding of EBV by healthy persons accounts for most of the spread of infectious mononucleosis, despite the fact that it is not a highly contagious disease.

Clinical Course

The onset of infectious mononucleosis usually is insidious. The incubation period lasts 4 to 8 weeks.⁸ A prodromal period, which lasts for several days, follows and is characterized by malaise, anorexia, and chills. The prodromal period precedes the onset of fever, pharyngitis, and lymphadenopathy. Occasionally, the disorder comes on abruptly with a high fever. Most persons seek medical attention for severe pharyngitis, which usually is most severe on days 5 to 7 and persists for 7 to 14 days. Rarely severe toxic pharyngotonsillitis may cause airway obstruction.

The lymph nodes are typically enlarged throughout the body, particularly in the cervical, axillary, and groin areas. Hepatitis and splenomegaly are common manifestations of the disease and are thought to be immune mediated. Hepatitis is characterized by hepatomegaly, nausea, anorexia, and jaundice. Although discomforting, it usually is a benign condition that resolves without causing permanent liver damage. The spleen may be enlarged two to three times its normal size, and rupture of the spleen is an infrequent complication. A rash that resembles rubella develops in 10% to 15% of cases.⁷ In less than 1% of cases, mostly in the adult age group, complications of the central nervous system (CNS) develop. These complications include cranial nerve palsies, encephalitis, meningitis, transverse myelitis, and Guillain-Barré syndrome.

The peripheral blood usually shows an increase in the number of leukocytes, with a white blood cell count between 12,000 and 18,000/ μ L, 95% of which are lymphocytes.⁷ The rise in white blood cells begins during the first week and continues during the second week of the infection; the white blood cell count returns to normal around the fourth week. Although leukocytosis is common, leukopenia may be seen in some persons during the first 3 days of the illness. Atypical lymphocytes are common, constituting more than 20% of the total lymphocyte count. Heterophil antibodies usually appear during the second or third week and decline after the acute illness has subsided. However, they may be detectable for as long as 9 months after onset of the disease.

Most persons with infectious mononucleosis recover without incident. The acute phase of the illness usually lasts 2 to 3 weeks, after which recovery occurs rapidly. Some degree of debility and lethargy may persist for 2 to 3 months. Treatment is primarily symptomatic and supportive. It includes bed rest and analgesics such as aspirin to relieve the fever, headache, and sore throat.8

In summary, neutropenia, a marked reduction in the number of circulating neutrophils, is one of the major disorders of the white blood cells. It can be acquired or congenital and can result from a combination of mechanisms. Severe neutropenia can occur as a complication of lymphoproliferative diseases, in which neoplastic cells crowd out neutrophil precursor cells, or of radiation therapy or treatment with cytotoxic drugs, which destroy neutrophil precursor cells. Neutropenia also may be encountered as an idiosyncratic reaction to various drugs. Because the neutrophil is essential to the cellular stage of inflammation, severe and often lifethreatening infections are common in persons with neutropenia.

Infectious mononucleosis is a self-limited lymphoproliferative disorder caused by EBV, a member of the herpesvirus family. The highest incidence of infectious mononucleosis is found in adolescents and young adults, and it is seen more frequently in the upper socioeconomic classes of developed countries. The virus is usually transmitted in the saliva. The disease is characterized by fever, generalized lymphadenopathy, sore throat, and the appearance in the blood of atypical lymphocytes and several antibodies, including the well-known heterophil antibodies that are used in the diagnosis of infectious mononucleosis. Most persons with infectious mononucleosis recover without incident. Treatment is largely symptomatic and supportive.

NEOPLASTIC DISORDERS OF HEMATOPOIETIC AND LYMPHOID ORIGIN

The neoplastic disorders of hematopoietic and lymphoid origin represent the most important of the white blood cell disorders. They include three somewhat overlapping categories: the lymphomas (Hodgkin's disease and non-Hodgkin's lymphoma), the leukemias, and plasma cell dyscrasias (multiple myeloma).

Malignant Lymphomas

The lymphomas, Hodgkin's disease and non-Hodgkin's lymphoma, represent solid tumors derived from neoplastic lymphoid tissue cells (i.e., lymphocytes or histiocytes) and their precursors or derivatives. The seventh most common cancer in the United States, the lymphomas are among the most studied human tumors and among the most curable.

KEY CONCEPTS

MALIGNANT LYMPHOMAS

- The lymphoma represent malignancies of cells derived from lymphoid cells and tissues.
- Hodgkin's disease is a group of cancers characterized by Reed-Sternberg cells that begins as a malignancy in a single lymph node and then spreads to contiguous lymph nodes.
- Non-Hodgkin's lymphomas represent a group of heterogeneous lymphocytic cancers that are multicentric in origin and spread to various tissues throughout the body, including the bone marrow.
- Both types of lymphomas are characterized by manifestations related to uncontrolled lymph node and lymphoid tissue growth, bone marrow involvement, and constitutional symptoms (fever, fatigue, weight loss) related to the rapid growth of abnormal lymphoid cells and tissues.

Hodgkin's Disease

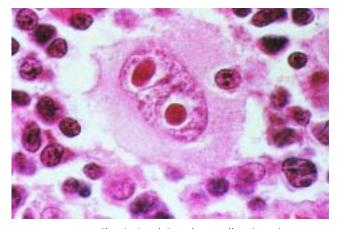
Hodgkin's disease is a specialized form of lymphoma that features the presence of an abnormal cell called a Reed-Sternberg cell.9 It was estimated that approximately 7000 new cases of Hodgkin's disease would be diagnosed in 2002, with 1400 deaths.¹⁰ Distribution of the disease is bimodal; the incidence rises sharply after 10 years of age, peaks in the early 20s, and then declines until 50 years of age. After 50 years of age, the incidence again increases steadily with age. The younger adult group consists equally of men and women, but after age 50 years, the incidence is higher among men.¹¹

The cause of Hodgkin's disease is unknown. Although exposure to carcinogens or viruses as well as genetic and immune mechanisms have been proposed as causes, none have been proven to be involved in the pathogenesis of Hodgkin's disease.

Hodgkin's disease is characterized by painless and progressive enlargement of a single node or group of nodes. It is believed to originate in one area of the lymphatic system, and if unchecked, it spreads throughout the lymphatic network. The initial lymph node involvement typically is above the level of the diaphragm. An exception is in elderly persons, in whom the subdiaphragmatic lymph nodes may be the first to be involved. Involvement of the retroperitoneal lymph nodes, liver, spleen, and bone marrow occurs after the disease becomes generalized.

A distinctive tumor cell (the Reed-Sternberg cell) is considered to be the true neoplastic element in Hodgkin's disease (Fig. 11-4).^{9,12} These malignant proliferating cells may invade almost any area of the body and may produce a wide variety of signs and symptoms. The spleen is involved in one third of the cases at the time of diagnosis.12

Manifestations. A common finding in Hodgkin's disease is the presence of painless lymph node enlargement, involving a single lymph node or groups of lymph nodes. The cervical



■ FIGURE 11-4 ■ Classic Reed-Sternberg cell. Mirror-image nuclei contain large eosinophilic nucleoli. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 1144]. Philadelphia: Lippincott Williams & Wilkins)

and mediastinal nodes are involved most frequently. Less commonly, the axillary, inguinal and retroperitoneal nodes are initially involved.¹²

Persons with Hodgkin's disease are commonly designated as stage A if they lack constitutional symptoms and stage B if significant weight loss, fever, or night sweats are present. Approximately 40% of persons with Hodgkin's disease exhibit the "B" symptoms.¹² Other symptoms such as fatigue, pruritus, and anemia are indicative of disease spread. In the advanced stages of Hodgkin's disease, the liver, lungs, digestive tract, and, occasionally, the CNS may be involved. As the disease progresses, the rapid proliferation of abnormal lymphocytes leads to an immunologic defect, particularly in cell-mediated responses, rendering the person more susceptible to viral, fungal, and protozoal infections. Anergy, or the failure to develop a positive response to skin tests, such as the tuberculin test, is common early in the course of the disease. An increased neutrophil count and mild anemia are often noted.

Diagnosis and Treatment. A definitive diagnosis of Hodgkin's disease requires that the Reed-Sternberg cell be present in a biopsy specimen of lymph node tissue.^{12,13} Computed tomographic (CT) scans of the abdomen commonly are used in screening for involvement of abdominal and pelvic lymph nodes. Radiologic visualization of the abdominal and pelvic lymph structures can be achieved through the use of bipedal lymphangiography. In this diagnostic test, radiopaque dye is injected into the lymphatic channels of the lower leg, enabling visualization of the iliac and para-aortic nodes. Nuclear studies, such as a gallium scan in which the tumor takes up the radionuclide, or a staging laparotomy to detect abdominal nodes and inspect the liver may be done.

The staging of Hodgkin's disease is of great clinical importance because the choice of treatment and the prognosis ultimately are related to the distribution of the disease. Staging is determined by the number of lymph nodes that are involved, whether the lymph nodes are on one or both sides of the diaphragm, and whether there is disseminated disease involving the bone marrow and liver. Irradiation and chemotherapy are used in treating the disease. Most people with localized disease are treated with radiation therapy. As the accuracy of staging techniques, delivery of radiation, and curative efficacy of combination chemotherapy regimens have improved, the survival of people with Hodgkin's disease also has improved. With modern treatment methods an overall cure rate of 70% can be achieved.

Non-Hodgkin's Lymphomas

The non-Hodgkin's lymphomas are a heterogeneous group of solid tumors composed of neoplastic lymphoid cells. The heterogeneity reflects the potential for malignant transformation at any stage of B- and T-lymphocyte differentiation. The non-Hodgkin's lymphomas occur three times more frequently than does Hodgkin's disease. In 2002, approximately 54,000 new cases were diagnosed in the United States, and approximately 24,400 deaths resulted from these disorders.¹⁰ Non-Hodgkin's lymphomas are the fifth most common malignancy in men and the sixth in women,¹³ and the disease's incidence and associated mortality rates have increased considerably during the last several decades.¹⁴

The etiology of most of the non-Hodgkin's lymphomas is unknown. A viral cause is suspected in at least some of the lymphomas. There is evidence of EBV infection in 95% of people with Burkitt's lymphoma, which is endemic to some parts of Africa.9,12 By contrast, the virus is found in only a small percentage of Burkitt's lymphoma occurring in the United States and other nonendemic areas.¹² A second virus, the human T cell/lymphoma virus (HTLV-1), which is endemic in the southwestern islands of Japan, has been associated with adult T-cell leukemia/lymphoma. Evidence of infection has been demonstrated in 90% of adult T-cell leukemia/lymphoma cases in Japan.9,12 Non-Hodgkin's lymphomas also are seen with increased frequency in persons with acquired immunodeficiency syndrome, in those who have received chronic immunosuppressive therapy after kidney or liver transplantation, and in individuals with acquired or congenital immunodeficiencies.

As neoplasms of the immune system, the non-Hodgkin's lymphomas can originate in either the T cells or B cells.^{9,12} Most (80% to 85%) are of B-cell origin with the remainder being largely of T-cell origin. All of these variants have the potential to spread to various tissues throughout the body, especially the liver, spleen, and bone marrow. Non-Hodgkin's lymphomas commonly are divided into three groups, depending on the grade of the tumor: low-grade lymphomas, which are predominantly B-cell tumors; intermediate-grade lymphomas; and high-grade lymphomas, which are largely immunoblastic (B-cell), lymphoblastic (T-cell), Burkitt's, and non-Burkitt's lymphomas.¹²

Clinical Course. The clinical course of non-Hodgkin's lymphomas depends upon the lymphoma type and the stage of the disease. For example, the small lymphocyte tumors, which account for about 4% of all non-Hodgkin's lymphomas, are low-grade tumors associated with mild symptoms and prolonged survival. The diffuse large B-cell lymphomas, which account for about 20% of all non-Hodgkin's lymphomas, are particularly aggressive tumors that are rapidly fatal if not treated. However, with intensive combination chemotherapy, complete remission can be achieved in 60% to 80% of cases.

The most frequently occurring clinical manifestation in the non-Hodgkin's lymphomas is painless, superficial lymphadenopathy. There may be noncontiguous nodal spread of the disease with more frequent involvement of the gastrointestinal tract, liver, testes, and bone marrow. Frequently, there is increased susceptibility to bacterial, viral, and fungal infections associated with hypogammaglobulinemia and a poor humoral antibody response, rather than the impaired cellular immunity seen with Hodgkin's disease.

Leukemic transformation with high peripheral lymphocytic counts occurs in a small percentage of persons with non-Hodgkin's lymphoma.9

Diagnosis and Treatment. As with Hodgkin's disease, a lymph node biopsy is used to confirm the diagnosis of non-Hodgkin's lymphoma. Bone marrow biopsy, blood studies, abdominal computed tomographic scans, and nuclear medicine studies often are used to determine the stage of the disease.

For early-stage disease, radiation therapy is used as a single treatment. However, because most people present with latestage disease, combination chemotherapy, combined adjuvant radiation therapy, or both are recommended. For rapidly progressive intermediate- or high-grade lymphomas, CNS prophylaxis is achieved with high doses of chemotherapeutic agents that can cross the blood-brain barrier or cranial irradiation. Bone marrow and peripheral stem cell transplantation are being investigated as potentially curative treatment modalities for people with highly resistant forms of the disease.

Leukemias

The leukemias are malignant neoplasms of cells originally derived from hematopoietic stem cells. They are characterized by diffuse replacement of bone marrow with unregulated, proliferating, immature neoplastic cells. In most cases, the leukemic cells spill out into the blood, where they are seen in large numbers. The term leukemia (i.e., "white blood") was first used by Virchow to describe a reversal of the usual ratio of red blood cells to white blood cells. The leukemic cells may also infiltrate the liver, spleen, lymph nodes, and other tissues throughout the body, causing enlargement of these organs.

Leukemia strikes approximately 31,000 persons in the United States each year. In 2002, approximately 30,800 new cases were diagnosed, and approximately 21,700 persons died of this disease.¹⁰ More children are stricken with leukemia than with any other form of cancer, and it is the leading cause of death in children between the ages of 3 and 14 years. Although leukemia commonly is thought of as a childhood disease, it strikes more adults than children.

The causes of leukemia are unknown. The incidence of leukemia among persons who have been exposed to high levels of radiation is unusually high. The number of cases of leukemia reported in the most heavily exposed survivors of the atomic blasts at Hiroshima and Nagasaki during the 20-year period from 1950 to 1970 was nearly 30 times the expected rate.15 An increased incidence of leukemia also is associated with exposure to benzene and the use of antitumor drugs (*i.e.*, mechlorethamine, procarbazine, cyclophosphamide, chloramphenicol, and the epipodophyllotoxins).¹⁶ Leukemia may occur as a second cancer after aggressive chemotherapy for other cancers, such as Hodgkin's disease.¹⁷ The existence of a

KEY CONCEPTS

LEUKEMIAS

- Leukemias are malignant neoplasms arising from the transformation of a single blood cell line derived from hematopoietic stem cells.
- The leukemias are classified as acute and chronic lymphocytic (lymphocytes) or myelogenous (granulocytes, monocytes) leukemias, according to their cell lineage.
- Because leukemic cells are immature and poorly differentiated, they proliferate rapidly and have a long life span; they do not function normally; they interfere with the maturation of normal blood cells; and they circulate in the bloodstream, cross the bloodbrain barrier, and infiltrate many body organs.

genetic predisposition to acute leukemia is suggested by the increased leukemia incidence among a number of congenital disorders, including Down syndrome, von Recklinghausen's disease, and Fanconi's anemia. In individuals with Down syndrome, the incidence of acute leukemia is 10 times that of the general population.¹⁸ In addition, there are numerous reports of multiple cases of acute leukemia occurring within the same family.

Classification

The leukemias commonly are classified according to their predominant cell type (i.e., lymphocytic or myelogenous) and whether the condition is acute or chronic. Biphenotypic leukemias demonstrate characteristics of both lymphoid and myeloid lineages. A rudimentary classification system divides leukemia into four types: acute lymphocytic (lymphoblastic) leukemia, chronic lymphocytic leukemia, acute myelogenous (myeloblastic) leukemia, and chronic myelogenous leukemia. The lymphocytic leukemias involve immature lymphocytes and their progenitors that originate in the bone marrow but infiltrate the spleen, lymph nodes, CNS, and other tissues. The myelogenous leukemias, which involve the pluripotent myeloid stem cells in bone marrow, interfere with the maturation of all blood cells, including the granulocytes, erythrocytes, and thrombocytes.

Acute Leukemias

The acute leukemias usually have a sudden and stormy onset of signs and symptoms related to depressed bone marrow function (Table 11-2). Acute lymphocytic leukemia (ALL) is the most common leukemia in childhood, comprising 80% to 85% of leukemia cases.¹⁹ The peak incidence occurs between 2 and 4 years of age. Acute myelogenous leukemia (AML) is chiefly an adult disease; although it is also seen in children and young adults. The incidence steadily increases after middle age. AML appears to be increasing among the elderly.¹²

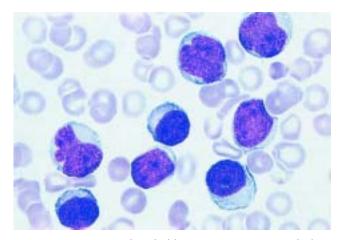
ALL encompasses a group of neoplasms composed of immature precursor B or T lymphocytes (Fig. 11-5). Most cases (about 85%) of ALL are of pre-B-cell origin.9 Approximately 90% of persons with ALL have nonrandom chromosome ab-

Clinical Manifestations	Pathologic Basis	
Bone marrow depression		
Malaise, easy fatigability	Anemia	
Fever	Infection or increased metabolism by neoplastic cells	
Bleeding	Decreased thrombocytes	
Petechiae		
Ecchymosis		
Gingival bleeding		
Epistaxis		
Bone pain and tenderness upon palpation	Subperiosteal bone infiltration, bone marrow expansion, and bone resorption	
Headache, nausea, vomiting, papilledema,	Leukemic infiltration of central nervous system	
cranial nerve palsies, seizures, coma		
Abdominal discomfort	Generalized lymphadenopathy, hepatomegaly, splenomegaly due to leukemic cell infiltration	
Increased vulnerability to infections	Immaturity of the white cells and ineffective immune function	
Hematologic abnormalities	Physical and metabolic encroachment of leukemia cells on red blood cell and thrombo	
Anemia	cyte precursors	
Thrombocytopenia		
Hyperuricemia and other metabolic disorders	Abnormal proliferation and metabolism of leukemic cells	

*Manifestations vary with the type of leukemia.

normalities. The AMLs are an extremely heterogeneous group of disorders. Some arise from the pluripotent stem cells in which myeloblasts predominate, and others arise from the monocyte-granulocyte precursor, which is the cell of origin for myelomonocytic leukemia. Of all the leukemias, AML is most strongly linked with toxins and underlying congenital and hematologic disorders. It is the type of leukemia associated with Down syndrome.

Clinical Manifestations. Although ALL and AML are distinct disorders, they typically present with similar clinical features. The warning signs and symptoms of acute leukemia are fatigue,



■ FIGURE 11-5 ■ Acute lymphoblastic anemia (L2 ALL). The lymphoblasts in the peripheral blood contain irregular and indented nuclei with prominent nucleoli and a moderate amount of cytoplasm. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 1129]. Philadelphia: Lippincott-Raven)

pallor, weight loss, repeated infections, easy bruising, nosebleeds, and other types of hemorrhage. These features often appear suddenly in children.

Persons with acute leukemia usually present for medical evaluation within 3 months of the onset of symptoms. Both ALL and AML are characterized by fatigue resulting from anemia; low-grade fever, night sweats, and weight loss caused by the rapid proliferation and hypermetabolism of the leukemic cells; bleeding caused by a decreased platelet count; and bone pain and tenderness caused by bone marrow expansion.^{19,20} Infection results from neutropenia, with the risk of infection becoming high as the neutrophil count falls to less than 500 cells/µL. Generalized lymphadenopathy, splenomegaly, and hepatomegaly caused by infiltration of leukemic cells occur in all acute leukemias but are more common in ALL. In addition to the common manifestations of acute leukemia (e.g., fatigue, weight loss, fever, easy bruising), infiltration of malignant cells in the skin, gums, and other soft tissue is particularly common in the monocytic form of AML.

CNS involvement is more common in ALL than AML, and is more common in children than adults. Signs and symptoms of CNS involvement include cranial nerve palsies, headache, nausea, vomiting, papilledema, and occasionally seizures and coma.

Leukostasis is a condition in which the circulating blast count is markedly elevated (usually 100,000 cells/ μ L). The high number of circulating leukemic blasts increases blood viscosity and predisposes to the development of leukoblastic emboli with obstruction of small vessels in the pulmonary and cerebral circulations. Plugging of the pulmonary vessels leads to vessel rupture and infiltration of lung tissue, resulting in sudden shortness of breath and progressive dyspnea. Cerebral leukostasis leads to diffuse headache and lethargy, which can progress to confusion and coma. Once identified, leukostasis requires immediate and effective treatment to lower the blast count rapidly. Diagnosis and Treatment. A definitive diagnosis of acute leukemia is based on blood and bone marrow studies; it requires the demonstration of leukemic cells in the peripheral blood, bone marrow, or extramedullary tissue. Laboratory findings reveal the presence of immature (blasts) white blood cells in the circulation and bone marrow, where they may constitute 60% to 100% of the cells. As these cells proliferate and begin to crowd the bone marrow, the development of other blood cell lines in the marrow is suppressed. Consequently, there is a loss of mature myeloid cells, such as erythrocytes, granulocytes, and platelets. Anemia is almost always present, and the platelet count is decreased.

Chemotherapy and selective irradiation (e.g., CNS irradiation) are used in the treatment of acute leukemia. Remission is defined as eradication of leukemic cells as detectable by conventional technology. Chemotherapy includes induction therapy designed to elicit a remission, intensification therapy to produce a further reduction in leukemic cells after a remission is achieved, and maintenance therapy to maintain the remission. Because systemic chemotherapeutic agents cannot cross the blood-brain barrier and eradicate leukemic cells that have entered the CNS, CNS irradiation is administered concurrent with systemic chemotherapy.^{19,20} The long-term effects of treatment on childhood cancer survivors is discussed in Chapter 5.

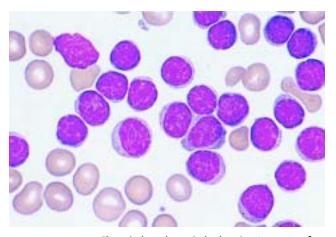
Massive destruction of malignant cells can occur during the initial phase of treatment. This phenomenon, known as tumor lysis syndrome, can lead to life-threatening metabolic disorders, including hyperkalemia, hyperphosphatemia, hyperuricemia, hypomagnesemia, hypocalcemia, and acidosis, with the potential for causing acute renal failure. Prophylactic aggressive hydration with alkaline solutions and administration of allopurinol to reduce uric acid levels is given to counteract these effects.

Bone marrow transplantation may be considered for persons with ALL and AML who experience no response to other forms of therapy. There has been recent interest in the use of stem cell transplantation in ALL²¹ and AML.²²

Chronic Leukemias

Chronic leukemias have a more insidious onset than do acute leukemias and may be discovered during a routine medical examination by a blood count. Chronic lymphocytic leukemia (CLL) is mainly a disorder of older persons; fewer than 10% of those who have the disease are younger than 50 years. Men are affected twice as frequently as women. Chronic myelogenous leukemia (CML) accounts for 15% to 20% of all leukemias in adults. It is predominantly a disorder of adults between the ages of 30 and 50 years, but it can affect children as well. The incidence is slightly higher in men than women.

CLL is a disorder characterized by the proliferation and accumulation of relatively mature lymphocytes that are immunologically incompetent (Fig. 11-6). In the United States,



■ FIGURE 11-6 ■ Chronic lymphocytic leukemia. A smear of peripheral blood shows numerous small-to-medium-sized lymphocytes. (Rubin E., Farber J.L. [1999]. Pathology [3rd ed., p. 1125]. Philadelphia: Lippincott Williams & Wilkins)

more than 95% of cases of CLL are of B-cell origin. The leukemic B cells fail to respond to antigenic stimulation; thus, persons with CLL have hypogammaglobulinemia. Infections remain a major cause of morbidity and mortality.

CML is a myeloproliferative disorder that involves expansion of all bone marrow elements. CML is associated in all cases with the presence of the Ph (Philadelphia) chromosome, representing a reciprocal translocation of the long arm of chromosome 22 to the long arm of chromosome 9.9,23 In about 95% of persons with CML, the Ph chromosome can be identified in granulocytic, erythroid, and megakaryocytic precursors, as well as B cells, and in some cases, T cells.9 Although CML originates in the pluripotent stem cells, granulocyte precursors remain the dominant cell type.

Clinical Course. Both CLL and CML have an insidious onset. However, the two types of chronic leukemias differ in their manifestations and clinical course.

CLL typically follows a slow, chronic course. The clinical signs and symptoms are largely related to the progressive infiltration of neoplastic lymphocytes in the bone marrow and extramedullary tissue and to secondary immunologic defects. Affected persons often have no symptoms at the time of diagnosis, and lymphocytosis is noted on a complete blood count obtained for another, unrelated disorder. Fatigue, reduced exercise tolerance, enlargement of superficial lymph nodes, or splenomegaly usually reflect a more advanced stage. As the disease progresses, lymph nodes gradually increase in size and new nodes are involved, sometimes in unusual areas such as the scalp, orbit, pharynx, pleura, gastrointestinal tract, liver, prostate, and gonads. Severe fatigue, recurrent or persistent infections, pallor, edema, thrombophlebitis, and pain are also experienced. As the malignant cell population increases, the proportion of normal marrow precursors is reduced until only lymphocytes remain in the marrow.²⁰

Typically CML follows a triphasic course: a chronic phase of variable length, a short accelerated phase, and a terminal blast crisis phase. The onset of the chronic phase is usually slow with nonspecific symptoms such as weakness and weight loss. The most characteristic laboratory finding at the time of

202 Unit Three: Alterations in the Hematologic System

presentation is leukocytosis with immature granulocyte cell types in the peripheral blood. Anemia and, eventually, thrombocytopenia develop. Anemia causes weakness, easy fatigability, and exertional dyspnea. Splenomegaly is often present at the time of diagnosis; hepatomegaly is less common; and lymphadenopathy is relatively uncommon. Although persons in the early chronic phase of CML generally have no symptoms, without effective treatment most will enter the accelerated phase within 3 to 5 years.

The accelerated phase is characterized by enlargement of the spleen and progressive symptoms. Splenomegaly often causes a feeling of abdominal fullness and discomfort. An increase in basophil count and more immature cells in the blood or bone marrow confirm transformation to the accelerated phase. During this phase, constitutional symptoms such as low-grade fever, night sweats, bone pain, and weight loss develop because of rapid proliferation and hypermetabolism of the leukemic cells. Bleeding and easy bruising may arise from dysfunctional platelets. Generally the accelerated phase is short (6 to 12 months).¹⁸

The terminal or blast crisis phase represents evolution to acute leukemia and is characterized by an increasing number of myeloid precursors, especially blast cells. Constitutional symptoms become more pronounced during this period, and splenomegaly may increase significantly. Isolated infiltrates of leukemic cells can involve the skin, lymph nodes, bones, and CNS. With very high blast counts (100,000/ μ L), symptoms of leukostasis may occur. The prognosis for patients who are in the blast crisis phase is poor, with survival rates averaging 2 to 4 months.

Diagnosis and Treatment. The diagnosis of chronic leukemia is based on blood and bone marrow studies. The treatment varies with the type of leukemic cell, the stage of the disease, other health problems, and the person's age. Most early cases of CLL require no specific treatment. Reassurance that persons with the disorder can live a normal life for many years is important. Indications for chemotherapy include progressive fatigue, troublesome lymphadenopathy, anemia, and thrombocytopenia. Complications such as autoimmune hemolytic anemia or thrombocytopenia may require treatment with corticosteroids or splenectomy. Unlike CML, transformation to acute leukemia with blast crisis is rare, and many persons live longer than 10 years and die of unrelated causes.

The treatment of CML is often palliative. The median survival is 5 to 7 years, with fewer than 50% of persons alive at 5 years.¹⁸ Standard treatment includes the use of single-agent chemotherapy to control the disease in persons in the chronic phase of the disease. Interferon- α induces clinical remission in 70% to 80% of persons treated during the early chronic phase of the disease. During the blast crisis phase, combination therapy is administered, although response rates are low (20% to 30%) and remissions vary from 2 to 12 months. Allogeneic bone marrow or stem cell transplantation provides the only cure for CML.

Multiple Myeloma

Multiple myeloma is a plasma cell cancer of the osseous tissue and accounts for 10% to 15% of all hematologic malignancies.²⁴ In the course of its dissemination, it also may involve nonosseous sites. It is characterized by the uncontrolled proliferation of an abnormal clone of plasma cells, which secrete primarily IgG or IgA. Fewer than 3% of cases occur before the age of 40 years, with the median age of patients with multiple myeloma being 65 years.

The cause of multiple myeloma is unknown. It does not appear to be caused by previous exposure to toxic agents (*e.g.*, solvents such as benzene, paints, pesticides). Interestingly, an association with human herpesvirus 8 has been described, but the role of this virus in the pathogenesis of the disease remains to be established.²⁴

In multiple myeloma, there is an atypical proliferation of one of the immunoglobulins, called the M protein, a monoclonal antibody.²⁵ Although multiple myeloma is characterized by excessive production of monoclonal immunoglobulin, levels of normal immunoglobulins are usually depressed. This contributes a general susceptibility to bacterial infections. Cytokines are important in the pathogenesis of the disorder. The multiple myeloma plasma cell has a surface-membrane receptor for interleukin-6, which is known to be a growth factor for the disorder. Another important growth factor for the myeloma cell is interleukin-1, which has important osteoclast activity.²⁶

The main sites involved in multiple myeloma are the bones and bone marrow. In addition to the abnormal proliferation of marrow plasma cells, there is proliferation and activation of osteoclasts that leads to bone resorption and destruction (Fig. 11-7). This increased bone resorption predisposes the individual to pathologic fractures and hypercalcemia. Paraproteins secreted by the plasma cells may cause a hyperviscosity of body fluids and may break down into amyloid, a proteinaceous substance deposited between cells, causing heart failure and neuropathy.



FIGURE 11-7 Multiple myeloma. Multiple lytic lesions of the vertebrae are present. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 1148]. Philadelphia: Lippincott Williams & Wilkins)

In some forms of multiple myeloma, the plasma cells produce only Bence Jones proteins, abnormal proteins that consist of the light chains of the immunoglobulin molecule. Because of their low molecular weight, Bence Jones proteins are partially excreted in the urine. Many of these abnormal proteins are directly toxic to renal tubular structures, which may lead to tubular destruction and, eventually, to renal failure.

The malignant plasma cells also can form plasmacytomas (plasma cell tumors) in bone and soft tissue sites. The most common site of soft tissue plasmacytomas is the gastrointestinal tract. The development of plasmacytomas in bone tissue is associated with bone destruction and localized pain. Occasionally, the lesions may affect the spinal column, causing vertebral collapse and spinal cord compression.²⁴

Manifestations. Bone pain is one of the first symptoms to occur and one of the most common, occurring in approximately 80% of all individuals with diagnoses of multiple myeloma. Bone destruction also impairs the production of erythrocytes and leukocytes and predisposes the patient to anemia and recurrent infections. Many patients experience weight loss and weakness. Renal insufficiency occurs in 50% of patients. Neurologic manifestations caused by neuropathy or spinal cord compression also may be present.

Diagnosis and Treatment. Diagnosis is based on clinical manifestations, blood tests, and bone marrow examination. The hallmark of myeloma is the finding of paraproteins on serum protein electrophoresis. Bone radiographs are important in establishing the presence of bone lesions.

Although numerous treatments for multiple myeloma have been attempted since the early 1970s, the results have been disappointing. With standard chemotherapy, the median survival may reach 2 to 3 years. Multiple myeloma is a radiosensitive disease, but most radiation therapy is used primarily for palliation, specifically to treat lytic bone lesions and compression fractures and to decrease pain. Recently, thalidomide has been shown to induce responses in persons whose myeloma was refractory to conventional therapies.²⁴

In summary, the lymphomas (Hodgkin's disease and non-Hodgkin's lymphoma) represent malignant neoplasms of cells native to lymphoid tissue (*i.e.*, lymphocytes and histiocytes) and their precursors or derivatives. Hodgkin's disease is characterized by painless and progressive enlargement of a single node or group of nodes. It is believed to originate in one area of the lymphatic system and, if unchecked, spreads throughout the lymphatic network. Non-Hodgkin's lymphomas are multicentric in origin and spread early to various tissues throughout the body, especially the liver, spleen, and bone marrow.

Leukemias are malignant neoplasms of the hematopoietic stem cells with diffuse replacement of bone marrow. Leukemias are classified according to cell type (*i.e.*, lymphocytic or myelogenous) and whether the disease is acute or chronic. The lymphocytic leukemias involve immature lymphocytes and their progenitors that originate in the bone marrow but infiltrate the spleen, lymph nodes, CNS, and other tissues. The myelogenous leukemias involve the pluripotent myeloid stem cells in bone marrow and interfere with the maturation of all blood cells, including the granulocytes, erythrocytes, and thrombocytes. The acute leukemias (*i.e.*, ALL, which primarily affects children, and AML, which primarily affects adults) have a sudden and stormy onset, with symptoms of depressed bone marrow function (anemia, fatigue, bleeding, and infections); bone pain; and generalized lymphadenopathy, splenomegaly, and hepatomegaly. The chronic leukemias, which largely affect adults, have a more insidious onset. CLL often has the most favorable clinical course, with many persons living long enough to die of unrelated causes. The course of CML is slow and progressive, with transformation to a course resembling that of AML.

Multiple myeloma results in the uncontrolled proliferation of immunoglobulin-secreting plasma cells, usually a single clone of IgG- or IgA-producing cells, that results in increased bone resorption, leading to pathologic bone lesions.

REVIEW QUESTIONS

■ Trace the development of the different blood cells from their origin in the pluripotent bone marrow stem cell to their circulation in the bloodstream.

Explain the signs and symptoms of neutropenia in terms of the function of the neutrophil.

Explain the manifestations of infectious mononucleosis in terms of the immune systems response to infection by the Epstein-Barr virus.

Contrast the signs and symptoms of Hodgkin's and non-Hodgkin's lymphoma based on the differences in involvement of lymphoid tissues and immune cells.

One of the manifestations of Hodgkin's disease is the inability of persons with tuberculosis to display a positive response to the tuberculin skin test. Explain.

Relate the constitutional symptoms of lymphomas (*e.g.,* fever, general malaise, and fatigue) to the pathophysiology of Hodgkin's disease and non-Hodgkin's lymphoma.

Explain the manifestations of leukemia in terms of altered cell differentiation.

Describe the pathophysiologic basis for the following complications of leukemia: leukostasis, tumor lysis syndrome, hyperuricemia, and blast crisis.

Relate abnormal proliferation of an abnormal plasma cell clone to the manifestations of multiple myeloma.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- 1. Guyton A.C., Hall J.E. (2000). *Textbook of medical physiology* (10th ed., pp. 392–401). Philadelphia: W.B. Saunders.
- 2. Metcalf D. (1999). Cellular hematopoiesis in the twentieth century. *Seminars in Hematology* 36 (Suppl. 7), 5–12.

204 Unit Three: Alterations in the Hematologic System

- Alexander W.S. (1998). Cytokines in hematopoiesis. International Reviews of Immunology 16, 651–682.
- 4. Curnutte J.T., Coates T.D. (2000). Disorders of phagocyte function and number. In Hoffman R., Benz E.K., Shattil S.J., et. al. (Eds.), *Hematology: Basic principles and practice* (3rd ed., pp. 720–762). New York: Churchill Livingstone.
- Boxer L.A. (2000). Leukopenia. In Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.), *Nelson textbook of pediatrics* (16th ed., pp. 624– 625). Philadelphia: W.B. Saunders.
- Sullivan J.L. (2000). Infectious mononucleosis and other Epstein-Barr virus-associated diseases. In Hoffman R., Benz E.K., Shattil S.J., et al. (Eds.), *Hematology: Basic principles and practice* (3rd ed., pp. 812–821). New York: Churchill Livingstone.
- Samuelson J. (1999). Infectious diseases. In Cotran R.S., Kumar V., Collins T. (Eds.), *Pathologic basis of disease* (6th ed., pp. 371–373). Philadelphia: W.B. Saunders.
- Godshall S.E., Krichner J.T. (2000). Infectious mononucleosis: Complexities of a common syndrome. *Postgraduate Medicine* 107, 175–186.
- Aster J., Kumar V., Samuelson J. (1999). White cells, lymph nodes, spleen, and thymus. In Cotran R.S., Kumar V., Collins T. (Eds.), *Pathologic basis of disease* (6th ed., pp. 629–672). Philadelphia: W.B. Saunders.
- 10. American Cancer Society. (2002). *Cancer facts and figures 2002*. Atlanta: American Cancer Society.
- 11. DeVita V.T., Mauch P.M., Harris N.L. (1997). Hodgkin's disease. In DeVita V.T., Hellman S., Rosenberg S.A. (Eds.), *Cancer: Principles and practice of oncology* (5th ed., pp. 2242–2283). Philadelphia: Lippincott-Raven.
- Bonner H., Bagg A., Cossman J. (1999). The blood and lymphoid organs. In Rubin E., Farber J.L. (Eds.), *Pathology* (3rd ed., pp. 1117– 1150). Philadelphia: Lippincott Williams & Wilkins.
- Cheson B.D. (2001). Hodgkin's disease and non-Hodgkin's lymphomas. In Lenbard R.E., Jr., Osteen R.T., Gansler T. (Eds.), *The American Cancer Society's clinical oncology* (pp. 497–516). Atlanta: American Cancer Society.

- 14. Groves F.D., Linet M.S., Travis L.B., et al. (2000). Cancer surveillance series: Non-Hodgkin's lymphoma incidence by histologic subtype in the United States from 1978 through 1995. *Journal of the National Cancer Institute* 92, 1240–1251.
- 15. Jablon S., Kato H. (1972). Studies of the mortality of A-bomb survivors. *Radiation Research* 50, 649–698.
- Scheinberg D.A., Maslak P., Weiss M. (1997). Acute leukemias. In DeVita V.T., Hellman S., Rosenberg S.A. (Eds.), *Cancer: Principles and practice of oncology* (5th ed., pp. 2293–2321). Philadelphia: Lippincott-Raven.
- Kaldor J.M., Day N.E., Clarke E.A., et al. (1990). Leukemia following Hodgkin's disease. *New England Journal of Medicine* 322, 1–6.
- Miller K.B., Grodman H.M. (2001). Leukemia. In Lenbard R.E., Jr., Osteen R.T., Gansler T. (Eds.), *The American Cancer Society's clinical oncology* (pp. 527–551). Atlanta: American Cancer Society.
- Khouri I., Sanchez F.G., Deisseroth A. (1997). Leukemias. In DeVita V.T., Hellman S., Rosenberg S.A. (Eds.), *Cancer: Principles and practice* of oncology (5th ed., pp. 2285–2293). Philadelphia: Lippincott-Raven.
- Callaghan M.E. (1996). Leukemia. In McCorkle R., Grant M., Frank-Stromborg M., et al. (Eds.), *Cancer nursing: A comprehensive textbook* (2nd ed., pp. 752–771). Philadelphia: W.B. Saunders.
- 21. Pui C., Evans W.E. (1998). Acute lymphoblastic leukemia. New England Journal of Medicine 339, 605–615.
- 22. Lowenberg B., Downing J.R., Burnett A. (1999). Acute myeloid leukemia. *New England Journal of Medicine* 341, 1051–1062.
- Thijsen S.F.T., Schuurhuis G.J., van Oostveen J.W., et al. (1999). Chronic myeloid leukemia from basics to bedside. *Leukemia* 13, 1646–1674.
- Rosenthal D.S., Schnipper L.E., McCaffrey R.P., et al. (2001). Multiple myeloma. In Lenbard R.E., Jr., Osteen R.T., Gansler T. (Eds.), *The American Cancer Society's clinical oncology* (pp. 516–525). Atlanta: American Cancer Society.
- Triko G. (2000). Multiple myeloma and other plasma cell disorders. In Hoffman R., Benz E.K., Shattil S.J., et al. (Eds.), *Hematology: Basic principles and practice* (3rd ed., pp. 1398–1416). New York: Churchill Livingstone.
- Bataille R., Harousseau J. (1997). Multiple myeloma. New England Journal of Medicine 336, 1657–1664.

205

Alterations in Hemostasis and Blood Coagulation

Mechanisms of Hemostasis Vessel Spasm Formation of the Platelet Plug **Blood Coagulation Clot Retraction Clot Dissolution** Hypercoagulability States **Increased Platelet Function** Increased Clotting Activity **Bleeding Disorders Platelet Defects** Thrombocytopenia **Impaired Platelet Function Coagulation Defects** Impaired Synthesis of Coagulation Factors Hemophilia A Von Willebrand Disease **Disseminated Intravascular Coagulation** Vascular Disorders

CHAPTER

he term *hemostasis* refers to the stoppage of blood flow. The normal process of hemostasis is regulated by a complex array of activators and inhibitors that maintain blood fluidity and prevent blood from leaving the vascular compartment. Hemostasis is normal when it seals a blood vessel to prevent blood loss and hemorrhage. It is abnormal when it causes inappropriate blood clotting or when clotting is insufficient to stop the flow of blood from the vascular compartment. Disorders of hemostasis fall into two main categories: the inappropriate formation of clots within the vascular system (*i.e.*, thrombosis) and the failure of blood to clot in response to an appropriate stimulus (*i.e.*, bleeding).

MECHANISMS OF HEMOSTASIS

Hemostasis is divided into five stages: vessel spasm, formation of the platelet plug, blood coagulation or development of an insoluble fibrin clot, clot retraction, and clot dissolution (Fig. 12-1).

Vessel Spasm

Vessel spasm is initiated by endothelial injury and caused by local and humoral mechanisms. A spasm constricts the vessel and reduces blood flow. It is a transient event that usually lasts less than 1 minute. Thromboxane A_2 (TXA₂), a prostaglandin released from the platelets, contributes to the vasoconstriction. A second prostaglandin, prostacyclin, released from the vessel endothelium, produces vasodilation and inhibits platelet aggregation.

Formation of the Platelet Plug

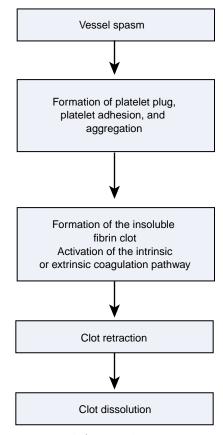
The platelet plug, the second line of defense, is initiated as platelets come in contact with the vessel wall. Small breaks in the vessel wall are often sealed with a platelet plug and do not require the development of a blood clot.

Platelets, also called *thrombocytes*, are large fragments from the cytoplasm of bone marrow cells called *megakaryocytes*. They are enclosed in a membrane but have no nucleus and cannot reproduce. Their cytoplasmic granules release mediators for hemostasis. Although they lack a nucleus, they have many of the characteristics of a whole cell. They have mitochondria and enzyme systems for producing adenosine triphosphate (ATP) and adenosine diphosphate (ADP), and they have the enzymes needed for synthesis of prostaglandins, which are required for their function in hemostasis. Platelets also produce a growth factor that causes vascular endothelial cells, smooth muscle cells, and fibroblasts to proliferate and grow.

The life span of a platelet is only 8 to 9 days. A protein called *thrombopoietin* causes proliferation and maturation of megakaryocytes.¹ The sources of thrombopoietin include the liver, kidney, smooth muscle, and bone marrow, which controls platelet production. Its production and release are regulated by the number of platelets in the circulation. The newly formed

Dr Murtadha Alshareifi e-Library

Unit Three: Alterations in the Hematologic System



■ FIGURE 12-1 ■ Steps in hemostasis.

platelets that are released from the bone marrow spend as long as 8 hours in the spleen before they are released into the blood.

The cell membrane of the platelet is important to its function. The outside of the platelet membrane is coated with glycoproteins that repulse adherence to the normal vessel endothelium, while causing adherence to injured areas of the vessel wall, particularly the subendothelial layer.² The platelet membrane also has glycoprotein receptors that bind fibrinogen and link platelets together. Glycoprotein receptor antagonists have been developed and are selectively used in the treatment

KEY CONCEPTS

HEMOSTASIS

- Hemostasis is the orderly, stepwise process for stopping bleeding that involves vasospasm, formation of a platelet plug, and the development of a fibrin clot.
- The blood clotting process requires the presence of platelets produced in the bone marrow, von Willebrand factor generated by the vessel endothelium, and clotting factors synthesized in the liver, using vitamin K.
- The final step of the process involves fibrinolysis or clot dissolution, which prevents excess clot formation.

of acute coronary myocardial infarction (see Chapter 17). In addition, the platelet membrane contains large amounts of phospholipids that play a role in activating several points in the blood-clotting process.

Platelet plug formation involves adhesion and aggregation of platelets. Platelet adhesion also requires a protein molecule called *von Willebrand factor* (vWF). vWF, which is produced by the endothelial cells of blood vessels, performs two important functions: it aids in platelet adhesion, and it circulates in the blood as a carrier protein for coagulation factor VIII.

Platelets are attracted to a damaged vessel wall, become activated, and change from smooth disks to spiny spheres, exposing receptors on their surfaces. Adhesion to the vessel subendothelial layer occurs when the platelet receptor binds to vWF at the injury site, linking the platelet to exposed collagen fibers (Fig. 12-2A). The process of adhesion is controlled by local hormones and substances released by platelet granules. As the platelets adhere to the collagen fibers on the damaged vessel wall, they begin to release large amounts of ADP and TXA₂. Platelet aggregation and formation of a loosely organized platelet plug occur as the ADP and TXA₂ cause nearby platelets to become sticky and adhere to the original platelets. Stabilization of the platelet plug occurs as the coagulation

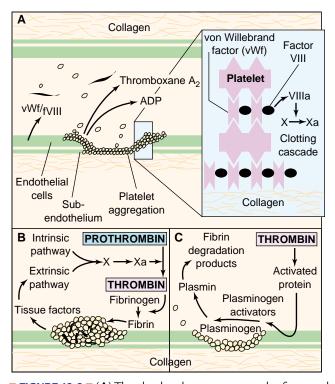


FIGURE 12-2 (A) The platelet plug occurs seconds after vessel injury. Von Willebrand's factor, released from the endothelial cells, binds to platelet receptors, causing *adhesion* of platelets to the exposed collagen. Platelet *aggregation* is induced by release of thromboxane A_2 and adenosine diphosphate. (B) Coagulation factors, activated on the platelet surface, lead to the formation of thrombin and fibrin, which stabilize the platelet plug. (C) Control of the coagulation process and clot dissolution are governed by thrombin and plasminogen activators. Thrombin activates protein C, which stimulates the release of plasminogen activators in turn promote the formation of plasmin, which digests the fibrin strands.

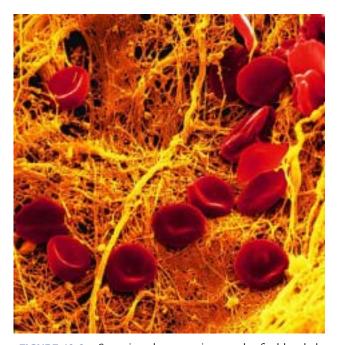
pathway is activated on the platelet surface and fibrinogen is converted to fibrin, thereby creating a fibrin meshwork that cements together the platelets and other blood components (see Fig. 12-2B).

Defective platelet plug formation causes bleeding in persons who are deficient in platelet receptor sites or vWF. In addition to sealing vascular breaks, platelets play an almost continuous role in maintaining normal vascular integrity. They may supply growth factors for the endothelial cells and arterial smooth muscle cells. Persons with platelet deficiency have increased capillary permeability and sustain small skin hemorrhages from the slightest trauma or change in blood pressure.

Blood Coagulation

Blood coagulation is controlled by many substances that promote clotting (*i.e.*, procoagulation factors) or inhibit it (*i.e.*, anticoagulation factors). Each of the procoagulation factors, identified by Roman numerals, performs a specific step in the coagulation process. The action of one coagulation factor or proenzyme is designed to activate the next factor in the sequence (*i.e.*, cascade effect). Because most of the inactive procoagulation factors are present in the blood at all times, the multistep process ensures that a massive episode of intravascular clotting does not occur by chance. It also means that abnormalities of the clotting process occur when one or more of the factors are deficient or when conditions lead to inappropriate activation of any of the steps.

The chemical events in the blood coagulation process involve a number of essential steps that result in the conversion of fibrinogen, a circulating plasma protein, to the fibrin strands that enmesh platelets, blood cells, and plasma to form the clot (Fig. 12-3). The initiation of the clotting process oc-



■ FIGURE 12-3 ■ Scanning electron micrograph of a blood clot (×3600). The fibrous bridges that form a meshwork between red blood cells are fibrin fibers. (© Oliver Meckes, Science Source/ Photo Researchers)

curs by way of the intrinsic or the extrinsic coagulation pathways (Fig. 12-4). The intrinsic pathway, which is a relatively slow process, begins in the blood itself. The extrinsic pathway, which is a much faster process, begins with tissue or vessel trauma and the subsequent release of a complex of several factors, called *tissue factor*, or *tissue thromboplastin*. The terminal steps in both pathways are the same: the activation of factor X, the conversion of prothrombin to thrombin, and the conversion of fibrinogen to fibrin. Prothrombin is an unstable plasma protein, which is easily split into smaller parts, one of which is thrombin. Thrombin, in turn, acts as an enzyme to convert fibrinogen to fibrin.

Both the intrinsic and extrinsic pathways are needed for normal hemostasis, and many interrelations exist between them. Each system is activated when blood passes out of the vascular system. The intrinsic system is activated as blood comes in contact with collagen in the injured vessel wall and the extrinsic system when blood is exposed to tissue extracts. Bleeding, when it occurs because of defects in the extrinsic system, usually is not as severe as that which results from defects in the intrinsic pathway.

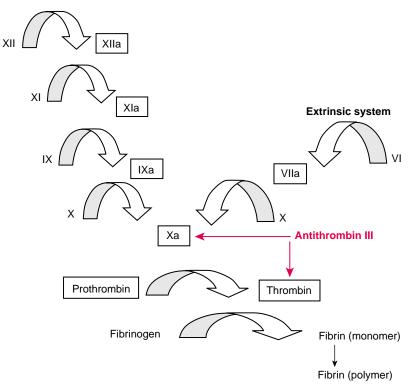
With few exceptions, almost all the blood-clotting factors are synthesized in the liver. Vitamin K is required for the synthesis of prothrombin, factors VII, IX, X, and protein C. Calcium (factor IV) is required in all but the first two steps of the clotting process. The body usually has sufficient amounts of calcium for these reactions. Inactivation of the calcium ion prevents blood from clotting when it is removed from the body. The addition of citrate to blood stored for transfusion purposes prevents clotting by chelating ionic calcium. Another chelator, EDTA, is often added to blood samples used for analysis in the clinical laboratory.

Coagulation is regulated by several natural anticoagulants. Antithrombin III inactivates coagulation factors and neutralizes thrombin, the last enzyme in the pathway for the conversion of fibrinogen to fibrin. When antithrombin III is complexed with naturally occurring heparin, its action is accelerated and provides protection against uncontrolled thrombus formation on the endothelial surface. Protein C, a plasma protein, acts as an anticoagulant by inactivating factors V and VIII. Protein S, another plasma protein, accelerates the action of protein C. Plasmin breaks down fibrin into fibrin degradation products that act as anticoagulants. It has been suggested that some of these natural anticoagulants may play a role in the bleeding that occurs with disseminated intravascular coagulation (DIC; discussed later).

The anticoagulant drugs heparin and warfarin are used to prevent venous thrombi and thromboembolic disorders, such as deep vein thrombosis and pulmonary embolism. Heparin is naturally formed by basophilic mast cells located at the precapillary junctions in tissues throughout the body. These cells continuously secrete small amounts of heparin, which is released into the circulation. Pharmacologic preparations of heparin, extracted from animal tissues, are available for treatment of coagulation disorders. Heparin binds to antithrombin III, causing a conformational change that increases the ability of antithrombin III to inactivate factor Xa, thrombin, and other clotting factors. By promoting the inactivation of clotting factors, heparin ultimately suppresses the formation of fibrin. Heparin is unable to cross the membranes of the gastrointestinal tract and must be given by injection. Warfarin acts by

208 Unit Three: Alterations in the Hematologic System





■ FIGURE 12-4 ■ Intrinsic and extrinsic coagulation pathways. The terminal steps in both pathways are the same. Calcium, factors X and V, and platelet phospholipids combine to form prothrombin activator, which then converts prothrombin to thrombin. This interaction causes conversion of fibrinogen into the fibrin strands that create the insoluble blood clot. Prothrombin and factors VII, IX, and X require vitamin K for synthesis.

decreasing prothrombin and other procoagulation factors. It alters vitamin K such that it reduces its availability to participate in synthesis of the vitamin K-dependent coagulation factors in the liver. Warfarin is readily absorbed after oral administration. Its maximum effect takes 36 to 72 hours because of the varying half-lives of preformed clotting factors that remain in the circulation.

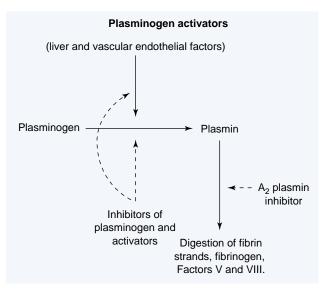
Clot Retraction

After the clot has formed, clot retraction, which requires large numbers of platelets, contributes to hemostasis by squeezing serum from the clot and joining the edges of the broken vessel.

Clot Dissolution

The dissolution of a blood clot begins shortly after its formation; this allows blood flow to be re-established and permanent tissue repair to take place (see Fig. 12-2C). The process by which a blood clot dissolves is called *fibrinolysis*. As with clot formation, clot dissolution requires a sequence of steps controlled by activators and inhibitors (Fig. 12-5). Plasminogen, the proenzyme for the fibrinolytic process, normally is present in the blood in its inactive form. It is converted to its active form, plasmin, by plasminogen activators formed in the vascular endothelium, liver, and kidneys. The plasmin formed from plasminogen digests the fibrin strands of the clot and certain clotting factors, such as fibrinogen, factor V, factor VIII, prothrombin, and factor XII. Circulating plasmin is rapidly inactivated by α_2 -plasmin inhibitor, which limits fibrinolysis to the local clot and prevents it from occurring in the entire circulation.

Two naturally occurring plasminogen activators are tissuetype plasminogen activator and urokinase-type plasminogen activator. The liver, plasma, and vascular endothelium are the major sources of physiologic activators. These activators are released in response to a number of stimuli, including vasoactive drugs, venous occlusion, elevated body temperature, and exercise. The activators are unstable and rapidly inactivated by inhibitors synthesized by the endothelium and the liver. For this reason, chronic liver disease may cause altered



■ FIGURE 12-5 ■ Fibrinolytic system and its modifiers. The *solid* lines indicate activation, and the *broken lines* indicate inactivation.

fibrinolytic activity. A major inhibitor, plasminogen activator inhibitor-1, in high concentrations has been associated with deep vein thrombosis, coronary artery disease, and myocar-dial infarction.³

In summary, hemostasis is designed to maintain the integrity of the vascular compartment. The process is divided into five phases: vessel spasm, which constricts the size of the vessel and reduces blood flow; platelet adherence and formation of the platelet plug; formation of the fibrin clot, which cements together the platelet plug; clot retraction, which pulls the edges of the injured vessel together; and clot dissolution, which involves the action of plasmin that dissolves the clot and allows blood flow to be re-established and tissue healing to take place. Blood coagulation requires the stepwise activation of coagulation factors, carefully controlled by activators and inhibitors.

HYPERCOAGULABILITY STATES

Hypercoagulability represents hemostasis in an exaggerated form and predisposes to thrombosis. Arterial thrombi caused by turbulence are composed of platelet aggregates, and venous thrombi caused by stasis of flow are largely composed of platelet aggregates and fibrin complexes that result from excess coagulation. There are two general forms of hypercoagulability states: conditions that create increased platelet function and conditions that cause accelerated activity of the coagulation system. Chart 12-1 summarizes conditions commonly associated with hypercoagulability states.

Increased Platelet Function

Increased platelet function predisposes to platelet adhesion, formation of a platelet or blood clot, and disturbance of blood flow. The causes of increased platelet function are disturbances in flow, endothelial damage, and increased sensitivity of platelets to factors that cause adhesiveness and aggregation. Athero-

CHART 12-1 Conditions Associated With Hypercoagulability States

Increased Platelet Function

Atherosclerosis Diabetes mellitus Smoking Elevated blood lipid and cholesterol levels Increased platelet levels

Accelerated Activity of the Clotting System

Pregnancy and the puerperium Use of oral contraceptives Postsurgical state Immobility Congestive heart failure Malignant diseases sclerotic plaques disturb flow, cause endothelial damage, and promote platelet adherence. Platelets that adhere to the vessel wall release growth factors that can cause proliferation of smooth muscle and thereby contribute to the development of atherosclerosis. Smoking, elevated levels of blood lipids and cholesterol, hemodynamic stress, diabetes mellitus, and immune mechanisms may cause vessel damage, platelet adherence, and, eventually, thrombosis. Some cancers and other diseases are associated with high platelet counts and the potential for thrombosis. The term *thrombocytosis* is used to describe platelet counts greater than 1,000,000/mm³. This occurs in some malignancies and inflammatory states and after splenectomy. Myeloproliferative disorders that result in excess platelet production may predispose to thrombosis or, paradoxically, bleeding when the rapidly produced platelets are defective.

Increased Clotting Activity

Increased clotting activity results from factors that increase the activation of the coagulation system, including stasis of blood flow and alterations in the coagulation components of the blood (*i.e.*, an increase in procoagulation factors or a decrease in anticoagulation factors). Stasis of blood flow causes the accumulation of activated clotting factors and platelets and prevents their interactions with inhibitors. Slow and disturbed flow is a common cause of venous thrombosis in the immobilized or postoperative patient. Heart failure also contributes to venous congestion and thrombosis. Elevated levels of estrogen tend to increase hepatic synthesis of many of the coagulation factors and decrease the synthesis of antithrombin III.⁴ The incidence of stroke, thromboemboli, and myocardial infarction is greater in women who use oral contraceptives, particularly after age 35 years, and in heavy smokers. Clotting factors are also increased during normal pregnancy. These changes, along with limited activity during the puerperium (immediate postpartum period), predispose to venous thrombosis. A hypercoagulability state is also common in cancer and sepsis. Many tumor cells are thought to release tissue factor molecules that, along with the increased immobility and sepsis seen in patients with malignant disease, contribute to increased risk of both venous and arterial thrombosis.

A reduction in anticoagulants such as antithrombin III, protein C, and protein S predisposes to venous thrombosis.⁵

KEY CONCEPTS

HYPERCOAGULABILITY STATES

- Hypercoagulability states increase the risk of clot or thrombus formation in either the arterial or venous circulations.
- Arterial thrombi are associated with conditions that produce turbulent blood flow and platelet adherence.
- Venous thrombi are associated with conditions that cause stasis of blood flow with increased concentrations of coagulation factors.

Deficiencies of these inhibitor proteins are uncommon inherited defects. It has been suggested that high circulating levels of homocysteine also predispose to venous and arterial thrombosis by activating platelets and altering antithrombotic mechanisms.⁶

Another cause of increased venous and arterial clotting is a condition known as the antiphospholipid syndrome.7 The syndrome is associated with a number of clinical manifestations, including multiple thromboses. The condition can manifest as a primary disorder in persons who exhibit only the manifestations of the hypercoagulable state or as a secondary disorder in persons with well-defined autoimmune disorders, such as systemic lupus erythematosus. Thrombosis may be precipitated by trauma, surgical conditions, use of drugs such as oral contraceptives, or abrupt withdrawal of anticoagulant drugs. Persons with the disorder have a history of having one or more of the following: deep venous thrombosis; arterial thrombosis, including stroke, myocardial infarction, or gangrene; or thrombocytopenia.7 Antiphospholipid antibody syndrome is also a cause of renal microangiopathy, resulting in renal failure caused by multiple capillary and arterial thromboses.

Women with the disorder often have a history of recurrent pregnancy losses within the fetal period (10 weeks or more gestation) because of ischemia and thrombosis of placental vessels.⁷ Pregnancies in women with antiphospholipid syndrome can also be complicated by premature delivery caused by pregnancy-associated hypertension and uteroplacental insufficiency.

In most persons with antiphospholipid syndrome, thrombotic events occur singly. However, recurrences may occur months or years after the initial event. Occasionally, someone will present with multiple almost simultaneous vascular occlusions involving different organ systems. The condition, sometimes termed *catastrophic antiphospholipid syndrome*, is a serious and sometimes fatal condition.

Exactly how antiphospholipid antibodies, which interact with cardiolipin, a mitochondrial phospholipid, produce thrombosis is unclear. Possible mechanisms include activation or injury of vascular endothelial cells, direct platelet activation, or inactivation of anticoagulation factors (*e.g.*, anti-thrombin III or protein C).

Treatment focuses on prophylaxis, treatment of acute thrombotic events, and prevention of future thrombotic events.⁷ Prophylaxis focuses on removal or reduction in factors that predispose to thrombosis, such as smoking and the use of estrogen-containing oral contraceptives by women. Treatment during a thrombotic event includes the use of anticoagulant medications (heparin and warfarin) and immune suppression in refractory cases. Aspirin and anticoagulant drugs may be used to prevent further thrombosis.

In summary, hypercoagulability causes excessive clotting and contributes to thrombus formation. It results from conditions that create increased platelet function or that cause accelerated activity of the coagulation system. Increased platelet function usually results from disorders such as atherosclerosis that damage the vessel endothelium and disturb blood flow or from conditions such as smoking that cause increased sensitivity of platelets to factors that promote adhesiveness and aggregation. Factors that cause accelerated activity of the coagulation system include blood flow stasis, resulting in an accumulation of coagulation factors, and alterations in the components of the coagulation system (*i.e.*, an increase in procoagulation factors or a decrease in anticoagulation factors). Another cause of increased venous and arterial clotting is the antiphospholipid syndrome, which can manifest as a primary disorder or as a secondary disorder in persons with autoimmune disorders, such as systemic lupus erythematosus.

BLEEDING DISORDERS

Bleeding disorders or impairment of blood coagulation can result from defects in any of the factors that contribute to hemostasis. Defects are associated with platelets, coagulation factors, and vascular integrity.

Platelet Defects

Platelets provide one of the first defenses against bleeding. Bleeding can occur as a result of a decrease in the number of circulating platelets or impaired platelet function. The depletion of platelets must be relatively severe (10,000 to 20,000/mm³, compared with the normal values of 150,000 to 400,000/mm³) before hemorrhagic tendencies or spontaneous bleeding become evident. Bleeding that results from platelet deficiency commonly occurs in small vessels and is characterized by petechiae (*i.e.*, pinpoint purplish-red spots) and purpura (*i.e.*, purple areas of bruising) on the arms and thighs. Bleeding from mucous membranes of the nose, mouth, gastrointestinal tract, and vagina is characteristic. Bleeding of the intracranial vessels is a rare danger with severe platelet depletion.

KEY CONCEPTS

BLEEDING DISORDERS

- Bleeding disorders are caused by defects associated with platelets, coagulation factors, and vessel integrity.
- Disorders of platelet plug formation include a decrease in platelet numbers due to inadequate platelet production (bone marrow dysfunction), excess platelet destruction (thrombocytopenia), abnormal platelet function (thrombocytopathia), or defects in von Willebrand factor.
- Impairment of the coagulation stage of hemostasis is caused by a deficiency in one or more of the clotting factors.
- Disorders of blood vessel integrity result from structurally weak vessels or vessel damage due to inflammation and immune mechanisms.

Thrombocytopenia

Thrombocytopenia represents a decrease in the number of circulating platelets (usually less than 100,000/mm³). It can result from decreased platelet production by the bone marrow, increased pooling of platelets in the spleen, or decreased platelet survival caused by immune or nonimmune mechanisms. Dilutional thrombocytopenia can result from massive transfusions because blood stored for more that 24 hours has virtually no platelets.

Decreased platelet production can result from suppression or failure of bone marrow function, such as occurs in aplastic anemia (see Chapter 13), or from replacement of bone marrow by malignant cells, such as occurs in leukemia (see Chapter 11). Infection with human immunodeficiency virus (HIV) suppresses the production of megakaryocytes. Radiation therapy and drugs such as those used in the treatment of cancer may suppress bone marrow function and reduce platelet production.

There may be normal production of platelets but excessive pooling of platelets in the spleen. The spleen normally sequesters approximately 30% to 40% of the platelets. However, as much as 80% of the platelets can be sequestered when the spleen is enlarged (splenomegaly). Splenomegaly occurs in cirrhosis with portal hypertension and in lymphomas.

Decreased platelet survival is an important cause of thrombocytopenia. In many cases, premature destruction of platelets is caused by antiplatelet antibodies or immune complexes. The antibodies can be directed against self-antigens (autoimmunity) or against nonself platelet antigens (from blood transfusions). Autoimmune thrombocytopenias include idiopathic thrombocytopenic purpura and HIV-associated thrombocytopenias. Decreased platelet survival may also occur as the result of mechanical injury associated with prosthetic heart valves.

Drug-Induced Thrombocytopenia. Some drugs, such as quinine, quinidine, and certain sulfa-containing antibiotics, may induce thrombocytopenia. These drugs act as a hapten (see Chapter 8) and induce antigen–antibody response and formation of immune complexes that cause platelet destruction by complement-mediated lysis. In persons with drug-associated thrombocytopenia, there is a rapid fall in platelet count within 2 to 3 days of resuming use of a drug or 7 or more days (*i.e.*, the time needed to mount an immune response) after starting use of a drug for the first time. The platelet count rises rapidly after the drug use is discontinued.

The anticoagulant drug heparin has been increasingly implicated in thrombocytopenia and, paradoxically, in thrombosis. The complications typically occur 5 days after the start of therapy and result from production of heparin-dependent antiplatelet antibodies that cause aggregation of platelets and their removal from the circulation. The antibodies often bind to vessel walls, causing injury and thrombosis. The newer, low-molecular-weight heparin has been shown to be effective in reducing the incidence of heparin-induced complications compared with the older, high-molecular-weight form of the drug.⁸

Idiopathic Thrombocytopenic Purpura. Idiopathic thrombocytopenic purpura, an autoimmune disorder, results in platelet antibody formation and excess destruction of platelets. The IgG antibody binds to two identified membrane glycoproteins while in the circulation. The platelets, which are made more susceptible to phagocytosis because of the antibody, are destroyed in the spleen.

Acute idiopathic thrombocytopenic purpura is more common in children and usually follows a viral infection. It is characterized by sudden onset of petechiae and purpura and is a self-limited disorder with no treatment. In contrast, the chronic form is usually seen in adults and seldom follows an infection. It is a disease of young people, with a peak incidence between the ages of 20 and 50 years, and is seen twice as often in women as in men. It may be associated with other immune disorders such as acquired immunodeficiency syndrome (AIDS) or systemic lupus erythematosus. The condition occasionally presents precipitously with signs of bleeding, often into the skin (i.e., purpura and petechiae) or oral mucosa. There is commonly a history of bruising, bleeding from gums, epistaxis (*i.e.*, nosebleeds), and abnormal menstrual bleeding. Because the spleen is the site of platelet destruction, splenic enlargement may occur.

Diagnosis usually is based on severe thrombocytopenia (platelet counts <20,000/mL), and exclusion of other causes. Treatment includes the initial use of corticosteroid drugs, often followed by splenectomy and the use of immunosuppressive agents.

Thrombotic Thrombocytopenic Purpura. Thrombotic thrombocytopenic purpura (TPP) is a combination of thrombocytopenia, hemolytic anemia, signs of vascular occlusion, fever, and neurologic abnormalities. The onset is abrupt, and the outcome may be fatal. Widespread vascular occlusions consist of thrombi in arterioles and capillaries of many organs, including the heart, brain, and kidneys. Erythrocytes become fragmented as they circulate through the partly occluded vessels and cause the hemolytic anemia. The clinical manifestations include purpura and petechiae and neurologic symptoms ranging from headache to seizures and altered consciousness.

Although TTP may have diverse causes, the initiating event seems to be widespread endothelial damage and activation of intravascular thrombosis. Toxins produced by certain strains of *Escherichia coli (e.g., E. coli O157:H7)* are a trigger for endothelial damage and an associated condition called the *hemolytic-uremic syndrome* (see Chapter 27).

Treatment for TTP includes *plasmapheresis*, a procedure that involves removal of plasma from withdrawn blood and replacement with fresh-frozen plasma. The treatment is continued until remission occurs. With plasmapheresis treatment, there is a complete recovery in 80% to 90% of cases.

Impaired Platelet Function

Impaired platelet function (also called *thrombocytopathia*) may result from inherited disorders of adhesion (*e.g.*, von Willebrand disease) or acquired defects caused by drugs, disease, or extracorporeal circulation. Defective platelet function is also common in uremia, presumably because of unexcreted waste products. Cardiopulmonary bypass also causes platelet defects and destruction.

Use of aspirin and other nonsteroidal anti-inflammatory drugs (NSAIDs) is the most common cause of impaired platelet function. Aspirin (acetylsalicylic acid) produces irreversible acetylation of platelet cyclooxygenase (COX), the enzyme required for TXA₂ synthesis. The antiplatelet effects of aspirin last

for the life of the platelet, usually approximately 8 to 9 days. Because aspirin prolongs bleeding time, it is usually recommended that aspirin use be avoided for a week before surgery. The antiplatelet effects of other NSAIDs, which do not contain the acetyl group, are reversible and last only for the duration of drug action.⁹ Because of aspirin's antiplatelet function, low doses of the drug (usually 81 mg daily) are commonly used in the prevention of heart attack and stroke.

Coagulation Defects

Impairment of blood coagulation can result from deficiencies of one or more of the known clotting factors. Deficiencies can arise because of defective synthesis, inherited defects, or increased consumption of the clotting factors. Bleeding that results from clotting factor deficiency typically occurs after injury or trauma. Large bruises, hematomas, or prolonged bleeding into the gastrointestinal or urinary tracts or joints are common.

Impaired Synthesis of Coagulation Factors

Coagulation factors V, VII, IX, X, XI, and XII; prothrombin; and fibrinogen are synthesized in the liver. In liver disease, synthesis of these clotting factors is reduced, and bleeding may result. Of the coagulation factors synthesized in the liver, factors VII, IX, and X and prothrombin require the presence of vitamin K for normal activity. In vitamin K deficiency, the liver produces the clotting factor, but in an inactive form. Vitamin K is a fatsoluble vitamin that is continuously being synthesized by intestinal bacteria. This means that a deficiency in vitamin K is not likely to occur unless intestinal synthesis is interrupted or absorption of the vitamin is impaired. Vitamin K deficiency can occur in the newborn infant before the establishment of the intestinal flora; it can also occur as a result of treatment with broad-spectrum antibiotics that destroy intestinal flora. Because vitamin K is a fat-soluble vitamin, its absorption requires bile salts. Vitamin K deficiency may result from impaired fat absorption caused by liver or gallbladder disease.

Hereditary defects have been reported for each of the clotting factors, but most are rare diseases. The bleeding disorders are hemophilia A, which affects 1 in 10,000 males, and von Willebrand disease, which occurs in 1% of the population.¹⁰ Factor IX deficiency (*i.e.*, hemophilia B) occurs in approximately 1 in 50,000 persons and is genetically and clinically similar to hemophilia A.

Circulating factor VIII is part of a complex molecule, bound to vWF. Factor VIII coagulant protein is the functional portion produced by the liver and endothelial cells. vWF, synthesized by the endothelium and megakaryocytes, binds and stabilizes factor VIII in the circulation by preventing proteolysis. It is also required for platelet adhesion to the subendothelial layer.

Hemophilia A

Hemophilia A, which is caused by a deficiency in factor VIII, is an X-linked recessive disorder that primarily affects males. Although it is a hereditary disorder, there is no family history of the disorder in approximately one third of newly diagnosed cases, suggesting that it has arisen as a new mutation in the factor VIII gene.¹⁰ Approximately 90% of persons with hemophilia produce insufficient quantities of the factor, and 10% produce a defective form. The percentage of normal factor VIII activity in the circulation depends on the genetic defect and determines the severity of hemophilia (i.e., 6% to 50% in mild hemophilia, 2% to 5% in moderate hemophilia, and 1% or less in severe forms of hemophilia).¹⁰ In mild or moderate forms of the disease, bleeding usually does not occur unless there is a local lesion or trauma such as surgery or dental procedures. The mild disorder may not be detected in childhood. In severe hemophilia, bleeding usually occurs in childhood (e.g., it may be noticed at the time of circumcision) and is spontaneous and severe. Characteristically, bleeding occurs in soft tissues, the gastrointestinal tract, and the hip, knee, elbow, and ankle joints. Joint bleeding usually begins when a child begins to walk. Often, a target joint is prone to repeated bleeding. The bleeding causes inflammation of the synovium, with acute pain and swelling. Without proper treatment, chronic bleeding and inflammation cause joint fibrosis and contractures, resulting in major disability. There is also the potential for lifethreatening intracranial hemorrhage.

Factor VIII replacement therapy is initiated when bleeding occurs or as prophylaxis with repeated bleeding episodes. The purpose is to limit the extent of tissue damage. Highly purified factor VIII concentrates prepared from human plasma are the usual replacement products for persons with severe hemophilia. Before blood was tested for infectious diseases, these products were prepared from multiple donor samples and carried a high risk of exposure to viruses for hepatitis and HIV. Donor screening and the development of effective virusinactivation procedures have effectively reduced the transmission of hepatitis viruses and HIV through clotting concentrates. Recombinant factor VIII, although expensive, is now available and should reduce the risk of transmitting HIV or other viruses. Desmopressin acetate (DDAVP, 1-desamino-8-D-arginine vasopressin) may be used to prevent bleeding in persons with mild hemophilia.¹¹ It stimulates the release of vWF (the carrier for factor VIII) from the endothelium, thus increasing factor VIII levels twofold to threefold for several hours.

The cloning of the factor VIII gene and progress in gene delivery systems have led to the hope that hemophilia A may be cured by gene therapy. Carrier detection and prenatal diagnosis can now be done by analysis of direct gene mutation or DNA linkage studies.

Von Willebrand Disease

Von Willebrand disease, which typically is diagnosed in adulthood, is the most common hereditary bleeding disorder. Transmitted as an autosomal trait, it is caused by a deficiency of or defect in vWF. This deficiency results in reduced platelet adhesion. There are many variants of the disease, and manifestations range from mild to severe. Because vWF carries factor VIII, its deficiency may also be accompanied by reduced levels of factor VIII and results in defective clot formation. Symptoms include bruising, excessive menstrual flow, and bleeding from the nose, mouth, and gastrointestinal tract. Many persons with the disorder receive a diagnosis when surgery or dental extraction results in prolonged bleeding. Most cases are mild and untreated.

In severe cases, factor VIII products that contain vWF are infused to replace the deficient clotting factors. The disorder also responds to desmopressin acetate (DDAVP), a synthetic analog of the hormone vasopressin, which stimulates the endothelial cells to release vWF and plasminogen activator. DDAVP can also be used to treat platelet dysfunction caused by uremia, heart bypass, and the effects of aspirin.¹¹

Disseminated Intravascular Coagulation

Disseminated intravascular coagulation is a paradox in the hemostatic sequence and is characterized by widespread intravascular coagulation and bleeding. It is not a primary disease but occurs as a complication of a wide variety of conditions. DIC begins with massive activation of the coagulation sequence as a result of unregulated generation of thrombin, resulting in systemic formation of fibrin. In addition, levels of all the major anticoagulants are reduced (Fig. 12-6). The generation of microthrombi results in vessel occlusion and tissue ischemia. Multiple organ failure may ensue. Clot formation consumes all available coagulation proteins and platelets, and severe hemorrhage results.

The disorder can be initiated by activation of the intrinsic or extrinsic pathways, both of which involve the formation of thrombin (see Fig. 12-4). Initiation of DIC through the extrinsic pathway, as occurs with trauma and cancer, begins with the liberation of tissue factor. The intrinsic pathway may be activated through extensive endothelial damage caused by viruses, infections, or immune mechanisms, or stasis of blood. Obstetric disorders that involve necrotic placental or fetal tissue commonly are associated with DIC. Other inciting clinical conditions include massive trauma, burns, sepsis, shock, meningococcemia, and malignant disease. The initiating factors in these conditions are multiple and often related. For example, in infections, particularly those caused by gramnegative bacteria, endotoxins released from the bacteria activate both the intrinsic and extrinsic pathways. In addition, endotoxins inhibit the anticoagulant activity of protein C.⁵ In obstetric conditions, tissue factor released from the necrotic placental or fetal tissue may enter the circulation. At the same time, the shock, hypoxia, and acidosis that often coexist with the obstetrical condition can also cause widespread endothelial injury.⁵ Chart 12-2 summarizes the conditions associated with DIC.

There is also evidence that the fibrinolytic system may be involved in the pathogenesis of DIC. It may be suppressed and thereby contribute to the formation of microthrombi, or it may be the source of fibrin degradation products that contribute to the bleeding that occurs. Finally, regardless of the inciting event, DIC may be a systemic inflammatory disorder with release of proinflammatory cytokines that mediate the derangement of coagulation and fibrin breakdown.¹²

Although coagulation and formation of microemboli initiate the events that characterize DIC, its acute manifestations usually are more directly related to the bleeding problems that occur. The bleeding may be present as petechiae, purpura, oozing from puncture sites, or severe hemorrhage. Cardiovascular shock is a common complication. Uncontrolled postpartum bleeding may indicate DIC. Microemboli may obstruct blood

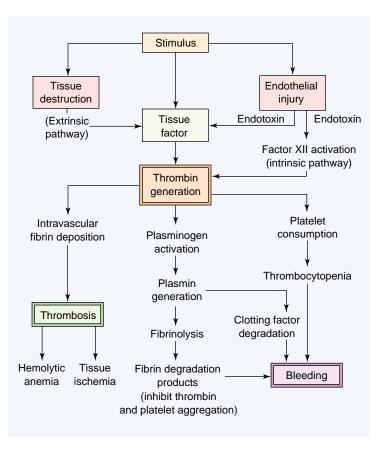


FIGURE 12-6 Pathophysiology of disseminated intravascular coagulation.

CHART 12-2 Conditions That Have Been Associated With DIC

Obstetric Conditions

Abruptio placentae Dead fetus syndrome Preeclampsia and eclampsia Amniotic fluid embolism

Cancers

Metastatic cancer Leukemia

Infections

Acute bacterial infections (*e.g.*, meningococcal meningitis) Acute viral infections Rickettsial infections (*e.g.*, Rocky Mountain spotted fever) Parasitic infection (*e.g.*, malaria)

Shock

Septic shock Severe hypovolemic shock

Trauma or Surgery

Burns Massive trauma Surgery involving extracorporeal circulation Snake bite Heatstroke

Hematologic Conditions

Blood transfusion reactions

causes a fragile wall; Cushing's disease, causing protein wasting and loss of vessel tissue support because of excess cortisol; and senile purpura (*i.e.*, bruising in elderly persons) caused by the aging process. Vascular defects also occur in the course of DIC as a result of the presence of microthrombi and corticosteroid therapy.

Vascular disorders are characterized by easy bruising and the spontaneous appearance of petechiae and purpura of the skin and mucous membranes. In persons with bleeding disorders caused by vascular defects, the platelet count and results of other tests for coagulation factors are normal.

In summary, bleeding disorders or impairment of blood coagulation can result from defects in any of the factors that contribute to hemostasis: platelets, coagulation factors, or vascular integrity. The number of circulating platelets can be decreased (i.e., thrombocytopenia), or platelet function can be impaired (i.e., thrombocytopathia). Impairment of blood coagulation can result from deficiencies of one or more of the known clotting factors. Deficiencies can arise because of defective synthesis (i.e., liver disease or vitamin K deficiency), inherited diseases (i.e., hemophilia A or von Willebrand disease), or increased consumption of the clotting factors (DIC). Bleeding may also occur from structurally weak vessels that result from impaired synthesis of vessel wall components (i.e., vitamin C deficiency, excessive cortisol levels such as in Cushing's disease, or the aging process) or from damage by genetic mechanisms (i.e., hemorrhagic telangiectasia) or the presence of microthrombi.

vessels and cause tissue hypoxia and necrotic damage to organ structures, such as the kidneys, heart, lungs, and brain. As a result, common clinical signs may be caused by renal, circulatory, or respiratory failure. A form of hemolytic anemia may develop as red cells are damaged as they pass through vessels partially blocked by thrombus.

The treatment of DIC is directed toward managing the primary disease, replacing clotting components, and preventing further activation of clotting mechanisms. Transfusions of fresh-frozen plasma, platelets, or fibrinogen-containing cryoprecipitate may correct the clotting factor deficiency. Heparin may be given to decrease blood coagulation, thereby interrupting the clotting process.

Vascular Disorders

Bleeding from small blood vessels may result from vascular disorders. These disorders may occur because of structurally weak vessel walls or because of damage to vessels by inflammation or immune responses. Among the vascular disorders that cause bleeding are hemorrhagic telangiectasia, an uncommon autosomal dominant disorder characterized by thinwalled, dilated capillaries and arterioles; vitamin C deficiency (*i.e.*, scurvy), resulting in poor collagen synthesis and failure of the endothelial cells to be cemented together properly, which

REVIEW QUESTIONS

Relate the function of the platelet to platelet clot formation in myocardial infarction and thrombocytopenia.

- Explain the physiologic basis for the antiplatelet effects of low-dose aspirin.
- Explain the mechanisms whereby immobility, dehydration, oral contraceptive medications, and the antiphospholipid syndrome predispose to blood clotting.
- State the mechanisms of drug-induced thrombocytopenia and idiopathic thrombocytopenia and the differing features in terms of onset and resolution of the disorders.
- Newborn infants often receive an injection of vitamin K shortly after birth. Explain the rationale for this treatment.
- Differentiate between the mechanisms of bleeding in hemophilia A and von Willebrand disease.
- Explain why desmopressin acetate (DDAVP, 1-desamino-8-Darginine vasopressin) is sometimes used in the treatment of mild hemophilia.
- Use information related to blood coagulation to explain the differing actions of the anticoagulant drugs heparin and warfarin.
- Describe the physiologic events that occur with acute disseminated intravascular coagulation.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- Kaushansky K. (1998). Thrombopoietin. New England Journal of Medicine 339, 746–754.
- Guyton A.C., Hall J.E. (2000). *Textbook of medical physiology* (10th ed., pp. 419–429). Philadelphia: W.B. Saunders.
- Kohler H.P., Grant P.J. (2000). Plasminogen-activator inhibitor type 1 and coronary artery disease. New England Journal of Medicine 342, 1792–1801.
- Chrousos G.P., Zoumakis E.N., Gravanis A. (2001). The gonadal hormones and inhibitors. In Katzung B.G. (Ed.), *Basic and clinical pharmacology* (8th ed., p. 683–684). Norwalk, CT: Appleton & Lange.

- 5. Alving B.M. (1993). The hypercoagulable states. *Hospital Practice* 28(2), 109–121.
- Mitchell R.N., Cotran R.S. (1999). Hemodynamic disorders, thrombosis, and shock. In Cotran R.S., Kumar V., Collins T. (Eds.), *Robbins pathologic basis of disease* (6th ed., p. 125–126). Philadelphia: W.B. Saunders.
- Levine J.L., Branch D.W., Rauch J. (2002). The antiphospholipid syndrome. *New England Journal of Medicine* 346 (10), 752–763.
- Warkentin T.E., Chong B.H., Greinacher A. (1998). Heparin-induced thrombocytopenia: Towards consensus. *Thrombosis and Haemostasis* 79, 1–7.
- George J.N., Shattil S.J. (2000). Acquired disorders of platelet function. In Hoffman R., Benz E.J., Shattil S.J., et al. (Eds.), *Hematology* (3rd ed., p. 2176). New York: Churchill Livingstone.
- Cotran R.S. (1999). Red cells and bleeding disorders. In Cotran R.S., Kumar V., Collins T. (Eds.), *Robbins pathologic basis of disease* (6th ed., pp. 633–642). Philadelphia: W.B. Saunders.
- Mannucci P.M. (1997). Desmopressin (DDAVP) in the treatment of bleeding disorders: The first 20 years. *Blood* 90, 2515–2521.
- 12. Levi M., ten Cate H. (1999). Disseminated intravascular coagulation. *New England Journal of Medicine* 341, 586–592.

Alterations in Red Blood Cells

The Red Blood Cell Red Cell Production Red Cell Destruction Red Cell Metabolism and Hemoglobin Oxidation Laboratory Tests Anemia **Blood Loss Anemia** Hemolytic Anemias Inherited Disorders of the Red Cell Membrane Hemoglobinopathies **Inherited Enzyme Defects Acquired Hemolytic Anemias** Anemias of Deficient Red Cell Production **Iron-Deficiency Anemia** Megaloblastic Anemias **Aplastic Anemia Chronic Disease Anemias** Polycythemia Age-Related Changes in Red Blood Cells **Red Cell Changes in the Neonate** Hyperbilirubinemia in the Neonate Hemolytic Disease of the Newborn **Red Cell Changes With Aging**

CHAPTER

Ithough the lungs provide the means for gas exchange between the external and internal environment, it is the hemoglobin in the red blood cells that transports oxygen to the tissues. The red blood cells also function as carriers of carbon dioxide and participate in acid-base balance. The function of the red blood cells, in terms of oxygen transport, is discussed in Chapter 19, and acid-base balance is covered in Chapter 6. This chapter focuses on the red blood cell, anemia, and polycythemia.

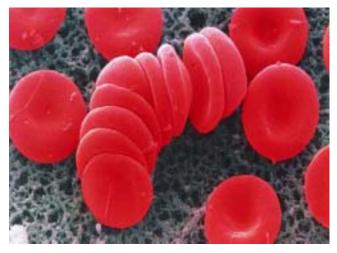
THE RED BLOOD CELL

The mature red blood cell, the erythrocyte, is a non-nucleated, biconcave disk (Fig. 13-1). This shape increases the surface area available for diffusion of oxygen and allows the cell to change in volume and shape without rupturing its membrane. A cyto-skeleton of proteins attached to the lipid bilayer provides this unique shape and flexibility. The biconcave form presents the plasma with a surface 20 to 30 times greater than if the red blood cell were an absolute sphere. The erythrocytes, 500 to 1000 times more numerous than other blood cells, are the most common type of blood cell.

The function of the red blood cell, facilitated by the hemoglobin molecule, is to transport oxygen to the tissues. Hemoglobin also binds some carbon dioxide and carries it from the tissues to the lungs. The hemoglobin molecule is composed of two pairs of structurally different polypeptide chains determined by genes (Fig. 13-2). Alterations in these genes can result in abnormal hemoglobins. Each of the four polypeptide chains is attached to a *heme* unit, which surrounds an atom of iron that binds oxygen. Thus, one molecule of hemoglobin can carry four molecules of oxygen.

The two major types of normal hemoglobin are adult hemoglobin (HbA) and fetal hemoglobin (HbF). HbA consists of a pair of α chains and a pair of β chains. HbF is the predominant hemoglobin in the fetus from the third through the ninth month of gestation. It has a pair of γ chains substituted for the β chains. Because of this chain substitution, HbF has a high affinity for oxygen. This facilitates the transfer of oxygen across the placenta. HbF is replaced within 6 months of birth with HbA.

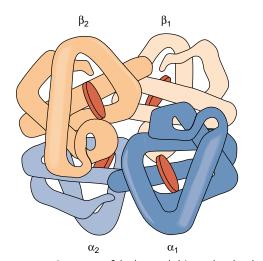
The rate at which hemoglobin is synthesized depends on the availability of iron for heme synthesis. A lack of iron results in relatively small amounts of hemoglobin in the red blood cells. The amount of iron in the body is approximately 35 to 50 mg/kg of body weight for males and less for females. Body iron is found in several compartments. Most of the functional iron (80%) is found in hemoglobin, with small amounts being found in the myoglobin of muscle, the cytochromes, and ironcontaining enzymes such as catalase. Approximately 15% to 20% is stored in the bone marrow, liver, spleen, and other organs. Iron in the hemoglobin compartment is recycled. As red



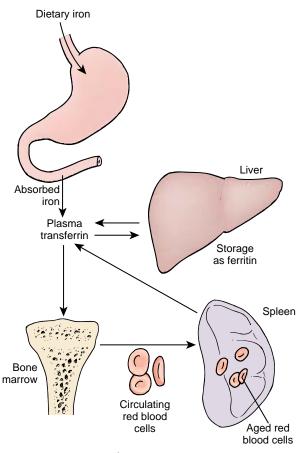
■ FIGURE 13-1 ■ Scanning micrograph of normal red blood cells shows their normal concave appearance (× 3000). (© Andrew Syred, Science Photo Lab, Science Source/Photo Researchers)

blood cells age and are destroyed in the spleen, iron from their hemoglobin is released into the circulation and transported to the bone marrow for use in production of new red cells or to the liver and other tissues for storage.

Dietary iron helps to maintain body stores. Iron, principally derived from meat, is absorbed in the small intestine, especially the duodenum (Fig. 13-3). When body stores of iron are diminished or red cell production is stimulated, absorption is increased. In iron overload, excretion of iron is accelerated. Normally, some iron is sequestered in the intestinal epithelial cells and is lost in the feces as these cells slough. The iron that is absorbed enters the circulation, where it immediately combines with a β -globulin, called *apotransferrin*, to form *transferrin*, which is then transported in the plasma. From the plasma, iron can be deposited in tissues such as the liver. There it is stored as *ferritin*, a protein–iron complex, which can easily return to the circulation. Serum



■ FIGURE 13-2 ■ Structure of the hemoglobin molecule, showing the four subunits.



■ FIGURE 13-3 ■ Iron cycle.

ferritin levels, which can be measured in the laboratory, provide an index of body iron stores. Clinically, decreased serum ferritin levels indicate the need for prescription of iron supplements. Transferrin can also deliver iron to the developing red cell in bone marrow by binding to membrane receptors. The iron is taken up by the developing red cell, where it is used in heme synthesis.

Red Cell Production

Erythropoiesis is the production of red blood cells. After birth, red cells are produced in the red bone marrow. Until age 5 years, almost all bones produce red cells to meet growth needs. After this period, bone marrow activity gradually declines. After 20 years of age, red cell production takes place mainly in the membranous bones of the vertebrae, sternum, ribs, and pelvis. With this reduction in activity, the red bone marrow is replaced with fatty yellow bone marrow.

The red cells are derived from the erythroblasts or red cell precursors, which are continuously being formed from the pluripotent stem cells in the bone marrow (Fig. 13-4). As they develop into mature red cells, the red cell precursors move through a series of divisions, each producing a smaller cell. Hemoglobin synthesis begins at an early stage and continues until the cell becomes a mature erythrocyte. During the maturation time, the red blood cell accumulates hemoglobin as the

KEY CONCEPTS

RED BLOOD CELLS

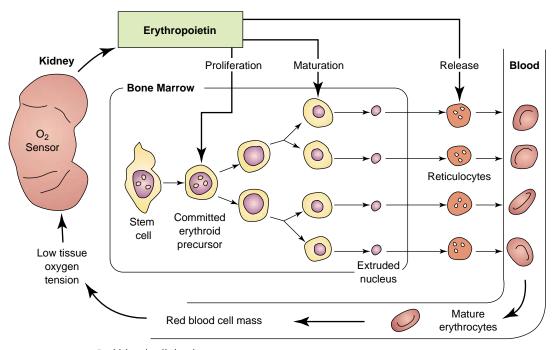
- The function of red blood cells, facilitated by the iron-containing hemoglobin molecule, is to transport oxygen from the lungs to the tissues.
- The production of red blood cells, which is regulated by erythropoietin, occurs in the bone marrow and requires iron, vitamin B₁₂, and folate.
- The red blood cell, which has a life span of approximately 120 days, is broken down in the spleen; the degradation products such as iron and amino acids are recycled.
- The heme molecule, which is released from the red blood cell during the degradation process, is converted to bilirubin and transported to the liver, where it is removed and rendered water soluble for elimination in the bile.

nucleus condenses and is finally lost. The period from stem cell to emergence of the reticulocyte in the circulation normally takes approximately 1 week. Maturation of reticulocyte to erythrocyte takes approximately 24 to 48 hours. During this process, the red cell loses its mitochondria and ribosomes, along with its ability to produce hemoglobin and engage in oxidative metabolism. Most maturing red cells enter the blood as reticulocytes. Approximately 1% of the body's total complement of red blood cells is generated from bone marrow each day, so the reticulocyte count serves as an index of the erythropoietic activity of the bone marrow.

Erythropoiesis is governed for the most part by tissue oxygen needs. Any condition that causes a decrease in the amount of oxygen that is transported in the blood produces an increase in red cell production. The oxygen content of the blood does not act directly on the bone marrow to stimulate red blood cell production. Instead, the decreased oxygen content is sensed by the kidneys, which then produce a hormone called *erythropoietin*. Normally, the kidneys produce approximately 90% of erythropoietin, with the remaining 10% being released by the liver. In the absence of erythropoietin, as in kidney failure, hypoxia has little or no effect on red blood cell production. Erythropoietin takes several days to effect a release of red blood cells from the bone marrow, and only after 5 days or more does red blood cell production reach a maximum.

Erythropoietin acts in the bone marrow by binding to receptors on committed stem cells. It functions on many levels to promote hemoglobin synthesis, increase production of membrane proteins, and cause differentiation of erythroblasts. Human erythropoietin can be produced by recombinant deoxyribonucleic acid (DNA) technology. It is used for management of anemia in cases of chronic renal failure, for treatment of chemotherapy-induced anemia in persons with malignancies, and treatment of anemia in persons with human immunodeficiency virus (HIV) infection who are being treated with zidovudine.

Because red blood cells are released into the blood as reticulocytes, the percentage of these cells is higher when there is a marked increase in red blood cell production. For example, in some severe anemias, the reticulocytes may account for as much as 30% of the total red cell count. In some situations, red cell production is so accelerated that numerous erythroblasts appear in the blood.



■ FIGURE 13-4 ■ Red blood cell development.

Red Cell Destruction

Mature red blood cells have a life span of approximately 4 months or 120 days. As the red blood cell ages, a number of changes occur. Metabolic activity in the cell decreases, and enzyme activity decreases; adenosine triphosphate (ATP) decreases, and the cell membrane becomes more fragile. Once the red cell membrane becomes fragile, the cell ruptures during passage through narrowed places in the circulation. Many of the red cells self-destruct in the small trabecular spaces in the red pulp of the spleen. The rate of red cell destruction (1% per day) normally is equal to red cell production, but in conditions such as hemolytic anemia, the cell's life span may be shorter.

The destruction of red blood cells is facilitated by a group of large phagocytic cells found in the spleen, liver, bone marrow, and lymph nodes. These phagocytic cells ingest the hemoglobin from the ruptured cells and break it down in a series of enzymatic reactions. During these reactions, the amino acids from the globulin chains and iron from the heme units are salvaged and reused (Fig. 13-5). The bulk of the heme unit is converted to bilirubin, the pigment of bile, which is insoluble in plasma and attaches to the plasma proteins for transport. Bilirubin is removed from the blood by the liver and conjugated with glucuronide to render it water soluble so that it can be excreted in the bile. The plasma-insoluble form of bilirubin is referred to as unconjugated bilirubin; the water-soluble form is referred to as conjugated bilirubin. Serum levels of conjugated and unconjugated bilirubin can be measured in the laboratory and are reported as direct and indirect, respectively. If red cell

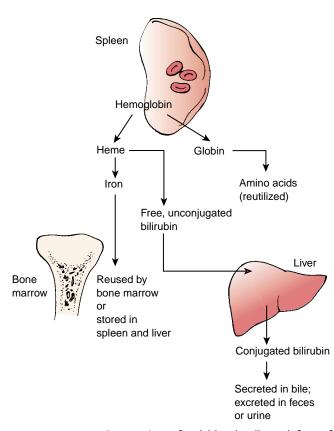


FIGURE 13-5 Destruction of red blood cells and fate of hemoglobin.

destruction and consequent bilirubin production are excessive, unconjugated bilirubin accumulates in the blood. This results in a yellow discoloration of the skin, called *jaundice*.

When red blood cell destruction takes place in the circulation, as in hemolytic anemia, the hemoglobin-binding protein called *haptoglobin*. Other plasma proteins, such as albumin, can also bind hemoglobin. With extensive intravascular destruction of red blood cells, hemoglobin levels may exceed the hemoglobin-binding capacity of haptoglobin. When this happens, free hemoglobin appears in the blood (*i.e.*, hemoglobinemia) and is excreted in the urine (*i.e.*, hemoglobinuria). Because excessive red blood cell destruction can occur in hemolytic transfusion reactions, urine samples are tested for free hemoglobin after a transfusion reaction.

Red Cell Metabolism and Hemoglobin Oxidation

The red blood cell, which lacks mitochondria, relies on glucose and the glycolytic pathway for its metabolic needs. The enzyme-mediated anaerobic metabolism of glucose generates the ATP needed for normal membrane function and ion transport. The depletion of glucose or the functional deficiency of one of the glycolytic enzymes leads to the premature death of the red blood cell. An offshoot of the glycolytic pathway is the production of 2,3-diphosphoglycerate (2,3-DPG), which binds to the hemoglobin molecule and reduces the affinity of hemoglobin for oxygen. This facilitates the release of oxygen at the tissue level. An increase in the concentration of 2,3-DPG occurs in conditions of chronic hypoxia such as chronic lung disease, anemia, and high altitudes.

The oxidation of hemoglobin—the combining of hemoglobin with oxygen-can be interrupted by certain chemicals (e.g., nitrates and sulfates) and drugs that oxidize hemoglobin to an inactive form. For example, the nitrite ion reacts with hemoglobin to produce methemoglobin, which has a low affinity for oxygen. Large doses of nitrites can result in high levels of methemoglobin, causing pseudocyanosis and tissue hypoxia. For example, sodium nitrate, which is used in curing meats, can produce methemoglobin when taken in large amounts. In nursing infants, the intestinal flora is capable of converting significant amounts of inorganic nitrate (e.g., from well water) to nitrite. A hereditary deficiency of glucose 6-phosphate dehydrogenase (G6PD; to be discussed) predisposes to oxidative denaturation of hemoglobin, with resultant red cell injury and lysis. Hemolysis occurs as the result of oxidative stress generated by either an infection or exposure to certain drugs.

Laboratory Tests

Red blood cells can be studied by means of a sample of blood (Table 13-1). In the laboratory, automated blood cell counters rapidly provide accurate measurements of red cell content and cell indices. The *red blood cell count* (RBC) measures the total number of red blood cells in 1 mm³ of blood. The *percentage of reticulocytes* (normally approximately 1%) provides an index of the rate of red cell production. The *hemoglobin* (grams per 100 mL of blood) measures the hemoglobin content of the blood. The major components of blood are the red cell mass and plasma volume. The *hematocrit* measures the volume of red

220 Unit Three: Alterations in the Hematologic System

TABLE 13-1	Standard Laboratory Values for Red Blood Cells						
Test		Normal Values	Significance				
Red blood cell count (RBC)							
Men		$4.2-5.4 imes 10^{6}/\mu L$	Number of red cells in the blood				
Women		$3.6-5.0 \times 10^{6}/\mu$ L					
Reticulocytes		1.0%–1.5% of total RBC	Rate of red cell production				
Hemoglobin							
Men		14–16.5 g/dL	Hemoglobin content of the blood				
Women		12–15 g/dL					
Hematocrit							
Men		40%–50%	Volume of cells in 100 mL of blood				
Women		37%–47%					
Mean corpuscular volume		85–100 fL/red cell	Size of the red cell				
Mean corpuscular hemoglobin concentration		31–35 g/dL	Concentration of hemoglobin in the red cell				
Mean cell hemog	obin	27–34 pg/cell	Red cell mass				

cell mass in 100 mL of plasma volume. To determine the hematocrit, a sample of blood is placed in a glass tube, which is then centrifuged to separate the cells and the plasma. The hematocrit may be deceptive because it varies with the quantity of extracellular fluid, rising with dehydration and falling with overexpansion of extracellular fluid volume.

Red cell indices are used to differentiate types of anemias by size or color of red cells. The *mean corpuscular volume* (MCV) reflects the volume or size of the red cells. The MCV falls in microcytic (small cell) anemia and rises in macrocytic (large cell) anemia. Some anemias are normocytic (*i.e.*, cells are of normal size or MCV). The *mean corpuscular hemoglobin concentration* (MCHC) is the concentration of hemoglobin in each cell. Hemoglobin accounts for the color of red blood cells. Anemias are described as *normochromic* (normal color or MCHC) or *hypochromic* (decreased color or MCHC). *Mean cell hemoglobin* (MCH) refers to the mass of the red cell and is less useful in classifying anemias.

A stained blood smear provides information about the size, color, and shape of red cells and the presence of immature or abnormal cells. If blood smear results are abnormal, examination of the bone marrow may be important. Marrow commonly is aspirated with a special needle from the posterior iliac crest or the sternum. The aspirate is stained and observed for number and maturity of cells and abnormal types.

In summary, the red blood cell provides the means for transporting oxygen from the lungs to the tissues. Red cells develop from stem cells in the bone marrow and are released as reticulocytes into the blood, where they become mature erythrocytes. Red blood cell production is regulated by the hormone erythropoietin, which is produced by the kidney in response to a decrease in oxygen levels. The life span of a red blood cell is approximately 120 days. Red cell destruction normally occurs in the spleen, liver, bone marrow, and lymph nodes. In the process of destruction, the iron from the hemoglobin is returned to the bone marrow for reuse in red cell production or taken to the liver or other tissues for storage. The heme portion of the hemoglobin molecule is converted to bilirubin. Bilirubin, which is insoluble in plasma, attaches to plasma proteins for transport in the blood. It is removed from the blood by the liver and conjugated to a water-soluble form so that it can be excreted in the bile.

The red blood cell, which lacks mitochondria, relies on glucose and the glycolytic pathway for its metabolic needs. The end-product of the glycolytic pathway, 2,3-DPG, increases the release of oxygen to the tissues during conditions of hypoxia by reducing hemoglobin's affinity for oxygen.

In the laboratory, automated blood cell counters rapidly provide accurate measurements of red cell content and cell indices. A stained blood smear provides information about the size, color, and shape of red cells and the presence of immature or abnormal cells. If blood smear results are abnormal, examination of the bone marrow may be important.

ANEMIA

Anemia is defined as an abnormally low hemoglobin level, number of circulating red blood cells, or both, resulting in diminished oxygen-carrying capacity of the blood. Anemia usually results from excessive loss (*i.e.*, bleeding) or destruction (*i.e.*, hemolysis) of red blood cells or from deficient red blood cell production because of a lack of nutritional elements or bone marrow failure.

Anemia is not a disease, but an indication of some disease process or alteration in body function. The manifestations of anemia can be grouped into three categories: (1) impaired oxygen transport and recruitment of compensatory mechanisms; (2) alterations in hemoglobin levels and red cell number and appearance; and (3) signs and symptoms associated with the pathologic process that is causing the anemia. The manifestations of anemia also depend on its severity, the rapidity of its development, and the affected person's age and health status. With rapid blood loss, circulatory shock and circulatory collapse may occur. Because the body adapts to slowly developing anemia, the amount of red cell mass lost may reach 50% without the occurrence of signs and symptoms.¹

In anemia, the oxygen-carrying capacity of hemoglobin is reduced, causing tissue hypoxia. Tissue hypoxia can give rise

KEY CONCEPTS

ANEMIA

- Anemia, which is a deficiency of red cells or hemoglobin, results from excessive loss (blood loss anemia), increased destruction (hemolytic anemia), or impaired production of red blood cells (irondeficiency, megaloblastic, and aplastic anemias).
- Blood loss anemia is characterized by loss of ironcontaining red blood cells from the body; hemolytic anemia involves destruction of red blood cells in the body with iron being retained in the body.
- Manifestations of anemia are caused by the decreased presence of hemoglobin in the blood (pallor), tissue hypoxia due to deficient oxygen transport (weakness and fatigue), and recruitment of compensatory mechanisms (tachycardia and palpitations) designed to increase oxygen delivery to the tissues.

to fatigue, weakness, dyspnea, and sometimes angina. Brain hypoxia results in headache, faintness, and dim vision. The redistribution of the blood from cutaneous tissues or a lack of hemoglobin causes pallor of the skin, mucous membranes, conjunctiva, and nail beds. Tachycardia and palpitations may occur as the body tries to compensate with an increase in cardiac output. A flow-type systolic heart murmur may result from the turbulence caused by a decrease in blood viscosity. Ventricular hypertrophy and high-output heart failure may develop in persons with severe anemia, particularly those with pre-existing heart disease. Erythropoiesis is accelerated and may be recognized by diffuse bone pain and sternal tenderness. The production of 2,3-DPG is a compensatory mechanism that reduces the hemoglobin affinity for oxygen, as evidenced by a shift to the right in the oxygen-hemoglobin saturation curve; this causes more oxygen to be released to the tissues rather than remaining bound to hemoglobin.

Blood Loss Anemia

With anemia caused by bleeding, iron and other components of the erythrocyte are lost from the body. Blood loss may be acute or chronic.

Acute blood loss is accompanied by a loss of vascular volume and carries with it a risk of hypovolemia and shock (see Chapter 18). The red cells are normal in size and color. Hemodilution caused by movement of fluid into the vascular compartment produces a fall in red blood cell count, hemoglobin, and hematocrit. The hypoxia that results from blood loss stimulates red cell production by the bone marrow. If the bleeding is controlled and sufficient iron stores are available, the red cell concentration returns to normal within 3 to 4 weeks.

Chronic blood loss does not affect blood volume but instead leads to iron-deficiency anemia when iron stores are depleted. Because of compensatory mechanisms, patients commonly have no symptoms until the hemoglobin level is less than 8 g/dL. The red cells that are produced have too little hemoglobin, giving rise to microcytic hypochromic anemia.

Hemolytic Anemia

Hemolytic anemia is characterized by the (1) premature destruction of red cells, (2) retention in the body of iron and the other products of hemoglobin destruction, and (3) marked increase in erythropoiesis within the bone marrow.² Almost all types of hemolytic anemia are distinguished by normocytic and normochromic red cells. Because of the red blood cell's shortened life span, the bone marrow usually is hyperactive, resulting in an increase in the number of reticulocytes in the circulating blood. As with other types of anemias, the person experiences easy fatigability, dyspnea, and other signs and symptoms of impaired oxygen transport. The person may also have an increase in serum bilirubin and mild jaundice.

In hemolytic anemia, red cell breakdown can occur in the vascular compartment, or it can result from phagocytosis within the reticuloendothelial system. Intravascular hemolysis occurs as a result of complement fixation in transfusion reactions, mechanical injury, or toxic factors. It is characterized by hemoglobinemia and hemoglobinuria. Extravascular hemolysis occurs when abnormal red cells are phagocytized in the spleen. A common example is sickle cell anemia.

The cause of hemolytic anemia can be intrinsic or extrinsic to the red blood cell. Intrinsic causes include defects of the red cell membrane, the various hemoglobinopathies, and inherited enzyme defects. Acquired forms of hemolytic anemia are caused by agents extrinsic to the red blood cell, such as drugs, bacterial and other toxins, antibodies, and physical trauma. Although all these factors can cause premature and accelerated destruction of red cells, they cannot all be treated in the same way. Some respond to splenectomy, others respond to treatment with corticosteroid hormones, and still others do not resolve until the primary disorder is corrected.

Inherited Disorders of the Red Cell Membrane

Hereditary spherocytosis, transmitted as an autosomal dominant trait, is the most common inherited disorder of the red cell membrane. The disorder is a deficiency of membrane proteins (*i.e.*, spectrin and ankyrin) that leads to gradual loss of the membrane surface during the life span of the red blood cell, resulting in a tight sphere instead of a concave disk. Although the spherical cell retains its ability to transport oxygen, it is poorly deformable and susceptible to destruction as it passes through the venous sinuses of the splenic circulation. Clinical signs are variable but typically include mild anemia, splenomegaly, jaundice, and bilirubin gallstones. A life-threatening aplastic crisis may occur when a sudden disruption of red cell production (in most cases from a viral infection) causes a rapid drop in hematocrit and the hemoglobin level. The disorder usually is treated with splenectomy to reduce red cell destruction.

Hemoglobinopathies

Hemoglobinopathies represent abnormalities in hemoglobin structure that can lead to accelerated red cell destruction. Two main types of hemoglobinopathies can cause red cell hemolysis: the abnormal substitution of an amino acid in the hemoglobin molecule, as in sickle cell anemia, and the defective synthesis of one of the polypeptide chains that form the globin portion of hemoglobin, as in the thalassemias.

Sickle Cell Disease (Anemia). Sickle cell disease is a chronic disorder resulting in organ failure and premature death. The disorder affects approximately 50,000 (0.1% to 0.2%) black Americans. Approximately 8% of black Americans carry the trait.²

Sickle cell disease results from a point mutation in the β chain of the hemoglobin molecule, with an abnormal substitution of a single amino acid, valine, for glutamic acid. Sickle hemoglobin (HbS) is transmitted by recessive inheritance and can manifest as sickle cell trait (*i.e.*, in heterozygotes with one HbS gene and one normal HbA gene) or sickle cell disease (*i.e.*, in homozygotes with two HbS genes). In the homozygote with sickle cell disease, almost all the hemoglobin is HbS. In the heterozygote with sickle cell trait, only approximately 40% of the hemoglobin is HbS.

In the homozygote, sickling occurs when the HbS becomes deoxygenated.³ The deoxygenated hemoglobin aggregates and polymerizes, creating a semisolid gel that changes the shape and deformability of the cell (Fig. 13-6). Sickling of red cells is initially a reversible process with oxygenation. HbS returns to its normal depolymerized state. However, with repeated episodes of deoxygenation, the cells remain permanently sickled. These sickled red cells are abnormally adhesive, attach to the vessel wall, and cause accumulation of more cells that obstruct blood flow in the microcirculation, leading to tissue hypoxia.³ Sickled cells have a rigid and nondeformable membrane, predisposing to premature destruction and hemolysis. Thus, the life span of the sickled cells is markedly reduced.

Perhaps the most important factor in promoting sickling is the amount of HbS and its interaction with other hemoglobin chains. The person with sickle cell trait who has less HbS has little tendency to sickle except during severe hypoxia and has virtually no symptoms. HbF does not interact with HbS or sickle; therefore, infants with sickle cell disease do not begin to experience the effects of the sickling until sometime



■ FIGURE 13-6 ■ Photograph of a sickled cell and a normal red blood cell. (© Dr. Gopal Murti, Science Photo Library, Science Source/Photo Researchers)

after 4 to 6 months of age, when the HbF has been replaced by HbS.

The factors associated with sickling and consequent blood vessel occlusion in persons with sickle cell disease include cold; stress; physical exertion; infection; illnesses that may cause hypoxia, acidosis, or dehydration; or even such trivial incidents as reduced oxygen tension induced by sleep. The rate of HbS polymerization is affected by the hemoglobin concentration in the cell. Thus dehydration, which increases the hemoglobin concentration, greatly facilitates sickling and vascular obstruction. Acidosis, or a fall in pH, which reduces the affinity of hemoglobin for oxygen, can increase sickling because it enhances the amount of deoxygenated HbS.

Persons with sickle cell disease experience problems associated with severe hemolytic anemia, chronic hyperbilirubinemia, and vaso-occlusion. Chronic hemolysis produces rather severe anemia, with hematocrit levels ranging from 18% to 30%.² The hyperbilirubinemia that results from the breakdown products of hemoglobin often leads to jaundice and the production of pigment stones in the gallbladder. Children with sickle cell disease may experience growth retardation and susceptibility to osteomyelitis.

Vaso-occlusion accounts for the most severe complications of sickle cell disease. An acute pain episode results from vessel occlusion and can affect almost any part of the body. Common sites obstructed by sickled cells include the abdomen, chest, bones, and joints. Multiple areas are frequently involved simultaneously, and symmetric involvement of both extremities is common. The frequency ranges from daily to yearly. Infarctions caused by sluggish blood flow can cause chronic damage to the liver, spleen, heart, kidneys, retina, and other organs. Acute chest syndrome is an atypical pneumonia resulting from pulmonary infarction. It affects approximately 40% of persons with sickle cell disease and is characterized by fever, chest pain, and cough.⁴ The syndrome can cause chronic respiratory insufficiency and is a leading cause of death in sickle cell disease. The most serious complication is stroke resulting from cerebral occlusion. Stroke associated with vessel occlusion occurs in children 1 to 15 years of age and may recur in two thirds of those afflicted.

The spleen is especially susceptible to damage by sickle cell hemoglobin. Because of the spleen's sluggish blood flow and low oxygen tension, hemoglobin is deoxygenated and causes ischemia. Splenic injury begins as early as 3 to 6 months of age with intense congestion and is usually asymptomatic.⁵ The congestion causes functional asplenia and predisposes the person to life-threatening infections by encapsulated organisms such as *Streptococcus pneumoniae*, *Haemophilus influenzae* type b, and *Klebsiella* species. Neonates and small children who have not had time to create antibodies to these organisms rely on the spleen for their removal.⁵

Most children with sickle cell disease are at risk for fulminant septicemia and death during the first 3 years of life, when bacteremia from encapsulated organisms occurs commonly even in normal children. Prophylactic penicillin should be begun as early as 2 months of age and continued until at least 5 years of age.⁶ Maintaining full immunization, including administration of *H. influenzae* and hepatitis B vaccines, is recommended. The National Institutes of Health Committee on Management of Sickle Cell Disease also recommends administration of the 7-valent pneumococcal vaccine (see Chapter 20) beginning at 2 to 6 months of age.⁶ The 7-valent vaccine should be followed by immunization with the 23-valent pneumococcal vaccine at 24 months of age or after.⁶

There is no known cure for sickle cell anemia, so treatment to reduce symptoms includes pain control, hydration, and management of complications. Persons with the disorder need to avoid situations that precipitate sickling episodes, such as infections, cold exposure, severe physical exertion, acidosis, and dehydration. Infections are aggressively treated, and blood transfusions may be warranted in a crisis or given chronically in severe disease.

In the United States, screening programs have been implemented to detect newborns with sickle cell disease and other hemoglobinopathies so that appropriate interventions can be implemented. Hydroxyurea is a promising new treatment for the prevention of complications. The drug allows synthesis of more HbF and less HbS, thereby decreasing sickling.^{4,6} Hematopoietic stem cell transplantation may offer the curative potential for sickle cell disease in the future.⁶

Thalassemias. In contrast to sickle cell anemia, the thalassemias result from absent or defective synthesis of the α or the β chains of hemoglobin. The β -thalassemias represent a defect in β -chain synthesis, and the α -thalassemias represent a defect in α -chain synthesis. The defect is inherited as a mendelian trait, and a person may be heterozygous for the trait and have a mild form of the disease or be homozygous and have the severe form of the disease. Like sickle cell anemia, the thalassemias occur with high degree of frequency in certain populations. The β -thalassemias, sometimes called *Cooley's anemia* or *Mediterranean anemia*, are most common in the Mediterranean populations of southern Italy and Greece, and the α -thalassemias are most common among Asians. Both α - and β -thalassemias are common in Africans and black Americans.

Two factors contribute to the anemia that occurs in the thalassemias: reduced hemoglobin synthesis and an imbalance in globin chain production. In α - and β -thalassemia, defective globin chain production leads to deficient hemoglobin production and the development of a hypochromic microcytic anemia. The unaffected type of chain continues to be synthesized, accumulates in the red cell, interferes with normal maturation, and contributes to red cell destruction and anemia. In the β -thalassemias, the excess α chains are denatured to form precipitates (*i.e.*, Heinz bodies) in the bone marrow red cell precursors.^{2,7} These Heinz bodies impair DNA synthesis and cause damage to the red cell membrane. Severely affected red cell precursors are destroyed in the bone marrow, and those that escape intramedullary death are at increased risk of destruction in the spleen.

The clinical manifestations of the β -thalassemias are based on the severity of the anemia. The presence of one normal gene in heterozygous persons (thalassemia minor) usually results in sufficient normal hemoglobin synthesis to prevent severe anemia.² Persons who are homozygous for the trait (thalassemia major) have severe, transfusion-dependent anemia evident at 6 to 9 months of age. Severe growth retardation affects children with the disorder. Increased hematopoiesis in response to erythropoietin causes bone marrow expansion, impairs bone growth, and causes bone abnormalities. The facial and cranial bones, in particular, tend to be enlarged and distorted.^{2,8} In addition, there is increased iron absorption, and splenomegaly and hepatomegaly resulting from increased red cell destruction.² Excess iron stores, which accumulate from increased dietary absorption and repeated transfusions, can deposit in the myocardium, liver, and pancreas and produce organ injury. Cardiac disease resulting from iron overload and secondary hemochromatosis is a frequent cause of death. Frequent transfusions prevent most of the complications, and iron chelation therapy can reduce the iron overload and extend life into the fifth decade.⁸ Bone marrow transplantation is a potential cure for some patients.

Synthesis of the α chains of the hemoglobin molecule is controlled by two pairs of genes; thus, α -thalassemia shows great variations in severity. Silent carriers who have deletion of a single α -thalassemia gene and those with deletion of two genes usually have no symptoms. The most severe form of α -thalassemia occurs in infants in whom all four α -thalassemia genes are deleted.^{8,9} Such a defect results in a hemoglobin molecule (Hb Bart's) that is formed exclusively from the chains of HbF. Hb Bart's, which has an extremely high oxygen affinity, cannot release oxygen in the tissues. Affected infants experience severe hypoxia and are stillborn or die shortly after birth unless intrauterine transfusions are given. Deletion of three of the four α -chain genes leads to unstable aggregates of β chains called hemoglobin H (HbH). The β chains are more soluble than the α chains, and their accumulation is less toxic to the red cells, so that senescent, rather than precursor, red cells are affected. Most persons with HbH have only mild to moderate hemolytic anemia, and manifestations of ineffective erythropoiesis (*i.e.*, bone marrow expansion and iron overload) are absent.

Inherited Enzyme Defects

The most common inherited enzyme defect that results in hemolytic anemia is a deficiency of G6PD. The gene that determines this enzyme is located on the X chromosome, and the defect is expressed only in males and homozygous females. There are many genetic variants of this disorder. The African variant has been found in 10% of black Americans.² The disorder makes red cells more vulnerable to oxidants and causes direct oxidation of hemoglobin to methemoglobin and the denaturing of the hemoglobin molecule to form Heinz bodies, which precipitate within the red blood cell. Hemolysis usually occurs as the damaged red blood cells move through the narrow vessels of the spleen, causing hemoglobinemia, hemoglobinuria, and jaundice. The hemolysis is short-lived, occurring 2 to 3 days after the trigger event. In blacks, the defect is mildly expressed and is not associated with chronic hemolytic anemia unless triggered by oxidant drugs, acidosis, or infection. The antimalarial drug primaquine, the sulfonamides, nitrofurantoin, aspirin, phenacetin, some chemotherapeutics, and other drugs cause hemolysis. Free radicals generated by phagocytes during infections also are possible triggers.

Acquired Hemolytic Anemias

Several acquired factors exogenous to the red blood cell produce hemolysis by direct membrane destruction or by antibodymediated lysis. Various drugs, chemicals, toxins, venoms, and infections such as malaria destroy red cell membranes. Hemolysis can also be caused by mechanical factors such as prosthetic heart valves, vasculitis, and severe burns. Obstructions in the microcirculation, as in disseminated intravascular coagulation, thrombotic thrombocytopenic purpura, and renal disease, may traumatize the red cells by producing turbulence and changing pressure gradients.

Many hemolytic anemias are immune mediated, caused by antibodies that damage the red cell membrane. Autoantibodies may be produced by a person in response to drugs and disease. Alloantibodies come from an exogenous source and are responsible for transfusion reactions and hemolytic disease of the newborn.

The autoantibodies that cause red cell destruction are of two types: warm-reacting antibodies of the immunoglobulin G (IgG) type, which are maximally active at 37°C, and coldreacting antibodies of the immunoglobulin M (IgM) type, which are optimally active at or near 4°C. The warm-reacting antibodies cause no morphologic or metabolic alteration in the red cell. Instead, they react with antigens on the red cell membrane, causing destructive changes that lead to spherocytosis, with subsequent phagocytic destruction in the spleen or reticuloendothelial system. They lack specificity for the ABO antigens but may react with the Rh antigens. The hemolytic reactions associated with the warm-reacting antibodies occur with an incidence of approximately 10 per 1 million. The reactions have a rapid onset, and persons usually have mild jaundice and manifestations of anemia. There are varied causes; approximately 50% are idiopathic, and 50% are drug induced or are related to cancers of the lymphoproliferative system (e.g., chronic lymphocytic leukemia, lymphoma) or collagen diseases (e.g., systemic lupus erythematosus).² The drug-related hemolytic anemias are usually benign conditions. The principle offenders are the antihypertensive drug α -methyldopa and the antiarrhythmic agent quinidine.8

The cold-reacting antibodies activate complement. Chronic hemolytic anemia caused by cold-reacting antibodies occurs with lymphoproliferative disorders and as an idiopathic disorder of unknown cause. The hemolytic process occurs in distal body parts, where the temperature may fall to less than 30°C. Vascular obstruction by red cells results in pallor, cyanosis of the body parts exposed to cold temperatures, and Raynaud's phenomenon (see Chapter 15). Hemolytic anemia caused by cold-reacting antibodies develops in only a few persons and is rarely severe.

The Coombs' test, or the antiglobulin test, is used to diagnose immune hemolytic anemias. It detects the presence of antibody or complement on the surface of the red cell.

Anemias of Deficient Red Cell Production

Anemia may result from the decreased production of erythrocytes by the bone marrow. A deficiency of nutrients for hemoglobin synthesis (iron) or DNA synthesis (cobalamin or folic acid) may reduce red cell production by the bone marrow. A deficiency of red cells also results when the marrow itself fails or is replaced by nonfunctional tissue.

Iron-Deficiency Anemia

Iron deficiency is a common worldwide cause of anemia affecting persons of all ages. The anemia results from dietary deficiency, loss of iron through bleeding, or increased demands. Because iron is a component of heme, a deficiency leads to decreased hemoglobin synthesis and consequent impairment of oxygen delivery. Body iron is used repeatedly. When red cells become senescent and are broken down, their iron is released and reused in the production of new red cells. Despite this efficiency, small amounts of iron are lost in the feces and need to be replaced by dietary uptake. Iron balance is maintained by the absorption of 0.5 to 1.5 mg daily to replace the 1 mg lost in the feces. The average Western diet supplies this amount. The absorbed iron is more than sufficient to supply the needs of most individuals but may be barely adequate in women and young children. Dietary deficiency of iron is uncommon in developed countries, except in certain populations. Most iron is derived from meat, and when meat is not available, as for deprived populations, or is not a dietary constituent, as for vegetarians, iron deficiency may occur.

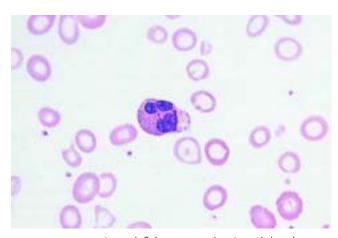
The usual reason for iron deficiency in adults is chronic blood loss because iron cannot be recycled to the pool. In men and postmenopausal women, blood loss may occur from gastrointestinal bleeding because of peptic ulcer, intestinal polyps, hemorrhoids, or cancer. Excessive aspirin intake may cause undetected gastrointestinal bleeding. In women, menstruation may account for an average of 1.5 mg of iron lost per day, causing a deficiency.¹⁰ Although cessation of menstruation removes a major source of iron loss in the pregnant woman, iron requirements increase at this time and deficiency is common. The expansion of the mother's blood volume requires approximately 500 mg of additional iron, and the growing fetus requires approximately 360 mg during pregnancy. During the postnatal period, lactation requires approximately 1.0 mg of iron daily.¹⁰

A child's growth places extra demands on the body. Blood volume increases, with a greater need for iron. Iron requirements are proportionally higher in infancy (3 to 24 months) than at any other age, although they are also increased in childhood and adolescence. In infancy, the two main causes of iron-deficiency anemia are low iron levels at birth because of maternal deficiency and a diet consisting mainly of cow's milk, which is low in absorbable iron. Adolescents are also susceptible to iron deficiency because of high requirements due to growth spurts, dietary deficiencies, and menstrual loss.¹¹

Iron deficiency anemia is characterized by low hemoglobin and hematocrit values, decreased iron stores, and low serum iron and ferritin levels. The red cells are decreased in number and are microcytic and hypochromic. Poikilocytosis (irregular shape) and anisocytosis (irregular size) are also present (Fig 13-7). The laboratory values indicate reduced MCHC and MCV. Membrane changes may predispose to hemolysis, causing further loss of red cells.

The manifestations of iron-deficiency anemia are related to lack of hemoglobin and impaired oxygen transport. Depending on the severity of the anemia, fatigability, palpitations, dyspnea, angina, and tachycardia may occur. Epithelial tissue atrophy is common and results in waxy pallor, brittle hair and nails, smooth tongue, sores in the corners of the mouth, and sometimes dysphagia and decreased acid secretion. A poorly understood symptom that sometimes is seen is pica, the bizarre compulsive eating of ice, dirt, or other abnormal substances.

The treatment of iron-deficiency anemia is directed toward controlling chronic blood loss, increasing dietary intake of iron, and administering supplemental iron. Ferrous sulfate, which is the usual oral replacement therapy, replenishes iron stores in several months. Parenteral iron (iron dextran) therapy



■ FIGURE 13-7 ■ Iron deficiency anemia. A peripheral smear shows hypochromic and microcytic erythrocytes. Poikilocytosis (irregular shape) and anisocytosis (irregular size) are often observed. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 1077]. Philadelphia: Lippincott-Raven)

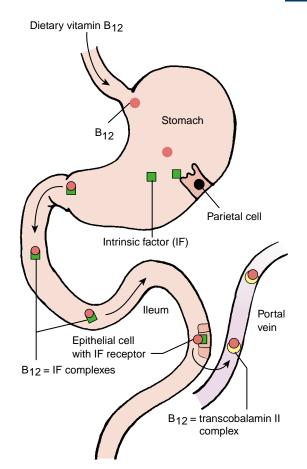
may be used when oral forms are not tolerated or are ineffective. Caution is required because of the possibility of severe hypersensitivity reactions.

Megaloblastic Anemias

Megaloblastic anemias are caused by abnormal nucleic acid synthesis that results in enlarged red cells (MCV >100 fL) and deficient nuclear maturation. Cobalamin (vitamin B_{12}) and folic acid deficiencies are the most common cause of megaloblastic anemias. Because megaloblastic anemias develop slowly, there are often few symptoms until the anemia is far advanced.

Cobalamin (Vitamin B_{12})-Deficiency Anemia. Vitamin B_{12} serves as a cofactor for two important reactions in humans. It is essential for the synthesis of DNA. When it is deficient, nuclear maturation and cell division, especially of the rapidly proliferating red cells, fail to occur. It is also involved in a reaction that prevents abnormal fatty acids from being incorporated into neuronal lipids. This abnormality may predispose to myelin breakdown and production of the neurologic complications of vitamin B_{12} deficiency.

Vitamin B₁₂ is found in all foods of animal origin. Dietary deficiency is rare and usually found only in strict vegetarians who avoid all dairy products as well as meat and fish. It is absorbed by a unique process. After release from the animal protein, vitamin B₁₂ is bound to intrinsic factor, a protein secreted by the gastric parietal cells (Fig. 13-8). The vitamin B₁₂-intrinsic factor complex travels to the ileum, where membrane receptors allow the binding of the complex and transport of B₁₂ across the membrane. From there it is bound to its carrier protein, transcobalamin II, which carries vitamin B₁₂ in the circulation to its storage and tissue sites. Any defects in this pathway may cause a deficiency. An important cause of vitamin B₁₂ deficiency is pernicious anemia, resulting from a hereditary atrophic gastritis. As discussed in Chapter 27, immunemediated chronic atrophic gastritis is a disorder that destroys the gastric mucosa, with loss of parietal cells and production of



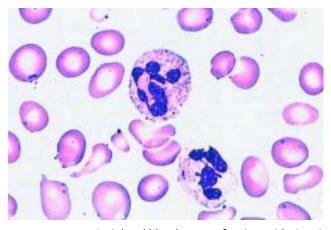
■ FIGURE 13-8 ■ Absorption of vitamin B₁₂.

antibodies that interfere with the binding of vitamin B_{12} to the intrinsic factor. Other causes of vitamin B_{12} deficiency anemia include gastrectomy, ileal resection, and malabsorption syndromes in which vitamin B_{12} and other vitamin B compounds are poorly absorbed.

The hallmark of vitamin B_{12} deficiency is megaloblastic anemia. When vitamin B_{12} is deficient, the red cells that are produced are abnormally large because of excess ribonucleic acid production of hemoglobin and structural protein (Fig. 13-9). They have flimsy membranes and are oval, rather than biconcave. These oddly shaped cells have a short life span that can be measured in weeks, rather than months. The MCV is elevated, and the MCHC is normal.

Neurologic changes that accompany the disorder are caused by deranged methylation of myelin protein. Demyelination of the dorsal and lateral columns of the spinal cord causes symmetric paresthesias of the feet and fingers, loss of vibratory and position sense, and eventual spastic ataxia. In more advanced cases, cerebral function may be altered. In some cases, dementia and other neuropsychiatric changes may precede hematologic changes.

Diagnosis of vitamin B_{12} deficiency is made by finding an abnormally low vitamin B_{12} serum level. The Schilling test, which measures the 24-hour urinary excretion of radiolabeled vitamin B_{12} administered orally, is used to document decreased absorption of vitamin B_{12} . Lifelong treatment consisting of



■ FIGURE 13-9 ■ Peripheral blood smear of patient with vitamin B₁₂ deficiency (pernicious anemia) that shows prominent megablastic anemia. The erythrocytes are large, often with oval shape, and are associated with poikilocytosis and teardrop shapes. The neutrophils are hypersegmented. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 1076]. Philadelphia: Lippincott-Raven)

intramuscular injections of vitamin B₁₂ reverses the anemia and improves the neurologic status.

Folic Acid-Deficiency Anemia. Folic acid is also required for DNA synthesis and red cell maturation, and its deficiency produces the same type of red cell changes that occur in vitamin B_{12} deficiency anemia (*i.e.*, increased MCV and normal MCHC). Symptoms are also similar, but the neurologic manifestations are not present.

Folic acid is readily absorbed from the intestine. It is found in vegetables (particularly the green leafy types), fruits, cereals, and meats. However, much of the vitamin is lost in cooking. The most common causes of folic acid deficiency are malnutrition or dietary lack, especially in the elderly or in association with alcoholism, and malabsorption syndromes such as sprue or other intestinal disorders. In neoplastic disease, tumor cells compete for folate, and deficiency is common. Some drugs used to treat seizure disorders (*e.g.*, primidone, phenytoin, phenobarbital) and triamterene, a diuretic, predispose to a deficiency by interfering with folic acid absorption. Methotrexate, a folic acid analog used in the treatment of cancer, impairs the action of folic acid by blocking its conversion to the active form.

Because pregnancy increases the need for folic acid 5- to 10-fold, a deficiency commonly occurs. Poor dietary habits, anorexia, and nausea are other reasons for folic acid deficiency during pregnancy. Studies also show an association between folate deficiency and neural tube defects.¹² The Public Health Service recommends that all women of childbearing age should take 400 micrograms (µg) of folic acid daily. It is estimated that as many as 50% of neural tube defects could be prevented.¹² The Institute of Medicine Panel on Folate and Other B Vitamins and Choline recently revised the recommended daily allowance (RDA) for pregnant women to 600 µg.¹³ To ensure adequate folate consumption, the U.S. Food and Drug Administration has issued a recommendation for the addition of folate to cereal grain products.¹⁴

Aplastic Anemia

Aplastic anemia (*i.e.*, bone marrow depression) describes a primary condition of bone marrow stem cells that results in a reduction of all three hematopoietic cell lines—red blood cells, white blood cells, and platelets—with fatty replacement of bone marrow. Pure red cell aplasia, in which only the red cells are affected, rarely occurs.

Anemia results from the failure of the marrow to replace senescent red cells that are destroyed and leave the circulation, although the cells that remain are of normal size and color. At the same time, because the leukocytes, particularly the neutrophils, and the thrombocytes have a short life span, a deficiency of these cells usually is apparent before the anemia becomes severe.

The onset of aplastic anemia may be insidious, or it may strike with suddenness and great severity. It can occur at any age. The initial presenting symptoms include weakness, fatigability, and pallor caused by anemia. Petechiae (*i.e.*, small, punctate skin hemorrhages) and ecchymoses (*i.e.*, bruises) often occur on the skin, and bleeding from the nose, gums, vagina, or gastrointestinal tract may occur because of decreased platelet levels. The decrease in the number of neutrophils increases susceptibility to infection.

Among the causes of aplastic anemia are exposure to high doses of radiation, chemicals, and toxins that suppress hematopoiesis directly, or through immune mechanisms. Chemotherapy and irradiation commonly result in bone marrow depression. Identified toxic agents include benzene, the antibiotic chloramphenicol, and the alkylating agents and antimetabolites used in the treatment of cancer (see Chapter 5). Aplastic anemia caused by exposure to chemical agents may be an idiosyncratic reaction because it affects only certain susceptible persons. It typically occurs weeks after use of a drug is initiated. Such reactions often are severe and sometimes irreversible and fatal. Aplastic anemia can develop in the course of many infections and has been reported most often as a complication of viral hepatitis, mononucleosis, and other viral illnesses, including acquired immunodeficiency syndrome (AIDS). In two thirds of cases, the cause is unknown, and these are called *idiopathic aplastic anemia.*

Therapy for aplastic anemia in the young and severely affected includes stem cell replacement by bone marrow or peripheral blood transplantation. For those who are not transplantation candidates, immunosuppressive therapy with lymphocyte immune globulin (*i.e.*, antithymocyte globulin) prevents suppression of proliferating stem cells, producing remission in as many as 50% of patients.¹⁵ Patients with aplastic anemia should avoid the offending agents and be treated with antibiotics for infection. Red cell transfusions to correct the anemia and platelets and corticosteroid therapy to minimize bleeding may also be required.

Chronic Disease Anemias

Anemia often occurs as a complication of chronic infections, inflammation, and cancer. Chronic diseases commonly associated with anemia include AIDS, osteomyelitis, rheumatoid arthritis, and Hodgkin's disease. It is theorized that the short life span, deficient red cell production, and low serum iron are caused by actions of macrophages and lymphocytes in response to cell injury. Macrophages sequester iron in the spleen and contribute to red cell destruction, and the lymphocytes release cytokines that suppress erythropoietin production and action.¹⁶ The mild to moderate anemia is usually reversed when the underlying disease is treated.

Chronic renal failure almost always results in a normocytic, normochromic anemia, primarily because of a deficiency of erythropoietin. Uremic toxins also interfere with the actions of erythropoietin and red cell production. They also cause hemolysis and bleeding tendencies, which contribute to the anemia. Until recently, dialysis and red cell transfusions constituted the only therapy. Recombinant erythropoietin injected several times each week for 10 or more weeks dramatically elevates the hemoglobin level and hematocrit to a range of 32% to 38% and eliminates the need for transfusions.¹⁷ Oral iron is usually required for a good response.

In summary, anemia is a condition of an abnormally low hemoglobin level, number of circulating red blood cells, or both. Anemia can result from excessive blood loss, red cell destruction caused by hemolysis, or deficient hemoglobin or red cell production. Blood loss anemia can be acute or chronic. With bleeding, iron and other components of the erythrocyte are lost from the body. Hemolytic anemia is characterized by the premature destruction of red cells, with retention in the body of iron and the other products of red cell destruction. Hemolytic anemia can be caused by defects in the red cell membrane, hemoglobinopathies (sickle cell anemia or thalassemia), or inherited enzyme defects (G6PD deficiency). Acquired forms of hemolytic anemia are caused by agents extrinsic to the red blood cell, such as drugs, bacterial and other toxins, antibodies, and physical trauma. Iron-deficiency anemia, which is characterized by decreased hemoglobin synthesis, can result from dietary deficiency, loss of iron through bleeding, or increased demands for red cell production. Vitamin B₁₂ and folic acid deficiency impair red cell production by interfering with DNA synthesis. Aplastic anemia is caused by bone marrow suppression and usually results in a reduction of white blood cells and platelets, as well as red blood cells.

The manifestations of anemia include those associated with impaired oxygen transport, recruitment of compensatory mechanisms, and the underlying process causing the anemia. commonly is seen in men between the ages of 40 and 60 years. In polycythemia vera, the manifestations are related to an increase in the red cell count, hemoglobin level, and hematocrit with increased blood volume and viscosity. Commonly reported symptoms include headache, dizziness, and some difficulty with hearing and vision because of decreased cerebral blood flow. Hypertension is common, the result of an increase in blood viscosity. Venous stasis gives rise to a plethoric appearance or dusky redness-even cyanosis-particularly of the lips, fingernails, and mucous membranes. Because of the increased concentration of blood cells, the person may experience itching and pain in the fingers or toes, and the hypermetabolism may induce night sweats and weight loss. Thrombosis, caused by increased blood viscosity and stagnation of blood flow, is a common complication of polycythemia vera and the major cause of morbidity and mortality.¹⁸

The goal of treatment in primary polycythemia is to reduce blood viscosity. This can be done by withdrawing blood by means of periodic phlebotomy to reduce red cell volume. Control of platelet and white cell counts is accomplished by suppressing bone marrow function with chemotherapy or radiation therapy.

Secondary polycythemia results from a physiologic increase in the level of erythropoietin, commonly as a compensatory response to hypoxia. The causes of secondary polycythemia due to hypoxia include living at high altitudes, chronic heart and lung disease, and smoking. Treatment of secondary polycythemia focuses on relieving hypoxia. For example, continuous low-flow oxygen therapy can be used to correct the severe hypoxia that occurs in some persons with chronic obstructive lung disease.

In summary, polycythemia describes a condition in which the red blood cell mass is increased. It can present as a relative, primary, or secondary disorder. Relative polycythemia results from a loss of vascular fluid and is corrected by replacing the fluid. Primary polycythemia, or polycythemia vera, is a proliferative disease of the bone marrow with an absolute increase in total red blood cell mass accompanied by elevated white cell and platelet counts. Secondary polycythemia results from increased erythropoietin levels caused by hypoxic conditions such as chronic heart and lung disease. Many of the manifestations of polycythemia are related to increased blood volume and viscosity that lead to hypertension and stagnation of blood flow.

POLYCYTHEMIA

Polycythemia is an abnormally high total red blood cell mass with a hematocrit greater than 54% in males and 51% in females.¹⁸ It is categorized as relative, primary, or secondary. In relative polycythemia, the hematocrit rises because of a loss of plasma volume without a corresponding decrease in red cells. This may occur with water deprivation, excess use of diuretics, or gastrointestinal losses. Relative polycythemia is corrected by increasing the vascular fluid volume.

Primary polycythemia, or polycythemia vera, is a proliferative disease of the pluripotent cells of the bone marrow characterized by an absolute increase in total red blood cell mass accompanied by elevated white cell and platelet counts. It most

AGE-RELATED CHANGES IN RED BLOOD CELLS

Red Cell Changes in the Neonate

At birth, changes in the red blood cell indices reflect the transition to extrauterine life and the need to transport oxygen from the lungs (Table 13-2). Hemoglobin concentrations at birth are high, reflecting the high synthetic activity in utero to provide adequate oxygen delivery.¹⁹ Toward the end of the first postnatal week, the hemoglobin concentration begins to decline, gradually falling to a minimum value at approximately age

8 Unit Three: Alterations in the Hematologic System

TABLE 13-2	Red Cell Values for Term Infants					
Age	RBC × 10 ⁶ /mL Mean ± SD	Hb (g/dL) Mean ± SD	Hct (%) Mean ± SD	MCV (fL) Mean ± SD		
Days						
1	5.14 ± 0.7	19.3 ± 2.2	61 ± 7.4	119 ± 9.4		
4	5.00 ± 0.6	18.6 ± 2.1	57 ± 8.1	114 ± 7.5		
7	$\textbf{4.86} \pm \textbf{0.6}$	17.9 ± 2.5	56 ± 9.4	118 ± 11.2		
Weeks						
1–2	4.80 ± 0.8	17.3 ± 2.3	54 ± 8.3	112 ± 19.0		
3–4	4.00 ± 0.6	14.2 ± 2.1	43 ± 5.7	105 ± 7.5		
8–9	3.40 ± 0.5	10.7 ± 0.9	31 ± 2.5	93 ± 12.0		
11–12	3.70 ± 0.3	11.3 ± 0.9	33 ± 3.3	88 ± 7.9		

Hb, hemoglobin; Hct, hematocrit; MCV, mean corpuscular volume.

(Adapted from Matoth Y., Zaizor R., Varsano I. [1971]. Postnatal changes in some red cell parameters. Acta Paediatrica Scandinavica 60, 317)

2 months. The red cell count, hematocrit, and MCV likewise fall. The factors responsible for the decline include reduced red cell production and plasma dilution caused by increased blood volume with growth. Neonatal red cells also have a shorter life span of 50 to 70 days and are thought to be more fragile than those of older persons. During the early neonatal period, there is also a switch from HbF to HbA. The amount of HbF in term infants varies from 53% to 95% of the total hemoglobin and decreases by approximately 3% per week after birth.²⁰ In an infant 6 months of age, HbF usually accounts for less than 2% of total hemoglobin. The switch to HbA provides greater unloading of oxygen to the tissues because HbA has a lower affinity for oxygen than does HbF. Infants who are small for gestational age, those born to mothers who have diabetes or smoke, and those who experienced hypoxia in utero have higher total hemoglobin levels, higher HbF levels, and a delayed switch to HbA.

A physiologic anemia of the newborn develops at approximately 2 months of age. It seldom produces symptoms and cannot be altered by nutritional supplements. Anemia of prematurity, an exaggerated physiologic response in low–birthweight infants, is thought to result from a poor erythropoietin response. The hemoglobin level rapidly declines after birth to a low of 7 to 10 g/dL at approximately 6 weeks of age. Signs and symptoms include apnea, poor weight gain, pallor, decreased activity, and tachycardia. In infants born before 33 weeks' gestation or those with hematocrits of less than 33%, the clinical features are more evident. One study suggests that the protein content of breast milk may not be sufficient for hematopoiesis in the premature infant. Protein supplementation significantly increases the hemoglobin concentrations between the ages of 4 and 10 weeks.¹⁹

Anemia at birth, characterized by pallor, congestive heart failure, or shock, usually is caused by hemolytic disease of the newborn. Bleeding from the umbilical cord, internal hemorrhage, congenital hemolytic disease, or frequent blood sampling are other possible causes of anemia. The severity of symptoms and presence of coexisting disease may warrant red cell transfusion.

Hyperbilirubinemia in the Neonate

Hyperbilirubinemia, an increased level of serum bilirubin, is a common cause of jaundice in the neonate. A benign, selflimited condition, it most often is related to the developmental state of the neonate. Rarely, cases of hyperbilirubinemia are pathologic and may lead to kernicterus and serious brain damage.

In the first week of life, approximately 60% of term and 80% of preterm neonates have jaundice.²¹ This physiologic jaundice appears in term infants on the second or third day of life. Under normal circumstances, the indirect bilirubin in umbilical cord blood is 1 to 3 mg/dL and rises at a rate of less than 5 mg/dL/24 hours, peaking at 5 to 6 mg/dL between the second and fourth days and decreasing to less than 2 mg/dL between the fifth and seventh days of life.²¹ The increase in bilirubin is related to the increased red cell breakdown and the inability of the immature liver to conjugate bilirubin. Premature infants exhibit a similar but slower rise in serum bilirubin level, perhaps because of poor hepatic uptake and reduced albumin binding of bilirubin. This generally results in higher bilirubin levels, with peak levels of 8 to 12 mg/dL being reached on the fifth to seventh day. Most neonatal jaundice resolves within 1 week and is untreated.

A search to determine the cause of the jaundice is usually made when (1) the jaundice appears during the first 24 to 36 hours of life or persists beyond 10 to 14 days, (2) serum bilirubin rises at a rate greater than 5 mg/dL/24 hours, (3) serum bilirubin is greater than 12 mg/dL in a full term infant or 10 to 14 mg/dL in preterm infants, (4) or the directreacting bilirubin is greater than 2 mg/dL at any time.

Many factors cause elevated bilirubin levels in the neonate: breast-feeding, hemolytic disease of the newborn, hypoxia, infections, and acidosis. Bowel or biliary obstruction and liver disease are less common causes. Associated risk factors include prematurity, Asian ancestry, and maternal diabetes. Breast milk jaundice occurs in approximately 2% of breast-fed infants.²¹ These neonates accumulate significant levels of unconjugated bilirubin 7 days after birth that reach maximum levels during the third week of life. It is thought that the breast milk contains fatty acids that inhibit bilirubin conjugation in the neonatal liver. A factor in breast milk is also thought to increase the absorption of bilirubin in the duodenum. This type of jaundice disappears if breast-feeding is discontinued. Breast-feeding can be resumed in 3 to 4 days without any hyperbilirubinemia ensuing.

Hyperbilirubinemia places the neonate at risk for the development of a neurologic syndrome called kernicterus. This condition is caused by the accumulation of unconjugated bilirubin in brain cells. Unconjugated bilirubin is lipid soluble, crosses the permeable blood-brain barrier of the neonate, and is deposited in cells of the basal ganglia, causing brain damage. The precise blood level of unconjugated bilirubin or duration of exposure necessary to produce toxic effects is largely unknown. The less mature the infant, the greater the susceptibility to kernicterus. Asphyxia and hyperosmolality may damage the blood-brain barrier and increase the risk of brain damage.²¹

Symptoms may appear 2 to 5 days after birth in term infants and as late as 7 days in premature infants. Lethargy, poor feeding, and short-term behavioral changes may be evident in mildly affected infants. Severe manifestations include rigidity, tremors, ataxia, and hearing loss. Extreme cases cause seizures and death. Most survivors of severe hyperbilirubinemia are seriously damaged and by 3 years of age exhibit involuntary muscle spasm, seizures, mental retardation, and deafness.

Hyperbilirubinemia in the neonate is treated with phototherapy or exchange transfusion. Phototherapy is more commonly used to treat infants with jaundice and reduce the risk of kernicterus. Exposure to fluorescent light in the blue range of the visible spectrum (420- to 470-nm wavelength) reduces bilirubin levels. Bilirubin in the skin absorbs the light energy and is converted to a structural isomer that is more water soluble and can be excreted in the stool and urine. Exchange transfusion is considered when signs of kernicterus are evident or hyperbilirubinemia is sustained or rising and unresponsive to phototherapy.

Hemolytic Disease of the Newborn

Erythroblastosis fetalis, or hemolytic disease of the newborn, occurs in Rh-positive infants of Rh-negative mothers who have been sensitized. The mother can produce anti-Rh antibodies from pregnancies in which the infants are Rh positive or by blood transfusions of Rh-positive blood. The Rh-negative mother usually becomes sensitized during the first few days after delivery, when fetal Rh-positive red cells from the placental site are released into the maternal circulation. Because the antibodies take several weeks to develop, the first Rh-positive infant of an Rh-negative mother usually is not affected. Infants with Rh-negative blood have no antigens on their red cells to react with the maternal antibodies and are not affected.

After an Rh-negative mother has been sensitized, the Rh antibodies from her blood are transferred to subsequent infants through the placental circulation. These antibodies react with the red cell antigens of the Rh-positive infant, causing agglutination and hemolysis. This leads to severe anemia with compensatory hyperplasia and enlargement of the bloodforming organs, including the spleen and liver, in the fetus. Liver function may be impaired, with decreased production of albumin causing massive edema, called hydrops fetalis. If blood levels of unconjugated bilirubin are abnormally high because of red cell hemolysis, there is danger of kernicterus developing in the infant, resulting in severe brain damage or death.

Several advances have served to significantly decrease the threat to infants born to Rh-negative mothers: prevention of sensitization, antenatal identification of the at-risk fetus, and intrauterine transfusion to the affected fetus. The injection of Rh immune globulin (i.e., gamma-globulin-containing Rh antibody) prevents sensitization in Rh-negative mothers who have given birth to Rh-positive infants if administered at 28 weeks' gestation and within 72 hours of delivery, abortion, genetic amniocentesis, or fetal-maternal bleeding. After sensitization has developed, the immune globulin is of no value. Since 1968, the year Rh immune globulin was introduced, the incidence of sensitization of Rh-negative women has dropped dramatically. Early prenatal care and screening of maternal blood continue to be important in reducing immunization. Efforts to improve therapy are aimed at production of monoclonal anti-D, the Rh antibody.

In the past, approximately 20% of erythroblastotic fetuses died in utero. Fetal Rh phenotyping can now be performed to identify at-risk fetuses during the first trimester using fetal blood or amniotic cells.²² Hemolysis in these fetuses can be treated by intrauterine transfusions of red cells through the umbilical cord. Exchange transfusions are administered after birth by removing and replacing the infant's blood volume with type O Rh-negative blood. The exchange transfusion removes most of the hemolyzed red cells and some of the total bilirubin, treating the anemia and hyperbilirubinemia.

Red Cell Changes With Aging

Aging is associated with red cell changes. Bone marrow cellularity declines with age, from approximately 50% cellularity at age 65 years to approximately 30% at age 75 years. The decline may reflect osteoporosis, rather than a decrease in hematopoietic cells.23

Hemoglobin levels decline after middle age. In studies of men older than 60 years of age, mean hemoglobin levels ranged from 15.3 to 12.4 g/dL, with the lowest levels found in the oldest persons. The decline is less in women, with mean levels ranging from 13.8 to 11.7 mg/dL.23 In most elderly persons with no symptoms, lower hemoglobin levels result from iron deficiency and anemia of chronic disease. Orally administered iron is poorly used in older adults, despite normal iron absorption. Underlying neoplasms also may contribute to anemia in this population.

In summary, hemoglobin concentrations at birth are high, reflecting the in utero need for oxygen delivery; toward the end of the first postnatal week, these levels begin to decline, gradually falling to a minimum value at approximately 2 months of age. During the early neonatal period, there is a shift from fetal to adult hemoglobin. Many infants have physiologic jaundice because of hyperbilirubinemia during the first week of life, probably related to increased red cell breakdown and the inability of the infant's liver to conjugate bilirubin. The term kernicterus describes elevated levels of lipid-soluble, unconjugated bilirubin, which can be toxic to brain cells.

Depending on severity, the condition is treated with phototherapy or exchange transfusions (or both). Hemolytic disease of the newborn occurs in Rh-positive infants of Rh-negative mothers who have been sensitized. It involves hemolysis of infant red cells in response to maternal Rh antibodies that have crossed the placenta. Administration of Rh immune globulin to the mother within 72 hours of delivery of an Rh-positive infant, abortion, or amniocentesis prevents sensitization.

Aging is associated with red cell changes. Bone marrow cellularity decreases, and there is a decrease in hemoglobin.

REVIEW QUESTIONS

Relate the function of the red blood cell to the manifestations of anemia.

Explain why the appearance of an increased number of reticulocytes in the blood is suggestive of blood loss and why a low ferritin level suggests the need for iron replacement therapy.

Explain why infants with sickle cell anemia usually do not display evidence of the disease until their hemoglobin F has been replaced with hemoglobin A.

Explain why fever, extreme exercise, and going to high altitudes produce sickling in persons with sickle cell anemia.

What is the common reason for the development of irondeficiency anemia in infancy, adolescence, and in pregnant women?

Describe the relation between vitamin B₁₂ deficiency and megaloblastic anemia. Explain why a person with sickle cell anemia who takes folic acid may not show evidence of megaloblastic anemia on laboratory tests but may have progressive neurologic changes caused by vitamin B₁₂ deficiency.

Explain why ecchymosis, signs of platelet deficiency, and decreased resistance to infection occur before signs of a decreased red blood cell count in persons with aplastic anemia.

Explain the pathogenesis of secondary polycythemia in persons with chronic lung disease.

Describe the pathogenesis of hemolytic disease of the newborn and compare the difference between conjugated and unconjugated bilirubin in terms of causing neurologic damage.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

 Beck W.S. (1991). Erythropoiesis and introduction to the anemias. In Beck W.S. (Ed.), *Hematology* (5th ed., pp. 27, 29). Cambridge, MA: MIT Press.

- Cotran R.S., Kumar V., Collins T. (Eds.). (1999). Robbins pathologic basis of disease (6th ed., pp. 610, 611, 614). Philadelphia: W.B. Saunders.
- 3. Beutler E. (1995). The sickle cell diseases and related disorders. In Beutler E., Lichtman A., Coller B.S., et al. (Eds.), *Williams' hematology* (5th ed., p. 616). New York: McGraw-Hill.
- Steinberg M.H. (1999). Management of sickle cell disease. New England Journal of Medicine 340, 1021–1030.
- Lane P. (1996). Sickle cell disease. Pediatric Clinics of North America 43, 639–666.
- National Institutes of Health. (2002). The management of sickle cell disease. NIH Publication No. 02-2117. (On-line.) Available at http://www.nhlbi.nih.gov/health/prof/blood/sickle/index.htm.
- Olivieri N.F. (1999). The β-thalassemias. New England Journal of Medicine 344, 99–109.
- Bonner H., Bagg A., Cossman J. (1999). The blood and lymphoid organs. In Rubin E., Farber J.L. (Eds.), *Pathology* (3rd ed., pp. 1066– 1087). Philadelphia: Lippincott Williams & Wilkins.
- Honig G.R. (2000). Hemoglobin disorders. In Behrman R.E., Kliegman R.M, Jenson H.B. (Eds.), *Nelson textbook of pediatrics* (16th ed., pp. 1478–1488). Philadelphia: W.B. Saunders.
- Brittenham G.M. (2000). Disorders of iron metabolism: Iron deficiency and overload. In Hoffman R., Benz E.J., Shattil S.J., et al. (Eds.), *Hematology: Basic principles and practice* (3rd ed., pp. 405, 413). New York: Churchill Livingstone.
- Schwartz E. (2000). Anemia of inadequate production. In Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.), Nelson textbook of pediatrics (16th ed., pp. 1469–1471). Philadelphia: W.B. Saunders.
- 12. Hoffbrand A.V., Herbert V. (1999). Nutritional anemias. Seminars in Hematology 36 (Suppl. 7), 13–23.
- Bailey L.B. (2000). New standard for dietary folate intake in pregnant women. American Journal of Clinical Nutrition 71 (Suppl.), 1304S– 1307S.
- Antony A.C. (2000). Megaloblastic anemias. In Hoffman R., Benz E.J., Shattil S.J., et al. (Eds.), *Hematology: Basic principles and practice* (3rd ed., p. 476). New York: Churchill Livingstone.
- Young N.S., Maciejewski J.P. (2000). Aplastic anemias. In Hoffman R., Benz E.J., Shattil S.J., et al. (Eds.), *Hematology: Basic principles and practice* (3rd ed., pp. 316, 318). New York: Churchill Livingstone.
- Means R.T. Jr. (1999). Advances in the anemia of chronic disease. International Journal of Hematology 70, 7–12.
- Caro J., Erslev A.J. (1995). Anemia of chronic renal failure. In Beutler E., Lichtman A., Coller B.S., et al. (Eds.), *Williams' hematology* (5th ed., p. 456). New York: McGraw-Hill.
- Linker C.A. (2002). Blood. In Tierney L.M. Jr., McPhee S.J., Papadakis M.A. (Eds.), *Current medical diagnosis & treatment* (41st ed., pp. 535– 537). New York: Lange Medical Books/McGraw-Hill.
- Brown M.S. (1988). Physiologic anemia of infancy: Nutritional factors and abnormal states. In Stockman J.A., Pochedly C. (Eds.), *Developmental and neonatal hematology* (pp. 252, 274). New York: Raven Press.
- Segel G.B. (1995). Hematology of the newborn. In Beutler E., Lichtman A., Coller B.S., et al. (Eds.), Williams' hematology (5th ed., p. 59). New York: McGraw-Hill.
- Stoll B.J., Kliegman R.M. (2000). The fetus and neonatal infant. In Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.), *Nelson textbook of pediatrics* (16th ed., pp. 513–525). Philadelphia: W.B. Saunders.
- Kramer K., Cohen H.J. (2000). Antenatal diagnosis of hematologic disorders. In Hoffman R., Benz E.J., Shattil S.J., et al. (Eds.), *Hematology: Basic principles and practice* (3rd ed., p. 2495). New York: Churchill Livingstone.
- 23. Williams W.J. (1995). Hematology in the aged. In Beutler E., Lichtman A., Coller B.S., et al. (Eds.), *Williams' hematology* (5th ed., p. 73). New York: McGraw-Hill.

Four

Alterations in the Cardiovascular System

CHAPTER

Structure and Function of the Cardiovascular System

Organization of the Circulatory System Pulmonary and Systemic Circulations Volume and Pressure Distribution Principles of Blood Flow Pressure, Flow, and Resistance **Cross-sectional Area and Velocity of Flow** Laminar and Turbulent Flow Wall Tension, Radius, and Pressure **Distention and Compliance** The Heart as a Pump Functional Anatomy of the Heart Pericardium Myocardium Endocardium Heart Valves and Fibrous Skeleton **Cardiac Conduction System Action Potentials** Absolute and Relative Refractory Periods Dysrhythmias and Conduction Disorders Electrocardiography Cardiac Cycle Ventricular Systole and Diastole Atrial Filling and Contraction

Regulation of Cardiac Performance Preload Afterload **Cardiac Contractility** Heart Rate **Blood Vessels and the Peripheral Circulation Blood Vessels** Vascular Smooth Muscle **Arterial System Aortic Pressure Pulses** Venous System Lymphatic System **Microcirculation and Local Control of Blood Flow** Capillaries Autoregulation **Collateral Circulation Neural Control of Circulatory Function Autonomic Regulation of Cardiac Function Autonomic Regulation of Vascular Function** Autonomic Neurotransmitters

he main function of the *circulatory system*, which consists of the heart and blood vessels, is transport. The circulatory system delivers oxygen and nutrients needed for metabolic processes to the tissues, carries waste products from cellular metabolism to the kidneys and other excretory organs for elimination, and circulates electrolytes and hormones

KEY CONCEPTS

FUNCTIONAL ORGANIZATION OF THE CIRCULATORY SYSTEM

- The circulatory system consists of the heart, which pumps blood; the arterial system, which distributes oxygenated blood to the tissues; the venous system, which collects deoxygenated blood from the tissues and returns it to the heart; and the capillaries, where exchange of gases, nutrients, and wastes occurs.
- The circulatory system is divided into two parts: the low-pressure pulmonary circulation, linking the transport function of the circulation with the gas exchange function of the lungs; and the high-pressure systemic circulation, providing oxygen and nutrients to the tissues.
- The circulation is a closed system, so the output of the right and left heart must be equal over time for effective functioning of the circulation.

needed to regulate body function. This process of nutrient delivery is carried out with exquisite precision so that the blood flow to each tissue of the body is exactly matched to tissue need.

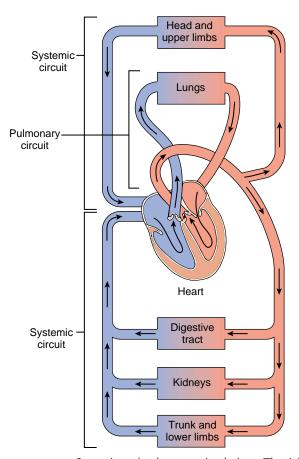
ORGANIZATION OF THE CIRCULATORY SYSTEM

Pulmonary and Systemic Circulations

The circulatory system can be divided into two parts: the *pul-monary circulation*, which moves blood through the lungs and creates a link with the gas exchange function of the respiratory system, and the *systemic circulation*, which moves blood throughout all the other tissues of the body (Fig. 14-1). The blood that is in the heart and pulmonary circulation is sometimes referred to as the *central circulation*, and that outside the central circulation.

The pulmonary circulation consists of the right heart, the pulmonary artery, the pulmonary capillaries, and the pulmonary veins. The large pulmonary vessels are unique in that the pulmonary artery is the only artery that carries deoxygenated venous blood and the pulmonary veins, the only veins that carry oxygenated arterial blood. The systemic circulation consists of the left heart, the aorta and its branches, the capillaries that supply the brain and peripheral tissues, and the systemic venous system and the vena cava. The veins from the lower portion of the body empty into the inferior vena cava and those from the head and upper extremities into the superior vena cava. Blood from both the inferior and superior vena cava empties into the right heart.

Although the pulmonary and systemic systems function similarly, they have some important differences. The pulmonary circulation is the smaller of the two and functions with a much lower pressure. Because the pulmonary circulation is located in the chest near to the heart, it functions as a low-



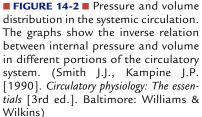
■ **FIGURE 14-1** ■ Systemic and pulmonary circulations. The right side of the heart pumps blood to the lungs, and the left side of the heart pumps blood to the systemic circulation.

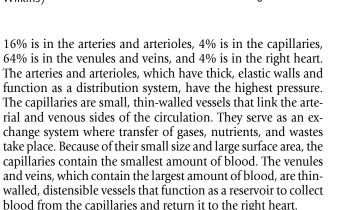
pressure system with a mean arterial pressure of approximately 12 mm Hg. The low pressure of the pulmonary circulation allows blood to move through the lungs more slowly, which is important for gas exchange. Because the systemic circulation must transport blood to distant parts of the body, often against the effects of gravity, it functions as a high-pressure system, with a mean arterial pressure of 90 to 100 mm Hg.

The heart, which propels the blood through the circulation, consists of two pumps in series—the right heart which propels blood through the lungs and the left heart which propels blood to all other tissues of the body. The effective function of the circulatory system requires that the outputs of both sides of the heart pump the same amount of blood over time. If the output of the left heart were to fall below that of the right heart, blood would accumulate in the pulmonary circulation. Likewise, if the right heart were to pump less effectively than the left heart, blood would accumulate in the systemic circulation.

Volume and Pressure Distribution

Blood flow in the circulatory system depends on a blood volume that is sufficient to fill the blood vessels and a pressure difference across the system that provides the force that is needed to move blood forward. As shown in Figure 14-2, approximately 4% of the blood at any given time is in the left heart,

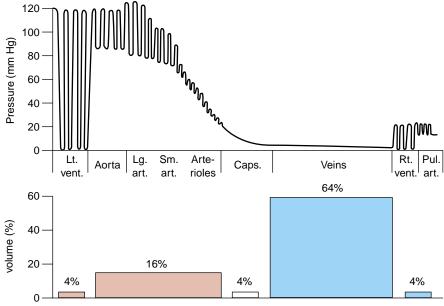




Total blood

Blood moves from the arterial to the venous side of the circulation along a pressure difference, moving from an area of higher pressure to one of lower pressure. The pressure distribution in the different parts of the circulation is almost an inverse of the volume distribution (see Fig. 14-2). The pressure in the arterial side of the circulation, which contains only approximately one sixth of the blood volume, is much greater than the pressure on the venous side of the circulation, which contains approximately two thirds of the blood. This pressure and volume distribution is due in large part to the structure and relative elasticity of the arteries and veins. It is the pressure difference between the arterial and venous sides of the circulation (approximately 84 mm Hg) that provides the driving force for flow of blood in the systemic circulation. The pulmonary circulation has a similar arterial-venous pressure difference, albeit of a lesser magnitude, that facilitates blood flow.

Because the pulmonary and systemic circulations are connected and function as a closed system, blood can be shifted from one circulation to the other. In the pulmonary circulation, the blood volume (approximately 450 mL in the adult) can vary from as low as 50% of normal to as high as 200% of normal. An increase in intrathoracic pressure, such as occurs when exhaling against a closed glottis, impedes venous return to the right heart. This can produce a transient shift from the central to the systemic circulation of as much as 250 mL of blood. Body position also affects the distribution of blood vol-



ume. In the recumbent position, approximately 25% to 30% of the total blood volume is in the central circulation. On standing, this blood is rapidly displaced to the lower part of the body because of the forces of gravity. Because the volume of the systemic circulation is approximately seven times that of the pulmonary circulation, a shift of blood from one system to the other has a much greater effect in the pulmonary than in the systemic circulation.

In summary, the circulatory system functions as a transport system that circulates nutrients and other materials to the tissues and removes waste products. The circulatory system can be divided into two parts: the systemic and the pulmonary circulation. The heart pumps blood throughout the system, and the blood vessels serve as tubes through which blood flows. The arterial system carries blood from the heart to the tissues, and the veins carry it back to the heart. The cardiovascular system is a closed system with a right and left heart connected in series. The systemic circulation, which is served by the left heart, supplies all the tissues except the lungs, which are served by the right heart and the pulmonary circulation. Blood moves throughout the circulation along a pressure gradient, moving from the high-pressure arterial system to the low-pressure venous system. In the circulatory system, pressure is inversely related to volume. The pressure on the arterial side of the circulation, which contains only approximately one sixth of the blood volume, is much greater than the pressure on the venous side of the circulation, which contains approximately two thirds of the blood.

PRINCIPLES OF BLOOD FLOW

The term *hemodynamics* describes the physical principles governing pressure, flow, and resistance as they relate to the circulatory system. The hemodynamics of the circulatory system are complex. The heart is an intermittent pump, and as a result, blood flow in the arterial circulation is pulsatile. The blood vessels are branched, distensible tubes of various dimensions. The blood is a suspension of blood cells, platelets, lipid globules, and plasma proteins. Despite this complexity, the function of the circulatory system can be explained by the principles of basic fluid mechanics that apply to nonbiologic systems, such as household plumbing systems.

Pressure, Flow, and Resistance

The most important factors governing the function of the circulatory system are *volume*, *pressure*, *resistance*, and *flow*. Optimal function requires a volume that is sufficient to fill the vascular compartment and a pressure that is sufficient to ensure blood flow to all body tissues.

Blood flow is determined by two factors: (1) a pressure difference between the two ends of a vessel or group of vessels and (2) the resistance that blood must overcome as it moves through the vessel or vessels (Fig. 14-3). The relation between pressure, resistance, and flow is expressed by the equation F = P/R, in which F is the blood flow, P is the difference in pressure between the two ends of the system, and R is the resistance to flow through the system. In the circulatory system, blood flow is represented by the cardiac output (CO).

The resistance that blood encounters as it flows through the peripheral circulation is referred to as the *peripheral vascular resistance* (PVR) or, sometimes, as the systemic vascular resistance. A helpful equation for understanding factors that affect blood flow ($F = \Delta P \times \pi \times r^4/8n \times L \times viscosity$) was derived by the French physician Poiseuille more than a century ago (see Fig. 14-3). It expands the previous equation, F = P/R, by relating flow to several determinants of resistance—radius, length, and viscosity. According to this equation, the two most important determinants of flow in the circulatory system are a difference in pressure (ΔP) and the vessel radius to the fourth power (r^4). Because flow is directly related to the fourth power of the radius, small changes in vessel radius can produce large changes in flow to an organ or tissue. For example, if the pressure remains constant, the rate of flow is 16 times greater in a

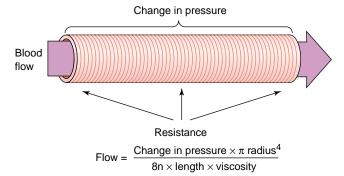
vessel with a radius of 2 mm $(2 \times 2 \times 2 \times 2)$ than in a vessel with a radius of 1 mm.

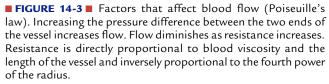
Blood flow is also affected by the viscosity of blood. Viscosity is the resistance to flow caused by the friction of molecules in a fluid. The viscosity of a fluid is largely related to its thickness. The more particles that are present in a solution, the greater the frictional forces that develop between the molecules. Unlike water that flows through plumbing pipes, blood is a nonhomogeneous liquid. It contains blood cells, platelets, fat globules, and plasma proteins that increase its viscosity. The red blood cells, which constitute 40% to 45% of the formed elements of the blood, largely determine the viscosity of the blood. Under special conditions, temperature may affect viscosity. There is a 2% rise in viscosity for each 1°C decrease in body temperature, a fact that helps explain the sluggish blood flow seen in persons with hypothermia. The length (L) of vessels does not usually change, and 8n is a constant that does not change.

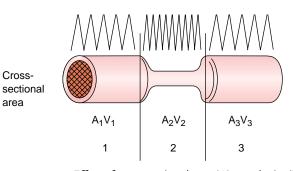
Cross-sectional Area and Velocity of Flow

Velocity is a distance measurement; it refers to the speed or linear movement with time (centimeters per second) with which blood flows through a vessel. *Flow* is a volume measurement (mL/second); it is determined by the cross-sectional area of a vessel and the velocity of flow (Fig. 14-4). When the flow through a given segment of the circulatory system is constant— as it must be for continuous flow—the velocity is inversely proportional to the cross-sectional area of the vessel (*i.e.*, the smaller the cross-sectional area, the greater the velocity of flow). This phenomenon can be compared with cars moving from a two-lane to a single-lane section of a highway. To keep traffic moving at its original pace, cars would have to double their speed in the single-lane section of the highway. So it is with flow in the circulatory system.

The linear velocity of blood flow in the circulatory system varies widely from 30 to 35 cm/second in the aorta to 0.2 to 0.3 mm/second in the capillaries. This is because even though each individual capillary is very small, the total cross-sectional area of all the systemic capillaries greatly exceeds the cross-sectional area of other parts of the circulation. As a result of this large surface area, the slower movement of blood allows ample







■ FIGURE 14-4 ■ Effect of cross-sectional area (A) on velocity (V) of flow. In section 1, velocity is low because of an increase in cross-sectional area. In section 2, velocity is increased because of a decrease in cross-sectional area. In section 3, velocity is again reduced because of an increase in cross-sectional area. Flow is assumed to be constant.

time for exchange of nutrients, gases, and metabolites between the tissues and the blood.

Laminar and Turbulent Flow

Blood flow normally is *laminar*, with the blood components arranged in layers so that the plasma is adjacent to the smooth, slippery endothelial surface of the blood vessel, and the blood cells, including the platelets, are in the center or axis of the bloodstream (Fig. 14-5). This arrangement reduces friction by allowing the blood layers to slide smoothly over one another, with the axial layer having the most rapid rate of flow.

Under certain conditions, blood flow switches from laminar to turbulent flow (see Fig. 14-5). Turbulent flow can be caused by a number of factors, including high velocity of flow, change in vessel diameter, and low blood viscosity. The tendency for turbulence to occur increases in direct proportion to the velocity of flow. Imagine the chaos as cars from a two- or three-lane highway converge on a single-lane section of the highway. The same type of thing happens in blood vessels that have been narrowed by disease processes, such as atherosclerosis. Low blood viscosity allows the blood to move faster and accounts for the transient occurrence of heart murmurs in some persons who are severely anemic. Turbulent flow may predispose to clot formation as platelets and other coagulation factors come in contact with the endothelial lining of the vessel. Turbulent flow often produces sounds that can be heard through the use of a stethoscope. For example, a heart murmur results from turbulent flow through a diseased heart valve.

Wall Tension, Radius, and Pressure

In a blood vessel, wall tension is the force in the vessel wall that opposes the distending pressure inside the vessel. The French astronomer and mathematician Pierre de Laplace described the relationship between wall tension, pressure, and the radius of a vessel or sphere more than 200 years ago. This relationship, which has come to be known as Laplace's law, can be expressed

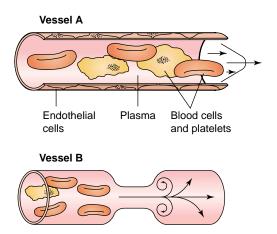
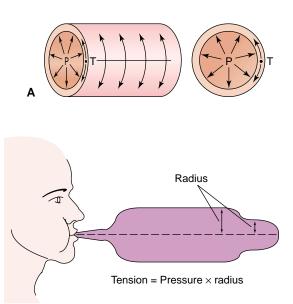


FIGURE 14-5 Laminar and turbulent flow in blood vessels. Vessel A shows streamlined or laminar flow in which the plasma layer is adjacent to the vessel endothelial layer and blood cells are in the center of the bloodstream. Vessel B shows turbulent flow in which the axial location of the platelets and other blood cells is disturbed.

by the equation, P = T/r, in which T is wall tension, P is the intraluminal pressure, and r is vessel radius (Fig. 14-6A). Accordingly, the internal pressure expands the vessel until it is exactly balanced by the tension in the vessel wall. The smaller the radius, the greater the pressure needed to balance the wall tension. Laplace's law can also be used to express the effect of the radius on wall tension (T = Pr). This correlation can be compared with a partially inflated balloon (Fig. 14-6B). Because the pressure is equal throughout, the tension in the part of the balloon with the smaller radius is less than the tension in the section with the larger radius (Fig. 14-6B). The same holds true for an arterial aneurysm in which the tension and risk of rupture increase as the aneurysm grows in size (see Chapter 16).

Laplace's law was later expanded to include wall thickness $(T = P \times r/wall thickness)$. Wall tension is inversely related to wall thickness, such that the thicker the vessel wall, the lower the tension, and vice versa. In hypertension, arterial vessel walls hypertrophy and become thicker, thereby reducing the tension and minimizing wall stress. Laplace's law can also be applied to the pressure required to maintain the patency of small blood vessels. Providing that the thickness of a vessel wall remains constant, it takes more pressure to overcome wall tension and keep a vessel open as its radius decreases in size. The critical closing pressure refers to the point at which vessels collapse so that blood can no longer flow through them. For example, in circulatory shock there is a decrease in blood volume and vessel radii, along with a drop in blood pressure. As a result, many of the small vessels collapse as blood pressure drops to the point where it can no longer overcome the wall tension. The collapse of peripheral veins often makes it difficult



■ **FIGURE 14-6** ■ Laplace's law relates pressure (P), tension (T), and radius in a cylindrical blood vessel (A). The pressure expanding the vessel is equal to the wall tension multiplied by the vessel radius. (B) Effect of the radius of a cylinder on tension. In a balloon, the tension in the wall is proportional to the radius because the pressure is the same everywhere inside the balloon. The tension is lower in the portion of the balloon with the smaller radius. (Rhoades R.A., Tanner G.A. [1996]. Medical physiology [p. 627]. Boston: Little, Brown)

R

to insert venous lines that are needed for fluid and blood replacement.

Distention and Compliance

Compliance refers to the total quantity of blood that can be stored in a given portion of the circulation for each millimeter rise in pressure. Compliance reflects the *distensibility* of the blood vessel. The distensibility of the aorta and large arteries allows them to accommodate the pulsatile output of the heart. The most distensible of all vessels are the veins, which can increase their volume with only slight changes in pressure, allowing them to function as a reservoir for storing large quantities of blood that can be returned to the circulation when it is needed. The compliance of a vein is approximately 24 times that of its corresponding artery, because it is eight times as distensible and has a volume three times as great.

In summary, blood flow is controlled by many of the same mechanisms that control fluid flow in nonbiologic systems. It is influenced by vessel length, pressure differences, vessel radius, blood viscosity, cross-sectional area, and wall tension. The rate of flow is directly related to the pressure difference between the two ends of the vessel and the vessel radius and inversely related to vessel length and blood viscosity. The cross-sectional area of a vessel influences the velocity of flow; as the cross-sectional area decreases, the velocity is increased, and vice versa. Laminar blood flow is flow in which there is layering of blood components in the center of the bloodstream. This reduces frictional forces and prevents clotting factors from coming in contact with the vessel wall. In contrast to laminar flow, turbulent flow is disordered flow, in which the blood moves crosswise and lengthwise in blood vessels. The relation between wall tension, transmural pressure, and radius is described by Laplace's law, which states the pressure needed to overcome wall tension becomes greater as the radius decreases. Wall tension is also affected by wall thickness; it increases as the wall becomes thinner and decreases as the wall becomes thicker.

THE HEART AS A PUMP

The heart is a four-chambered muscular pump approximately the size of a man's fist that beats an average of 70 times each minute, 24 hours each day, 365 days each year for a lifetime. In 1 day, this pump moves more than 1800 gallons of blood throughout the body, and the work performed by the heart over a lifetime would lift 30 tons to a height of 30,000 ft.

Functional Anatomy of the Heart

The heart is located between the lungs in the mediastinal space of the intrathoracic cavity in a loose-fitting sac called the *pericardium*. It is suspended by the great vessels, with its broader side (*i.e.*, base) facing upward and its tip (*i.e.*, apex) pointing downward, forward, and to the left (Fig.14-7).

The wall of the heart is composed of an outer epicardium, which lines the pericardial cavity; the myocardium or muscle

KEY CONCEPTS

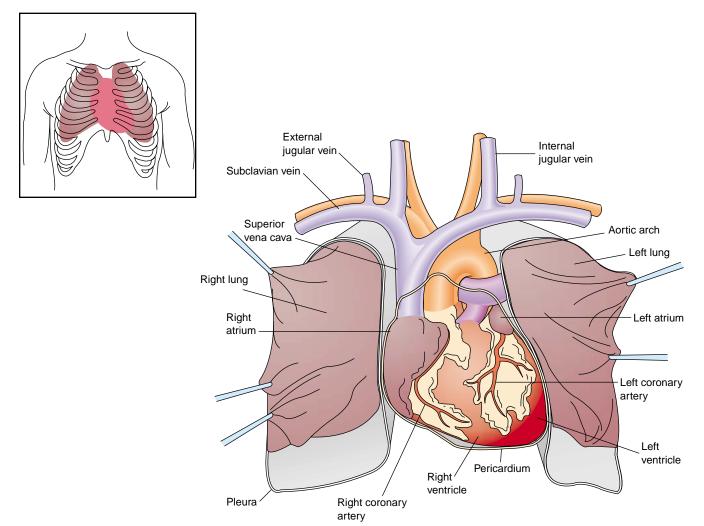
THE HEART

- The heart is a four-chambered pump consisting of two atria (the right atrium, which receives blood returning to the heart from the systemic circulation, and the left atrium, which receives oxygenated blood from the lungs) and two ventricles (a right ventricle, which pumps blood to the lungs, and a left ventricle, which pumps blood into the systemic circulation).
- Heart valves control the direction of blood flow from the atria to the ventricles (the atrioventricular valves), from the right side of the heart to the lungs (pulmonic valve), and from the left side of the heart to the systemic circulation (aortic valve).
- The cardiac cycle is divided into two major periods: systole, when the ventricles are contracting, and diastole, when the ventricles are relaxed and filling.
- The work and efficiency of the heart is determined by the volume of blood it pumps out (preload), the pressure that it must generate to pump the blood out of the heart (afterload), and the rate at which it performs these functions (heart rate).

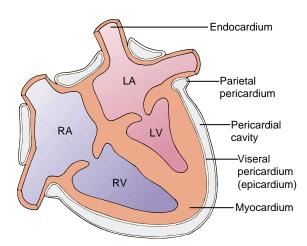
layer; and the smooth endocardium, which lines the chambers of the heart. A fibrous skeleton supports the valvular structures of the heart. The interatrial and interventricular septa divide the heart into a right and a left pump, each composed of two muscular chambers: a thin-walled atrium, which serves as a reservoir for blood coming into the heart, and a thick-walled ventricle, which pumps blood out of the heart. The increased thickness of the left ventricular wall results from the additional work this ventricle is required to perform.

Pericardium

The pericardium forms a fibrous covering around the heart, holding it in a fixed position in the thorax and providing physical protection and a barrier to infection. The pericardium consists of a tough outer fibrous layer and a thin inner serous layer. The outer fibrous layer is attached to the great vessels that enter and leave the heart, the sternum, and the diaphragm. The fibrous pericardium is highly resistant to distention; it prevents acute dilatation of the heart chambers and exerts a restraining effect on the left ventricle. The inner serous layer consists of a visceral layer and a parietal layer. The visceral layer, also known as the epicardium, covers the entire heart and great vessels and then folds over to form the parietal layer that lines the fibrous pericardium (Fig. 14-8). Between the visceral and parietal layers is the pericardial cavity, a potential space that contains 30 to 50 mL of serous fluid. This fluid acts as a lubricant to minimize friction as the heart contracts and relaxes.



■ FIGURE 14-7 ■ Anterior view of the heart and great vessels and their relationship to lungs and skeletal structures of the chest cage (*upper left box*).



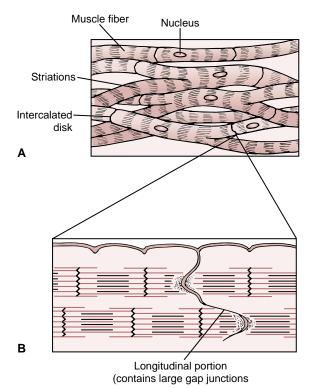
■ FIGURE 14-8 ■ Layers of the heart, showing the visceral pericardium, pericardial cavity, parietal pericardium, myocardium, and endocardium. RA, right atrium; LA, left atrium; RV, right ventricle; LV, left ventricle.

Myocardium

The myocardium, or muscular portion of the heart, forms the wall of the atria and ventricles. Cardiac muscle cells, like skeletal muscle, are striated and composed of *sarcomeres* that contain actin and myosin filaments (see Chapter 1). They are smaller and more compact than skeletal muscle cells and contain many large mitochondria, reflecting their continuous energy needs.

Cardiac muscle contracts much like skeletal muscle, except the contractions are involuntary and the duration of contraction is much longer. Unlike the orderly longitudinal arrangement of skeletal muscle fibers, cardiac muscle cells are arranged as an interconnecting latticework, with their fibers dividing, recombining, and then dividing again (Fig. 14-9). The fibers are separated from neighboring cardiac muscle cells by dense structures called *intercalated disks*. The intercalated disks, which are unique to cardiac muscle, contain gap junctions that serve as low-resistance pathways for passage of ions and electrical impulses from one cardiac cell to another (Fig. 14-9). Thus, the myocardium behaves as a single unit, or *syncytium*, rather than





■ FIGURE 14-9 ■ (A) Cardiac muscle fibers showing the branching structure. (B) Area indicated where cell junctions lie in the intercalated disks.

as a group of isolated units, as does skeletal muscle. When one myocardial cell becomes excited, the impulse travels rapidly so the heart can beat as a unit.

As in skeletal muscle, cardiac muscle contraction involves actin and myosin filaments, which interact and slide along one another during muscle contraction. However, compared with skeletal muscle cells, cardiac muscle cells have less well-defined sarcoplasmic reticulum for storing calcium, and the distance from the cell membrane to the myofibrils is shorter. Because less calcium can be stored in the muscle cells, cardiac muscle relies more heavily than skeletal muscle on an influx of extracellular calcium ions for contraction.

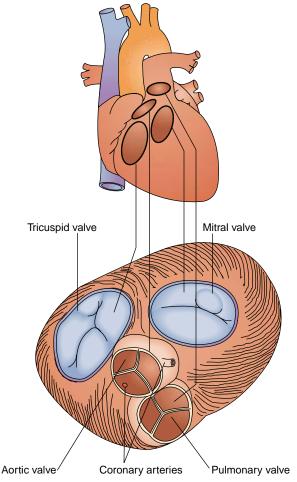
Endocardium

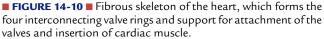
The endocardium is a thin, three-layered membrane that lines the heart and covers the valves. The innermost layer consists of smooth endothelial cells supported by a thin layer of connective tissue. The endothelial lining of the endocardium is continuous with the lining of the blood vessels that enter and leave the heart. The middle layer consists of dense connective tissue with elastic fibers. The outer layer, composed of irregularly arranged connective tissue cells, contains blood vessels and branches of the conduction system and is continuous with the myocardium.

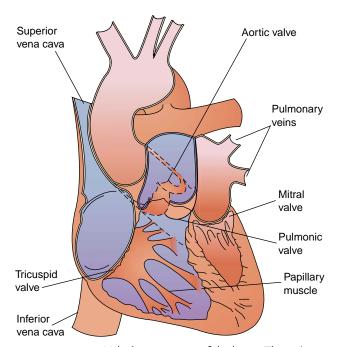
Heart Valves and Fibrous Skeleton

An important structural feature of the heart is its fibrous skeleton, which consists of four interconnecting valve rings and surrounding connective tissue. The fibrous skeleton separates the atria and ventricles and forms a rigid support for attachment of the valves and insertion of the cardiac muscle (Fig. 14-10). The tops of the valve rings are attached to the muscle tissue of the atria, pulmonary trunks, aorta, and valve rings. The bottoms are attached to the ventricular walls. For the heart to function effectively, blood flow must occur in a one-way direction, moving forward through the chambers of the right heart to the lungs and then through the chambers of the left heart to the systemic circulation. This unidirectional flow is provided by the heart's two atrioventricular (*i.e.*, tricuspid and mitral) valves and two semilunar (*i.e.*, pulmonic and aortic) valves (Fig. 14-11).

The atrioventricular (AV) valves control the flow of blood between the atria and the ventricles (Fig. 14-12). The thin edges of the AV valves form cusps, two on the left side of the heart (*i.e.*, *bicuspid* or *mitral valve*) and three on the right side (*i.e.*, *tricuspid valve*). The AV valves are supported by the papillary muscles, which project from the wall of the ventricles, and the chordae tendineae, which attach to the valve. Contraction of the papillary muscles at the onset of systole ensures closure by producing tension on the leaflets of the AV valves before the full force of ventricular contraction pushes against them. The chordae tendineae are cordlike structures that support the







■ FIGURE 14-11 ■ Valvular structures of the heart. The atrioventricular valves are in an open position, and the semilunar valves are closed. There are no valves to control the flow of blood at the inflow channels (*i.e.*, vena cava and pulmonary veins) to the heart.

AV valves and prevent them from everting into the atria during systole.

The *aortic* and *pulmonic* valves control the movement of blood out of the ventricles (Fig.14-12). Because of their half-moon shape, they often are referred to as the semilunar valves.

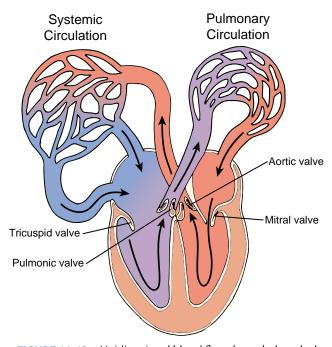


FIGURE 14-12 Unidirectional blood flow through the valvular structures of the heart.

The semilunar valves have three little teacup-shaped leaflets. These cuplike structures collect the *retrograde*, or backward, flow of blood that occurs toward the end of systole, enhancing closure. For the development of a perfect seal along the free edges of the semilunar valves, each valve cusp must have a triangular shape, which is facilitated by a nodular thickening at the apex of each leaflet (Fig. 14-13). The openings for the coronary arteries are located in the aorta just above the aortic valve.

There are no valves at the atrial sites (*i.e.*, venae cavae and pulmonary veins) where blood enters the heart. This means that excess blood is pushed back into the veins when the atria become distended. For example, the jugular veins typically become prominent in severe right-sided heart failure, whereas normally they are flat or collapsed. Likewise, the pulmonary venous system becomes congested when outflow from the left atrium is impeded.

Cardiac Conduction System

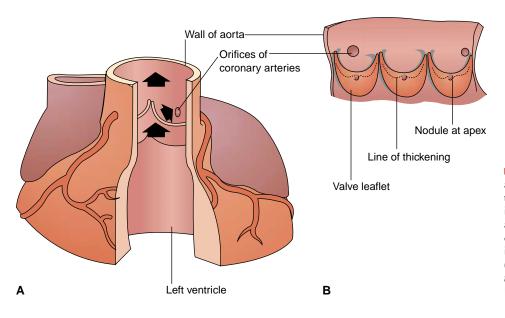
Heart muscle is unique among other muscles in that it is capable of generating and rapidly conducting its own action potentials (*i.e.*, electrical impulses). These action potentials result in excitation of muscle fibers throughout the myocardium. Impulse formation and conduction result in weak electrical currents that spread through the entire body. When electrodes are applied to various positions on the body and connected to an electrocardiograph machine, an electrocardiogram (ECG) can be recorded.

In certain areas of the heart, the myocardial cells have been modified to form the specialized cells of the conduction system (Fig. 14-14). Although most myocardial cells are capable of initiating and conducting impulses, it is this specialized conduction system that maintains the pumping efficiency of the heart. Specialized pacemaker cells generate impulses at a faster rate than do other types of heart tissue, and the conduction tissue transmits impulses at a faster rate than do other types of heart tissue. Because of these properties, the conduction system usually controls the rhythm of the heart.

The conduction system consists of the sinoatrial node (SA node), where the rhythmic impulse is generated; the internodal pathways, which conduct the impulse from the SA node to the atrioventricular (AV) node; the AV node, in which the impulse from the atria is delayed before passing to the ventricles; the AV bundle, which conducts the impulse from the atria to the ventricles; and the left and right bundles of the Purkinje system, which conduct the cardiac impulses to all parts of the ventricles.

The *sinoatrial (SA) node* has the fastest intrinsic rate of firing (60 to 100 beats per minute) and is normally the pacemaker of the heart. The heart essentially has two conduction systems: one that controls atrial activity and one that controls ventricular activity. The *AV node* connects the two conduction systems and provides one-way conduction between the atria and ventricles. Within the AV node, atrial fibers connect with very small junctional fibers of the node itself. The velocity of conduction through these fibers is very slow (approximately one half that of normal cardiac muscle), which greatly delays transmission of the impulse into the AV node. A further delay occurs as the impulse travels through the AV node into the transitional fibers and into the AV bundle, also called the *bundle of His.* This delay provides a mechanical advantage whereby the atria complete

240 Unit Four: Alterations in the Cardiovascular System



■ FIGURE 14-13 ■ Diagram of the aortic valve. (A) The position of the aorta at the base of the ascending aorta is indicated. (B) The appearance of the three leaflets of the aortic valve when the aorta is cut open and spread out, flat. (Cormack D.H. [1987]. Ham's histology [9th ed.]. Philadelphia: J.B. Lippincott)

their ejection of blood before ventricular contraction begins. Under normal circumstances, the AV node provides the only connection between the two conduction systems. The atria and ventricles would beat independently of each other if the transmission of impulses through the AV node were blocked.

The Purkinje fibers lead from the AV node through the AV bundle into the ventricles where they divide to form the *right and left bundle branches* that straddle the interventricular septum. The main trunk of the left bundle branch extends for approximately 1 to 2 cm before fanning out as it enters the septal

KEY CONCEPTS

CARDIAC CONDUCTION SYSTEM

- The cardiac conduction system consists of the SA node, which functions as the pacemaker of the heart; the AV node, which connects the atrial and ventricular conduction systems; and the AV bundle and large Purkinje fibers, which provide for rapid depolarization of the ventricles.
- Cardiac action potentials are controlled by three types of ion channels—the fast sodium channels, which are responsible for the spikelike onset of the action potential; the slower calcium-sodium channels, which are responsible for the plateau; and the potassium channels, which are responsible for the repolarization phase and return of the membrane to the resting potential.
- There are two types of cardiac action potentials: the fast response, which occurs in atrial and ventricular muscle cells and the Purkinje conduction system and uses the fast sodium channels; and the slow response of the SA and AV nodes, which uses the slow calcium channels.

area and divides further into two segments: the *left posterior* and *anterior fascicles*. The Purkinje system has very large fibers that allow for rapid conduction and almost simultaneous excitation of the entire right and left ventricles. This rapid rate of conduction is necessary for the swift and efficient ejection of blood from the heart.

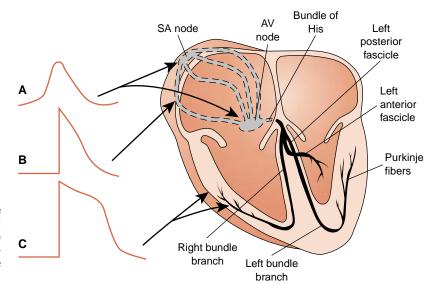
Action Potentials

A stimulus delivered to excitable tissues evokes an action potential that is characterized by a sudden change in voltage resulting from transient depolarization and subsequent repolarization. These action potentials are electrical currents involving the movement or flow of electrically charged ions at the level of the cell membrane (see Chapter 1).

The action potential of cardiac muscle is divided into five phases: phase 0—the upstroke or rapid depolarization; phase 1-early repolarization; phase 2-the plateau; phase 3rapid repolarization; and *phase* 4—the resting membrane potential (Fig. 14-15). Cardiac muscle has three types of membrane ion channels that contribute to the voltage changes that occur during these phases of the action potential. They are the (1) fast sodium channels, (2) slow calcium-sodium channels, and (3) potassium channels. During phase 0 in atrial and ventricular muscle and in the Purkinje system, opening of the fast sodium channels for a few ten-thousandths of a second is responsible for the spikelike onset of the action potential. The point at which the sodium gates open is called the depolarization threshold. When the cell has reached this threshold, a rapid influx of sodium ions to the interior of the membrane causes the membrane potential to shift from a resting membrane potential of approximately -90 mV to +20 mV.

Phase 1 occurs at the peak of the action potential and signifies inactivation of the fast sodium channels with an abrupt decrease in sodium permeability. *Phase 2* represents the plateau of the action potential. It is caused primarily by the slower opening of the calcium-sodium channels, which lasts for a few tenths of a second. Calcium ions entering the muscle during this phase of the action potential play a key role in the contractile process of the cardiac muscle fibers. These

241



■ FIGURE 14-14 ■ Conduction system of the heart and action potentials. (A) Action potential of sinoatrial (SA) and atrioventricular (AV) nodes; (B) atrial muscle action potential; (C) action potential of ventricular muscle and Purkinje fibers.

unique features of the phase 2 plateau cause the action potential of cardiac muscle to last 3 to 15 times longer than that of skeletal muscle and cause a corresponding increased period of contraction.

Phase 3 reflects final rapid repolarization and begins with the downslope of the action potential. During the phase 3 repolarization period, the slow channels close and the influx of calcium and sodium ceases. There is a sharp rise in potassium permeability, contributing to the rapid outward movement of potassium during this phase and facilitating the reestablishment of the resting membrane potential (–90 mV). At the conclusion of phase 3, distribution of sodium and potassium returns to the normal resting state. *Phase 4* is the resting membrane potential. During phase 4, the sodiumpotassium pump is activated, transporting sodium out of the cell and moving potassium back into the cell.

There are two main types of action potentials in the heart-the slow response and the fast response. The slow response, which is initiated by the slow calcium-sodium channels, is found in the SA node, which is the natural pacemaker of the heart, and the conduction fibers of the AV node (see Fig. 14-15). The fast response, which is characterized by opening of the fast sodium channels, occurs in the normal myocardial cells of the atria, the ventricles, and the Purkinje fibers. The fast-response cardiac cells do not normally initiate cardiac action potentials. Instead, these impulses originate in the specialized slow-response cells of the SA node and are conducted to the fast-response myocardial cells in the atria and ventricles, where they effect a change in membrane potential to the threshold level. On reaching threshold, the voltagedependent sodium channels open to initiate the rapid upstroke of the phase 1 action potential. The amplitude and the rate of rise of phase 1 are important to the conduction velocity of the fast response.

The hallmark of the pacemaker cells in the SA and AV nodes is a spontaneous phase 4 depolarization. The membrane permeability of these cells allows a slow inward leak of current to occur through the slow channels during phase 4. This leak continues until the threshold for firing is reached, at which point the cell spontaneously depolarizes. The rate of pacemaker cell discharge varies with the resting membrane potential and the slope of phase 4 depolarization (see Fig. 14-15). Catecholamines (*i.e.*, epinephrine and norepinephrine) increase the heart rate by increasing the slope or rate of phase 4 depolarization. Acetylcholine, which is released during vagal stimulation of the heart, slows the heart rate by decreasing the slope of phase 4.

Absolute and Relative Refractory Periods

The pumping action of the heart requires alternating contraction and relaxation. There is a period in the action potential curve during which no stimuli can generate another action potential (Fig. 14-16). This period, which is known as the absolute refractory period, includes phases 0, 1, 2, and part of phase 3. During this time, the cell cannot depolarize again under any circumstances. In skeletal muscle, the refractory period is very short compared with the duration of contraction, such that a second contraction can be initiated before the first is over, resulting in a summated tetanized contraction. In cardiac muscle, the absolute refractory period is almost as long as the contraction, and a second contraction cannot be stimulated until the first is over. The longer length of the absolute refractory period of cardiac muscle is important in maintaining the alternating contraction and relaxation that is essential to the pumping action of the heart and for the prevention of fatal dysrhythmias. When repolarization has returned the membrane potential to below the threshold potential, but not to the resting membrane potential, the cell is capable of responding to a greater-than-normal stimulus. This condition is referred to as the *relative refractory period*. After the relative refractory period there is a short period of time, called the supernormal excitatory period, during which a weak stimulus can evoke a response. It is during this period that many cardiac dysrhythmias develop.

Dysrhythmias and Conduction Disorders

Dysrhythmias represent disorders of cardiac rhythm. Cardiac dysrhythmias are commonly divided into two categories: supraventricular and ventricular dysrhythmias. The supraventricular dysrhythmias include those that are generated in the SA

242 Unit Four: Alterations in the Cardiovascular System

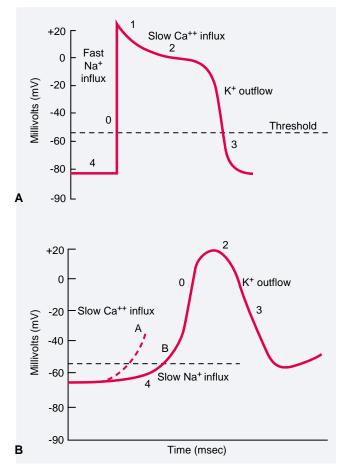
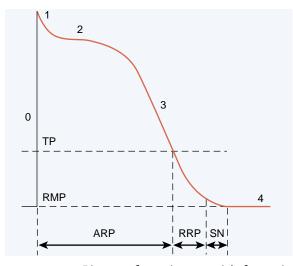


FIGURE 14-15 Changes in action potential recorded from a fast response in cardiac muscle cell (**A**) and from a slow response recorded in the sinoatrial and atrioventricular nodes (**B**). The phases of the action potential are identified by numbers: phase 4, resting membrane potential; phase 0, depolarization; phase 1, brief period of repolarization; phase 2, plateau; phase 3, repolarization. The slow response is characterized by a slow, spontaneous rise in the phase 4 membrane potential to threshold levels; it has a lesser amplitude and shorter duration than the fast response. Increased automaticity (*A*) occurs when the rate of phase 4 depolarization is increased.

node, atria, AV node, and junctional tissues. The ventricular dysrhythmias include those that are generated in the ventricular conduction system and the ventricular muscle. Because the ventricles are the pumping chambers of the heart, ventricular dysrhythmias (*e.g.*, ventricular tachycardia and fibrillation) are the most serious in terms of immediate life-threatening events.

Disorders of cardiac rhythm can range from excessively rapid heart rate (tachyarrhythmias) to an extremely slow heart rate (bradyarrhythmias). Conduction disorders disrupt the flow of impulses through the conduction system of the heart. *Heart block* occurs when the conduction of impulses is blocked, often in the AV node. Under normal conditions, the AV junction provides the only connection for transmission of impulses between the atrial and ventricular conduction systems; in complete heart block, the atria and ventricles beat independently of



■ FIGURE 14-16 ■ Diagram of an action potential of a ventricular muscle cell, showing the threshold potential (TP), resting membrane potential (RMP), absolute refractory period (ARP), relative refractory period (RRP), and supernormal (SN) period.

each other. The most serious effect of some forms of AV block is a slowing of heart rate to the extent that circulation to the brain is compromised. An *ectopic pacemaker* is an excitable focus outside the normally functioning SA node. A *premature ventricular contraction* (PVC) occurs when an ectopic pacemaker initiates a beat. The occurrence of frequent PVCs in the diseased heart predisposes one to the development of other, more serious dysrhythmias, including ventricular tachycardia and ventricular fibrillation.

Fibrillation is the result of disorganized current flow within the atria (atrial fibrillation) or ventricle (ventricular fibrillation). Fibrillation interrupts the normal contraction of the atria or ventricles. In ventricular fibrillation, the ventricle quivers but does not contract. When the ventricle does not contract, there is no cardiac output, and there are no palpable or audible pulses. Ventricular fibrillation is a fatal event unless treated with immediate defibrillation.

Electrocardiography

The ECG is a recording of the electrical activity of the heart. The electrical currents generated by the heart spread through the body to the skin, where they can be sensed by appropriately placed electrodes, amplified, and viewed on an oscilloscope or chart recorder. The deflection points of an ECG are designated by the letters P, Q, R, S, and T. Figure 14-17 depicts the electrical activity of the conduction system on an ECG tracing. The P wave represents the SA node and atrial depolarization; the QRS complex (*i.e.*, beginning of the Q wave to the end of the S wave) depicts ventricular depolarization; and the T wave portrays ventricular repolarization. The isoelectric line between the P wave and the Q wave represents depolarization of the AV node, bundle branches, and Purkinje system (Fig. 14-17). Atrial repolarization occurs during ventricular depolarization and is hidden in the QRS complex.

The horizontal axis of the ECG measures time in seconds, and the vertical axis measures the amplitude of the impulse in

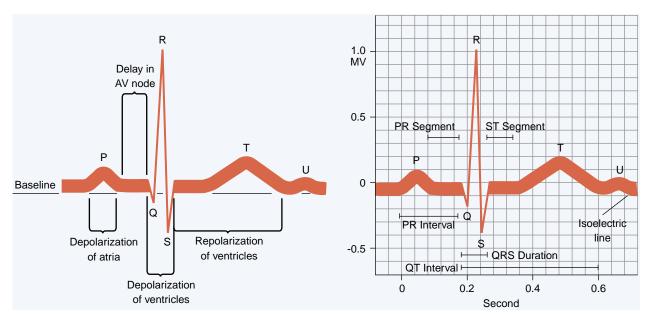


FIGURE 14-17 Diagram of the electrocardiogram (lead II) and representative depolarization and repolarization of the atria and ventricle. The P wave represents atrial depolarization, the QRS complex ventricular depolarization, and the T wave ventricular repolarization. Atrial repolarization occurs during ventricular depolarization and is hidden under the QRS complex.

millivolts (mV). Each heavy vertical line represents 0.2 second, and each thin line represents 0.04 second (see Fig. 14-17). The widths of ECG complexes are commonly referred to in terms of duration of time. On the vertical axis, each heavy horizontal line represents 0.5 mV. The connections of the ECG are arranged such that an upright deflection indicates a positive potential and a downward deflection indicates a negative potential.

The ECG records the potential difference in charge between two electrodes as the depolarization and repolarization waves move through the heart and are conducted to the skin surface. The shape of the recorder tracing is determined by the direction in which the impulse spreads through the heart muscle in relation to electrode placement. A depolarization wave that moves toward the recording electrode registers as a positive, or upward, deflection. Conversely, if the impulse moves away from the recording electrode, the deflection is downward, or negative. When there is no flow of charge between electrodes, the potential is zero, and a straight line is recorded at the baseline of the chart.

Conventionally, 12 leads are recorded for a diagnostic ECG, each providing a unique view of the electrical forces of the heart from a different position on the body's surface. Six limb leads view the electrical forces as they pass through the heart on the frontal or vertical plane. The electrodes for the limb leads are attached to the four extremities or representative areas on the body near the shoulders and lower chest or abdomen. Chest electrodes provide a view of the electrical forces as they pass through the heart on the horizontal plane. They are moved to different positions on the chest, including the right and left sternal borders and the left anterior surface. The right lower extremity lead is used as a ground electrode. When indicated, additional electrodes may be applied to other areas of the body, such as the back or right anterior chest.

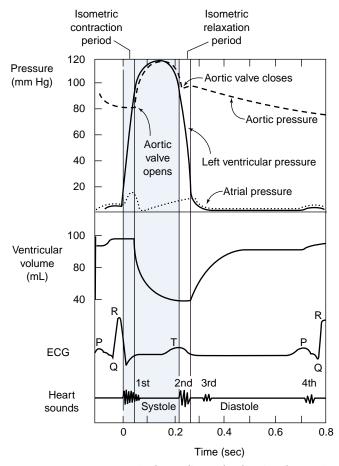
Cardiac Cycle

The term *cardiac cycle* is used to describe the rhythmic pumping action of the heart. The cardiac cycle is divided into two parts: *systole*, the period during which the ventricles are contracting, and *diastole*, the period during which the ventricles are relaxed and filling with blood. Simultaneous changes occur in atrial pressure, ventricular pressure, aortic or pulmonary artery pressure, ventricular volume, the electrocardiogram (ECG), and heart sounds during the cardiac cycle (Fig. 14-18).

Ventricular Systole and Diastole

Ventricular systole is divided into two periods: the isovolumetric contraction period and the ejection period. The isovolumetric contraction period, which begins with the closure of the AV valves and occurrence of the first heart sound, heralds the onset of systole. Immediately after closure of the AV valves, there is an additional 0.02 to 0.03 second during which the semilunar outlet (pulmonic and aortic) valves remain closed. During this period, the ventricular volume remains the same while the ventricles contract, producing an abrupt increase in pressure. The ventricles continue to contract until left ventricular pressure is slightly higher than aortic pressure, and right ventricular pressure is higher than pulmonary artery pressure. At this point, the semilunar valves open, signaling the onset of the ejection period. Approximately 60% of the stroke volume is ejected during the first quarter of systole, and the remaining 40% is ejected during the next two quarters of systole. Little blood is ejected from the heart during the last quarter of systole, although the ventricle remains contracted. At the end of systole, the ventricles relax, causing a precipitous fall in intraventricular pressures. As this occurs, blood from the large arteries flows back toward the ventricles, causing the aortic and

Unit Four: Alterations in the Cardiovascular System



■ FIGURE 14-18 ■ Events in the cardiac cycle, showing changes in aortic pressure, left ventricular pressure, atrial pressure, left ventricular volume, the electrocardiogram (ECG), and heart sounds.

pulmonic valves to snap shut—an event that is marked by the second heart sound.

The aortic pressure reflects changes in the ejection of blood from the left ventricle. There is a rise in pressure and stretching of the elastic fibers in the aorta as blood is ejected into the aorta at the onset of the ejection period. The aortic pressure continues to rise and then begins to fall during the last quarter of systole as blood flows out of the aorta into the peripheral vessels. The incisura, or notch, in the aortic pressure tracing represents closure of the aortic valve. The aorta is highly elastic and as such stretches during systole to accommodate the blood that is being ejected from the left heart during systole. During diastole, recoil of the elastic fibers in the aorta serves to maintain the arterial pressure.

Diastole is marked by ventricular relaxation and filling. After closure of the semilunar valves, the ventricles continue to relax for another 0.03 to 0.06 second. During this time, which is called the isovolumetric relaxation period, ventricular volume remains the same but ventricular pressure drops until it becomes less than atrial pressure. As this happens, the AV valves open, and the blood that has been accumulating in the atria during systole flows into the ventricles. Most of ventricular filling occurs during the first third of diastole, which is called the rapid filling period. During the middle third of diastole, inflow into the ventricles is almost at a standstill. The last third of diastole is marked by atrial contraction, which gives an additional thrust to ventricular filling. When audible, the third heart sound is heard during the rapid filling period of diastole as blood flows into a distended or noncompliant ventricle. A fourth heart sound, when present, occurs during the last third of diastole as the atria contract.

During diastole, the ventricles increase their volume to approximately 120 mL (i.e., the end-diastolic volume), and at the end of systole, approximately 50 mL of blood (i.e., the endsystolic volume) remains in the ventricles (Fig. 14-19). The difference between the end-diastolic and end-systolic volumes (approximately 70 mL) is called the stroke volume. The ejection fraction, which is the stroke volume divided by the enddiastolic volume, represents the fraction or percentage of the diastolic volume that is ejected from the heart during systole.

Atrial Filling and Contraction

Because there are no valves between the junctions of the central veins (i.e., venae cavae and pulmonary veins) and the atria, atrial filling occurs during both systole and diastole. During normal quiet breathing, right atrial pressure usually varies between -2 and +2 mm Hg. It is this low atrial pressure that maintains the movement of blood from the systemic veins into the right atrium and from the pulmonary veins into the left atrium.

Right atrial pressure is regulated by a balance between the ability of the right ventricle to move blood out of the right heart and the pressures that move blood from the venous circulation into the right atrium (venous return). When the heart pumps strongly, right atrial pressure is decreased and atrial filling is enhanced. Right atrial pressure is also affected by changes in intrathoracic pressure. It is decreased during inspiration when intrathoracic pressure becomes more negative, and it is increased during coughing or forced expiration when intrathoracic pressure becomes more positive. Venous return is a reflection of the amount of blood in the systemic circulation that is available for return to the right heart and the forces that move blood back to the right heart. Venous return is in-

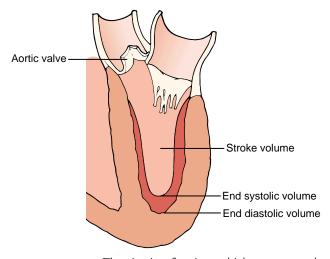


FIGURE 14-19 The ejection fraction, which represents the difference between left ventricular end-diastolic and end-systolic volumes.

creased when the blood volume is expanded or when right atrial pressure falls and is decreased in hypovolemic shock or when right atrial pressure rises.

Although the main function of the atria is to store blood as it enters the heart, these chambers also act as pumps that aid in ventricular filling. Atrial contraction occurs during the last third of diastole. Atrial contraction becomes more important during periods of increased activity when the diastolic filling time is decreased because of an increase in heart rate or when heart disease impairs ventricular filling. In these two situations, the cardiac output would fall drastically were it not for the action of the atria. It has been estimated that atrial contraction can contribute as much as 30% to cardiac reserve during periods of increased need, while having little or no effect on cardiac output during rest.

Regulation of Cardiac Performance

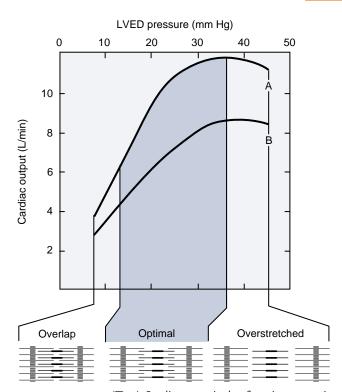
The efficiency and work of the heart as a pump often is measured in terms of *cardiac output* or the amount of blood the heart pumps each minute. The cardiac output (CO) is the product of the *stroke volume* (SV) and the *heart rate* (HR) and can be expressed by the equation: $CO = SV \times HR$. The cardiac output varies with body size and the metabolic needs of the tissues. It increases with physical activity and decreases during rest and sleep. The average cardiac output in normal adults ranges from 3.5 to 8.0 L/minute. In the highly trained athlete, this value can increase to levels as high as 32 L/minute during maximum exercise.

The *cardiac reserve* refers to the maximum percentage of increase in cardiac output that can be achieved above the normal resting level. The normal young adult has a cardiac reserve of approximately 300% to 400%. The heart's ability to increase its output according to body needs mainly depends on four factors: the *preload*, or ventricular filling; the *afterload*, or resistance to ejection of blood from the heart; *cardiac contractility*; and the *heart rate*. Cardiac performance is influenced by the work demands of the heart and the ability of the coronary circulation to meet its metabolic needs.

Preload

The preload represents the volume work of the heart. It is called the *preload* because it is the work imposed on the heart before the contraction begins. Preload represents the amount of blood that the heart must pump with each beat and is largely determined by the venous return to the heart and the accompanying stretch of the muscle fibers.

The increased force of contraction that accompanies an increase in ventricular end-diastolic volume is referred to as the *Frank-Starling mechanism* or Starling's law of the heart (Fig. 14-20). The anatomic arrangement of the actin and myosin filaments in the myocardial muscle fibers is such that the tension or force of contraction is greatest when the muscle fibers are stretched just before the heart begins to contract. The maximum force of contraction and cardiac output is achieved when venous return produces an increase in left ventricular end-diastolic filling (*i.e.*, preload) such that the muscle fibers are stretched about two and one-half times their normal resting length (Fig. 14-20, *curve B*). When the muscle fibers are stretched to this degree, there is optimal overlap of the actin and myosin filaments needed for maximal contraction.



■ FIGURE 14-20 ■ (Top) Starling ventricular function curve in normal heart. An increase in left ventricular end-diastolic (LVED) pressure produces an increase in cardiac output (*curve B*) by means of the Frank-Starling mechanism. The maximum force of contraction and increased stroke volume are achieved when diastolic filling causes the muscle fibers to be stretched about two and one half times their resting length. In *curve A*, an increase in cardiac contractility produces an increase in cardiac output without a change in LVED volume and pressure. (**Bottom**) Stretching of the actin and myosin filaments at the different LVED filling pressures.

The Frank-Starling mechanism allows the heart to adjust its pumping ability to accommodate various levels of venous return. Cardiac output is less when decreased filling causes excessive overlap of the actin and myosin filaments or when excessive filling causes the filaments to be pulled too far apart.

Afterload

The afterload is the pressure or tension work of the heart. It is the pressure that the heart must generate to move blood into the aorta. It is called the *afterload* because it is the work presented to the heart after the contraction has commenced. The systemic arterial blood pressure is the main source of afterload work on the left heart and the pulmonary arterial pressure is the main source of afterload work for the right heart. The afterload work of the left ventricle is also increased with narrowing (*i.e.*, stenosis) of the aortic valve. For example, in the late stages of aortic stenosis, the left ventricle may need to generate systolic pressures as great as 300 mm Hg to move blood through the diseased valve.

Cardiac Contractility

Cardiac contractility refers to the ability of the heart to change its force of contraction without changing its resting (*i.e.*, diastolic) length. The contractile state of the myocardial muscle is determined by biochemical and biophysical properties that govern the actin and myosin interactions in the myocardial cells. It is strongly influenced by the number of calcium ions that are available to participate in the contractile process.

An *inotropic* influence is one that modifies the contractile state of the myocardium independent of the Frank-Starling mechanism (see Fig. 14-20, *curve A*). For instance, sympathetic stimulation produces a positive inotropic effect by increasing the calcium that is available for interaction between the actin and myosin filaments. Hypoxia exerts a negative inotropic effect by interfering with the generation of adenosine triphosphate (ATP), which is needed for muscle contraction.

Heart Rate

The heart rate influences cardiac output and the work of the heart by determining the frequency with which the ventricle contracts and blood is ejected from the heart. Heart rate also determines the time spent in diastolic filling. While systole and the ejection period remain fairly constant across heart rates, the time spent in diastolic and filling of the ventricles becomes shorter as the heart rate increases. This leads to a decrease in stroke volume, and at high heart rates, may produce a decrease in cardiac output. One of the dangers of ventricular tachycardia is a reduction in cardiac output because the heart does not have time to fill adequately.

In summary, the heart is a four-chambered muscular pump that lies in the pericardial sac within the mediastinal space of the intrathoracic cavity. The wall of the heart is composed of an outer epicardium, which lines the pericardial cavity; a fibrous skeleton; the myocardium, or muscle layer; and the smooth endocardium, which lines the chambers of the heart. The four heart valves control the direction of blood flow.

The specialized cells of the heart's conduction system control the rhythmic contraction and relaxation of the heart. The SA node has the fastest inherent rate of impulse generation and acts as the pacemaker of the heart. Impulses from the SA node travel through the atria to the AV node and then to the AV bundle and the ventricular Purkinje system. The AV node provides the only connection between the atrial and ventricular conduction systems. The action potential of cardiac muscle is controlled by the (1) fast sodium channels, (2) slow calcium-sodium channels, and (3) potassium channels. Opening of the fast sodium channels is responsible for the rapid spikelike onset of the ventricular action potential; the slower opening calcium-sodium channels for the plateau of the action potential, and potassium channels for repolarization and return to the resting membrane potential. The absolute refractory period, which represents the time during which a normal cardiac impulse cannot re-excite an already excited area of cardiac muscle, is important in preventing disorders of cardiac rhythm that would disrupt the normal pumping ability of the heart. Disorders of the cardiac conduction system include dysrhythmias and conduction defects. Ventricular dysrhythmias are generally more serious than atrial dysrhythmias because they afford the potential for disrupting the pumping ability of the heart.

The cardiac cycle describes the pumping action of the heart. It is divided into two parts: systole, during which the ventricles contract and blood is ejected from the heart, and diastole, during which the ventricles are relaxed and blood is filling the heart. The stroke volume (approximately 70 mL) represents the difference between the end-diastolic volume (approximately 120 mL) and the end-systolic volume (approximately 50 mL). Atrial contraction occurs during the last third of diastole. Although the main function of the atria is to store blood as it enters the heart, atrial contraction acts to increase cardiac output during periods of increased activity when the filling time is reduced or in disease conditions in which ventricular filling is impaired.

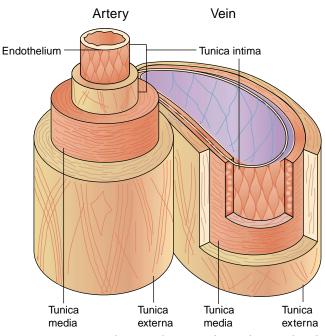
The heart's ability to increase its output according to body needs depends on the preload, or filling of the ventricles (*i.e.*, end-diastolic volume); the afterload, or resistance to ejection of blood from the heart; cardiac contractility, which is determined by the interaction of the actin and myosin filaments of cardiac muscle fibers; and the heart rate, which determines the frequency with which blood is ejected from the heart. The maximum force of cardiac contraction occurs when an increase in preload stretches muscle fibers of the heart to approximately two and one-half times their resting length (*i.e.*, Frank-Starling mechanism).

BLOOD VESSELS AND THE PERIPHERAL CIRCULATION

The vascular system functions in the delivery of oxygen and nutrients and removal of wastes from the tissues. It consists of arteries and arterioles, the capillaries, and the venules and veins. Blood vessels are dynamic structures that constrict and relax to adjust blood pressure and flow to meet the varying needs of the many different tissue types and organ systems. Structures such as the heart, brain, liver, and kidneys require a large and continuous flow to carry out their vital functions. In other tissues, such as the skin and skeletal muscle, the need for blood flow varies with the level of function. For example, there is a need for increased blood flow to the skin during fever and for increased skeletal muscle blood flow during exercise.

Blood Vessels

All blood vessels, except the capillaries, have walls composed of three layers, or coats, called *tunicae* (Fig. 14-21). The *tunica externa*, or *tunica adventitia*, is the outermost covering of the vessel. This layer is composed of fibrous and connective tissues that support the vessel. The *tunica media*, or middle layer, is largely a smooth muscle layer that constricts to regulate and control the diameter of the vessel. The *tunica intima*, or inner layer, has an elastic layer that joins the media and a thin layer of endothelial cells that lie adjacent to the blood. The endothelial layer provides a smooth and slippery inner surface for the vessel. This smooth inner lining, as long as it remains intact, prevents platelet adherence and blood clotting. The layers of the different types of blood vessels vary with vessel function.



■ FIGURE 14-21 ■ Medium-sized artery and vein, showing the relative thickness of the three layers.

The walls of the arterioles, which control blood pressure, have large amounts of smooth muscle. Veins are thin-walled, distensible, and collapsible vessels. Capillaries are single-cell– thick vessels designed for the exchange of gases, nutrients, and waste materials.

Vascular Smooth Muscle

Smooth muscle contracts slowly and generates high forces for long periods with low energy requirements; it uses only 1/10 to 1/300 the energy of skeletal muscle. These characteristics are important in structures, such as blood vessels, that must maintain their tone day in and day out.

Although vascular smooth muscle contains actin and myosin filaments, these contractile filaments are not arranged in striations as they are in skeletal and cardiac muscle. The smooth muscle fibers are instead linked together in a strong cablelike system that generates a circular pull as it contracts. In addition, smooth muscle has less well developed sarcoplasmic reticulum for storing intracellular calcium than do skeletal and cardiac muscle, and it has very few fast sodium channels. Instead, depolarization of smooth muscle relies largely on extracellular calcium, which enters through calcium channels in the muscle membrane. These channels respond to changes in membrane potential or receptor-activated responses to chemical mediators such as norepinephrine. Sympathetic nervous system control of vascular smooth muscle tone occurs by way of receptor-activated channels. In general, α -adrenergic receptors are excitatory and produce vasoconstriction, and β-adrenergic receptors are inhibitory and produce vasodilation. Calcium-channel blocking drugs cause vasodilation by blocking calcium entry through the calcium channels.

Arterial System

The arterial system consists of the large and medium-sized arteries and the arterioles. Arteries are thick-walled vessels with large amounts of elastic fibers. The elasticity of these vessels allows them to stretch during cardiac systole, when the heart contracts and blood is ejected into the circulation, and to recoil during diastole, when the heart relaxes. The arterioles, which are predominantly smooth muscle, serve as resistance vessels for the circulatory system. They act as control valves through which blood is released as it moves into the capillaries. Changes in the activity of sympathetic fibers that innervate these vessels cause them to constrict or to relax as needed to maintain blood pressure. The regulation of arterial blood pressure is discussed further in Chapter 16.

Aortic Pressure Pulse

The delivery of blood to the tissues of the body is dependent on pressure pulsations or waves of pressure that are generated by the intermittent ejection of blood from the left ventricle into the distensible aorta and large arteries of the arterial system. The aortic pressure pulse represents the energy that is transmitted from molecule to molecule along the length of the vessel (Fig. 14-22). In the aorta, this pressure pulse is transmitted at a velocity of 4 to 6 meters/second, which is approximately 20 times faster than the flow of blood. Therefore, the pressure pulse has no direct relation to blood flow and could occur if there was no flow at all. When taking a pulse, it is the pressure pulses that are felt, and it is the pressure pulses that produce the Korotkoff sounds heard during blood pressure measurement. The tip or maximum deflection of the pressure pulsation coincides with the systolic blood pressure, and the minimum point of deflection coincides with the diastolic pressure.

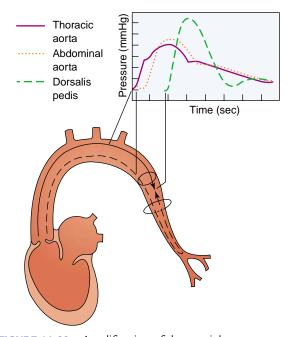
Both the pressure values and the conformation of the pressure wave changes as it moves though the peripheral arteries (Fig. 14-22). As the pressure wave moves out through the aorta into the arteries, it changes as it collides with reflected waves from the periphery. This is why the systolic pressure is higher in the medium-sized arteries than in the aorta even though the diastolic pressure is lower. After its initial amplification, the pressure pulse becomes smaller and smaller as it moves through the smaller arteries and arterioles, until it disappears almost entirely in the capillaries. This allows for continuous, rather than pulsatile, flow in the capillary beds.

Although the pressure pulses usually are not transmitted to the capillaries, there are situations in which this does occur. For example, injury to a finger or other area of the body often results in a throbbing sensation. In this case, extreme dilatation of the small vessels in the injured area produces a reduction in the dampening of the pressure pulse. Capillary pulsations also occur in conditions that cause exaggeration of aortic pressure pulses, such as aortic regurgitation (see Chapter 17).

Venous System

The veins and venules are thin-walled, distensible, and collapsible vessels. The venules collect blood from the capillaries, and the veins transport blood back to the heart. The veins are capable of enlarging and storing large quantities of blood, which can be made available to the circulation as needed. Even though the veins are thin walled, they are muscular. This allows

248 Unit Four: Alterations in the Cardiovascular System



■ FIGURE 14-22 ■ Amplification of the arterial pressure wave as it moves forward in the peripheral arteries. This amplification occurs as a forward-moving pressure wave merges with a backwardmoving reflected pressure wave. (Inset) The amplitude of the pressure pulse increases in the thoracic aorta, abdominal aorta, and dorsalis pedis.

them to contract or expand to accommodate varying amounts of blood. Veins are innervated by the sympathetic nervous system. When blood is lost from the circulation, the veins constrict as a means of maintaining intravascular volume.

The venous system is a low-pressure system, and when a person is in the upright position, blood flow in the venous system must oppose the effects of gravity. Valves in the veins of extremities prevent retrograde flow (Fig. 14-23), and with the

help of skeletal muscles that surround and intermittently compress the veins in a milking manner, blood is moved forward to the heart. Their pressure ranges from approximately 10 mm Hg at the end of the venules to approximately 0 mm Hg at the entrance of the vena cava into the heart. There are no valves in the abdominal or thoracic veins, and blood flow in these veins is heavily influenced by the pressure in the abdominal and thoracic cavities, respectively.

Lymphatic System

The lymphatic system, commonly called the *lymphatics*, serves almost all body tissues, except cartilage, bone, epithelial tissue, and tissues of the central nervous system. However, most of these tissues have prelymphatic channels that eventually flow into areas supplied by the lymphatics. Lymph is derived from interstitial fluids that flow through the lymph channels. It contains plasma proteins and other osmotically active particles that rely on the lymphatics for movement back into the circulatory system. The lymphatic system is also the main route for absorption of nutrients, particularly fats, from the gastrointestinal tract.

The lymphatic system is made up of vessels similar to those of the circulatory system. These vessels commonly travel along with an arteriole or venule or with its companion artery and vein. The terminal lymphatic vessels are made up of a single layer of connective tissue with an endothelial lining and resemble blood capillaries. The lymphatic vessels lack tight junctions and are loosely anchored to the surrounding tissues by fine filaments (Fig. 14-24). The loose junctions permit the entry of large particles, and the filaments hold the vessels open under conditions of edema, when the pressure of the surrounding tissues would otherwise cause them to collapse. The lymph capillaries drain into larger lymph vessels that ultimately empty into the right and left thoracic ducts (Fig. 14-25). The thoracic ducts empty into the circulation at the junctions of the subclavian and internal jugular veins.

Although the divisions are not as distinct as in the circulatory system, the larger lymph vessels show evidence of having intimal, medial, and adventitial layers similar to those of blood vessels. Contraction of the smooth muscle in the medial layer

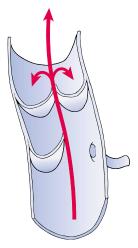
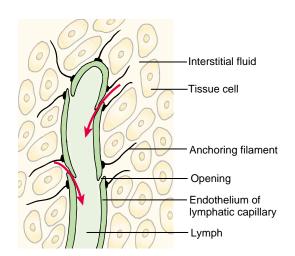
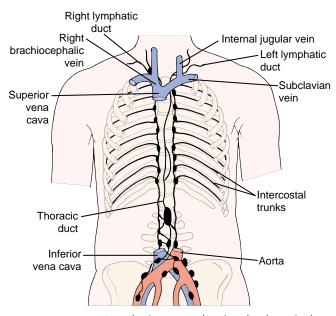
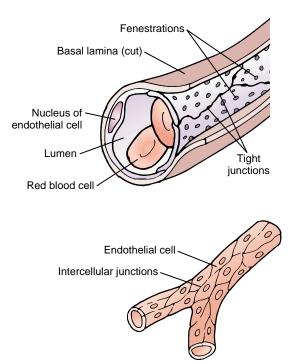


FIGURE 14-23 Portion of a femoral vein opened, to show the valves. The direction of flow is upward. Backward flow closes the valve.



■ FIGURE 14-24 ■ Details of a lymphatic capillary.





■ FIGURE 14-25 ■ Lymphatic system showing the thoracic duct and position of the left and right lymphatic ducts.

of the larger collecting lymph channels assists in propelling lymph fluid toward the thorax. External compression of the lymph channels by active and passive movements of body parts also aid in forward propulsion of lymph fluid. The rate of flow through the lymphatic system by way of all of the various lymph channels, approximately 120 mL per hour, is determined by the interstitial fluid pressure and the activity of lymph pumps.

The Microcirculation and Local Control of Blood Flow

The capillaries, venules, and arterioles of the circulatory system are collectively referred to as the *microcirculation*. It is here that exchange of gases, nutrients, and metabolites takes place between the tissues and the circulating blood. The lymphatic system represents an accessory system that removes excess fluid, proteins, and large particles from the interstitial spaces and returns them to the circulation. Because of their size, these particles cannot be reabsorbed into the capillaries.

Capillaries

Capillaries are microscopic, single-cell-thick vessels that connect the arterial and venous segments of the circulation. In each person, there are approximately 10 billion capillaries, with a total surface area of 500 to 700 m².

The capillary wall is composed of a single layer of endothelial cells surrounded by a basement membrane (Fig. 14-26). Intracellular junctions join the capillary endothelial cells; these are called the capillary pores. Lipid-soluble materials diffuse directly through the capillary cell membrane. Water and watersoluble materials leave and enter the capillary through the capillary pores. The size of the capillary pores varies with capillary function. In the brain, the endothelial cells are joined by tight junctions that form the blood-brain barrier. This prevents sub-

■ FIGURE 14-26 ■ Endothelial cells and intercellular junctions in a section of capillary.

stances that would alter neural excitability from leaving the capillary. In organs that process blood contents, such as the liver, capillaries have large pores so that substances can pass easily through the capillary wall. In the kidneys, the glomerular capillaries have small openings called *fenestrations* that pass directly through the middle of the endothelial cells. Fenestrated capillary walls are consistent with the filtration function of the glomerulus.

Blood enters the microcirculation through an arteriole, passes through the capillaries, and leaves by way of a small venule. The metarterioles serve as thoroughfare channels that link arterioles and capillaries (Fig. 14-27). Small cuffs of smooth muscle, the precapillary sphincters, are positioned at the arterial end of the capillary. The smooth muscle tone of the arterioles, venules, and precapillary sphincters serves to control blood flow through the capillary bed. Depending on venous pressure, blood flows through the capillary channels when the precapillary sphincters are open.

Blood flow through capillary channels, designed for exchange of nutrients and metabolites, is called nutrient flow. In some parts of the microcirculation, blood flow bypasses the capillary bed, moving through a connection called an arteriovenous shunt, which directly connects an arteriole and a venule. This type of blood flow is called non-nutrient flow because it does not allow for nutrient exchange. Non-nutrient channels are common in the skin and are important in terms of heat exchange and temperature regulation.

Autoregulation

Tissue blood flow is regulated on a minute-to-minute basis in relation to tissue needs and on a longer-term basis through the development of collateral circulation. Neural mechanisms

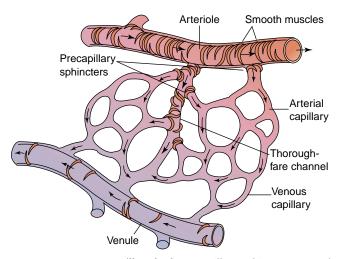


FIGURE 14-27 Capillary bed. Precapillary sphincters control the flow of blood through the capillary network. Thoroughfare channels (*i.e.*, arteriovenous shunts) allow blood to move directly from the arteriole into the venule without moving through nutrient channels of the capillary.

regulate the cardiac output and blood pressure needed to support these local mechanisms.

Local control of blood flow is governed largely by the nutritional needs of the tissue. For example, blood flow to organs such as the heart, brain, and kidneys remains relatively constant, although blood pressure may vary throughout a range of 60 to 180 mm Hg. The ability of the tissues to regulate their own blood flow throughout a wide range of pressures is called *autoregulation*. Autoregulation of blood flow is mediated by changes in blood vessel tone caused by changes in flow through the vessel or by local tissue factors, such as lack of oxygen or accumulation of tissue metabolites (*i.e.*, potassium, lactic acid, or adenosine, which is a breakdown product of ATP). Local control is particularly important in tissues such as skeletal muscle, which has blood flow requirements that vary according to the level of activity.

An increase in local blood flow is called *hyperemia*. The ability of tissues to increase blood flow in situations of increased activity, such as exercise, is called *functional hyperemia*. When the blood supply to an area has been occluded and then restored, local blood flow through the tissues increases within seconds to restore the metabolic equilibrium of the tissues. This increased flow is called *reactive hyperemia*. The transient redness seen on an arm after leaning on a hard surface is an example of reactive hyperemia. Local control mechanisms rely on a continuous flow from the main arteries; therefore, hyperemia cannot occur when the arteries that supply the capillary beds are narrowed. For example, if a major coronary artery becomes occluded, the opening of channels supplied by that vessel cannot restore blood flow.

Tissue Factors Contributing to Local Control of Blood Flow. Vasoactive substances, formed in tissues in response to a need for increased blood flow, also aid in the local control of blood flow. The most important of these are histamine, serotonin (*i.e.*, 5-hydroxytryptamine), the kinins, and the prostaglandins. *Histamine* increases blood flow. Most blood vessels contain histamine in mast cells and nonmast cell stores; when these tissues are injured, histamine is released. In certain tissues, such as skeletal muscle, the activity of the mast cells is mediated by the sympathetic nervous system; when sympathetic control is withdrawn, the mast cells release histamine. Vasodilation then results from increased histamine and the withdrawal of vasoconstrictor activity.

Serotonin is liberated from aggregating platelets during the clotting process; it causes vasoconstriction and plays a major role in the control of bleeding. Serotonin is found in brain and lung tissues, and there is some speculation that it may be involved in the vascular spasm associated with some allergic pulmonary reactions and migraine headaches.

The *kinins* (*i.e.*, kallidins and bradykinin) are liberated from the globulin kininogen, which is present in body fluids. The kinins cause relaxation of arteriolar smooth muscle, increase capillary permeability, and constrict the venules. In exocrine glands, the formation of kinins contributes to the vasodilation needed for glandular secretion.

Prostaglandins are synthesized from constituents of the cell membrane (*i.e.*, the long-chain fatty acid *arachidonic acid*). Tissue injury incites the release of arachidonic acid from the cell membrane, which initiates prostaglandin synthesis. There are several prostaglandins (*e.g.*, E_2 , F_2 , D_2), which are subgrouped according to their solubility; some produce vasoconstriction and some produce vasodilation. As a rule of thumb, those in the E group are vasodilators, and those in the F group are vasoconstrictors. The adrenal glucocorticoid hormones produce an anti-inflammatory response by blocking the release of arachidonic acid, preventing prostaglandin synthesis.

Endothelial Control of Vasodilation and Vasoconstriction. The *endothelium*, which lies between the blood and the vascular smooth muscle, serves as a physical barrier for vasoactive substances that circulate in the blood. Once thought to be nothing more than a single layer of cells that line blood vessels, it is now known that the endothelium plays an active role in controlling vascular function. In capillaries, which are composed of a single layer of endothelial cells, the endothelium is active in transporting cell nutrients and wastes. In addition to its function in capillary transport, the vascular endothelium removes vasoactive agents such as norepinephrine from the blood, and it produces enzymes that convert precursor molecules to active products (*e.g.*, angiotensin I to angiotensin II in lung vessels).

One of the important functions of the endothelial cells in the small arteries and arterioles is to synthesize and release factors that control vessel dilation. Of particular importance was the discovery, first reported in the early 1980s, that the intact endothelium was able to produce a factor that caused relaxation of vascular smooth muscle. This factor was originally named *endothelium-derived relaxing factor* and is now known to be *nitric oxide*. Many other cell types produce nitric oxide. In these tissues, nitric oxide has other functions, including modulation of nerve activity in the nervous system.

The normal endothelium maintains a continuous release of nitric oxide, which is formed from L-arginine through the action of an enzyme called *nitric oxide synthase* (Fig. 14-28). The production of nitric oxide can be stimulated by a variety of

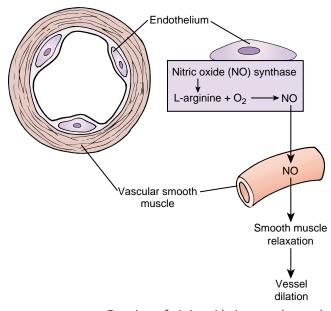


FIGURE 14-28 Function of nitric oxide in smooth muscle relaxation.

endothelial agonists, including acetylcholine, bradykinin, histamine, and thrombin. Shear stress on the endothelium, resulting from an increase in blood flow or blood pressure, also stimulates nitric oxide production and vessel relaxation. Nitric oxide also inhibits platelet aggregation and secretion of platelet contents, many of which cause vasoconstriction. The fact that nitric oxide is released into the vessel lumen (to inactivate platelets) and away from the lumen (to relax smooth muscle) suggests that it protects against both thrombosis and vasoconstriction. Nitroglycerin, which is used in treatment of angina, produces its effects by releasing nitric oxide in vascular smooth muscle of the target tissues.

The endothelium also produces a number of vasoconstrictor substances, including angiotensin II, vasoconstrictor prostaglandins, and a family of peptides called endothelins. There are at least three endothelins. Endothelin-1, made by human endothelial cells, is the most potent endogenous vasoconstrictor known. Receptors for endothelins also have been identified.

Collateral Circulation

Collateral circulation is a mechanism for the long-term regulation of local blood flow. In the heart and other vital structures, anastomotic channels exist between some of the smaller arteries. These channels permit perfusion of an area by more than one artery. When one artery becomes occluded, these anastomotic channels increase in size, allowing blood from a patent artery to perfuse the area supplied by the occluded vessel. For example, persons with extensive obstruction of a coronary blood vessel may rely on collateral circulation to meet the oxygen needs of the myocardial tissue normally supplied by that vessel. As with other long-term compensatory mechanisms, the recruitment of collateral circulation is most efficient when obstruction to flow is gradual, rather than sudden.

In summary, the walls of all blood vessels, except the capillaries, are composed of three layers: the tunica externa, tunica media, and tunica intima. The layers of the vessel vary with its function. Arteries are thick-walled vessels with large amounts of elastic fibers. The walls of the arterioles, which control blood pressure, have large amounts of smooth muscle. Veins are thin-walled, distensible, and collapsible vessels. Venous flow is designed to return blood to the heart. It is a low-pressure system and relies on venous valves and the action of muscle pumps to offset the effects of gravity. Capillaries are single-cell-thick vessels designed for the exchange of gases, nutrients, and waste materials.

The delivery of blood to the tissues of the body is dependent on pressure pulses that are generated by the intermittent ejection of blood from the left ventricle into the distensible aorta and large arteries of the arterial system. The combination of distensibility of the arteries and their resistance to flow reduces the pressure pulsations, therefore, blood flow is almost constant by the time blood reaches the capillaries. Two major factors affect the pressure pulsations: (1) the stroke volume output of the heart, and (2) and the compliance of the arterial system into which the blood is ejected. A large stroke volume and/or decrease in arterial compliance produce an increase in pulse pressure.

The mechanisms that control blood flow are designed to ensure adequate delivery of blood to the capillaries in the microcirculation, where the exchange of cellular nutrients and wastes occurs. Local control is governed largely by the needs of the tissues and is regulated by local tissue factors such as lack of oxygen and the accumulation of metabolites. Hyperemia is a local increase in blood flow that occurs after a temporary occlusion of blood flow. It is a compensatory mechanism that decreases the oxygen debt of the deprived tissues. Collateral circulation is a mechanism for long-term regulation of local blood flow that involves the development of collateral vessels.

NEURAL CONTROL OF CIRCULATORY FUNCTION

The neural control of the circulatory system occurs primarily through the sympathetic and parasympathetic divisions of the autonomic nervous system (ANS). The ANS contributes to the control of cardiovascular function through modulation of cardiac (*i.e.*, heart rate and cardiac contractility) and vascular (*i.e.*, peripheral vascular resistance) function.

The neural control centers for the integration and modulation of cardiac function and blood pressure are located bilaterally in the medulla oblongata. The medullary cardiovascular neurons are grouped into three distinct pools that lead to sympathetic innervation of the heart and blood vessels and parasympathetic innervation of the heart. The first two, which control sympathetic-mediated acceleration of heart rate and blood vessel tone, are called the vasomotor center. The third, which controls parasympathetic-mediated slowing of heart rate, is called the *cardioinhibitory center*. These brain stem centers receive information from many areas of the nervous system, including the hypothalamus. The arterial baroreceptors and chemoreceptors provide the medullary cardiovascular center with continuous information regarding changes in blood pressure (see Chapter 16).

Autonomic Regulation of Cardiac Function

The heart is innervated by the parasympathetic and sympathetic nervous systems. Parasympathetic innervation of the heart is achieved by means of the vagus nerve. The parasympathetic outflow to the heart originates from the vagal nucleus in the medulla. The axons of these neurons pass to the heart in the cardiac branches of the vagus nerve. The effect of vagal stimulation on heart function is largely limited to heart rate, with increased vagal activity producing a slowing of the pulse. Sympathetic outflow to the heart and blood vessels arises from neurons located in the reticular formation of the brain stem. The axons of these neurons exit the thoracic segments of the spinal cord to synapse with the postganglionic neurons that innervate the heart. Cardiac sympathetic fibers are widely distributed to the SA and AV nodes and the myocardium. Increased sympathetic activity produces an increase in the heart rate and the velocity and force of cardiac contraction.

Autonomic Regulation of Vascular Function

The sympathetic nervous system serves as the final common pathway for controlling the smooth muscle tone of the blood vessels. Most of the sympathetic preganglionic fibers that control vessel function originate in the vasomotor center of the brain stem, travel down the spinal cord, and exit in the thoracic and lumbar (T1-L2) segments. The sympathetic neurons that supply the blood vessels maintain them in a state of tonic activity, so that even under resting conditions, the blood vessels are partially constricted. Vessel constriction and relaxation are accomplished by altering this basal input. Increasing sympathetic activity causes constriction of some vessels, such as those of the skin, the gastrointestinal tract, and the kidneys. Blood vessels in skeletal muscle are supplied by both vasoconstrictor and vasodilator fibers. Activation of sympathetic vasodilator fibers causes vessel relaxation and provides the muscles with increased blood flow during exercise. Although the parasympathetic nervous system contributes to the regulation of heart function, it has little or no control over blood vessels.

Autonomic Neurotransmitters

The actions of the ANS are mediated by chemical neurotransmitters. *Acetylcholine* is the postganglionic neurotransmitter for parasympathetic neurons and *norepinephrine* is the main neurotransmitter for postganglionic sympathetic neurons. Sympathetic neurons also respond to epinephrine, which is released into the bloodstream by the adrenal medulla. The neurotransmitter *dopamine* can also act as a neurotransmitter for some sympathetic neurons. The synthesis, release, and inactivation of the autonomic neurotransmitters are discussed in Chapter 36. In summary, the neural control centers for the regulation of cardiac function and blood pressure are located in the reticular formation of the lower pons and medulla of the brain stem, where the integration and modulation of ANS responses occur. These brain stem centers receive information from many areas of the nervous system, including the hypothalamus. Both the parasympathetic and sympathetic nervous systems innervate the heart. The parasympathetic nervous system functions in regulating heart rate through the vagus nerve, with increased vagal activity producing a slowing of heart rate. The sympathetic nervous system has an excitatory influence on heart rate and contractility, and it serves as the final common pathway for controlling the smooth muscle tone of the blood vessels.

REVIEW QUESTIONS

Compare the functions and distribution of blood flow and blood pressure in the systemic and pulmonary circulations.

■ Use Laplace's law to explain the collapse of small blood vessels during shock caused by a loss of blood volume and to explain why atherosclerotic disease is usually a silent disorder until the disease has produced a 75% reduction in vessel radius.

Describe the cardiac conduction system and relate it to the mechanical functioning of the heart.

Characterize the four phases of a cardiac action potential and differentiate between the fast and slow responses.

Draw a figure of the cardiac cycle, incorporating volume, pressure, phonocardiographic, and electrocardiographic changes that occur during atrial and ventricular systole and diastole.

Explain the concepts of *preload* and *afterload* in terms of venous return to the heart, aortic valvular stenosis, and hypertension.

State the formula for calculating the cardiac output and explain the effects that venous return, cardiac contractility, and heart rate have on cardiac output.

Describe the cardiac reserve and relate it to the Frank-Starling mechanism.

Compare the structure and function of arteries, arterioles, veins, and capillaries.

Define autoregulation and characterize mechanisms responsible for short-term and long-term regulation of blood flow.

Describe the distribution of the sympathetic and parasympathetic nervous system in innervation of the circulatory system and their effects on heart rate and cardiac contractility.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

- Berne R.M., Levy M.N. (2000). Principles of physiology (3rd ed., pp. 201–275). St. Louis: C.V. Mosby.
- Berne R.M., Levy M.N. (2001). Cardiovascular physiology (8th ed.). St. Louis: C.V. Mosby.
- Feletou M., Vanhoutte P.M. (1999). The alternative: EDHF. Journal of Molecular and Cellular Cardiology 31, 15–22.
- Ganong W.F. (1999). *Review of medical physiology* (19th ed., 493–601). Stamford, CT: Appleton & Lange.
- Guyton A.C., Hall J.E. (2000). *Medical physiology* (10th ed., pp. 144–222). Philadelphia: W.B. Saunders.
- McCormack D.H. (1987). *Ham's histology* (9th ed., p. 448). Philadelphia: J.B. Lippincott.
- Rhoades R.A, Tanner G.A. (1996). *Medical physiology* (pp. 207–301). Boston: Little, Brown.
- Smith J.J., Kampine J.P. (1989). Circulatory physiology (3rd ed.). Baltimore: Williams & Wilkins.
- Vanhoutte P.M. (1999). How to assess endothelial function in human blood vessels. *Journal of Hypertension* 17, 1047–1058.

CHAPTER 15

Alterations in Blood Flow

Disorders of the Arterial Circulation Hyperlipidemia

Lipoproteins Hypercholesterolemia **Atherosclerosis Risk Factors** Pathology and Pathogenesis **Clinical Manifestations** Aneurysms and Dissection **Aortic Aneurysms Aortic Dissection** The Vasculitides Arterial Disease of the Extremities **Acute Arterial Occlusion** Atherosclerotic Occlusive Disease **Thromboangiitis Obliterans** Raynaud's Disease and Phenomenon **Disorders of the Venous Circulation** Venous Circulation of the Lower Extremities Varicose Veins **Chronic Venous Insufficiency** Venous Thrombosis **Disorders of Blood Flow Caused by Extravascular Forces Compartment Syndrome Pressure Ulcers** Mechanisms of Development Prevention Staging and Treatment

B lood vessels function in the delivery of oxygen and nutrients to the tissues and in the removal of waste products from the tissues. Unlike disorders of the respiratory system or central circulation that cause hypoxia and impair oxygenation of tissues throughout the body, the effects of blood vessel disease usually are limited to local tissues supplied by a particular vessel or group of vessels.

Disturbances in blood flow can result from pathologic changes in the vessel wall (*i.e.*, atherosclerosis, vasculitides), acute vessel obstruction caused by thrombus or embolus, vasospasm (*i.e.*, Raynaud's phenomenon), abnormal vessel dilation (*i.e.*, arterial aneurysms or varicose veins), or compression of blood vessels by extravascular forces (*i.e.*, tumors, edema, or firm surfaces such as those associated with pressure ulcers).

DISORDERS OF THE ARTERIAL CIRCULATION

The arterial system distributes blood to all the tissues in the body. There are three types of arteries: large arteries, including the aorta and its distal branches; medium-size arteries, such as the coronary and renal arteries; and small arteries and arterioles that pass through the tissues. Each of these different types of arteries tends to be affected by different disease processes.

Pathology of the arterial system affects body function by impairing blood flow. The effect of impaired blood flow on the body depends on the structures involved and the extent of altered flow. The term *ischemia* (*i.e.*, holding back of blood) denotes a reduction in arterial flow to a level that is insufficient to meet the oxygen demands of the tissues. *Infarction* refers to an area of ischemic necrosis in an organ produced by occlusion of its blood supply. The discussion in this section focuses on hyperlipidemia, atherosclerosis, arterial aneurysms, vasculitides, and arterial disease of the extremities.

Hyperlipidemia

Hyperlipidemia with elevated cholesterol levels is a major cause of atherosclerosis with its attendant risk of heart attack and stroke. An estimated 41.3 million Americans have high

KEY CONCEPTS

DISORDERS OF THE ARTERIAL CIRCULATION

- The arterial system delivers oxygen and nutrients to the tissues. Disorders of the arterial circulation produce ischemia owing to narrowing of blood vessels, thrombus formation associated with platelet adhesion, and weakening of the vessel wall.
- Atherosclerosis is a progressive disease characterized by the formation of fibrofatty plaques in the intima of large and medium-sized arteries, producing a decrease in blood flow due to a narrowing of the vessel lumen.
- Aneurysms represent an abnormal localized dilatation of an artery due to a weakness in the vessel wall. As the aneurysm increases in size, the tension in the wall of the vessel increases, predisposing it to rupture.

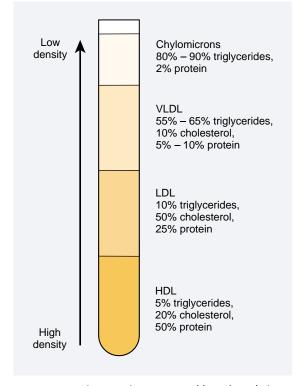
serum cholesterol levels that could contribute to a heart attack, stroke, or other cardiovascular event associated with atherosclerosis.¹

Lipoproteins

Because cholesterol and triglyceride are insoluble in plasma, they are encapsulated by special fat-carrying proteins called *lipoproteins* for transport in the blood. There are five types of lipoproteins, classified by their densities as measured by ultracentrifugation: chylomicrons, very–low-density lipoprotein (VLDL), intermediate-density lipoprotein (IDL), low-density lipoprotein (LDL), and high-density lipoprotein (HDL) (see Fig. 15-1).

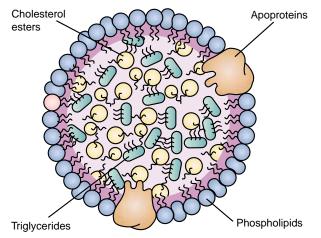
Each type of lipoprotein consists of a large molecular complex of lipids combined with proteins called *apoproteins*.^{2,3} The major lipid constituents are cholesterol esters, triglycerides, nonesterified cholesterol, and phospholipids. The insoluble cholesterol esters and triglycerides are located in the hydrophobic core of the lipoprotein macromolecule, surrounded by the soluble phospholipids, nonesterified cholesterol, and apoproteins (Fig. 15-2). Nonesterified cholesterol and phospholipids provide a negative charge that allows the lipoprotein to be soluble in plasma. The apoproteins control the interactions and ultimate metabolic fate of the lipoproteins. Some of the apoproteins activate the lipolytic enzymes that facilitate the removal of lipids from the lipoproteins; others serve as a reactive site that cellular receptors can recognize and use in the endocytosis and metabolism of the lipoproteins.

There are two sites of lipoprotein synthesis: the small intestine and the liver. The chylomicrons, which are the largest of the lipoprotein molecules, are synthesized in the wall of the small intestine. They are involved in the transport of dietary triglycerides and cholesterol that have been absorbed from the gastrointestinal tract. Chylomicrons transfer their triglycerides to the cells of adipose and skeletal muscle tissue. The remnant chylomicron particles, which contain cholesterol, are then taken up by the liver, and the cholesterol is used in the synthesis of VLDL or excreted in the bile.



■ FIGURE 15-1 ■ Lipoproteins are named based on their protein content, which is measured as density. Because fats are less dense than proteins, as the proportion of triglycerides decreases, the density increases.

The liver synthesizes and releases VLDL and HDL. The VLDLs contain large amounts of triglycerides and lesser amounts of cholesterol esters.⁴ They provide the primary pathway for transport of the triglycerides produced in the liver, as opposed to those obtained from the diet. Like chylomicrons, VLDLs carry their triglycerides to fat and muscle cells, where the triglycerides are removed. The resulting IDL fragments are reduced in triglyceride content and enriched in cholesterol. They



■ FIGURE 15-2 ■ General structure of a lipoprotein. The cholesterol esters and triglycerides are located in the hydrophobic core of the macromolecule, surrounded by phospholipids and apoproteins.

are taken to the liver and recycled to form VLDL, or converted to LDL in the vascular compartment. The pathways for triglyceride and cholesterol transport are shown in Figure 15-3.

LDL, sometimes called the bad cholesterol, is the main carrier of cholesterol. The IDLs are the main source of LDL. The LDL is removed from the circulation by either LDL receptors or by scavenger cells such as monocytes or macrophages. Approximately 70% of LDL is removed by way of the LDL receptordependent pathway.⁴ Although LDL receptors are widely distributed, approximately 75% are located on hepatocytes; thus the liver plays an extremely important role in LDL metabolism. Tissues with LDL receptors can control their cholesterol intake by adding or removing LDL receptors.

The scavenger cells, such as the monocytes and macrophages, have receptors that bind LDL that has been oxidized or chemically modified. The amount of LDL that is removed by the "scavenger pathway" is directly related to the plasma cholesterol level. When there is a decrease in LDL receptors or when LDL levels exceed receptor availability, the amount of LDL that is removed by scavenger cells is greatly increased. The uptake of LDL by macrophages in the arterial wall can result in the accumulation of insoluble cholesterol esters, the formation of foam cells, and the development of atherosclerosis.

HDL is synthesized in the liver and often is referred to as the good cholesterol. Epidemiological studies show an inverse relation between HDL levels and the development of atherosclerosis.⁵ It is thought that HDL, which is low in cholesterol and rich in surface phospholipids, facilitates the clearance of cholesterol from atheromatous plagues and transports it back to the liver, so that it can be excreted in the bile. HDL also is believed to inhibit the uptake of LDL into the arterial wall. It has been observed that regular exercise and moderate alcohol consumption increase HDL levels. Smoking and diabetes, which are in themselves risk factors for atherosclerosis, are associated with decreased levels of HDL.⁴

Hypercholesterolemia

The Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults has published a classification system for hyperlipidemia that list the laboratory values for optimal to very high levels of LDL cholesterol, desirable to high levels of total cholesterol, and low and high levels of HDL cholesterol (see Table 15-1).6 The NCEP recommends that all adults 20 years of age and older should have a fasting lipoprotein profile (total cholesterol, LDL cholesterol, HDL cholesterol, and triglycerides) measured once every 5 years.6 If testing is done in the nonfasting state, only the total cholesterol and HDL are considered useful. A follow-up lipoprotein profile should be done on persons with nonfasting total cholesterol levels ≥200 mg/dL or HDL levels <40 mg/dL. Lipoprotein measurements are particularly important in persons at high risk for the development of coronary heart disease (CHD).

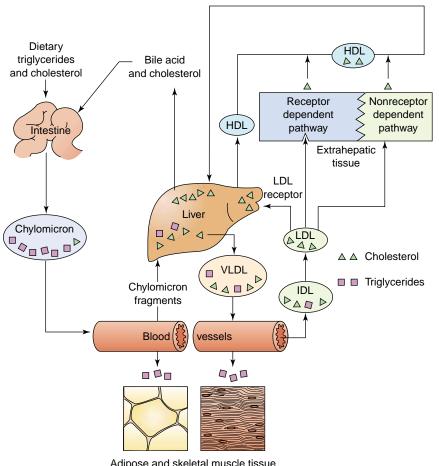


FIGURE 15-3 Schematic representation of the exogenous and endogenous pathways for triglyceride and cholesterol transport.

Adipose and skeletal muscle tissue

TABLE 15-1 NCEP A	dult Treatment
Panel III Classification of LDL, Total, and HDL Cholesterol	
Cholesterol Level (mg/dL)	Classification
LDL Cholesterol	
<100	Optimal
100–129	Near optimal/above optimal
130–159	Borderline high
160–189	High
≥190	Very high
Total Cholesterol	
<200	Desirable
200–239	Borderline high
≥240	High
HDL Cholesterol	
<40	Low
≥60	High

National Institutes of Health Expert Panel. (2001). Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). (NIH publication no. 01-3670). Bethesda, MD: National Institutes of Health.

Causes. Three factors-genetics, nutrition, and metabolic diseases-contribute to an increase in blood lipid levels. Many types of hyperlipidemia have a genetic basis. There may be a defective synthesis of the apoproteins, a lack of LDL receptors, defective LDL receptors, or defects in the intracellular handling of cholesterol.² The LDL receptor is deficient or defective in the genetic disorder known as familial hypercholesterolemia. This autosomal dominant type of hyperlipoproteinemia results from a mutation in the gene specifying the receptor for LDL. Because most of the circulating cholesterol is removed by receptordependent mechanisms, blood cholesterol levels are markedly elevated in persons with this disorder. The disorder is probably one of the most common of all mendelian disorders; the frequency of heterozygotes is 1 in 500 persons in the general population.² Although heterozygotes commonly have an elevated cholesterol level from birth, they do not experience symptoms until adult life, when xanthomas (i.e., cholesterol deposits) develop along the tendons and atherosclerosis occurs (Fig. 15-4). Myocardial infarction before the age of 40 years is common. Homozygotes are much more severely affected; they develop cutaneous xanthomas in childhood and may experience myocardial infarction by the age of 20 years.²

Secondary causes of hyperlipoproteinemia include diets high in saturated fats and cholesterol, obesity caused by highcalorie intake, and diabetes mellitus. Diets that are high in triglycerides and saturated fats increase cholesterol synthesis and suppress LDL receptor activity. Excess ingestion of cholesterol reduces the formation of LDL receptors and thereby decreases LDL removal. In diabetes mellitus, metabolic derangements cause an elevation of lipoproteins.7

The management of hyperlipidemia focuses on dietary and lifestyle modifications; when these are unsuccessful, pharmacologic treatment may be necessary. Lifestyle modification includes increased emphasis on physical activity, dietary measures to reduce LDL cholesterol levels, and weight reduction for people who are overweight. The aim of dietary therapy is to reduce total and LDL cholesterol levels and increase HDL levels. Three dietary elements affect serum cholesterol and its lipoprotein fractions: excess calorie intake, saturated fats, and cholesterol. High-calorie diets increase the production of VLDL, with triglyceride elevation and high conversion of VLDL to LDL. Saturated fats and cholesterol in the diet tend to increase LDL cholesterol levels.

The third report of the NCEP continues to identify reduction in LDL cholesterol as the primary target for cholesterol-lowering therapy, particularly in people at risk for CHD.⁶ Lipid-lowering drugs work mainly by decreasing cholesterol absorption from the intestine, decreasing cholesterol synthesis by the liver, or



FIGURE 15-4 Xanthomas in the skin and tendons (A, C, D). Arcus lipoides represents the deposition of lipids in the peripheral cornea (**B**). (Rubin E., Farber J.L. [1999]. Pathology [3rd ed., p. 506]. Philadelphia: Lippincott Williams & Wilkins)

inhibiting VLDL release. Drugs that act directly to decrease cholesterol levels also have the beneficial effect of further lowering cholesterol levels by stimulating the production of additional LDL receptors. Because many of these drugs have significant adverse effects, they usually are used only in persons with significant hyperlipidemia that cannot be controlled by other means, such as diet.

Atherosclerosis

Atherosclerosis is a type of arteriosclerosis or hardening of the arteries. The term *atherosclerosis*, which comes from the Greek words *atheros* (meaning "gruel" or "paste") and *sclerosis* (meaning "hardness"), denotes the formation of fibrofatty lesions in the intimal lining of the large and medium-size arteries such as the aorta and its branches, the coronary arteries, and the large vessels that supply the brain. Atherosclerosis contributes to more mortality and more serious morbidity than any other disorder in the western world.⁴ The major complications of atherosclerosis, including ischemic heart disease, stroke, and peripheral vascular disease, account for more than 40% of the deaths in the United States.⁸

Risk Factors

The cause or causes of atherosclerosis have not been determined with certainty. However, epidemiologic studies have identified predisposing risk factors, which are listed in Chart 15-1.^{2,4} Some risk factors can be affected by a change in health behavior and others cannot. Risk factors that cannot be changed in-

CHART 15-1 Risk Factors in Coronary Heart Disease Other Than Low-Density Lipoproteins

Positive Risk Factors

Age

Men: ≥45 years

Women: ≥55 years or premature menopause without estrogen replacement therapy

Family history of premature coronary heart disease (definite myocardial infarction or sudden death before 55 years of age in father or other male first-degree relative, or before 65 years of age in mother or other female first-degree relative)

Current cigarette smoking

Hypertension (≥140/90 mm Hg* or on antihypertensive medication)

Low HDL cholesterol (<40 mg/dL*) Diabetes mellitus

Negative Risk Factor

High HDL cholesterol (≥60 mg/dL)

HDL, high-density lipoprotein.

*Confirmed by measurements on several occasions.

(Modified from National Institute of Health Expert Panel [2001]. Third Report of the National Cholesterol Program [NCEP] Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults [Adult Treatment Panel III]. [NIH Publication No. 01-3670]. Bethesda, MD: National Institutes of Health.) clude age, male gender, and family history of premature coronary heart disease.

The tendency to develop atherosclerosis appears to run in families. Persons who come from families with a strong history of heart disease or stroke caused by atherosclerosis are at greater risk for developing atherosclerosis than are those with a negative family history. Several genetically determined alterations in lipoprotein and cholesterol metabolism have been identified, and it seems likely that others will be identified in the future. The incidence of atherosclerosis also increases with age. Other factors being equal, men are at greater risk for coronary heart disease than are premenopausal women, probably because of the protective effects of natural estrogens. After menopause, the incidence of atherosclerotic-related diseases in women increases, and by the 7th to 8th decade of life, the frequency of myocardial infarction in the two sexes tends to equalize.⁴

The major risk factors that can be affected by a change in health behaviors include hyperlipidemia, cigarette smoking, hypertension, and diabetes mellitus. These risk factors can often be modified or controlled by a change in diet, exercise, health care practices, or medications. The presence of hyperlipidemia is the strongest risk factor for atherosclerosis in persons younger than 45 years of age. Both primary and secondary hyperlipidemia increase the risk. Cigarette smoking is closely linked with coronary heart disease and sudden death. Cessation of smoking reduces the risk substantially. High blood pressure produces mechanical stress on the vessel endothelium. It is a major risk factor for atherosclerosis in all age groups and may be as important or more important than hypercholesterolemia after the age of 45 years. Both systolic and diastolic pressures are important in increasing risk. Diabetes mellitus (type 2) typically develops in middle-aged persons and those who are overweight. Diabetes elevates blood lipid levels and otherwise increases the risk of atherosclerosis (see Chapter 32). Controlling other risk factors, such as hypertension and hypercholesterolemia, is particularly important in persons with diabetes.

Other factors, known as "soft" risk factors, are not as convincing as the established risk factors. These include insufficient physical activity, a stressful lifestyle, and obesity. These "soft" risk factors commonly are linked with the established and other contributing risk factors. For example, obesity and physical inactivity often are observed in the same person. Both conditions are reported to bring about elevations in blood lipid levels. Likewise, major risk factors such as cigarette smoking are closely associated with stress and personality patterns.

There are a number of other less well-established risk factors for atherosclerosis, including high serum homocysteine levels, elevated serum C-reactive protein, and infectious agents.^{9,10} Homocysteine is derived from the metabolism of dietary methionine, an amino acid that is abundant in animal protein. Homocysteine inhibits elements of the anticoagulant cascade and is associated with endothelial damage, which is thought to be an important first step in the development of atherosclerosis.¹⁰ Factors tending to increase plasma levels of homocysteine include lower serum levels of folate and vitamins B₆ and B₁₂; genetic defects in homocysteine metabolism; renal impairment; malignancies, increasing age; male gender; and female menopause.¹¹ *C-reactive protein* (CRP) is a serum marker for systemic inflammation. Several prospective studies have indicated that elevated CRP levels are associated with vascular disbus formation.¹⁰ Recently, there has been increased interest in the possible connection between infectious agents (*Chlamydia pneumoniae*, *herpesvirus hominis, cytomegalovirus*) and the development of vascular disease. The presence of these organisms in atheromatous lesions has been demonstrated by immunocytochemistry, but no cause-and-effect relationship has been established. The organisms may play a role in atherosclerotic development by initiating and enhancing the inflammatory response.¹²

Pathology and Pathogenesis

The lesions associated with atherosclerosis are of three types: the fatty streak, the fibrous atheromatous plaque, and the complicated lesion. The latter two are responsible for the clinically significant manifestations of the disease.

Fatty streaks are thin, flat yellow intimal discolorations that progressively enlarge by becoming thicker and slightly elevated as they grow in length (Fig. 15-5). Histologically, they consist of macrophages and smooth muscle cells that have become distended with lipid to form foam cells. Fatty streaks are present in children, often in the first year of life.^{2,4} This occurs regardless of geographic setting, gender, or race. They increase in number until about age 20 years, and then they remain static or regress. There is controversy about whether fatty streaks, in and of themselves, are precursors of atherosclerotic lesions.

The *fibrous atheromatous plaque* is the basic lesion of clinical atherosclerosis. It is characterized by the accumulation of intracellular and extracellular lipids, proliferation of vascular smooth muscle cells, and formation of scar tissue. The lesions begin as a gray to pearly white elevated thickening of the vessel intima with a core of extracellular lipid (mainly cholesterol, which usually is complexed to proteins) covered by a fibrous cap of connective tissue and smooth muscle (Fig. 15-6). As the lesions increase in size, they encroach on the lumen of the artery and eventually may occlude the vessel or predispose to throm-



■ FIGURE 15-6 ■ Fatty plaque of atherosclerosis. The aorta shows discrete, raised, tan plaques. Focal ulcerations are also present. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 497]. Philadel-phia: Lippincott Williams & Wilkins)

bus formation, causing a reduction of blood flow. Because blood flow is related to the fourth power of the radius, reduction in blood flow becomes more severe as the disease progresses.

The more advanced *complicated lesions* are characterized by hemorrhage, ulceration, and scar tissue deposits. Thrombosis is the most important complication of atherosclerosis. It is



■ FIGURE 15-5 ■ Fatty streaks of atherosclerosis. The aorta of a young man shows numerous fatty streaks on the luminal surface when stained with Sudan red, shown on the left. The unstained specimen is shown on the right. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 496]. Philadelphia: Lippincott Williams & Wilkins)

caused by slowing and turbulence of blood flow in the region of the plaque and ulceration of the plaque. The thrombus may cause occlusion of small vessels in the heart and brain. Aneurysms may develop in arteries weakened by extensive plaque formation.

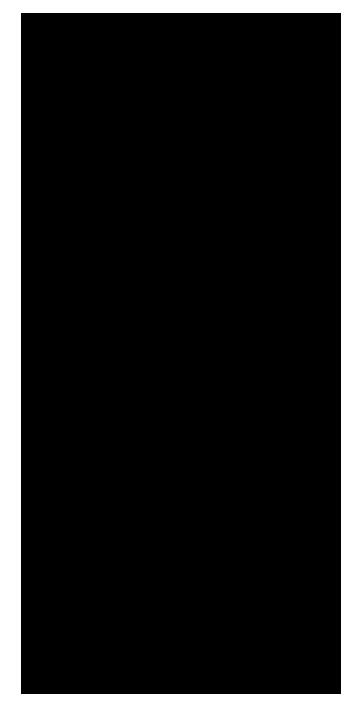
Although the risk factors associated with atherosclerosis have been identified through epidemiologic studies, many unanswered questions remain regarding the mechanisms by which these risk factors contribute to the development of atherosclerosis. There is increasing evidence that atherosclerosis is at least partially the result of (1) endothelial injury with leukocyte (lymphocyte and monocyte) adhesion and platelet adherence, (2) smooth muscle cell emigration and proliferation, (3) lipid engulfment of activated macrophages, and (4) subsequent development of an atherosclerotic plaque with lipid core (Fig. 15-7).

The vascular endothelial layer, which consists of a single layer of cells with cell-to-cell attachments, normally serves as a selective barrier that protects the subendothelial layers by interacting with blood cells and other blood components. One hypothesis of plaque formation suggests that injury to the endothelial vessel layer is the initiating factor in the development of atherosclerosis.^{2,4} A number of factors are regarded as possible injurious agents, including products associated with smoking, immune mechanisms, and mechanical stress, such as that associated with hypertension. The fact that atherosclerotic lesions tend to form where vessels branch or where there is turbulent flow suggests that hemodynamic factors also play a role.

Hyperlipidemia, particularly LDL with its high cholesterol content, is also believed to play an active role in the pathogenesis of the atherosclerotic lesion. Interactions between the endothelial layer of the vessel wall and white blood cells, particularly the macrophages (blood monocytes), normally occur throughout life; these interactions increase when blood cholesterol levels are elevated. One of the earliest responses to elevated cholesterol levels is the attachment of monocytes to the endothelium.¹⁰ The monocytes have been observed to emigrate through the cell-to-cell attachments of the endothelial layer into the subendothelial spaces, where they are transformed into macrophages. Activated macrophages release free radicals that oxidize LDL. Oxidized LDL is toxic to the endothelium, causing endothelial loss and exposure of the subendothelial tissue to blood components. This leads to platelet adhesion and aggregation and fibrin deposition. Platelets and activated macrophages release various factors that are thought to promote growth factors that modulate the proliferation of smooth muscle cells and deposition of extracellular matrix in the lesions.^{2,4} Activated macrophages also ingest oxidized LDL to become foam cells, which are present in all stages of atherosclerotic plaque formation. Lipids released from necrotic foam cells accumulate to form the lipid core of unstable plaques.

Clinical Manifestations

The clinical manifestations of atherosclerosis depend on the vessels involved and the extent of vessel obstruction. Atherosclerotic lesions produce their effects through narrowing of the vessel and production of ischemia; sudden vessel obstruction caused by plaque hemorrhage or rupture; thrombosis and formation of emboli resulting from damage to the vessel endothelium; and aneurysm formation caused by weakening of the vessel wall.⁴ In larger vessels such as the aorta, the important complications are those of thrombus formation and weakening of the vessel



wall. In medium-size arteries such as the coronary and cerebral arteries, ischemia and infarction caused by vessel occlusion are more common. Although atherosclerosis can affect any organ or tissue, the arteries supplying the heart, brain, kidneys, lower extremities, and small intestine are most frequently involved.

Aneurysms

An *aneurysm* is an abnormal localized dilatation of a blood vessel. Aneurysms can occur in different types of arterial vessels, but they are most common in the aorta. Aneurysms can assume

several forms and may be classified according to their cause, location, and anatomic features (Fig. 15-8). A *berry aneurysm* consists of a small spherical dilatation of the vessel at a bifurcation.² This type of aneurysm usually is found in the circle of Willis in the cerebral circulation. A *fusiform aneurysm* involves the entire circumference of the vessel and is characterized by a gradual and progressive dilatation of the vessel. These aneurysms, which vary in diameter (as large as 20 cm) and length, may involve the entire ascending and transverse portions of the thoracic aorta or may extend over large segments of the abdominal aorta.² A *saccular aneurysm* extends over part of the circumference of the vessel and appears saclike. An *aortic dissection* is a false aneurysm resulting from a tear in the intimal layer of the vessel that allows blood to enter the vessel wall, dissecting its layers to create a blood-filled cavity.

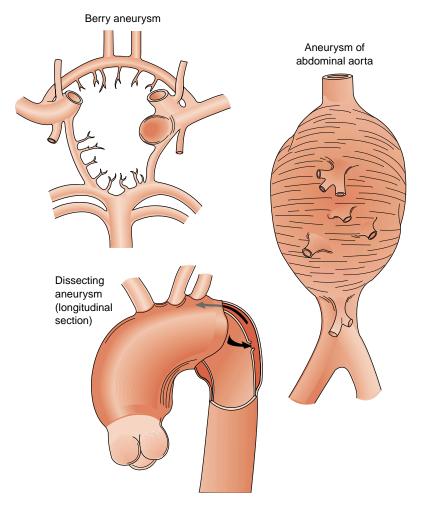
The weakness that leads to aneurysm formation may be caused by several factors, including congenital defects, trauma, infections, and atherosclerosis. Once initiated, the aneurysm grows larger as the tension in the vessel wall increases. As an aneurysm increases in diameter, the tension in the wall of the vessel increases in direct proportion to its increased size. If untreated, the aneurysm may rupture because of the increased tension. Even an unruptured aneurysm can cause damage by exerting pressure on adjacent structures and interrupting blood flow.

Aortic Aneurysms

Aortic aneurysms may involve any part of the thoracic or abdominal aorta. Aortic aneurysms occur more often in men than women. The two most common causes of aortic aneurysms are atherosclerosis and degeneration of the vessel media. Half of the people with aortic aneurysms have hypertension.² Populationbased studies suggest that as many as 9% of persons older than 65 years have unsuspected and asymptomatic abdominal aortic aneurysms and that ruptured abdominal aortic aneurysms cause at least 15,000 deaths each year in the United States.¹³

Most abdominal aortic aneurysms are located below the level of the renal artery and involve the bifurcation of the aorta and proximal end of the common iliac arteries.^{2,4} They can involve any part of the vessel circumference (saccular) or extend to involve the entire circumference (fusiform). The greatest threat associated with abdominal aortic aneurysms is that of rupture. The risk of rupture rises from less than 2% for small abdominal aneurysms (those less than 4 cm in diameter) to 5% to 10% per year for aneurysms larger than 5 cm in diameter.⁴

The clinical manifestations of abdominal aortic aneurysms depend on its size and location. Most abdominal aneurysms are asymptomatic and are often discovered during a routine physical examination. Sometimes, the first evidence of an abdominal aneurysm is associated with vessel rupture. Because an aneurysm is of arterial origin, a pulsating mass may provide the



■ FIGURE 15-8 ■ Three forms of aneurysms—berry aneurysm in the circle of Willis, fusiform-type aneurysm of the abdominal aorta, and aortic dissection.

262 Unit Four: Alterations in the Cardiovascular System

first evidence of the disorder. As the aneurysm expands, it may compress the lumbar nerve roots, causing lower back pain that radiates to the posterior aspects of the legs. The aneurysm may extend to and produce obstruction of the renal, iliac, mesenteric arteries, or vertebral arteries that supply the spinal cord. Stasis of blood favors thrombus formation along the wall of the vessel (Fig. 15-9), and peripheral emboli may develop, causing symptomatic arterial insufficiency.

Aneurysms of the thoracic aorta are less common than abdominal aneurysms. They account for less than 10% of aortic aneurysms and may present with substernal, back, or neck pain. There also may be dyspnea, stridor, or a brassy cough caused by pressure on the trachea. Hoarseness may result from pressure on the recurrent laryngeal nerve, and there may be difficulty swallowing because of pressure on the esophagus.¹⁴

Diagnostic methods include use of ultrasound imaging, computed tomographic (CT) scans, and magnetic resonance imaging (MRI). Surgical repair, in which the involved section of the aorta is replaced with a synthetic graft of woven Dacron, frequently is the treatment of choice.¹⁵

Aortic dissection

Aortic dissection is an acute, life-threatening condition. It involves hemorrhage into the vessel wall with longitudinal tearing or separation (*i.e.*, dissection) of the vessel wall to form



■ FIGURE 15-9 ■ Atherosclerotic aneurysm of the abdominal aorta. The aneurysm has been opened longitudinally to reveal a large thrombus in the lumen. The aorta and common iliac arteries display complicated lesions of atherosclerosis. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 521]. Philadelphia: Lippincott Williams & Wilkins)

a blood-filled channel. Unlike atherosclerotic aneurysms, aortic dissection often occurs without evidence of previous vessel dilatation. They can originate anywhere along the length of the aorta. Two thirds of dissections involve the ascending aorta.^{14,15} The second most common site is the thoracic aorta just distal to the origin of the subclavian artery.

Aortic dissections are caused by conditions that weaken or cause degenerative changes in the elastic and smooth muscle of the layers of the aorta. They are most common in the 40- to 60-year-old age group and more prevalent in men than in women.⁴ There are two risk factors that predispose to a dissection: hypertension and degeneration of the medial layer of the vessel wall. Aortic dissections also are associated with connective tissue diseases, such as Marfan's syndrome. Other factors that predispose to aortic dissection are congenital defects of the aortic valve (*i.e.*, bicuspid or unicuspid valve structures) and aortic coarctation.

Aortic dissections can originate at different sites in the aorta. When the ascending aorta is involved, expansion of the wall of the aorta may impair closure of the aortic valve. There also is risk of aortic rupture, with blood moving into the pericardium and compressing the heart. Although the length of dissection varies, it is possible for the abdominal aorta to be involved with progression into the renal, iliac, or femoral arteries. Partial or complete occlusion of the arteries that arise from the aortic arch or the intercostal or lumbar arteries may lead to stroke, ischemic peripheral neuropathy, or impaired blood flow to the spinal cord.

A major symptom of an aortic dissection is the abrupt presence of excruciating pain, described as tearing or ripping. The location of the pain may point to the site of dissection.⁴ Pain associated with dissection of the ascending aorta frequently is located in the anterior chest, and pain associated with dissection of the descending aorta often is located in the back. In the early stages, blood pressure typically is moderately or markedly elevated. Later, the blood pressure and the pulse rate become unobtainable in one or both arms as the dissection disrupts arterial flow to the arms. Syncope, hemiplegia, or paralysis of the lower extremities may occur because of occlusion of blood vessels that supply the brain or spinal cord. Heart failure may develop when the aortic valve is involved.

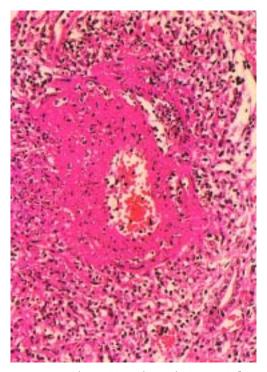
Diagnosis of aortic dissection is based on history and physical examination. Aortic angiography, transesophageal echocardiography, CT scans, and MRI studies aid in the diagnosis. The treatment may be medical or surgical. Aortic dissection is a life-threatening emergency situation; persons with a probable diagnosis are stabilized medically even before the diagnosis is confirmed. Two important factors that participate in propagating the dissection are high blood pressure and the steepness of the pulse wave. Without intervention, these forces continue to cause extension of the dissection. Thus, medical treatment focuses on control of hypertension and the use of drugs that lessen the force of systolic blood ejection from the heart. Surgical intervention may be indicated when there is threat of rupture or compromise of major aortic branches.

The Vasculitides

The vasculitides, which are a group of vascular disorders that cause inflammatory injury of blood vessels (vasculitis), are a common pathway for tissue and organ involvement in many different disease conditions.^{16,17} Vessels of any type (arteries, veins, and capillaries) in virtually any organ can be affected. Clinical manifestations often include fever, myalgia, arthralgia, and malaise. Vasculitis may result from direct injury to the vessel, infectious agents, or immune processes, or may be secondary to other disease states such as systemic lupus erythematosus. Physical agents such as cold (*i.e.*, frostbite), irradiation (*i.e.*, sunburn), mechanical injury, and toxins may secondarily cause vessel damage, often leading to necrosis of the vessels.

The vasculitides are commonly classified based on etiology, pathologic findings, and prognosis. One classification system divides the conditions into three groups: (1) small vessel, (2) medium-size vessel, and (3) large vessel vasculitides.^{2,4} The small vessel vasculitides are involved in a number of different diseases, most of which are mediated by type III immune complex hypersensitivity reaction (see Chapter 10). They commonly involve the skin and are often a complication of an underlying disease (*i.e.*, vasculitis associated with neoplasms or connective tissue disease) and exposure to environmental agents (*i.e.*, serum sickness and urticarial vasculitis).

Medium-size vessel vasculitides produce necrotizing damage to medium-size muscular arteries of major organ systems. This group includes polyarteritis nodosa, Kawasaki's disease (discussed in Chapter 17), and thromboangiitis obliterans (discussed in the section on arterial diseases of the extremities). Polyarteritis nodosa is an uncommon acute multisystem inflammatory disease of small and medium-size blood vessels of the kidney, liver, intestine, peripheral nerves, skin, and muscles (Fig. 15-10). The usual course of the disease is progressive with various signs and symptoms according to the pattern of organ



■ FIGURE 15-10 ■ Polyarteritis nodosa. The intense inflammatory cell infiltration in the arterial wall and surrounding connective tissue is mediated by fibrinoid necrosis and disruption of the vessel wall. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 515]. Philadelphia: Lippincott Williams & Wilkins)

involvement. Most cases were fatal before corticosteroid and immunosuppressant agents became available for use in treatment of the disorder.⁴

Large vessel vasculitides involve large elastic arteries; they are called *giant cell arterides* because they involve infiltration of the vessel wall with giant cells and mononuclear cells. Giant cell (temporal) arteritis, the most common of the large vessel vasculitides, is an acute and chronic inflammation of large to small size arteries. It mainly affects arteries of the head especially the temporal arteries—but may include the vertebral and ophthalmic arteries. About half of persons with the disease have an accompanying pain and stiffness of the shoulder and hip (polymyalgia rheumatica, see Chapter 43). The most common clinical presentation is with headache and tenderness over the superficial temporal artery. Diagnosis is by biopsy of the artery. Diagnosis followed by treatment with corticosteroid drugs is important because involvement of the ophthalmic artery can cause blindness.⁴

Arterial Disease of the Extremities

Disorders of the circulation in the extremities often are referred to as *peripheral vascular disorders*. In many respects, the disorders that affect arteries in the extremities are the same as those affecting the coronary and cerebral arteries in that they produce ischemia, pain, impaired function, and in some cases infarction and tissue necrosis. Not only are the effects similar, but the pathologic conditions that impair circulation in the extremities are identical. This section focuses on acute arterial occlusion of the extremities, atherosclerotic occlusive disease, thromboangiitis obliterans, and Raynaud's disease and phenomenon.

Acute Arterial Occlusion

Acute arterial occlusion is a sudden event that interrupts arterial flow to the affected tissues or organ. Most acute arterial occlusions are the result of an embolus or a thrombus. Although much less common than emboli and thrombi, trauma or arterial spasm caused by arterial cannulation can be another cause of acute arterial occlusion.

An embolus is a freely moving particle such as a blood clot that breaks loose and travels in the larger vessels of the circulation until lodging in a smaller vessel and occluding blood flow. Most arterial emboli arise in the heart and are caused by conditions that cause blood clots to develop on the wall of a heart chamber or valve surface. Arterial emboli usually are a complication of heart disease: ischemic heart disease with or without infarction, atrial fibrillation, or rheumatic heart disease. Prosthetic heart valves can be another source of emboli. Other types of emboli are fat emboli that originate from bone marrow of fractured bones or amniotic fluid emboli that develop during childbirth.

A thrombus is a blood clot that forms on the wall of a vessel and continues to grow until reaching a size that obstructs blood flow. These thrombi often arise as the result of rupture of the fibrous cap of an atherosclerotic plaque.

The signs and symptoms of acute arterial occlusion depend on the artery involved and the adequacy of the collateral circulation. Emboli tend to lodge in bifurcations of the major arteries, including the aorta and iliac, femoral, and popliteal arteries. Occlusion in an extremity causes sudden onset of acute pain with numbness, tingling, weakness, pallor, and coldness. There often is a sharp line of demarcation between the oxygenated tissue above the line of obstruction and that below the line of obstruction. Pulses are absent below the level of the occlusion. These changes are followed rapidly by cyanosis, mottling, and loss of sensory, reflex, and motor function. Tissue death occurs unless blood flow is restored.

Diagnosis of acute arterial occlusion is based on signs of impaired blood flow. It uses visual assessment, palpation of pulses, and methods to assess blood flow. Treatment of acute arterial occlusion is aimed at restoring blood flow. Thrombolytic therapy (*i.e.*, streptokinase or tissue plasminogen activator) may be used in an attempt to dissolve the clot. Anticoagulant therapy (*i.e.*, heparin) usually is given to prevent extension of the embolus. Application of heat and cold should be avoided, and the extremity should be protected from injury resulting from hard surfaces and overlying bedclothes. An embolectomy—surgical removal of the embolus—may be indicated.

Atherosclerotic Occlusive Disease

Atherosclerosis is an important cause of peripheral vascular disease and is seen most commonly in the vessels of the lower extremities. The condition is sometimes referred to as *arteriosclerosis obliterans*. The superficial femoral and popliteal arteries are the most commonly affected vessels. When lesions develop in the lower leg and foot, the tibial, common peroneal, or pedal vessels are the arteries most commonly affected. The disease is seen most commonly in men in their 60s and 70s.^{18,19} The risk factors for this disorder are similar to those for atherosclerosis. Cigarette smoking contributes to the progress of the atherosclerosis of the lower extremities and to the development of symptoms of ischemia. Persons with diabetes mellitus experience more extensive and rapidly progressive vascular disease than do individuals without diabetes.

As with atherosclerosis in other locations, the signs and symptoms of vessel occlusion are gradual. Usually, there is at least a 50% narrowing of the vessel before symptoms of ischemia arise. The primary symptom of chronic obstructive arterial disease is intermittent claudication or pain with walking.¹⁸ Typically, persons with the disorder report calf pain caused by ischemia of the gastrocnemius muscle, which has the highest oxygen consumption of any muscle group in the leg during walking. Some persons may report a vague aching feeling or numbness, rather than pain. Other signs of ischemia include atrophic changes and thinning of the skin and subcutaneous tissues of the lower leg and diminution in the size of the leg muscles. The foot often is cool, and the popliteal and pedal pulses are weak or absent. Limb color blanches with elevation of the leg because of the effects of gravity on perfusion pressure and becomes deep red when the leg is in the dependent position because of an autoregulatory increase in blood flow and a gravitational increase in perfusion pressure.

When blood flow is reduced to the extent that it no longer meets the minimal needs of resting muscle and nerves, ischemic pain at rest, ulceration, and gangrene develop. As tissue necrosis develops, there typically is severe pain in the region of skin breakdown, which is worse at night with limb elevation and is improved with standing.¹⁸

Diagnostic methods include inspection of the limbs for signs of chronic low-grade ischemia, such as subcutaneous atrophy, brittle toenails, hair loss, pallor, coolness, or dependent rubor. Palpation of the femoral, popliteal, posterior tibial, and dorsalis pedis pulses allows for an estimation of the level and degree of obstruction. Blood pressures may be taken at various levels on the leg to determine the level of obstruction. A Doppler ultrasound stethoscope may be used for detecting pulses and measuring blood pressure. Ultrasound imaging, radionuclide imaging, and contrast angiography also may be used as diagnostic methods.¹⁸

The tissues of extremities affected by atherosclerosis are easily injured and slow to heal. Treatment includes measures directed at protection of the affected tissues and preservation of functional capacity. Walking (slowly) to the point of claudication usually is encouraged because it increases collateral circulation.

Surgery (*i.e.*, femoropopliteal bypass grafting using a section of saphenous vein) may be indicated in severe cases. In persons with diabetes, the peroneal arteries between the knees and ankles commonly are involved, making revascularization difficult. Thromboendarterectomy with removal of the occluding core of atherosclerotic tissue may be done if the section of diseased vessel is short. Percutaneous transluminal angioplasty, in which a balloon catheter is inserted into the area of stenosis and the balloon inflated to increase vessel diameter, is another form of treatment.^{18,19}

Thromboangiitis Obliterans

Thromboangiitis obliterans (*e.g.*, Buerger's disease) is a vasculitis affecting the medium-size arteries, usually the plantar and digital vessels in the foot and lower leg. Arteries in the arm and hand also may be affected. Although primarily an arterial disorder, the inflammatory process often extends to involve adjacent veins and nerves. It usually is a disease of men between the ages of 25 and 40 years who are heavy cigarette smokers, but it can occur in women. The pathogenesis of the disorder remains speculative, although cigarette smoking and in some instances tobacco chewing seem to be involved. It has been suggested that the tobacco may trigger an immune response in susceptible persons or it may unmask a clotting defect, either of which could incite an inflammatory reaction of the vessel wall.²⁰

Pain is the predominant symptom of the disorder. It usually is related to distal arterial ischemia. During the early stages of the disease, there is intermittent claudication in the arch of the foot and the digits. In severe cases, pain is present even when the person is at rest. The impaired circulation increases sensitivity to cold. The peripheral pulses are diminished or absent, and there are changes in the color of the extremity. In moderately advanced cases, the extremity becomes cyanotic when the person assumes a dependent position, and the digits may turn reddish blue even when in a nondependent position. With lack of blood flow, the skin assumes a thin, shiny look, and hair growth and skin nutrition suffer. Chronic ischemia causes thick, malformed nails. If the disease continues to progress, tissues eventually ulcerate and gangrenous changes arise that may necessitate amputation.

Diagnostic methods are similar to those for atherosclerotic disease of the lower extremities. It is essential that the person stop smoking cigarettes or using tobacco. Other treatment measures are of secondary importance and focus on methods for producing vasodilation and preventing tissue injury. Sympathectomy may be done to alleviate the vasospastic manifestations of the disease.

Raynaud's Disease and Phenomenon

Raynaud's disease or phenomenon is a functional disorder caused by intense vasospasm of the arteries and arterioles in the fingers and, less often, the toes. The disorder is divided into two types: the primary type, called *Raynaud's disease*, occurs without demonstrable cause, and the secondary type, called *Raynaud's phenomenon*, is associated with other disease states or known causes of vasospasm.^{21,22}

Vasospasm implies an excessive vasoconstrictor response to stimuli that normally produce only moderate vasoconstriction. In contrast to other regional circulations that are supplied by vasodilator and vasoconstrictor fibers, the cutaneous vessels of the fingers and toes are innervated only by sympathetic vasoconstrictor fibers. In these vessels, vasodilation depends on withdrawal of sympathetic stimulation. Cooling of specific body parts such as the head, neck, and trunk produces a sympathetic-mediated reduction in digital blood flow, as does emotional stress.

Raynaud's disease is usually seen in otherwise healthy young women. It often is precipitated by exposure to cold or by strong emotions and usually is limited to the fingers. It also follows a more benign course than Raynaud's phenomenon, seldom causing tissue necrosis. The cause of vasospasm in primary Raynaud's disease is unknown. Hyperreactivity of the sympathetic nervous system has been suggested as a contributing cause.²¹ Raynaud's phenomenon is associated with previous vessel injury, such as frostbite, occupational trauma associated with the use of heavy vibrating tools, collagen diseases, neurologic disorders, and chronic arterial occlusive disorders. Another occupation-related cause is the exposure to alternating hot and cold temperatures such as that experienced by butchers and food preparers.²¹ Raynaud's phenomenon often is the first symptom of collagen diseases. It occurs in persons with scleroderma and those with systemic lupus erythematosus.

In Raynaud's disease and Raynaud's phenomenon, ischemia caused by vasospasm causes changes in skin color that progress from pallor to cyanosis, a sensation of cold, and changes in sensory perception, such as numbness and tingling. The color changes usually are first noticed in the tips of the fingers, later moving into one or more of the distal phalanges (Fig.15-11). After the ischemic episode, there is a period of



FIGURE 15-11 Raynaud's phenomenon. The tips of the fingers show marked pallor. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 514]. Philadelphia: Lippincott Williams & Wilkins)

hyperemia with intense redness, throbbing, and paresthesia. The period of hyperemia is followed by a return to normal color. Although all of the fingers usually are affected symmetrically, the involvement may affect only one or two digits. In some cases, only a portion of the digit is affected.

In severe, progressive cases usually associated with Raynaud's phenomenon, trophic changes may develop. The nails may become brittle, and the skin over the tips of the affected fingers may thicken. Nutritional impairment of these structures may give rise to arthritis. Ulceration and superficial gangrene of the fingers, although infrequent, may occur.

The initial diagnosis is based on history of vasospastic attacks supported by other evidence of the disorder. Immersion of the hand in cold water may be used to initiate an attack as an aid to diagnosis. Laser-Doppler velocimetry may be used to quantify digital blood flow during changes in temperature. Serial computed thermography (finger skin temperature) also may be a useful tool in diagnosing the extent of disease. Raynaud's disease is differentiated from Raynaud's phenomenon by excluding secondary disorders known to cause vasospasm.²²

Treatment measures are directed toward eliminating factors that cause vasospasm and protecting the digits from trauma during an ischemic episode. Abstinence from smoking and protection from cold are priorities. Avoidance of emotional stress is another important factor in controlling the disorder because anxiety and stress may precipitate a vascular spasm in predisposed persons. Vasoconstrictor medications, such as the decongestants contained in allergy and cold preparations, should be avoided. Drugs with a vasodilating action (*e.g.*, calcium channel blockers and α -adrenergic receptor blocking agents) may be indicated, particularly if episodes are frequent. Surgical interruption of sympathetic nerve pathways (sympathectomy) may be used for persons with severe symptoms.²²

In summary, the arterial system distributes blood to all the tissues of the body, and lesions of the arterial system exert their effects through ischemia or impaired blood flow. Hyperlipidemia with elevated cholesterol levels play a major role in the development of atherosclerotic disorders of the arterial system. Because cholesterol and triglycerides are insoluble in plasma, they transported as lipoproteins. Elevated levels of LDLs, which carry large amounts of cholesterol, are a major risk factor for atherosclerosis. The HDLs, which are protective, remove cholesterol from the tissues and carry it back to the liver for disposal. LDL receptors play a major role in removing cholesterol from the blood; persons with reduced numbers of LDL receptors are at particularly high risk for the development of atherosclerosis.

Atherosclerosis affects large and medium-size arteries, such as the coronary and cerebral arteries. It has an insidious onset, and its lesions usually are far advanced before symptoms appear. Risk factors associated with its development include factors such as heredity, sex, and age, which cannot be controlled; factors such as smoking, high blood pressure, high serum cholesterol levels, and diabetes, which can be modified; and other contributing factors such as obesity, lack of exercise, and stress.

Aneurysms are localized areas of vessel dilation caused by weakness of the arterial wall. Abdominal aortic aneurysms are the most common type of aneurysm. They are characterized by gradual and progressive enlargement of the abdominal aorta. The most serious consequence of abdominal aneurysms is rupture. Aortic dissection is an acute, life-threatening condition. It involves hemorrhage into the vessel wall with longitudinal tearing (dissection) of the vessel wall to form a blood-filled channel.

The vasculitides are a group of vascular disorders characterized by inflammation and necrosis of blood vessels in various tissues and organs. They can be caused by injury, infectious agents, or immune processes, or can occur secondary to other disease conditions, such as systemic lupus erythematosus. Occlusive disorders interrupt arterial flow of blood and interfere with the delivery of oxygen and nutrients to the tissues. Occlusion of flow can result from a thrombus, embolus, vessel compression, vasospasm, or structural changes in the vessel. Peripheral arterial diseases affect blood vessels outside the heart and thorax. They include atherosclerotic occlusive disease, thromboangiitis obliterans, and Raynaud's disease or phenomenon, caused by vessel spasm.

DISORDERS OF THE VENOUS CIRCULATION

Veins are low-pressure, thin-walled vessels that rely on the ancillary action of skeletal muscle pumps and changes in abdominal and intrathoracic pressure to return blood to the heart. Unlike the arterial system, the venous system is equipped with valves that prevent retrograde flow of blood. Although its structure enables the venous system to serve as a storage area for blood, it also renders the system susceptible to problems related to stasis and venous insufficiency. This section focuses on three common problems of the venous system: varicose veins, venous insufficiency, and venous thrombosis.

Venous Circulation of the Lower Extremities

The venous system in the legs consists of two components: the superficial veins (*i.e.*, saphenous vein and its tributaries) and the deep venous channels (Fig. 15-12). Perforating or commu-

KEY CONCEPTS

DISORDERS OF THE VENOUS CIRCULATION

- The venous system is a low-pressure system that relies on the pumping action of the skeletal muscles to move blood forward and the presence of venous valves to prevent retrograde flow.
- Varicose veins are dilated and tortuous veins that result from a sustained increase in pressure that causes the venous valves to become incompetent, allowing for reflux of blood and vein engorgement.
- Thrombophlebitis refers to thrombus formation and the accompanying inflammatory response that occurs in veins. Deep vein thrombosis may be a precursor to pulmonary embolism.

nicating veins connect these two systems. Blood from the skin and subcutaneous tissues in the leg collects in the superficial veins and is then transported across the communicating veins into the deeper venous channels for return to the heart. Venous valves prevent the retrograde flow of blood and play an important role in the function of the venous system. Although these valves are irregularly located along the length of the veins, they almost always are found at junctions where the communicating veins merge with the larger deep veins and where two veins meet. The number of venous valves differs somewhat from one person to another, as does the structural competence, factors that may help to explain the familial predisposition to development of varicose veins.

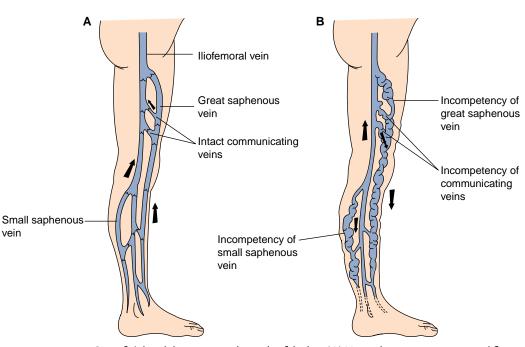
The action of the leg muscles assists in moving venous blood from the lower extremities back to the heart. When a person walks, the action of the leg muscles serves to increase flow in the deep venous channels and return venous blood to the heart (Fig. 15-13). The function of the so-called muscle pump, located in the gastrocnemius and soleus muscles of the lower extremities, can be compared with the pumping action of the heart.²³ During muscle contraction, which is similar to systole, valves in the communicating channels close to prevent backward flow of blood into the superficial system, as blood in the deep veins is moved forward by the action of the contracting muscles. During relaxation, which is similar to diastole, the communicating valves open, allowing blood from the superficial veins to move into the deep veins.

Varicose Veins

Varicose, or dilated, tortuous veins of the lower extremities are common and often lead to secondary problems of venous insufficiency (Fig. 15-12). Varicose veins are described as being primary or secondary. Primary varicose veins originate in the superficial saphenous veins, and secondary varicose veins result from impaired flow in the deep venous channels. Approximately 80% to 90% of venous blood from the lower extremities is transported through the deep channels. The development of secondary varicose veins becomes inevitable when flow in these deep channels is impaired or blocked. The most common cause of secondary varicose veins is deep vein thrombosis (DVT). Other causes include congenital or acquired arteriovenous fistulas, congenital venous malformations, and pressure on the abdominal veins caused by pregnancy or a tumor.

Primary varicose veins is more common after 50 years of age and in obese persons, and it occurs more often in women than men, probably because of venous stasis caused by pregnancy.² More than 50% of persons with primary varicose veins have a family history of the disorder, suggesting that heredity may play a role. Prolonged standing and increased intra-abdominal pressure are important contributing factors in the development of primary varicose veins. One of the most important factors in the elevation of venous pressure is the hydrostatic effect associated with the standing position. When a person is in the erect position, the full weight of the venous columns of blood is transmitted to the leg veins. The effects of gravity are compounded in persons who stand for long periods without using their leg muscles to assist in pumping blood back to the heart.

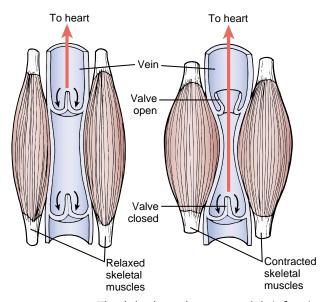
Because there are no valves in the inferior vena cava or common iliac veins, blood in the abdominal veins must be supported by the valves located in the external iliac or femoral



■ FIGURE 15-12 ■ Superficial and deep venous channels of the leg. (A) Normal venous structures and flow patterns. (B) Varicosities in the superficial venous system are the result of incompetent valves in the communicating veins. The arrows in both views indicate the direction of blood flow. (Modified from Abramson D.I. [1974]. *Vascular disorders of the extremities* [2nd ed.]. Philadelphia: Lippincott Williams & Wilkins)

veins. When intra-abdominal pressure increases, as it does during pregnancy, or when the valves in these two veins are absent or defective, the stress on the saphenofemoral junction is increased. Lifting also increases intra-abdominal pressure and decreases flow of blood through the abdominal veins. Occupations that require repeated heavy lifting predispose to development of varicose veins.

Prolonged exposure to increased pressure causes the venous valves to become incompetent so they no longer close properly. When this happens, the reflux of blood causes fur-



■ FIGURE 15-13 ■ The skeletal muscle pumps and their function in promoting blood flow in the deep and superficial calfvessels of the leg.

ther venous enlargement, pulling the valve leaflet apart and causing valvular incompetence in sections of adjacent distal veins. Another consideration in the development of varicose veins is the fact that the superficial veins have only subcutaneous fat and superficial fascia for support, but the deep venous channels are supported by muscle, bone, and connective tissue. Obesity reduces the support provided by the superficial fascia and tissues, increasing the risk for development of varicose veins.

The signs and symptoms associated with primary varicose veins vary. Most women with superficial varicose veins complain of their unsightly appearance. In many cases, aching in the lower extremities and edema, especially after long periods of standing, may occur. The edema usually subsides at night when the legs are elevated. When the communicating veins are incompetent, symptoms are more common.

The diagnosis of varicose veins often can be made after physical inspection. The Doppler ultrasonic flow probe also may be used to assess the flow in the large vessels. Angiographic studies using a radiopaque contrast medium also are used to assess venous function.

After the venous channels have been repeatedly stretched and the valves rendered incompetent, little can be done to restore normal venous tone and function. Ideally, measures should be taken to prevent the development and progression of varicose veins. These measures center on avoiding activities such as continued standing that produce prolonged elevation of venous pressure. Treatment measures for varicose veins focus on improving venous flow and preventing tissue injury. When correctly fitted, elastic support stockings or leggings compress the superficial veins and prevent distention. The most precise control is afforded by prescription stockings, measured to fit properly. These stockings should be applied before the standing position is assumed, when the leg veins are empty.

268 Unit Four: Alterations in the Cardiovascular System

Sclerotherapy, which often is used in the treatment of small residual varicosities, involves the injection of a sclerosing agent into the collapsed superficial veins to produce fibrosis of the vessel lumen. Surgical treatment consists of removing the varicosities and the incompetent perforating veins, but it is limited to persons with patent deep venous channels.

Chronic Venous Insufficiency

The term *venous insufficiency* refers to the physiologic consequences of DVT, valvular incompetence, or a combination of both conditions. The most common cause is DVT, which causes deformity of the valve leaflets, rendering them incapable of closure. In the presence of valvular incompetence, effective unidirectional flow of blood and emptying of the deep veins cannot occur. The muscle pumps also are ineffective, often driving blood in retrograde directions. Secondary failure of the communicating and superficial veins subjects the subcutaneous tissues to high pressures.

With venous insufficiency, there are signs and symptoms associated with impaired blood flow. In contrast to the ischemia caused by arterial insufficiency, venous insufficiency leads to tissue congestion, edema, and eventual impairment of tissue nutrition.^{23,24} The edema is exacerbated by long periods of standing. Necrosis of subcutaneous fat deposits occurs, followed by skin atrophy. Brown pigmentation of the skin caused by hemosiderin deposits resulting from the breakdown of red blood cells is common. Secondary lymphatic insufficiency occurs, with progressive sclerosis of the lymph channels in the face of increased demand for clearance of interstitial fluid.

In advanced venous insufficiency, impaired tissue nutrition causes stasis dermatitis and the development of stasis or venous ulcers (Fig. 15-14).² Stasis dermatitis is characterized by the presence of thin, shiny, bluish brown, irregularly pigmented desquamative skin that lacks the support of the underlying subcutaneous tissues. Minor injury leads to relatively painless ulcerations that are difficult to heal. The lower part of the leg is particularly prone to development of stasis dermatitis and venous ulcers. Most lesions are located medially over the ankle and lower leg, with the highest frequency just above the medial malleolus. Persons with long-standing venous insufficiency may experience stiffening of the ankle joint and loss of muscle mass and strength.

Venous Thrombosis

The term *venous thrombosis*, or *thrombophlebitis*, describes the presence of thrombus in a vein and the accompanying inflammatory response in the vessel wall. Thrombi can develop in the superficial or the deep veins. DVT most commonly occurs in the lower extremities. DVT of the lower extremity is a serious disorder, complicated by pulmonary embolism (see Chapter 21), recurrent episodes of DVT, and development of chronic venous insufficiency.²⁵

In 1846, Virchow described the triad that has come to be associated with venous thrombosis: stasis of blood, increased blood coagulability, and vessel wall injury.²⁶ Risk factors for DVT are summarized in Chart 15-2. *Stasis of blood* occurs with immobility of an extremity or the entire body. Bed rest and immobilization are associated with decreased blood flow, venous pooling in the lower extremities, and increased risk of DVT.



■ FIGURE 15-14 ■ Varicose veins of the legs. Severe varicosities of the superficial leg veins have led to stasis dermatitis and secondary ulcerations. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 525]. Philadelphia: Lippincott Williams & Wilkins)

CHART 15-2 Risk Factors Associated With Venous Thrombosis*

Venous Stasis

Bed rest Immobility Spinal cord injury Acute myocardial infarction Congestive heart failure Shock Venous obstruction

Hyperreactivity of Blood Coagulation

Stress and trauma Pregnancy Childbirth Oral contraceptive use Dehydration Cancer

Vascular Trauma

Indwelling venous catheters Surgery Massive trauma or infection Fractured hip Orthopedic surgery

*Many of these disorders involve more than one mechanism.

Persons who are immobilized by a hip fracture, joint replacement, or spinal cord injury are particularly vulnerable to DVT. The risk of DVT is increased in situations of impaired cardiac function. This may account for the relatively high incidence in persons with acute myocardial infarction and congestive heart failure. Elderly persons are more susceptible than younger persons, probably because disorders that produce venous stasis occur more frequently in older persons. Long airplane travel poses a particular threat in persons predisposed to DVT because of prolonged sitting and increased blood viscosity caused by dehydration.

Hypercoagulability and conditions that increase the concentration or activation of clotting factors predispose to DVT. The postpartum state is associated with increased levels of fibrinogen, prothrombin, and other coagulation factors. The use of oral contraceptives appears to increase coagulability and predispose to venous thrombosis, especially in women older than 30 years and in those who smoke. Certain cancers are associated with increased clotting tendencies, and although the reason for this is largely unknown, substances that promote blood coagulation may be released from the tissues because of the cancerous growth.

Vessel injury can result from a trauma situation or from surgical intervention. It also may occur secondary to infection or inflammation of the vessel wall. Persons undergoing hip surgerv and total hip replacement are at particular risk because of trauma to the femoral and iliac veins, and in the case of hip replacement, thermal damage from heat generated by the polymerization of the acrylic cement that is used in the procedure.²⁵ Venous catheters are another source of vascular injury.

Many persons with DVT are asymptomatic. The person may have experienced a pulmonary emboli with no previous signs or symptoms. When present, the most common signs and symptoms of DVT are those related to the inflammatory process: pain, swelling, and deep muscle tenderness. Fever, general malaise, and an elevated white blood cell count and sedimentation rate are accompanying indications of inflammation. There may be tenderness and pain along the vein. Swelling may vary from minimal to maximal. The site of thrombus formation determines the location of the physical findings. The most common site is in the venous sinuses in the soleus muscle and posterior tibial and peroneal veins. Swelling in these cases involves the foot and ankle, although it may be slight or absent. Calf pain and tenderness are common. Femoral vein thrombosis with calf thrombosis produces pain and tenderness in the distal thigh and popliteal area. Thrombi in ileofemoral veins produce the most profound manifestations, with swelling, pain, and tenderness of the entire extremity. With DVT in the calf veins, active dorsiflexion of the foot produces calf pain (Homans' sign).

The risk of pulmonary embolism emphasizes the need for early detection and treatment of DVT. Several tests are useful for this purpose: ascending venography, impedance plethysmography, and ultrasonography. The goals of treatment for DVT are to prevent formation of additional thrombi, prevent extension and embolization of existing thrombi, and to minimize venous valve damage. Anticoagulation therapy is used to treat and prevent venous thrombosis. A 15- to 20-degree elevation of the legs prevents stasis. Surgical removal of the thrombus may be undertaken in selected cases. Surgical interruption of the vena cava may be done in persons at high risk for experiencing pulmonary emboli. This procedure involves ligating the vena cava with a suture or clamp or in creating a filter-like insertion to prevent large clots from moving through the vessel.

When possible, venous thrombosis should be prevented in preference to being treated. Early ambulation after childbirth and surgery is one measure that decreases the risk of thrombus formation. Exercising the legs and wearing support stockings improve venous flow. A further precautionary measure is to avoid assuming body positions that favor venous pooling. Antiembolism stockings of the proper fit and length should be used routinely in persons at risk for DVT. Another strategy used for immobile persons at risk for experiencing DVT is a sequential pneumatic compression device. This consists of a plastic sleeve that encircles the legs and provides alternating periods of compression on the lower extremity. When properly used, these devices enhance venous emptying to augment flow and reduce stasis.

In summary, the storage function of the venous system renders it susceptible to venous insufficiency, stasis, and thrombus formation. Varicose veins occur with prolonged distention and stretching of the superficial veins owing to venous insufficiency. Varicosities can arise because of defects in the superficial veins (*i.e.*, primary varicose veins) or because of impaired blood flow in the deep venous channels (i.e., secondary varicose veins). Venous insufficiency reflects chronic venous stasis resulting from valvular incompetence. It is associated with stasis dermatitis and stasis or venous ulcers. Venous thrombosis describes the presence of thrombus in a vein and the accompanying inflammatory response in the vessel wall. It is associated with vessel injury, stasis of venous flow, and hypercoagulability states. Thrombi can develop in the superficial or the deep veins (*i.e.*, DVT). Thrombus formation in deep veins is a precursor to venous insufficiency and embolus formation.

DISORDERS OF BLOOD FLOW CAUSED BY EXTRAVASCULAR FORCES

Blood flow occurs along a pressure gradient, moving from the arterial to the venous side of the circulation. For blood to move through the vessels of the systemic circulation, arterial pressure must be greater than venous pressure, and the arterial and venous pressures must be greater than the external pressure of the surrounding tissues. Injury or infections that cause tissue swelling can compromise blood flow, particularly in parts of the body where the skin or other supporting tissues cannot expand to accommodate the increased volume. In other situations, external pressure may compress the tissues and the blood vessels. Two conditions resulting from compromised blood flow caused by extravascular forces are compartment syndrome and pressure ulcers.

Compartment Syndrome

The muscles and nerves of an extremity are enclosed in a tough, inelastic fascial envelope called a muscle compartment (Fig. 15-15). Compartment syndrome describes a condition of in-

KEY CONCEPTS

DISORDERS OF BLOOD FLOW DUE TO EXTRAVASCULAR FORCES

- Blood flow requires that arterial pressure is greater than venous pressure and that arterial, venous, and capillary pressures are greater than the pressure surrounding the vessels.
- Compartment syndrome describes a condition of compromised blood flow resulting from pressure in an anatomic space that cannot expand. Causes include decreases in compartment size (*i.e.*, cast) or increases in compartment volume (*i.e.*, internal bleeding or edema).
- Pressure ulcers are ischemic lesions of the skin and underlying tissues caused by compression of blood vessels due to external pressure, such as that exerted by the weight of the body on the bed or chair surface. Prevention of pressure ulcers is preferable to treatment. Frequent position changes and meticulous skin care are essential components of prevention.

creased pressure in a limited anatomic space, usually a muscle compartment, that impairs circulation and produces ischemic tissue injury.²⁷ If the pressure in the compartment is sufficiently high, tissue circulation is compromised, causing death of nerve and muscle cells. Permanent loss of function may occur.

The amount of pressure required to produce a compartment syndrome depends on many factors, including the duration of the pressure elevation, the metabolic rate of the tissues, vascular tone, and local blood pressure. Less tissue pressure is required to stop circulation when hypotension or vasoconstriction is present. Intracompartmental pressures of 30 to 40 mm Hg (normal is approximately 6 mm Hg) are considered sufficient to impair capillary blood flow.²⁸

Compartment syndrome can result from: (1) a decrease in compartment size or (2) an increase in the volume of its contents (Chart 15-3). Among the causes of decreased compartment size are constrictive dressings and casts, closure of fascial defects, and thermal injuries or frostbite. In persons with circumferential third-degree burns, the inelastic and constricting eschar decreases the size of the underlying compartments. Burns also are associated with the formation of massive edema and an increase in compartment volume. The combination of the two problems may lead to necrosis of the underlying neuromuscular tissues. Frostbite produces neuromuscular injury for similar reasons.

An increase in compartment volume can be caused by trauma, vascular injury and bleeding, infiltration of intravenous infusions, postischemic swelling, and venous obstruction. One of the most important causes of compartment syndrome is bleeding and edema caused by fractures and osteotomies (see Chapter 42). Contusions and soft tissue injury also are common causes of compartment syndrome. Bleeding can occur as a complication of arterial punctures, particularly in persons with bleeding disorders or those who are receiving anticoagulant drugs. Infiltration of intravenous fluids also can restrict compartment size and cause compartment ischemia and postischemic swelling. Increased compartment volume may follow ischemic events, such as arterial occlusion, that are of sufficient duration to produce capillary damage, causing increased capillary permeability and edema. During unattended coma caused by drug overdose or carbon monoxide poisoning, high compartment pressures are produced when an extremity is compressed by the weight of the overlying head or torso. Exercise may produce acute or chronic elevations in compartment pressure.

The most important symptom of compartment syndrome is unrelenting pain, usually described as a deep, throbbing sensation, that is greater than that expected for the primary problem, such as fracture or contusion. Pain with passive stretch

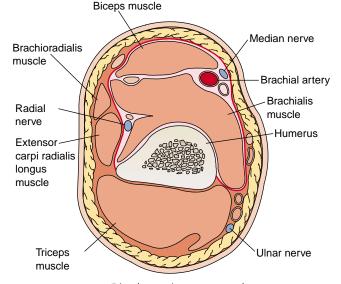


FIGURE 15-15 Distal anterior arm muscle compartments, showing the location of fascia, muscles, nerves, and blood vessels.

CHART 15-3 Causes of Compartment Syndrome	
Decreased Compartment Size	
Constrictive dressings and casts Infiltration of intravenous fluids Thermal injury and frostbite Surgical closure of fascial defects	
Increased Compartment Volume	
Fractures and orthopedic surgery Trauma and bleeding Postischemic injury Severe exercise Prolonged immobilization with limb compression (e.g., drug overdose) Thermal injury and frostbite Intravenous infiltration	

is a common finding. Tenseness and tenderness of the involved compartment are specific symptoms of compartment syndrome. The skin over the compartment may become taut, shiny, warm, and red. Paresthesias progressing to anesthesia occur secondary to nerve involvement. Muscle weakness results from muscle ischemia.

It is important that persons at risk for compartment syndrome be identified and that proper assessment methods be instituted. Assessment should include pain assessment, examination of sensory (*i.e.*, light touch and two-point discrimination) and motor function (*i.e.*, movement and muscle strength), test of passive stretch, and palpation of the muscle compartments. Peripheral pulses frequently are normal in the presence of compartment syndrome because the major arteries are located outside the muscle compartments. Although edema may make it difficult to palpate the pulse, the increased compartment pressure seldom is sufficient to occlude flow in a major artery. Doppler methods usually confirm the existence of a pulse. Direct measurements of tissue pressure can be obtained using a needle or wick catheter inserted into the muscle compartment. This method is particularly useful in persons who are unresponsive and in those with nerve deficits. Compartment decompression is recommended when pressures rise to 30 mm Hg.

Treatment consists of reducing compartmental pressures. This entails cast splitting or removal of restrictive dressings. These procedures often are sufficient to relieve most of the underlying pressure and symptoms. When an extremity is elevated, its arterial pressure falls because of the effects of gravity. Therefore, limb elevation is not recommended when compartment syndrome is suspected. When compartment syndrome cannot be relieved by conservative measures, a fasciotomy may become necessary. During this procedure, the fascia is incised longitudinally and separated so that the compartment volume can expand and blood flow can be re-established. Because of potential problems with wound infection and closure, this procedure is performed as a last resort.

Pressure Ulcers

Pressure ulcers are ischemic lesions of the skin and underlying structures caused by external pressure that impairs the flow of blood and lymph. Pressure ulcers often are referred to as decubitus ulcers or bedsores. The word decubitus comes from the Latin term meaning "lying down." However, a pressure ulcer may result from pressure exerted in the seated or the lying position. Pressure ulcers are most likely to develop over a bony prominence, but they may occur on any part of the body that is subjected to external pressure, friction, or shearing forces.

Several subpopulations are at particular risk, including persons with quadriplegia, elderly persons with restricted activity and hip fractures, and persons in the critical care setting. The prevention and treatment of pressure ulcers is a public health issue and is addressed in *Healthy People 2010*, a national public health policy statement, which has set a target of a 50% decrease in prevalence of pressure ulcers in nursing home residents.29

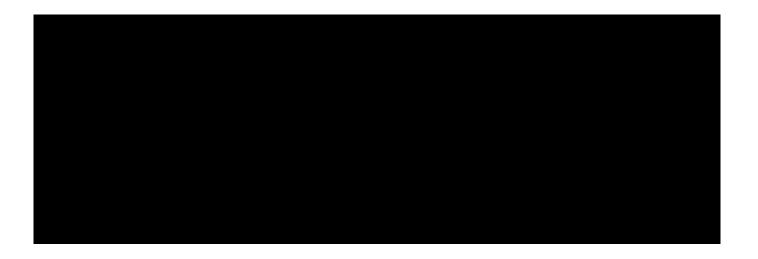
Mechanisms of Development

Two factors contribute to the development of pressure ulcers: (1) external pressure that compresses blood vessels and (2) friction and shearing forces that tear and injure blood vessels.

External pressure that exceeds capillary pressure interrupts blood flow in the capillary beds. When the pressure between a bony prominence and a support surface exceeds the normal capillary filling pressure of approximately 32 mm Hg, capillary flow essentially is obstructed.³⁰ If this pressure is applied constantly for 2 hours, oxygen deprivation coupled with an accumulation of metabolic end products leads to irreversible tissue damage. The same amount of pressure causes more damage when it is distributed over a small area than when it is distributed over a larger area. Approximately 7 lb of pressure per square inch of tissue surface is sufficient to obstruct blood flow.³¹

Whether a person is sitting or lying down, the weight of the body is borne by tissues covering the bony prominences. More than 90% of pressure ulcers are located on the lower part of the body, most often over the sacrum, the coccygeal areas, the ischial tuberosities, and the greater trochanter. Pressure over a bony area is transmitted from the surface to the underlying dense bone, compressing all of the intervening tissue. As a result, the greatest pressure occurs at the surface of the bone and dissipates outward in a conelike manner toward the surface of the skin (Fig. 15-16). Thus, extensive underlying tissue damage can be present when a small superficial skin lesion is first noticed.

Altering the distribution of pressure from one skin area to another prevents tissue injury. Pressure ulcers most commonly



occur in persons with conditions such as spinal cord injury in which normal sensation and movement to effect redistribution of body weight are impaired. Normally, persons unconsciously shift their weight to redistribute pressure on the skin and underlying tissues. For example, during the night, people turn in their sleep, preventing ischemic injury of tissues that overlie the bony prominences that support the weight of the body; the same is true for sitting for any length of time. The movements needed to shift the body weight are made unconsciously, and only when movement is restricted do people become aware of discomfort.

Shearing forces are caused by the sliding of one tissue layer over another with stretching and angulation of blood vessels, causing injury and thrombosis. For persons who are bedridden, injury commonly occurs when the head of the bed is elevated, causing the torso to slide down toward the foot of the bed. When this happens, friction and perspiration cause the skin and superficial fascia to remain fixed against the bed linens while the deep fascia and skeleton slide downward. The same thing can happen when a person sitting up in a chair slides downward. Another source of shearing forces is pulling, rather than lifting, the bedridden person up in bed. In this case, the skin remains fixed to the sheet while the fascia and muscles are pulled upward.

Prevention

The prevention of pressure ulcers is preferable to treatment. In 1992, a special panel of the Agency for Health Care Policy and Research (AHCPR; now the Agency for Healthcare Research and Quality), the Panel for the Prediction and Prevention of Pressure Ulcers in Adults, released the Clinical Practice Guidelines for Pressure Ulcers in Adults.³² The panel recommended four overall goals: (1) identifying at-risk persons who need preventative measures and the specific factors placing them at risk; (2) maintaining and improving tissue tolerance to prevent injury; (3) protecting against the adverse effects of external mechanical forces (*i.e.*, pressure, friction, and shear); and (4) reducing the incidence of pressure ulcers through educational programs.³² A 1994 publication of the AHCPR made specific recommendations for assessment of the person with pressure ulcers, management of tissue load, ulcer care, managing bacterial colonization and infection, operative repair, and education and quality control.³³

Risk factors identified as contributing to the development of pressure ulcers were those related to sensory perception (*i.e.*, ability to respond meaningfully to pressure-related discomfort), level of skin moisture, urine and fecal continence, nutrition and hydration status, mobility, circulatory status, and presence of shear and friction forces.

Methods for preventing pressure ulcers include frequent position change, meticulous skin care, and frequent and careful observation to detect early signs of skin breakdown. Adequate hydration of the stratum corneum appears to protect the skin against mechanical insult.³² The prevention of dehydration improves the circulation. It also decreases the concentration of urine, thereby minimizing skin irritation in persons who are incontinent, and it reduces urinary problems that contribute to incontinence. Maintenance of adequate nutrition is important. Anemia and malnutrition contribute to tissue breakdown and delay healing after tissue injury has occurred.

Staging and Treatment

Pressure ulcers can be staged using four categories.^{32,33} Stage I ulcers are characterized by a defined area of persistent redness in lightly pigmented skin or an area of persistent redness with blue or purple hues in darker pigmented skin. Stage II ulcers represent a partial-thickness loss of skin involving epidermis or dermis, or both. The ulcer is superficial and presents clinically as an abrasion, a blister, or a shallow crater. Stage III ulcers represent a full-thickness skin loss involving damage and necrosis of subcutaneous tissue that may extend down to but not through underlying fascia. The ulcer manifests as a deep crater with or without undermining of adjacent tissue. Stage IV ulcers involve full-thickness skin loss and necrosis with extensive destruction or damage to the underlying subcutaneous tissues that may extend to involve muscle, bone, and supporting structures (e.g., tendon or joint capsule).

After skin breakdown has occurred, special treatment measures are needed to prevent further ischemic damage, reduce bacterial contamination and infection, and promote healing. Treatment methods are selected based on the stage of the ulcer.^{30,33} Stage I ulcers usually are treated with frequent turning and measures to remove pressure. Stage II or III ulcers with little exudate are treated with petroleum gauze, or semipermeable or occlusive dressings to maintain a moist healing environment. Stage III ulcers usually require debridement (*i.e.*, removal of necrotic tissue and eschar). This can be done surgically, with wet-to-dry dressings, or through the use of proteolytic enzymes.³⁰ Stage IV wounds often require packing to obliterate dead space and are covered with nonadherent dressings. Stage IV ulcers may require surgical interventions, such as skin grafts or myocutaneous flaps.

In summary, blood flow to the tissues is dependent upon a pressure gradient between the arterial and venous side of the circulation and a transmural pressure (*i.e.*, internal minus external) that holds the vessel open. Under certain conditions, such as compartment syndrome and pressure ulcers, increases in external pressures can exceed intravascular pressure and interrupt blood flow.

Compartment syndrome is a condition of increased pressure in a muscle compartment that compromises blood flow and potentially leads to death of nerve and muscle tissue. It can result from a decrease in compartment size (*e.g.*, constrictive dressings, closure of fascial defects, thermal injury, frostbite) or an increase in compartment volume (*e.g.*, postischemic swelling, fractures, contusion and soft tissue trauma, bleeding caused by vascular injury, venous congestion).

Pressure ulcers are caused by ischemia of the skin and underlying tissues. They result from external pressure, which disrupts blood flow, or shearing forces, which cause stretching and injury to blood vessels. Pressure ulcers are divided into four stages, according to the depth of tissue involvement. The prevention of pressure ulcers is preferable to treatment. The goals of prevention should include identifying at-risk persons who need prevention along with the specific factors placing them at risk; maintaining and improving tissue tolerance to pressure to prevent injury; and protecting against the adverse effects of external mechanical forces (*i.e.*, pressure, friction, and shear).

REVIEW QUESTIONS

Characterize the role of LDL, HDL, and LDL receptors in the pathogenesis of atherosclerosis.

List risk factors in atherosclerosis and relate them to the possible mechanisms associated with the development of atherosclerosis.

Compare the mechanisms and manifestations of ischemia associated with atherosclerotic peripheral vascular disease and Raynaud's phenomenon.

Compare the pathology and manifestations of thoracic or abdominal and aortic dissection.

Describe the pathology of venous insufficiency and relate it to the development of stasis dermatitis and venous ulcers.

Cite risk factors associated with venous thrombosis and describe the manifestation of the disorder and its treatment.

Characterize the impairment of blood flow caused by external compression of blood vessels in compartment syndrome and relate it to five possible causes of compartment syndrome.

Explain how pressure and shearing forces contribute to the development of pressure ulcers.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- 1. American Heart Association. (2002). Cholesterol statistics for professionals. [On-line]. Available: www.americanheart.org/cholesterol.
- Rubin E., Farber J.L. (Eds.) (1999). Pathology (3rd ed., pp. 491–509, 520–522). Philadelphia: Lippincott Williams & Wilkins.
- Beisiegel U. (1998). Lipoprotein metabolism. European Heart Journal 19 (Suppl. A), A20–A23.
- Schoen F.J., Cotran R.S. (2003). Blood vessels. In Kumar V., Cotran R.S., Robbins S.L., et al. (Eds.), *Robbins basic pathology* (7th ed., pp. 325–360). Philadelphia: W.B. Saunders.
- Steinberg D., Gotto A.M. (1999). Preventing coronary artery disease by lowering cholesterol levels. *Journal of the American Medical Association*, 282, 2043–2050.
- National Institutes of Health Expert Panel. (2001). Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). (NIH publication no. 01-3670.) Bethesda, MD: National Institutes of Health.
- Verges B.L. (1999). Dyslipidemia in diabetes mellitus: Review of the main lipoprotein abnormalities and their consequences on the development of atherogenesis. *Diabetes and Metabolism* 25 (Suppl. 3), 32–40.
- 8. American Heart Association. (2002). Heart facts 2002: All Americans. [On-line]. Available: www.americanheart.org.
- Ross R. (1999). Mechanisms of disease: Atherosclerosis—an inflammatory disease. New England Journal of Medicine 340, 115–126.

- Kullo I.J., Gau G.T., Tajik A.J. (2000). Novel risk factors for atherosclerosis. *Mayo Clinic Proceedings* 75, 369–380.
- Hankey G.J., Eikelboom J.W. (1999). Homocysteine and vascular disease. *Lancet* 354, 407–413.
- Fong I.W. (2000). Emerging relations between infectious diseases and coronary artery disease and atherosclerosis. *Canadian Medical Association Journal* 163, 49–56.
- Thompson R.W. (2002). Detection and management of small aortic aneurysms. New England Journal of Medicine 346(19), 1484–1486.
- Creager M.A., Halperin J.L., Whittemore A.D. (1992). Aneurysm disease of the aorta and its branches. In Loscalzo J., Creager M.A., Dzau V.J. (Eds.), Vascular medicine: A textbook of vascular biology and diseases (pp. 903–923). Boston: Little, Brown.
- Coady M.A., Rizzo J.A., Goldstein L.J., et al. (1999). Natural history, pathogenesis, and etiology of thoracic aortic aneurysms and dissections. *Cardiology Clinics of North America* 17, 615–635.
- Savage C.O.S., Harper L., Cockwell P., et al. (2000). Vasculitis. Clinical review: ABC of arterial and vascular disease. *British Medical Journal* 320, 1325–1328.
- 17. Gross W.L., Trabandt A., Reinhold-Keller E. (2000). Diagnosis and evaluation of vasculitis. *Rheumatology* 39, 245–252.
- Bartholomew J.R., Gray B.H. (1999). Large artery occlusive disease. Rheumatic Disease Clinics of North America 25, 669–686.
- Carter S.A. (1999). Peripheral arterial disease. Canadian Journal of Cardiology 15 (Suppl. G), 106G–109G.
- 20. Tanaka K. (1998). Pathology and pathogenesis of Buerger's disease. International Journal of Cardiology 66 (Suppl. 1), S237–S242.
- 21. Belch J. (1997). Raynaud's phenomenon. *Cardiovascular Research* 33, 25–30.
- 22. Ho M., Belch J. (1998). Raynaud's phenomenon: State of the art. *Scandinavian Journal of Rheumatology* 27, 319–322.
- Alguire P.C., Mathes B.M. (1997). Chronic venous insufficiency and venous ulceration. *Journal of General Internal Medicine* 12, 374–383.
- Gorman W.P., Davis K.R., Donnelly R. (2000). Swollen lower limb— 1: General assessment and deep vein thrombosis: Clinical review: ABC of arterial and venous disease. *British Medical Journal* 320, 1453–1456.
- 25. Virchow R. (1846). Weinere untersuchungen uber die verstropfung der lungenrarterie und ihre folgen. *Beitrage zur Experimentelle Pathologie und Physiologie* 2, 21.
- Weinmann E.E., Salzman E.W. (1994). Deep-vein thrombosis. New England Journal of Medicine 331, 1630–1641.
- Kalb R.L. (1999). Preventing the sequelae of compartment syndrome. Hospital Practice 34 (1), 105–107.
- Matsen F. (1975). Compartment syndrome: A unified concept. Clinical Orthopaedics and Related Research 113, 8–13.
- 29. National Institutes of Health. *Healthy people 2010*. [On-line]. Available: www.health.gov/healthypeople.
- Patterson J.A., Bennett R.G. (1995). Prevention and treatment of pressure sores. *Journal of the American Geriatric Society* 43, 919–927.
- Beland I., Passos J.Y. (1981). Clinical nursing (4th ed., p. 1112). New York: Macmillan.
- 32. Panel for the Prediction and Prevention of Pressure Ulcers in Adults. (1992). Pressure ulcers in adults: Prediction and prevention. Clinical practice guideline no. 3. AHCPR publication no. 92-0047. Rockville, MD: Agency for Health Care Policy and Research, Public Health Service, U.S. Department of Health and Human Services.
- Bergstrom N., Bennett M.A., Carlson C.E., et al. (1994). Treatment of pressure ulcers. Clinical practice guideline no. 15. AHCPR publication no. 95-0652. Rockville, MD: U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research.

CHAPTER **16**

Alterations in Blood Pressure

Control of Blood Pressure Determinants of Blood Pressure

Systolic Blood Pressure **Diastolic Blood Pressure Pulse Pressure** Mean Arterial Pressure **Mechanisms of Blood Pressure Regulation Short-Term Regulation Long-Term Regulation Hypertension Essential Hypertension** Mechanisms of Blood Pressure Elevation **Contributing Factors** Manifestations **Diagnosis and Treatment** Systolic Hypertension Secondary Hypertension **Renal Hypertension** Disorders of Adrenocorticosteroid Hormones Pheochromocytoma Coarctation of the Aorta Malignant Hypertension Hypertension During Pregnancy Classification **Diagnosis and Treatment** Hypertension in Children **Diagnosis and Treatment** Hypertension in the Elderly **Diagnosis and Treatment Orthostatic Hypotension Classification** Causes **Reduced Blood Volume Drug-Induced Hypotension**

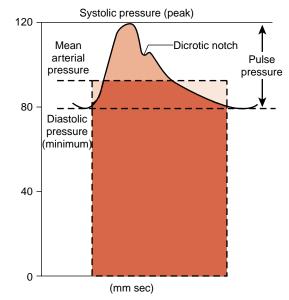
Aging Bed Rest Disorders of Autonomic Nervous System Function Diagnosis and Treatment

Blood pressure is probably one of the most variable but best regulated functions of the body. The purpose of the control of blood pressure is to keep blood flow constant to vital organs such as the heart, brain, and kidneys. Without constant flow to these organs, death ensues within seconds, minutes, or days. Although a decrease in flow produces an immediate threat to life, the continuous elevation of blood pressure that occurs with hypertension is a contributor to premature death and disability due to its effect on the heart, blood vessels, and kidneys.

CONTROL OF BLOOD PRESSURE

The arterial blood pressure reflects the rhythmic ejection of blood from the left ventricle into the aorta. It rises as the left ventricle contracts and falls as it relaxes. The contour of the arterial pressure tracing shown in Figure 16-1 is typical of the pressure changes that occur in the large arteries of the systemic circulation. There is a rapid rise in the pulse contour during left ventricular contraction, followed by a slower rise to peak pressure. Approximately 70% of the blood that leaves the left ventricle is ejected during the first one third of systole; this accounts for the rapid rise in the pressure contour. The end of systole is marked by a brief downward deflection and formation of the dicrotic notch, which occurs when ventricular pressure falls below that in the aorta. The sudden closure of the aortic valve is associated with a small rise in pressure due to continued contraction of the aorta and other large vessels against the closed valve. As the ventricles relax and blood flows into the peripheral vessels during diastole, the arterial pressure falls rapidly at first and then declines slowly as the driving force decreases.

In healthy adults, the highest pressure, called the *systolic pressure*, ideally is less than 120 mm Hg, and the lowest pressure, called the *diastolic pressure*, is less than 80 mm Hg. The



■ FIGURE 16-1 ■ Intra-arterial pressure tracing made from the brachial artery. Pulse pressure is the difference between systolic and diastolic pressures. The darker area represents the mean arterial pressure, which can be calculated by using the formula of mean arterial pressure = diastolic pressure + pulse pressure/3.

difference between the systolic and diastolic pressure (approximately 40 mm Hg) is the *pulse pressure*. The *mean arterial pressure* (approximately 90 to 100 mm Hg), depicted by the darker area under the pressure tracing in Figure 16-1, represents the average pressure in the arterial system during ventricular contraction and relaxation.

Determinants of Blood Pressure

The systolic and diastolic components of blood pressure are determined by the cardiac output and the peripheral vascular resistance and can be expressed as a product of the two (blood pressure = cardiac output × peripheral vascular resistance). The cardiac output is the product of the stroke volume (amount of blood ejected from the heart with each beat) and the heart rate. The peripheral vascular resistance reflects changes in the radius of the arterioles as well as the viscosity or thickness of the blood. The arterioles often are referred to as the *resistance vessels* because they can selectively constrict or relax to control the resistance to outflow of blood into the capillaries. The body maintains its blood pressure by adjusting the cardiac output to compensate for changes in peripheral vascular resistance, and it changes the peripheral vascular resistance to compensate for changes in cardiac output.

In hypertension and disease conditions that affect blood pressure, changes in blood pressure are influenced by the stroke volume, the rapidity with which blood is ejected from the heart, the elastic properties of the aorta and large arteries and their ability to accept various amounts of blood as it is ejected from the heart, and the properties of the resistance blood vessels that control the runoff of blood into the smaller vessels and capillaries that connect the arterial and venous circulations.

KEY CONCEPTS

DETERMINANTS OF BLOOD PRESSURE

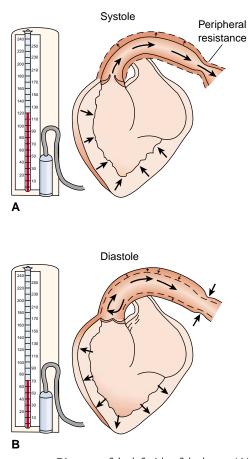
- The arterial blood pressure represents the pressure of the blood as it moves through the arterial system. It reaches its peak (systolic pressure) as blood is ejected from the heart during systole and its lowest level (diastolic pressure) as the heart relaxes during diastole.
- Blood pressure is determined by the cardiac output (stroke volume × heart rate) and the resistance that the blood encounters as it moves through the peripheral vessels (peripheral vascular resistance).
- The systolic blood pressure is largely determined by the characteristics of the stroke volume being ejected from the heart and the ability of the aorta to stretch and accommodate the stroke volume.
- The diastolic pressure is largely determined by the energy that is stored in the aorta as its elastic fibers are stretched during systole and by the resistance to the runoff of blood from the peripheral blood vessels.

Systolic Blood Pressure

The systolic blood pressure reflects the rhythmic ejection of blood into the aorta (Fig. 16-2). As blood is ejected into the aorta, it stretches the vessel wall and produces a rise in aortic pressure. The extent to which the systolic pressure rises or falls with each cardiac cycle is determined by the amount of blood ejected into the aorta with each heart beat (*i.e.*, stroke volume), the velocity of ejection, and the elastic properties of the aorta. Systolic pressure increases when there is a rapid ejection of a large stroke volume or when the stroke volume is ejected into a rigid aorta. The elastic walls of the aorta normally stretch to accommodate the varying amounts of blood that are ejected into the aorta; this prevents the pressure from rising excessively during systole and maintains the pressure during diastole. In some elderly persons, the elastic fibers of the aorta lose some of their elasticity, and the aorta becomes more rigid. When this occurs, the aorta is less able to stretch and buffer the pressure that is generated as blood is ejected into the aorta, resulting in an elevated systolic pressure.

Diastolic Blood Pressure

The diastolic blood pressure is maintained by the energy that has been stored in the elastic walls of the aorta during systole (see Fig. 16-2). The level at which the diastolic pressure is maintained depends on the elastic properties of the aorta and large arteries and their ability to stretch and store energy, the resistance of the arterioles that control the outflow of blood into the microcirculation, and the competency of the aortic valve. The small diameter of the arterioles contributes to their effectiveness as resistance vessels because it takes more force to push blood through a smaller vessel than a larger vessel. When



■ FIGURE 16-2 ■ Diagram of the left side of the heart. (A) Systolic blood pressure represents the ejection of blood into the aorta during ventricular systole; it reflects the stroke volume, the distensibility of the aorta, and the velocity with which blood is ejected from the heart. (B) Diastolic blood pressure represents the pressure in the arterial system during diastole; it is largely determined by the peripheral vascular resistance.

there is an increase in peripheral vascular resistance, as with sympathetic stimulation, diastolic blood pressure rises. Closure of the aortic valve at the onset of diastole is essential to the maintenance of the diastolic pressure. When there is incomplete closure of the aortic valve, as in aortic regurgitation (see Chapter 17), the diastolic pressure drops as blood flows backward into the left ventricle, rather than moving forward into the arterial system.

Pulse Pressure

The pulse pressure is the difference between the systolic and diastolic pressures. It reflects the pulsatile nature of arterial blood flow and is an important component of blood pressure. During the rapid ejection period of ventricular systole, the volume of blood that is ejected into the aorta exceeds the amount that exits the arterial system. The pulse pressure reflects this difference. The pulse pressure rises when additional amounts of blood are ejected into the arterial circulation, and it falls when the resistance to outflow is decreased. In hypovolemic shock, the pulse pressure declines because of a decrease in stroke volume and systolic pressure. This occurs despite an increase in peripheral vascular resistance, which maintains the diastolic pressure.

Mean Arterial Pressure

The mean arterial blood pressure represents the average blood pressure in the systemic circulation. The mean arterial pressure can be estimated by adding one third of the pulse pressure to the diastolic pressure (*i.e.*, diastolic blood pressure + pulse pressure/3). Hemodynamic monitoring equipment in intensive and coronary care units measures or computes mean arterial pressure automatically. Because it is a good indicator of tissue perfusion, the mean arterial pressure often is monitored, along with systolic and diastolic blood pressures, in critically ill patients.

Mechanisms of Blood Pressure Regulation

Although different tissues in the body are able to regulate their own blood flow, it is necessary for the arterial pressure to remain relatively constant as blood shifts from one area of the body to another. The method by which the arterial pressure is regulated depends on whether short-term or long-term adaptation is needed. The mechanisms of blood pressure regulation are illustrated in Figure 16-3.

Short-Term Regulation

The mechanisms for short-term regulation of blood pressure, those occurring over minutes or hours, are intended to correct temporary imbalances in blood pressure, such as occur during physical exercise and changes in body position. These mechanisms also are responsible for maintenance of blood pressure at survival levels during life-threatening situations. The short-term regulation of blood pressure relies mainly on neural and hormonal mechanisms, the most rapid of which are the neural mechanisms.

Neural Mechanisms. The neural control center for the regulation of blood pressure is located in the reticular formation of the lower pons and medulla of the brain where integration and modulation of autonomic nervous system (ANS) responses occur.1 This area of the brain contains the vasomotor and cardiac control centers and is often collectively referred to as the cardiovascular center. The cardiovascular center transmits parasympathetic impulses to the heart through the vagus nerve and transmits sympathetic impulses to the heart and blood vessels through the spinal cord and peripheral sympathetic nerves. Vagal stimulation of the heart produces a slowing of heart rate, whereas sympathetic stimulation produces an increase in heart rate and cardiac contractility. Blood vessels are selectively innervated by the sympathetic nervous system. Increased sympathetic activity produces constriction of the small arteries and arterioles with a resultant increase in peripheral vascular resistance.

The ANS control of blood pressure is mediated through intrinsic circulatory reflexes, extrinsic reflexes, and higher neural control centers. The *intrinsic reflexes*, including the *baroreflex* and *chemoreceptor-mediated reflex*, are located in the circulatory system and are essential for rapid and short-term regulation of blood pressure. The sensors for extrinsic reflexes are found outside the circulation. They include blood pressure responses associated with factors such as pain and cold. The neural path-

FIGURE 16-3 Mechanisms of blood pressure regulation. The *solid lines* represent the mechanisms for renal and baroreceptor control of blood pressure through changes in cardiac output and peripheral vascular resistance. The *dashed lines* represent the stimulus for regulation of blood pressure by the baroreceptors and the kidneys.

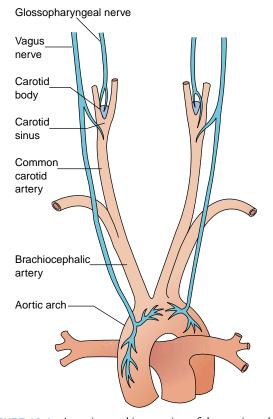
ways for these reactions are more diffuse, and their responses are less consistent than those of the intrinsic reflexes. Many of these responses are channeled through the hypothalamus, which plays an essential role in the control of sympathetic nervous system responses. Among higher-center responses are those due to changes in mood and emotion.

The *baroreceptors* are pressure-sensitive receptors located in the walls of blood vessels and the heart. The carotid and aortic baroreceptors are located in strategic positions between the heart and the brain (Fig. 16-4). They respond to changes in the stretch of the vessel wall by sending impulses to cardiovascular centers in the brain stem to effect appropriate changes in heart rate and vascular smooth muscle tone. For example, the fall in blood pressure that occurs on moving from the lying to the standing position produces a decrease in the stretch of the baroreceptors with a resultant increase in heart rate and sympathetically induced vasoconstriction that causes an increase in peripheral vascular resistance.

The *arterial chemoreceptors* are sensitive to changes in the oxygen, carbon dioxide, and hydrogen ion content of the blood. They are located in the carotid bodies, which lie in the bifurcation of the two common carotids, and in the aortic bodies of the aorta (see Fig.16-4). Because of their location, these chemoreceptors are always in close contact with the arterial blood. Although the main function of the chemoreceptors is to regulate ventilation, they also communicate with the cardiovascular center and can induce widespread vasoconstriction. When the arterial pressure drops below a critical level, the chemoreceptors are stimulated because of diminished oxygen supply and a buildup of carbon dioxide and hydrogen ions. In persons with chronic lung disease, systemic and pulmonary hypertension may develop because of hypoxemia (see Chapter 21).

Humoral Mechanisms. A number of hormones and humoral mechanisms contribute to blood pressure regulation, including the renin-angiotensin-aldosterone mechanism and vasopressin. Other humoral substances such as epinephrine, a sympathetic neurotransmitter released from the adrenal gland, has the effect of directly stimulating an increase in heart rate, cardiac contractility, and vascular tone.

The *renin-angiotensin-aldosterone* system plays a central role in blood pressure regulation. Renin is an enzyme that is synthesized, stored, and released by the kidneys in response to an increase in sympathetic nervous system activity or a decrease

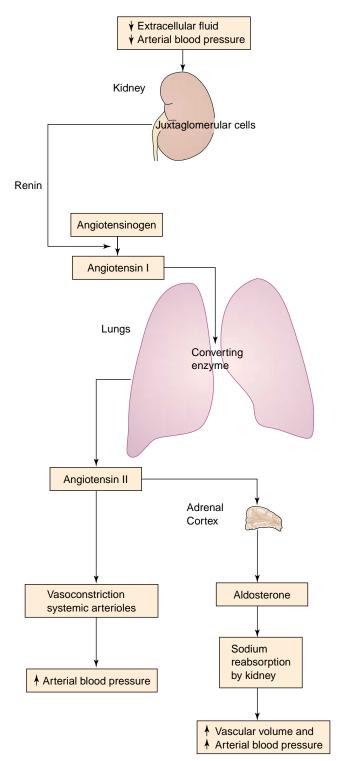


■ FIGURE 16-4 ■ Location and innervation of the aortic arch and carotid sinus baroreceptors and the carotid body chemoreceptors.

in blood pressure, extracellular fluid volume, or extracellular sodium concentration. Most of the renin that is released leaves the kidney and enters the bloodstream, where it acts enzymatically to convert an inactive circulating plasma protein called *angiotensinogen* to angiotensin I (Fig. 16-5). Angiotensin I travels to the small blood vessels of the lung, where it is converted to angiotensin II by the angiotensin-converting enzyme that is present in the endothelium of the lung vessels. Although angiotensin II has a half-life of several minutes, renin persists in the circulation for 30 minutes to 1 hour and continues to cause production of angiotensin II during this time.

Angiotensin II functions in both the short-term and longterm regulation of blood pressure. It is a strong vasoconstrictor, particularly of arterioles and to a lesser extent of veins. The vasoconstrictor response produces an increase in peripheral vascular resistance (and blood pressure) and functions in the shortterm regulation of blood pressure. A second major function of angiotensin II, stimulation of aldosterone secretion from the adrenal gland, contributes to the long-term regulation of blood pressure by increasing salt and water retention by the kidney. It also acts directly on the kidney to decrease the elimination of salt and water.

Vasopressin, also known as antidiuretic hormone (ADH), is released from the posterior pituitary gland in response to decreases in blood volume and blood pressure, an increase in the osmolality of body fluids, and other stimuli. The antidiuretic actions of vasopressin are discussed in Chapter 6. Vasopressin has a direct vasoconstrictor effect on blood vessels, particularly those of the splanchnic circulation that supplies the abdomi-



■ FIGURE 16-5 ■ Control of blood pressure by the reninangiotensin-aldosterone system. Renin enzymatically converts the plasma protein angiotensinogen to angiotensin I; angiotensinconverting enzyme in the lung converts angiotensin I to angiotensin II; and angiotensin II produces vasoconstriction and increases salt and water retention through direct action on the kidney and through increased aldosterone secretion by the adrenal cortex.

nal viscera. However, long-term increases in vasopressin cannot maintain volume expansion or hypertension, and it does not enhance hypertension produced by sodium-retaining hormones or other vasoconstricting substances. It has been suggested that vasopressin plays a permissive role in hypertension through its fluid-retaining properties or as a neurotransmitter that serves to modify ANS function.

Long-Term Regulation

Long-term mechanisms control the daily, weekly, and monthly regulation of blood pressure. Although the neural and hormonal mechanisms involved in the short-term regulation of blood pressure act rapidly, they are unable to maintain their effectiveness over time. Instead, the long-term regulation of blood pressure is largely vested in the kidneys and regulation of extracellular fluid volume.1 Blood pressure is normally regulated around an equilibrium point, which represents the normal pressure for a given individual. Accordingly, when the body contains too much extracellular fluid, the arterial pressure rises and the rate at which water (*i.e.*, pressure diuresis) and salt (*i.e.*, pressure natruresis) are excreted by the kidney is increased.¹ When blood pressure returns to its equilibrium point, water and salt excretion return to normal. A fall in blood pressure due to a decrease in extracellular fluid volume has the opposite effect. In persons with hypertension, renal control mechanisms are often altered such that the equilibrium point for blood pressure regulation is maintained at a higher level of salt and water elimination.

There are several ways that extracellular fluid volume serves to regulate blood pressure. One is through a direct effect on cardiac output and another is indirect resulting from autoregulation of blood flow and its effect on peripheral vascular resistance. Autoregulatory mechanisms function in distributing blood flow to the various tissues of the body, according to their metabolic needs (see Chapter 14). When the blood flow to a specific tissue bed is excessive, local blood vessels constrict, and when the flow is deficient the local vessels dilate. In situations of increased blood volume and cardiac output, all of the tissues of the body are exposed to the same increase in flow. This results in a generalized constriction of arterioles and an increase in the peripheral vascular resistance.

The role that the kidneys play in blood pressure regulation is emphasized by the fact that many hypertension medications produce their blood pressure-lowering effects by increasing salt and water elimination.

In summary, the arterial blood pressure that moves blood through the circulatory system reflects the alternating contraction and relaxation of the left heart. The systolic pressure denotes the peak pressure that occurs during ventricular contraction, and the diastolic pressure denotes the lowest point that occurs during relaxation. The pulse pressure is the difference between these two pressures, and the mean arterial pressure reflects the average pressure throughout the cardiac cycle.

The level to which the arterial blood pressure rises during systole and falls during diastole is determined by the cardiac output (stroke volume × heart rate) and the peripheral vascular resistance. Systolic pressure is determined primarily by the characteristics of the stroke volume and elastic properties of the aorta and large arteries and their ability to stretch and

accommodate the varying amounts of blood that is ejected from the heart. Diastolic pressure is determined largely by the smaller arteries and their ability to maintain the peripheral vascular resistance and accept the runoff of blood from the larger arteries.

Normally, blood pressure is regulated at levels sufficient to ensure adequate tissue perfusion. Blood pressure regulation requires the use of short-term and long-term mechanisms. Short-term regulation, which occurs over minutes and hours, involves ANS responses associated with baroreceptor and chemoreceptor stimulation and hormonal mechanisms such as the renin-angiotensin-aldosterone system and vasopressin. Long-term mechanisms are involved in the daily, weekly, and monthly regulation of blood pressure. They are largely vested in the kidney and involve regulation of the extracellular fluid volume through the elimination of water (pressure diuresis) and salt (pressure natruresis).

HYPERTENSION

Hypertension, or high blood pressure, is probably the most common of all health problems in adults and is the leading risk factor for cardiovascular disorders. In the United States, approximately 25% of all adults older than 18 years of age have high blood pressure.² It has been estimated that as many as 60 million adults have cardiovascular disease. In 50 million of these people (83%), the cause of cardiovascular disease is high blood pressure. Hypertension is more common in younger men compared with younger women, in blacks compared with whites, in persons from lower socioeconomic groups, and in older persons. Men have higher blood pressures than do women up until the time of menopause, at which point women quickly lose their protection.²

Hypertension commonly is divided into the categories of primary and secondary hypertension. In primary, or *essential*, hypertension, which accounts for 90% to 95% of all hypertension, the chronic elevation in blood pressure occurs without evidence of other disease. In secondary hypertension, the elevation of blood pressure results from some other disorder, such as kidney disease. Malignant hypertension, as the name implies, is an accelerated form of hypertension.

Essential Hypertension

The sixth report of the Joint National Committee on Detection, Evaluation, and Treatment of High Blood Pressure (JNC-VI) of the National Institutes of Health was published in 1997.² The JNC-VI report classifies blood pressure as optimal, normal, high-normal, and hypertensive (Table 16-1). Hypertension is further divided into stages 1, 2, and 3 based on systolic and diastolic blood pressure measurements. For adults with diabetes mellitus, the blood pressure goal has been lowered to less than 130/80 mm Hg.³

Mechanisms of Blood Pressure Elevation

Several factors, including hemodynamic, neural, humoral, and renal mechanisms, are thought to interact in producing long-term elevations in blood pressure. As with other disease

KEY CONCEPTS

HYPERTENSION

- Essential hypertension is characterized by a chronic elevation in blood pressure that occurs without evidence of other disease and secondary hypertension by an elevation of blood pressure that results from some other disorder, such as kidney disease.
- The pathogenesis of essential hypertension is thought to reside with the kidney and its role regulating vascular volume through salt and water elimination; the renin-angiotensin-aldosterone system through its effects on blood vessel tone, regulation of renal blood flow, and salt metabolism; and the sympathetic nervous system, which regulates the tone of the resistance vessels.
- Uncontrolled hypertension increases the work demands on the heart, resulting in left ventricular hypertrophy and heart failure, and on the vessels of the arterial system, leading to atherosclerosis, kidney disease, and stroke.

conditions, it is probable that there is not a single cause of essential hypertension or that the condition is a single disease. Because arterial blood pressure is the product of cardiac output and peripheral vascular resistance, all forms of hypertension involve hemodynamic mechanisms—an increase in either cardiac output or peripheral vascular resistance, or a combination of the two. Other factors, such as sympathetic

for Adults 18 Years of Age and Older*

nervous system activity, kidney function in terms of salt and water retention, the electrolyte composition of the intracellular and extracellular fluids, and humoral influences such as the renin-angiotensin-aldosterone mechanism, play an active or permissive role in regulating the hemodynamic mechanisms that control blood pressure.

Contributing Factors

Risk Factors. Although the cause or causes of essential hypertension are largely unknown, several constitutional factors have been implicated as contributing to its development. These risk factors include family history of hypertension, race, diabetes mellitus, and age-related increases in blood pressure. The inclusion of heredity as a contributing factor in the development of hypertension is supported by the fact that hypertension is seen most frequently among persons with a family history of hypertension. The inherited predisposition does not seem to rely on other risk factors, but when they are present, the risk apparently is additive. Hypertension not only is more prevalent in African Americans than whites, it is more severe, tends to occur earlier, and often is not treated early enough or aggressively enough. Blacks also tend to develop greater cardiovascular and renal damage at any level of pressure.⁴ Diabetes mellitus and hypertension are closely interrelated disorders that share similar genetic and life-style factors. Also, persons with both hypertension and diabetes have a greater risk of target organ damage. Maturation and growth are known to cause predictable increases in blood pressure from infancy to adolescence. In adult life, the systolic pressure continues to slowly rise, and hypertension becomes more common with increasing age.

Lifestyle Factors. Lifestyle factors can contribute to the development of hypertension by interacting with the constitutional risk factors. These lifestyle factors include high sodium intake,

TABLE 16-1

Category*	Systolic Blood Pressure (mm Hg)		Diastolic Blood Pressure (mm Hg)	Follow-up Recommended		
Optimal [†]	<120	and	<80	Recheck in 2 years		
Normal	<130	and	<85	Recheck in 2 years		
High-Normal	130–139	or	85–89	Recheck in 1 year		
Hypertension [‡]						
Stage 1	140–159	or	90–99	Confirm within 2 months		
Stage 2	160–179	or	100–109	Evaluate or refer to source of care within 1 month		
Stage 3	≥180	or	≥110	Evaluate or refer to source of care immediately or withir 1 week, depending on clinical situation		

Classification and Follow-up of Blood Pressure Measurements

*Not taking antihypertensive drugs and not acutely ill. When systolic and diastolic pressures fall into different categories, the higher category should be selected to classify the individual's blood pressure status. For instance, 160/92 should be classified as stage 2 and 174/120 should be classified as stage 3. Isolated systolic hypertension (ISH) is defined as systolic blood pressure ≥140 mm Hg and diastolic blood pressure <90 mm Hg and staged appropriately (*e.g.*, 170/82 mm Hg is defined as stage 2 ISH). In addition to classifying stages of hypertension on the basis of average blood pressure levels, clinicians should specify presence or absence of target-organ disease and additional risk factors. This specificity is important for risk classification and treatment.

[†]Optimal blood pressure with respect to cardiovascular risk is below 120/80 mm Hg. However, unusually low readings should be evaluated for clinical significance.

[‡]Based on the average of two or more readings taken at each of two or more visits after an initial screening. (Adapted from the National Heart, Lung, and Blood Institute. [1997]. *The sixth report of the National Committee on Detection,*

Evaluation, and Treatment of High Blood Pressure [pp. 11, 13]. NIH publication no. 98-4080. Bethesda, MD: National

Institutes of Health. Available: http://www.nhlbi.nih.gov/guidelines/hypertension/jncintro.htm)

excessive calorie intake and obesity, physical inactivity, and excessive alcohol consumption. Oral contraceptive drugs also may increase blood pressure in predisposed women. Although stress can raise blood pressure acutely, there is less evidence linking it to chronic elevations in blood pressure. Dietary fats and cholesterol are independent risk factors for cardiovascular disease, but there is no evidence that they raise blood pressure. Smoking, although not identified as a primary risk factor in hypertension, is an independent risk factor in coronary heart disease and should be avoided.

Increased salt intake has long been implicated as an etiologic factor in the development of hypertension.^{2,5} Just how increased salt intake contributes to the development of hypertension is still unclear. It may be that salt causes an elevation in blood volume, increases the sensitivity of cardiovascular or renal mechanisms to sympathetic nervous system stimuli, or exerts its effects through some other mechanisms such as the renin-angiotensin-aldosterone system.

Excessive weight commonly is associated with hypertension. It has been suggested that fat distribution might be a more critical indicator of hypertension risk than actual overweight. The waist-to-hip ratio commonly is used to differentiate central or upper body obesity (*i.e.*, fat cell deposits in the abdomen) from peripheral or lower body obesity with fat cell deposits in the buttocks and legs (see Chapter 29).⁶

Regular alcohol consumption can play a role in the development of hypertension. The effect is seen with different types of alcoholic beverages, in men and women, and in a variety of ethnic groups.^{7,8} Systolic pressure is usually more markedly affected than diastolic pressure. Blood pressure may improve or return to normal when alcohol consumption is decreased or eliminated.

Oral contraceptives cause a mild increase in blood pressure in many women and overt hypertension in approximately 5%.⁹ Why this occurs is largely unknown, although it has been suggested that estrogen and progesterone are responsible for the effect. Various contraceptive drugs contain different amounts and combinations of estrogen and progestational agents, and these differences may contribute to the occurrence of hypertension in some women but not others. Fortunately, the hypertension associated with oral contraceptives usually disappears after use of the drug has been discontinued, although it may take as long as 6 months for this to happen. However, in some women the blood pressure may not return to normal; they may be at risk for hypertension. The risk of hypertension-associated cardiovascular complications is found primarily in women older than 35 years of age and in those who smoke.²

Manifestations

Essential hypertension is typically an asymptomatic disorder. When symptoms do occur they are usually related to the longterm effects of hypertension on other organ systems of the body including the kidneys, heart, eyes, and blood vessels.

The 1997 JNC-VI report uses the term *target organ disease* to describe the cardiac, cerebrovascular, peripheral arterial disease, kidney, and retinal complications associated with hypertension (Chart 16-1).² The excess morbidity and mortality related to hypertension is progressive over the whole range of systolic and diastolic pressures. Target organ damage varies markedly among persons with similar levels of hypertension.

Hypertension is a major risk factor for atherosclerosis; it predisposes to all major atherosclerotic cardiovascular disorders,

CHART 16-1 Components of Target Organ Damage/Cardiovascular Disease in Persons With Hypertension

Heart disease

Left ventricular hypertrophy
Angina/prior myocardial infarction
Prior coronary revascularization
Heart failure
Stroke or transient ischemic attack
Nephropathy
Peripheral arterial disease

Retinopathy

Adapted from National Heart, Lung, and Blood Institute. (1997). The sixth report of the National Committee on Prevention, Detection, and Treatment of High Blood Pressure (p. 16). NIH publication no. 98-4080. Bethesda, MD: National Institutes of Health.

including heart failure, stroke, coronary artery disease, and peripheral artery disease. The risk of coronary artery disease and stroke depends to a great extent on other risk factors, such as obesity, smoking, and elevated cholesterol levels. If all else is favorable, the risk of a coronary event in persons with mild hypertension is no greater than in the average population of the same age. However, if a cluster of risk factors exists, the risk is greatly increased. The same is true for stroke. The risk of stroke in persons with hypertension occurs over an eightfold range, depending on the number of associated risk factors.¹⁰ Cerebrovascular complications are more closely related to systolic than diastolic hypertension.

Hypertension increases the workload of the left ventricle by increasing the pressure against which the heart must pump as it ejects blood into the systemic circulation. As the workload of the heart increases, the left ventricular wall hypertrophies to compensate for the increased pressure work. Despite its compensatory function, left ventricular hypertrophy is a major risk factor for ischemic heart disease, cardiac dysrhythmias, sudden death, and congestive heart failure. The prevalence of left ventricular hypertrophy increases with age and is highest in persons with blood pressures greater than 160/95 mm Hg.¹¹ Hypertensive left ventricular hypertrophy usually regresses with therapy. Regression is most closely related to systolic pressure reduction and does not appear to reflect the particular type of medication used.

Hypertension also can lead to nephrosclerosis, a common cause of renal insufficiency (see Chapter 23). Hypertensive kidney disease is more common in blacks than whites. Hypertension also plays an important role in accelerating the course of other types of kidney disease, particularly diabetic nephropathy.³ Because of the risk of diabetic nephropathy, the American Diabetes Association recommends that persons with diabetes maintain their blood pressure at levels less than 130/80 mm Hg (see Chapter 32).

Diagnosis and Treatment

Unlike disorders of organ structure that are diagnosed by methods such as x-rays and tissue examination, hypertension and other blood pressure disorders are determined by repeated blood pressure measurement. The diagnosis of hypertension in a person who is not taking antihypertensive medications should be based on the average of at least two or more blood pressure readings taken at each of two or more visits after an initial screening visit.² Laboratory tests, x-ray films, and other diagnostic tests usually are done to exclude secondary hypertension and determine the presence or extent of target organ disease.

Clinically, blood pressure measurements are usually obtained by the auscultatory method, which uses a sphygmomanometer and a stethoscope. Accuracy of blood pressure measurement requires that persons taking the pressure are adequately trained in blood pressure measurement, that the equipment is properly maintained and calibrated, and the cuff bladder is appropriate for the arm size.¹² The width of the bladder should be at least 40% of arm circumference and the length at least 80% of arm circumference. Undercuffing (using a cuff with a bladder that is too small) can cause an overestimation of blood pressure.¹³ This is because a cuff that is too small results in an uneven distribution of pressure across the arm, such that a greater cuff pressure is needed to occlude blood flow. Likewise, overcuffing (using a cuff with a bladder that is too large) can cause an underestimation of blood pressure.

The main objective for treatment of essential hypertension is to achieve and maintain arterial blood pressure of less than 140/90 mm Hg, with the goal of preventing morbidity and mortality. For persons with secondary hypertension, efforts are made to correct or control the disease condition causing the hypertension. Antihypertensive medications and other measures supplement the treatment for the underlying disease. The JNC-VI report contains a treatment algorithm for hypertension that includes lifestyle modification and, when necessary, guidelines for the use of pharmacologic agents to achieve and maintain systolic pressure below 140 mm Hg and diastolic pressure below 90 mm Hg.²

Lifestyle Modification. Lifestyle modification includes reduction in sodium intake, maintenance of adequate potassium intake, weight reduction if overweight, regular aerobic physical activity, and modification of alcohol intake. A reduction in dietary saturated fats and cholesterol is recommended for overall cardiovascular health. Smoking cessation should be encouraged for people who smoke. For persons with stage 1 hypertension (see Table 16-1), an attempt to control blood pressure with weight loss and other lifestyle modifications should be tried for at least 3 to 6 months before initiating pharmacologic treatment.⁴

The American Heart Association recommends the daily salt intake for adults in the general population should not exceed 6 g/day.¹⁴ Because many prepared foods are high in sodium, it was recommended that persons consult package labels for the sodium content of canned foods, frozen foods, soft drinks, and other foods and beverages to reduce sodium intake adequately. High dietary potassium intake may protect against the development of hypertension or improve blood pressure control in people with hypertension. Therefore, an adequate intake of potassium (approximately 90 mmol per day), preferably from food sources such as fresh fruits and vegetables, is recommended.²

Weight reduction of as little as 4.5 kg (10 lb) can produce a decrease in blood pressure in a large proportion of overweight people with hypertension.² Because of alcohol's association

with high blood pressure, the JNC-VI report recommends restriction of alcohol consumption to no more than 1 oz (30 mL) ethanol per day (equal to 2 oz of 100-proof whiskey, 10 oz of wine, or 24 oz of beer) or 0.5 oz per day for women or lighter weight people.²

A regular program of aerobic physical exercise (*e.g.*, walking, biking, swimming) is protective, especially for those at increased risk for cardiovascular disease because of hypertension. Exercise may have additional indirect benefits, such as weight loss or motivation for changing other risk factors.

Pharmacologic Treatment. The decision to initiate pharmacologic treatment is based on the severity of the hypertension, the presence of target organ disease, and the existence of other conditions and risk factors. Drug selection is based on the stage of hypertension. Among the drugs used in the treatment of hypertension are diuretics, β -adrenergic–blocking drugs, angiotensin-converting enzyme (ACE) inhibitors or angiotensin II receptor blockers, the calcium channel-blocking drugs, central α_2 -adrenergic agonists, α_1 -adrenergic receptor blockers, and vasodilators.

The physiologic mechanisms whereby the different hypertension drugs produce a reduction in blood pressure differ among agents. Diuretics lower blood pressure initially by decreasing vascular volume (by suppressing renal reabsorption of sodium and increasing salt and water excretion) and cardiac output. With continued therapy, a reduction in peripheral resistance becomes a major mechanism of blood pressure reduction. β-Adrenergic–blocking drugs are effective in treating hypertension because they decrease heart rate, cardiac output, and renin release by the kidney. The ACE inhibitors act by inhibiting the conversion of angiotensin I to angiotensin II, thus decreasing angiotensin II levels and reducing its effect on vasoconstriction, aldosterone levels, intrarenal blood flow, and the glomerular filtration rate. The calcium channel blockers decrease peripheral vascular resistance by inhibiting the movement of calcium into arterial smooth muscle cells. The centrally acting α_2 -adrenergic agonists act in a negative-feedback manner to decrease sympathetic outflow from the central nervous system. The α_1 -adrenergic receptor antagonists block α_1 receptors on vascular smooth muscle causing vasodilation and a reduction in peripheral vascular resistance. The direct-acting smooth muscle vasodilators promote a decrease in peripheral vascular resistance by producing relaxation of vascular smooth muscle, particularly of the arterioles.

Pharmacologic treatment of hypertension usually follows a stepwise approach.² It is usually initiated with a low dose of a single drug. The dose is slowly increased at a schedule dependent on the person's age, needs, and desired response.² If the response to the initial drug is not adequate, one of three approaches can be used: the dose can be increased if the initial dose was below the maximum recommended; a drug with a different mode of action can be added; or the initial drug can be discontinued and another substituted. Combining drugs with different modes of action often allows smaller doses to be used to achieve blood pressure control, while minimizing the dose-dependent side effects from any one drug. In treating stage 3 or 4 hypertension, it often is necessary to add a second or third drug after a short interval if the treatment goal is not achieved.

Systolic Hypertension

Essential hypertension may be classified as systolic/diastolic hypertension in which both the systolic and diastolic pressures are elevated; as diastolic hypertension in which the diastolic pressure is selectively elevated; or as systolic hypertension in which the systolic pressure is selectively elevated. The JNC-VI report defined systolic hypertension as a systolic pressure of 140 mm Hg or greater and a diastolic pressure less than 90 mm Hg and indicated a need for increased recognition and control of isolated systolic hypertension.² Historically, diastolic hypertension was thought to confer a greater risk for cardiovascular events than systolic hypertension.² However, there is mounting evidence that elevated systolic blood pressure is at least as important, if not more so, than diastolic hypertension.^{15,16}

There are two aspects of systolic hypertension that confer increased risk of cardiovascular events: one is the actual elevation in systolic pressure, and the other is the disproportionate rise in pulse pressure. Elevated pressures during systole favor the development of left ventricular hypertrophy, increased myocardial oxygen demands, and eventual left heart failure. At the same time, the absolute or relative lowering of diastolic pressure is a limiting factor in coronary perfusion because coronary perfusion is greatest during diastole. Elevated levels of pulse pressure produce greater stretch of arteries, causing damage to the elastic elements of the vessel and thus predisposing to aneurysms and development of the endothelial cell damage that leads to atherosclerosis and thrombosis.

Secondary Hypertension

Only 5% to 10% of hypertensive cases are classified as secondary hypertension (*i.e.*, hypertension due to another disease condition). Unlike essential hypertension, many of the conditions causing secondary hypertension can be corrected or cured by surgery or specific medical treatment. Secondary hypertension tends to be seen more commonly in persons younger than 30 years and those older than 50 years of age.¹⁷ Among the most common causes of secondary hypertension are kidney disease (*i.e.*, renovascular hypertension), adrenal cortical disorders, pheochromocytoma, and coarctation of the aorta. Cocaine and cocaine-like substances also can cause significant hypertension.

Renal Hypertension

With the dominant role that the kidney assumes in blood pressure regulation, it is not surprising that the largest single cause of secondary hypertension is renal disease. Most acute kidney disorders result in decreased urine formation, retention of salt and water, and hypertension. This includes acute glomerulonephritis, acute renal failure, and acute urinary tract obstruction. Hypertension also is common among persons with chronic pyelonephritis, polycystic kidney disease, diabetic nephropathy, and end-stage renal disease, regardless of cause.

Renovascular hypertension refers to hypertension caused by reduced renal blood flow and activation of the reninangiotensin-aldosterone mechanism. It is the most common cause of secondary hypertension, accounting for 1% to 2% of all cases of hypertension. The reduced renal blood flow that occurs with renovascular disease causes the affected kidneys to release excessive amounts of renin, increasing circulating levels of angiotensin II. One or both of the kidneys may be affected. Manifestations of renovascular hypertension include hypokalemia (caused by increased aldosterone levels), the presence of an abdominal bruit, and a duration of hypertension of less than 1 year (to help to distinguish renovascular hypertension from essential hypertension). Because renal blood flow depends on the increased blood pressure generated by the reninangiotensin system, administration of ACE inhibitors can cause a rapid decline in renal function.

Disorders of Adrenocorticosteroid Hormones

Excess production of aldosterone caused by adrenocortical hyperplasia or adenoma (primary hyperaldosteronism) and excess levels of glucocorticoid (Cushing's disease or syndrome) tend to raise the blood pressure (see Chapter 31). These hormones produce hypertension through increased salt and water retention by the kidney. For persons with primary hyperaldosteronism, a salt-restricted diet often produces a reduction in blood pressure. Because aldosterone acts on the distal renal tubule to increase sodium absorption in exchange for potassium elimination in the urine, persons with hyperaldosteronism usually have decreased potassium levels. Potassium-sparing diuretics, such as spironolactone, which is an aldosterone antagonist, often are used in the medical care of persons with the disorder.

Pheochromocytoma

A pheochromocytoma is a tumor of chromaffin tissue, which contains sympathetic nerve cells. The tumor is most commonly located in the adrenal medulla but can arise in other sites, such as sympathetic ganglia, where there is chromaffin tissue.¹⁸ Although only 0.1% to 0.5% of persons with hypertension have an underlying pheochromocytoma, the disorder can cause serious hypertensive crises. Eight percent to 10% of the tumors are malignant.

Like adrenal medullary cells, the tumor cells of a pheochromocytoma produce and secrete the catecholamines epinephrine and norepinephrine. The hypertension that develops is the result of a massive release of these catecholamines. Their release may be paroxysmal, rather than continuous, causing periodic episodes of headache, excessive sweating, and palpitations. Headache is the most common symptom and can be quite severe. Nervousness, tremor, facial pallor, weakness, fatigue, and weight loss occur less frequently. Marked variability in blood pressure between episodes is typical. Some persons with pheochromocytoma have paroxysmal episodes of hypertension, sometimes to dangerously high levels; others may have sustained hypertension; and some may even be normotensive.

Coarctation of the Aorta

Coarctation of the aorta represents a narrowing of the aorta just distal to the origin of the subclavian arteries (see Chapter 17).¹⁹ The ejection of a large stroke volume into a narrowed aorta with limited ability to accept the runoff results in an increase in systolic blood pressure and blood flow to the upper part of the body. Blood pressure in the lower extremities may be normal, although it frequently is low. It has been suggested that the increase in cardiac output and maintenance of the pressure to the lower part of the body is achieved through the reninangiotensin-aldosterone mechanism in response to a decrease in renal blood flow. Pulse pressure in the legs almost always is narrowed, and the femoral pulses are weak. Because the aortic

capacity is diminished, there usually is a marked increase in pressure (measured in the arms) during exercise, when the stroke volume and heart rate are exaggerated. It is important that blood pressure be measured in both arms and one leg when coarctation of the aorta is suspected. A 20 mm Hg or higher pressure in the arms than in the legs suggests coarctation of the aorta.

Malignant Hypertension

A small number of persons with secondary hypertension develop an accelerated and potentially fatal form of the disease malignant hypertension. This usually is a disease of younger persons, particularly young African-American men, women with hypertension of pregnancy, and persons with renal and collagen diseases.

Malignant hypertension is characterized by sudden marked elevations in blood pressure, with diastolic values above 120 mm Hg, renal disorders, vascular changes, and retinopathy. There may be intense arterial spasm of the cerebral arteries with hypertensive encephalopathy. Cerebral vasoconstriction probably is an exaggerated homeostatic response designed to protect the brain from excesses of blood pressure and flow. The regulatory mechanisms often are insufficient to protect the capillaries, and cerebral edema frequently develops. As it advances, papilledema (*i.e.*, swelling of the optic nerve at its point of entrance into the eye) ensues, giving evidence of the effects of pressure on the optic nerve and retinal vessels. The patient may have headache, restlessness, confusion, stupor, motor and sensory deficits, and visual disturbances. In severe cases, convulsions and coma follow.

Prolonged and severe exposure to exaggerated levels of blood pressure in malignant hypertension injures the walls of the arterioles, and intravascular coagulation and fragmentation of red blood cells may occur. The renal blood vessels are particularly vulnerable to hypertensive damage. Renal damage caused by vascular changes probably is the most important prognostic determinant in malignant hypertension. Elevated levels of blood urea nitrogen and serum creatinine, metabolic acidosis, hypocalcemia, and proteinuria provide evidence of renal impairment. The complications associated with a hypertensive crisis demand immediate and rigorous medical treatment in an intensive care unit with continuous monitoring of arterial blood pressure. With proper therapy, the death rate from this cause can be markedly reduced, as can the potential for additional episodes. Because chronic hypertension is associated with autoregulatory changes in cerebral blood flow, care is taken to avoid excessively rapid decreases in blood pressure, which can lead to cerebral hypoperfusion and brain injury.

Hypertension During Pregnancy

Hypertensive disorders complicate 6% to 8% of pregnancies. They are the second leading cause, after embolism, of maternal mortality in the United States, accounting for almost 15% of such deaths. Hypertensive disorders also contribute to stillbirths and neonatal morbidity and mortality. The incidence of hypertensive disorders of pregnancy increases with maternal age and is more common in African-American women.²⁰

Classification

The National Institutes of Health Working Group Report on High Blood Pressure in Pregnancy published a revised classification system for high blood pressure in pregnancy that included preeclampsia-eclampsia, preeclampsia superimposed on chronic hypertension, chronic hypertension, and gestational hypertension²¹ (Table 16-2).

Preeclampsia-Eclampsia. Preeclampsia-eclampsia is a pregnancy-specific syndrome. It is defined as an elevation in blood pressure and proteinuria developing after the 20th week of gestation. It is defined as an elevation in blood pressure (systolic >140 or diastolic >90 mm Hg) and proteinuria (≥ 0.3 g/24 hours) developing after the 20th week of gestation. The presence of systolic pressure ≥ 160 mm Hg, diastolic pressure ≥ 110 mm Hg; proteinuria (≥ 0.0 g/24 h); increased serum creatinine (>1.2 mg); platelet counts <100,000 cells/mm³; elevated liver enzymes; persistent headache or cerebral or visual disturbances; and persistent epigastric pain serve to reinforce the diagnosis.²¹ Preeclampsia may occur in women who already are hypertensive, in which case the prognosis for the mother and fetus tends

TABLE 16-2	Classif	Classification of High Blood Pressure in Pregnancy					
Classification		Description					
Gestational hypertension		Blood pressure elevation, without proteinuria, that is detected for the first time during midpregnancy and returns to normal by 12 weeks postpartum.					
Chronic hypertens	sion	Blood pressure ≥140 mm Hg systolic or ≥90 mm Hg diastolic that is present and observable before the 20th week of pregnancy. Hypertension that is diagnosed for the first time during pregnancy and does not resolve after pregnancy also is classified as chronic hypertension.					
Preeclampsia-eclampsia		Pregnancy-specific syndrome of blood pressure elevation (blood pressure >140 mm Hg systolic or >90 Hg diastolic) that occurs after the first 20 weeks of pregnancy and is accompanied by proteinuria (urinary excretion of 0.3 g protein in a 24-hour specimen).					
Preeclampsia supe on chronic hype		Chronic hypertension (blood pressure ≥140 mm Hg systolic or ≥90 mm Hg diastolic prior to 20th week of pregnancy) with superimposed proteinuria and with or without signs of the preeclampsia syndrome					

(Developed using information from National Institutes of Health. [2000]. *Working group report on high blood pressure in pregnancy*. NIH publication no. 00-3029. Bethesda, MD: Author. Available: http://www.nhlbi.gov/health/prof/heart/ hbp/hbp_preg.htm)

to be worse than for either condition alone.²¹ Eclampsia is the occurence, in a woman with preeclampsia, of seizures that cannot be attributed to other causes.

Preeclampsia occurs primarily during first pregnancies and during subsequent pregnancies in women with multiple fetuses, diabetes mellitus, or coexisting renal disease. It is associated with a condition called a *hydatidiform mole* (*i.e.*, abnormal pregnancy caused by a pathologic ovum, resulting in a mass of cysts). Women with chronic hypertension who become pregnant have an increased risk of preeclampsia and adverse neonatal outcomes, particularly when associated with proteinuria early in pregnancy.²²

Pregnancy-induced hypertension is thought to involve a decrease in placental blood flow, leading to the release of toxic mediators that alter the function of endothelial cells in blood vessels throughout the body, including those of the kidney, brain, liver, and heart.²² The endothelial changes result in signs and symptoms of preeclampsia and, in more severe cases, of intravascular clotting and hypoperfusion of vital organs. There is risk for development of disseminated intravascular coagulation, cerebral hemorrhage, hepatic failure, and acute renal failure. Thrombocytopenia is the most common hematologic complication of preeclampsia. Platelet counts of less than below 100,000/mm³ signal serious disease. The cause of thrombocytopenia has been ascribed to platelet deposition at the site of endothelial injury. The renal changes that occur with preeclampsia include a decrease in glomerular filtration rate and renal blood flow. Sodium excretion may be impaired, although this is variable. Edema may or may not be present. Some of the most severe forms of preeclampsia occur in the absence of edema. Even when there is extensive edema, the plasma volume usually is lower than that seen in normal pregnancy. Liver damage, when it occurs, may range from mild hepatocellular necrosis with elevation of liver enzymes to the more ominous hemolysis, elevated liver function tests, and low platelet count (HELLP) syndrome that is associated with significant maternal mortality. Eclampsia, the convulsive stage of preeclampsia, is a significant cause of maternal mortality. The pathogenesis of eclampsia remains unclear and has been attributed to both increased coagulability and fibrin deposition in the cerebral vessels.

Chronic Hypertension. Chronic hypertension is considered in the same manner as hypertension that is unrelated to the pregnancy. It is defined as a history of high blood pressure before pregnancy, identification of hypertension before 20 weeks of pregnancy, and hypertension that persists after pregnancy. In women with chronic hypertension, blood pressure often decreases in early pregnancy and increases during the last trimester (3 months) of pregnancy, resembling preeclampsia. Women with chronic hypertension are at increased risk for the development of preeclampsia.

Gestational Hypertension. Gestational hypertension represents a blood pressure elevation without proteinuria that is detected for the first time after midpregnancy. It includes women with preeclampsia syndrome who have not yet manifested proteinuria, as well as women who do not have the syndrome. The final determination that a woman does not have the preeclampsia syndrome is made only postpartum. If preeclampsia has not developed and blood pressure has returned to normal by 12 weeks postpartum, the condition is considered to be gestational hypertension. If blood pressure elevation persists, a diagnosis of chronic hypertension is made.

Diagnosis and Treatment

Early prenatal care is important in the detection of high blood pressure during pregnancy. It is recommended that all pregnant women, including those with hypertension, refrain from alcohol and tobacco use. Salt restriction usually is not recommended during pregnancy because pregnant women with hypertension tend to have lower plasma volumes than do normotensive pregnant women and because the severity of hypertension may reflect the degree of volume contraction. The exception is women with preexisting hypertension who have been following a salt-restricted diet.

In women with preeclampsia, delivery of the fetus is curative. The timing of delivery becomes a difficult decision in preterm pregnancies because the welfare of both the mother and the infant must be taken into account. Bed rest is a traditional therapy. Antihypertensive medications, when required, must be carefully chosen because of their potential effects on uteroplacental blood flow and on the fetus. For example, the ACE inhibitors can cause injury and even death of the fetus when given during the second and third trimesters of pregnancy.

Hypertension in Children

Blood pressure is known to increase from infancy to late adolescence. The average systolic pressure at 1 day of age is approximately 70 mm Hg and increases to approximately 85 mm Hg at 1 month of age.^{23,24} Systolic blood pressure continues to increase with physical growth to about 120 mm Hg at the end of adolescence. During the preschool years, blood pressure begins to follow a pattern that tends to be maintained as the children grow older. This pattern continues into adolescence and adulthood, suggesting the roots of essential hypertension are established early in life. A familial influence on blood pressure often can be identified early in life. Children of parents with high blood pressure tend to have higher blood pressures than do children with normotensive parents.

Blood pressure norms for children are based on age, height, and gender-specific percentiles. The 1996 Task Force updated earlier recommendations by including height as a variable in determination of blood pressure (Table 16-3).²⁵ The Task Force recommended continued classification of blood pressure into three ranges: normal (*i.e.*, systolic and diastolic pressures below the 90th percentile for age, height, and gender); high normal (*i.e.*, systolic or diastolic blood pressures between the 90th and 95th percentile for age, height, and gender); and high blood pressures, or hypertension (*i.e.*, systolic and diastolic blood pressures, or normal the 95th percentile for age, height, and gender); and high blood pressures equal to or greater than the 95th percentile for age, height, and gender on at least three occasions).

Secondary hypertension is the most common form of high blood pressure in infants and children. In later childhood and adolescence, essential hypertension is more common. Approximately 75% to 80% of secondary hypertension in children is caused by kidney abnormalities.²⁵ Coarctation of the aorta is another cause of hypertension in children and adolescents. Endocrine causes of hypertension such as pheochromocytoma and adrenal cortical disorders are rare. Hypertension in infants is associated most commonly with high umbilical catheterization and renal artery obstruction caused by thrombosis.²⁶ Most

286 Unit Four: Alterations in the Cardiovascular System

TABLE 16-3

The 90th and 95th Percentiles of Systolic and Diastolic Blood Pressure for Boys and Girls 1 to 16 years of Age by Percentiles for Height*

	Age (yr)	Height Percentile for Boys				Height Percentile for Girls				
Blood Pressure										
Percentiles		5th	25th	75th	95th	5th	25th	75th	95th	
Systolic Pressure										
90th	1	94	97	100	102	97	99	102	104	
95th		98	101	104	106	101	103	105	107	
90th	3	100	103	107	109	100	102	104	106	
95th		104	107	111	113	104	105	108	110	
90th	6	105	108	111	114	104	106	109	111	
95th		109	112	115	117	108	110	112	114	
90th	10	110	113	117	119	112	114	116	118	
95th		114	117	121	123	116	117	120	122	
90th	13	117	120	124	126	118	119	122	124	
95th		121	124	128	130	121	123	126	128	
90th	16	125	128	132	134	122	123	126	128	
95th		129	132	136	138	125	127	130	132	
Diastolic Pressure										
90th	1	50	52	54	55	53	53	55	56	
95th		55	56	58	59	57	57	59	60	
90th	3	59	60	62	63	61	61	63	64	
95th		63	64	66	67	65	65	67	68	
90th	6	67	69	70	72	67	68	69	71	
95th		72	73	75	76	71	72	73	75	
90th	10	73	74	76	78	73	73	75	76	
95th		77	79	80	82	77	77	79	80	
90th	13	75	76	78	80	76	77	78	80	
95th		79	81	83	84	80	81	82	84	
90th	16	79	80	82	83	79	79	81	82	
95th		83	84	86	87	83	83	85	86	

* The height percentiles were determined with standard growth curves. (Data adapted from National Heart, Lung, and Blood Institute. [1996]. Update of the 1987 Task Force Report of the Second Task Force on High Blood Pressure in

Children and Adolescents. A working group report from the National High Blood Pressure Education Program.

Pediatrics 98, 653-654. Available: http://www.nhibi.nih.gov/health/prof/heart/hbp/hbp_ped.htm)

cases of essential hypertension are associated with obesity or a family history of hypertension.

A number of drugs of abuse, therapeutic agents, and toxins also may increase blood pressure. Alcohol should be considered as a risk factor in adolescents. Oral contraceptives may be a cause of hypertension in adolescent females. The nephrotoxicity of the drug cyclosporine, an immunosuppressant used in transplant therapy, may cause hypertension in children after bone marrow, heart, kidney, or liver transplantation. The coadministration of corticosteroid drugs appears to increase the incidence of hypertension.

Diagnosis and Treatment

The Task Force recommended that children 3 years of age through adolescence should have their blood pressure taken once each year. Repeated measurements over time, rather than a single isolated determination, are required to establish consistent and significant observations. Children with high blood pressure should be referred for medical evaluation and treatment as indicated. Treatment includes nonpharmacologic methods and, if necessary, pharmacologic therapy. The Task Force suggested use of the stepped-care approach for drug treatment of children who require antihypertensive medications.²⁵



The prevalence of hypertension in the elderly population (65 to 74 years of age) of the United States ranges from 60% for whites to 71% for African Americans.²⁷ The most common type of hypertension in the elderly is isolated systolic hypertension, in which systolic pressure is elevated while diastolic pressure remains within normal range. A Clinical Advisory Statement intended to advance and clarify the JNC-VI guidelines on the importance of systolic pressure in older Americans, issued by the Coordinating Committee of the National High Blood Pressure Education Program in 2000, reaffirmed the importance of lifelong maintenance of a blood pressure of 140/90 mm Hg or less.28

Among the aging processes that contribute to an increase in blood pressure are a stiffening of the large arteries, particularly the aorta; decreased baroreceptor sensitivity; increased peripheral vascular resistance; and decreased renal blood flow. Systolic blood pressure rises almost linearly between 30 and 84 years of age, whereas diastolic pressure rises until 50 years of age and then levels off or decreases.²⁹ This rise in systolic pressure is thought to be related to increased stiffness of the large arteries.

With aging, the elastin fibers in the walls of the arteries are gradually replaced by collagen fibers that render the vessels stiffer and less compliant. Differences in the central and peripheral arteries relate to the larger vessels containing more elastin, whereas the peripheral resistance vessels have more smooth muscle and less elastin. Because of increased wall stiffness, the aorta and large arteries are less able to buffer the increase in systolic pressure that occurs as blood is ejected from the left heart, and they are less able to store the energy needed to maintain the diastolic pressure. As a result, the systolic pressure increases, the diastolic pressure remains unchanged or actually decreases, and the pulse pressure or difference between the systolic pressure and diastolic pressure widens.

Isolated systolic hypertension (systolic pressure \geq 140 mm Hg and diastolic < 90 mm Hg) is recognized as an important risk factor for cardiovascular morbidity and mortality in older persons.²⁹ Stroke is two to three times more common in elderly hypertensive people than in age-matched normotensive subjects. The treatment of hypertension in the elderly has beneficial effects in terms of reducing the incidence of cardiovascular events such as stroke. The Systolic Hypertension in the Elderly Program (SHEP) showed a reduction of 36% in stroke and a 27% reduction in myocardial infarction in persons who were treated for hypertension compared with those who were not.³⁰

Diagnosis and Treatment

The recommendations for measurement of blood pressure in the elderly are similar to those for the rest of the population. Blood pressure variability is particularly prevalent among older persons, so it is especially important to obtain multiple measurements on different occasions to establish a diagnosis of hypertension. The effects of food, position, and other environmental factors also are exaggerated in older persons. Although sitting has been the standard position for blood pressure measurement, it is recommended that blood pressure also be taken in the supine and standing positions in the elderly. In some elderly persons with hypertension, a silent interval, called the auscultatory gap, may occur between the end of the first and beginning of the third phases of the Korotkoff sounds, providing the potential for underestimating the systolic pressure, sometimes by as much as 50 mm Hg. Because the gap occurs only with auscultation, it is recommended that a preliminary determination of systolic blood pressure be made by palpation and the cuff be inflated 30 mm Hg above this value for auscultatory measurement of blood pressure. In some older persons, the indirect measurement using a blood pressure cuff and the Korotkoff sounds has been shown to give falsely elevated reading compared with the direct intra-arterial method. This is because excessive cuff pressure is needed to compress the rigid vessels of some older persons. Pseudohypertension should be suspected in older persons with hypertension in whom the radial or brachial artery remains palpable but pulseless at higher cuff pressures.

The treatment of hypertension in the elderly is similar to that for younger age groups. However, blood pressure should be reduced slowly and cautiously. When possible, appropriate lifestyle modification measures should be tried first. Antihypertensive medications should be prescribed carefully because the older person may have impaired baroreflex sensitivity and renal function. Usually, medications are initiated at smaller doses, and doses are increased more gradually. There is also the dangers of adverse drug interactions in older persons,

who may be taking multiple medications, including over-thecounter drugs.

In summary, hypertension (systolic pressure \geq 140 mm Hg and/or diastolic pressure >90 mm Hg) is one of the most common cardiovascular disorders. It may occur as a primary disorder (*i.e.*, essential hypertension) or as a symptom of some other disease (i.e., secondary hypertension). Essential hypertension may be classified as systolic/diastolic hypertension in which both the systolic and diastolic pressures are elevated; as diastolic hypertension in which the diastolic pressure is selectively elevated; or as systolic hypertension in which the systolic pressure is selectively elevated. The incidence of essential hypertension increases with age; the condition is seen more frequently among African Americans, and it is linked to a family history of high blood pressure, obesity, and increased salt intake.

Causes of secondary hypertension include renal disorders and adrenal cortical disorders, such as hyperaldosteronism and Cushing's disease, which increase salt and water retention; pheochromocytomas, which increase catecholamine levels; and coarctation of the aorta, which produces a decrease in leg blood pressures and compensatory increase in arm pressures.

Uncontrolled hypertension increases the risk of heart disease, renal complications, retinopathy, and stroke. Treatment of essential hypertension focuses on nonpharmacologic methods such as weight reduction, reduction of sodium intake, regular physical activity, modification of alcohol intake, and smoking cessation. Among the drugs used in the treatment of hypertension are diuretics, adrenergic inhibitors, vasodilators, ACE inhibitors, and calcium channel-blocking drugs.

Hypertension that occurs during pregnancy can be divided into four categories: chronic hypertension, preeclampsiaeclampsia, chronic hypertension with superimposed preeclampsia-eclampsia, and gestational hypertension. Preeclampsia-eclampsia, which is hypertension that develops after 20 weeks' gestation, is accompanied by proteinuria, and poses a particular threat to the mother and the fetus. Blood pressure norms for children are based on age, height, and gender-specific percentiles. Secondary hypertension is the most common form of high blood pressure in infants and children. In later childhood and adolescence, essential hypertension is more common. Isolated systolic hypertension is the most common type of hypertension in the elderly. It represents the effects of aging on the distensibility of the aorta and its ability to stretch and accommodate blood being ejected from the left heart during systole. Untreated systolic hypertension is recognized as an important risk factor for stroke and other cardiovascular morbidity and mortality in older persons.

ORTHOSTATIC HYPOTENSION

Orthostatic or postural hypotension is an abnormal drop in blood pressure on assumption of the standing position. In the absence of normal circulatory reflexes or blood volume, blood pools in the lower part of the body; when the standing position is assumed, cardiac output falls, and blood flow to the brain is inadequate. Dizziness, syncope (*i.e.*, fainting), or both may occur.

After the assumption of the upright posture from the supine position, approximately 500 to 700 mL of blood is momentarily shifted to the lower part of the body, with an accompanying decrease in central blood volume and arterial pressure.³¹ Normally, this decrease in blood pressure is transient, lasting through several cardiac cycles, because the baroreceptors located in the thorax and carotid sinus area sense the decreased pressure and initiate reflex constriction of the veins and arterioles and an increase in heart rate, which brings blood pressure back to normal. Within a few minutes of a change to the standing position, blood levels of antidiuretic hormone and sympathetic neuromediators increase as a secondary means of ensuring maintenance of normal blood pressure in the standing position. Muscle movement in the lower extremities also aids venous return to the heart by pumping blood out of the legs.

In persons with healthy blood vessels and normal autonomic nervous system function, cerebral blood flow usually is not reduced on assumption of the upright position unless arterial pressure falls below 70 mm Hg. The strategic location of the arterial baroreceptors between the heart and brain is designed to ensure that the arterial pressure is maintained within a range sufficient to prevent a reduction in cerebral blood flow.

Classification

Although there is no firm agreement on the definition of orthostatic hypotension, many authorities consider a drop in systolic pressure of 20 mm Hg or more or a drop in diastolic blood pressure of 10 mm Hg or more as diagnostic of the condition.³² Some authorities regard the presence of orthostatic symptoms (*e.g.*, dizziness, syncope) as being more relevant than the numeric decrease in blood pressure.³³

Causes

A wide variety of conditions, acute and chronic, are associated with orthostatic hypotension. These include reduced blood volume, drug-induced hypotension, altered vascular responses associated with aging, bed rest, and autonomic nervous system dysfunction.

Reduced Blood Volume

Orthostatic hypotension often is an early sign of reduced blood volume or fluid deficit. When blood volume is decreased, the vascular compartment is only partially filled; although cardiac output may be adequate when a person is in the recumbent position, it often decreases to the point of causing weakness and fainting when the person assumes the standing position. Common causes of orthostatic hypotension related to hypovolemia are excessive use of diuretics, excessive diaphoresis, loss of gastrointestinal fluids through vomiting and diarrhea, and loss of fluid volume associated with prolonged bed rest.

Drug-Induced Hypotension

Antihypertensive drugs and psychotropic drugs are the most common cause of chronic orthostatic hypotension. In most cases, the orthostatic hypotension is well tolerated. If postural hypotension is severe enough to cause light-headedness or dizziness, it is recommended that the dosage of the drug be reduced or a different drug be used.³³

Aging

Weakness and dizziness on standing are common complaints of elderly persons. Orthostatic hypotension is associated with systolic hypertension, major electrocardiographic abnormalities, and carotid artery stenosis. Because cerebral blood flow primarily depends on systolic pressure, patients with impaired cerebral circulation may experience symptoms of weakness, ataxia, dizziness, and syncope when their arterial pressure falls even slightly. This may happen in older persons who are immobilized for brief periods or whose blood volume is decreased owing to inadequate fluid intake or overzealous use of diuretics.

Postprandial blood pressure often decreases in elderly persons.³⁴ The greatest postprandial changes occur after a highcarbohydrate meal. Although the mechanism responsible for these changes is not fully understood, it is thought to result from glucose-mediated impairment of baroreflex sensitivity and increased splanchnic blood flow mediated by insulin and vasoactive gastrointestinal hormones.

Bed Rest

Prolonged bed rest promotes a reduction in plasma volume, a decrease in venous tone, failure of peripheral vasoconstriction, and weakness of the skeletal muscles that support the veins and assist in returning blood to the heart. Physical deconditioning follows even short periods of bed rest. After 3 to 4 days, the blood volume is decreased. Loss of vascular and skeletal muscle tone is less predictable but probably becomes maximal after approximately 2 weeks of bed rest. Orthostatic intolerance is a recognized problem of space flight, a potential risk after re-entry into the earth's gravitational field.

Disorders of Autonomic Nervous System Function

The sympathetic nervous system plays an essential role in adjustment to the upright position. Sympathetic stimulation increases heart rate and cardiac contractility and causes constriction of peripheral veins and arterioles. Orthostatic hypotension caused by altered autonomic function is common in peripheral neuropathies associated with diabetes mellitus, after injury or disease of the spinal cord, or as the result of a cerebral vascular accident in which sympathetic outflow from the brain stem is disrupted. The American Autonomic Society and the American Academy of Neurology have distinguished three forms of primary autonomic nervous system dysfunction: pure autonomic failure, defined as a sporadic, idiopathic cause of persistent orthostatic hypotension and other manifestations of autonomic failure such as urinary retention, impotence, or decreased sweating; Parkinson's disease with autonomic failure; and multiple-system atrophy (Shy-Drager syndrome).35 The Shy-Drager syndrome usually develops in middle to late life and manifests as orthostatic hypotension associated with uncoordinated movements, urinary incontinence, constipation, and other signs of neurologic deficits referable to the corticospinal, extrapyramidal, corticobulbar, and cerebellar systems.

Diagnosis and Treatment

Orthostatic hypotension can be assessed with the blood pressure cuff. A reading should be made when the patient is supine, immediately after assumption of the seated or upright position,

and after 2 to 3 minutes following assumption of the standing position. A tilt table also can be used for this purpose. With a tilt table, the recumbent patient can be moved to a head-up position without voluntary movement when the table is tilted. The tilt table also has the advantage of rapidly and safely returning persons with a profound postural drop in blood pressure to the horizontal position. Persons with a drop in blood pressure to orthostatic levels should be evaluated to determine the cause and seriousness of the condition. A history should be done to elicit information about symptoms, particularly dizziness and history of syncope and falls; medical conditions, particularly those such as diabetes mellitus that predispose to orthostatic hypotension; use of prescription and over-the-counter drugs; and symptoms of autonomic nervous system dysfunction, such as impotence or bladder dysfunction. A physical examination should document blood pressure in both arms and the heart rate while the patient is in the supine, sitting, and standing positions and should note the occurrence of symptoms. Noninvasive, 24-hour ambulatory blood pressure monitoring may be used to determine blood pressure responses to other stimuli of daily life, such as food ingestion and exertion.

Treatment of orthostatic hypotension usually is directed toward alleviating the cause or, if this is not possible, helping people learn ways to cope with the disorder and prevent falls and injuries. Medications that predispose to postural hypotension should be avoided. Other measures include preventing or correcting the fluid deficit and avoidance of situations that encourage excessive vasodilatation (*e.g.*, drinking alcohol, exercising vigorously in a warm environment). Measures designed to help persons prevent symptomatic orthostatic drops in blood pressure include gradual ambulation (*i.e.*, sitting on the edge of the bed for several minutes and moving the legs to initiate skeletal muscle pump function before standing). Elastic support hose or an abdominal support garment may help prevent pooling of blood in the lower extremities and abdomen.

Pharmacologic treatment may be used when nonpharmacologic methods are unsuccessful. A number of types of drugs can be used for this purpose.³² Mineralocorticoids can be used to reduce salt and water loss. Vasopressin-2 receptor agonists (desmopressin as a nasal spray) may be used to reduce nocturnal polyuria. Sympathomimetic drugs that act directly on the resistance vessels or on the capacitance vessels may be used. Many of these agents have undesirable side effects.

In summary, orthostatic hypotension refers to an abnormal decrease in systolic and diastolic blood pressures that occurs on assumption of the upright position. An important consideration in orthostatic hypotension is the occurrence of dizziness and syncope. Among the factors that contribute to its occurrence are decreased fluid volume, medications, aging, defective function of the autonomic nervous system, and the effects of immobility. Diagnosis of orthostatic hypotension includes blood pressure measurement in the supine and upright positions, a history of symptomatology, medication use, and disease conditions that contribute to a postural drop in blood pressure. Treatment includes correcting the reversible causes and assisting the person to compensate for the disorder to prevent falls and injuries.

REVIEW QUESTIONS

Explain how cardiac output and peripheral vascular resistance interact in determining systolic and diastolic blood pressure.

Explain the difference between the essential and secondary forms of hypertension.

Cite some of the major contributing factors in the development of hypertension.

Cite the risks of hypertension in terms of target organ damage.

Explain the changes in blood pressure that accompany normal pregnancy and describe the four types of hypertension that can occur during pregnancy.

Cite the criteria for the diagnosis of high blood pressure in children.

Define systolic hypertension, characterize the effect of increased systolic and pulse pressure on the production of target organ damage, and explain why systolic hypertension is more common in the elderly.

Define the term orthostatic hypotension.

Explain how fluid deficit, medications, aging, disorders of the autonomic nervous system, and bed rest contribute to the development of orthostatic hypotension

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- 1. Guyton A., Hall J.E. (2000). *Medical physiology* (10th ed., pp. 185–186, 195–205). Philadelphia: W.B. Saunders.
- National Heart, Lung, and Blood Institute. (1997). The sixth report of the Joint National Committee on Detection, Evaluation, and Treatment of High Blood Pressure. Archives of Internal Medicine 157, 2413–2443.
- 3. American Diabetes Association. (2001). Summary of revisions for the 2001 clinical practice recommendations. *Diabetes Care* 24 (Suppl. 1), 1.
- Grim C.E., Henry J.P., Myers H. (1995). High blood pressure in blacks: Salt, slavery, survival, stress, and racism. In Laragh J.H., Brenner B.M. (Eds.), *Hypertension: Pathophysiology, diagnosis, and management* (pp. 171–207). New York: Raven Press.
- Kotchen T.A., McCarron D.A. (1998). Dietary electrolytes and blood pressure: A statement for healthcare professionals from the American Heart Association Nutrition Committee. *Circulation* 98, 613–617.
- Cassano P.A., Segal M.R., Vokonas P.S., et al. (1990). Body fat distribution, blood pressure, and hypertension: A prospective study of men in the normative aging study. *Annals of Epidemiology* 1, 33–48.
- Fuchs F.D., Chambless L.E., Whelton P.K., et al. (2001). Alcohol consumption and the incidence of hypertension. *Hypertension* 37, 1242– 1250.
- Marmot M.G., Elliott P., Shipley M.J., et al. (1994). Alcohol and blood pressure: The INTERSALT study. *British Medical Journal* 308, 1263– 1267.
- 9. Kaplan N.M. (1995). The treatment of hypertension in women. Archives of Internal Medicine 155, 563–567.
- 10. Kannel W.B. (1996). Blood pressure as a cardiovascular risk factor. *Journal of the American Medical Association* 275, 1571–1576.

290 Unit Four: Alterations in the Cardiovascular System

- Frohlich E.D., Chobanian A.B., Devereux R.G., et al. (1992). The heart in hypertension. *New England Journal of Medicine* 327, 998–1008.
- Grim C.E., Grim C.M. (2001). Accurate and reliable blood pressure measurement in the clinic and home: The key to hypertension control. In Hollenberg N. (Ed.), *Hypertension: Mechanisms and management* (3rd ed., pp. 315–324). Philadelphia: Current Medicine.
- O'Brien E. (1996). Review: A century of confusion; which bladder for accurate blood pressure measurement? *Journal of Human Hypertension* 10, 565–572.
- Klotchen T.A., McCarron D.A. (1998). Dietary electrolytes and blood pressure: A statement for health care professionals from the American Heart Association Nutrition Committee. *Circulation* 98, 613–617.
- Black H.R., Kuller L.H., O'Rourke M.F., et al. (1999). The first report of the Systolic and Pulse Pressure (SYPP) working group. *Journal of Hypertension* 17 (Suppl. 5), S3–S14.
- Alderman M.H. (1999). A new model of risk implications of increasing pulse pressure and systolic blood pressure in cardiovascular disease. *Journal of Hypertension* 17 (Suppl. 5), S23–S28.
- 17. Ram C.V. (1994). Secondary hypertension: Workup and correction. *Hospital Practice* 4, 137–155.
- Venkata C., Ram S., Fierro-Carrion G.A. (1995). Pheochromocytoma. Seminars in Nephrology 15, 126–137.
- Roa P.S. (1995). Coarctation of the aorta. Seminars in Nephrology 15, 87–105.
- Chames M.C., Sibal B.M. (2001). When chronic hypertension complicates pregnancy. *Contemporary OB/GYN Archive* April 2 [On-line.] Available: http://ahgyn.pdf.net/public.htm. Accessed May 14, 2001.
- 21. Gifford R.W. Jr. (Chair). (2000). *National High Blood Pressure Working Group report on high blood pressure in pregnancy*. NIH publication no. 00-3029. Bethesda, MD: National Institutes of Health.
- 22. Sibai B.M., Lindheimer M., Hauth J., et al. (1998). Risk factors for preeclampsia, abruptio placentae, and adverse neonatal outcomes among women with chronic hypertension. *New England Journal of Medicine* 339, 667–671.
- Sinaiko A.R. (1996). Hypertension in children (Review). New England Journal of Medicine 335, 1968–1973.
- Bartosh S.M., Aronson A.J. (1999). Childhood hypertension. *Pediatric Clinics of North America* 46, 235–251.

- 25. National Heart, Lung and Blood Institute. (1996). Update of the 1987 Task Force Report of the Second Task Force on High Blood Pressure in Children and Adolescents: A working group report from the National High Blood Pressure Education Program. *Pediatrics* 88, 649–658.
- Behrman R.E., Kliegman R.M., Arvin A.M. (2000). Systemic hypertension. In Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.), *Nelson textbook of pediatrics* (16th ed., pp. 1450–1455). Philadelphia: W.B. Saunders.
- National High Blood Pressure Education Program Working Group. (1994). National High Blood Pressure Education Working Group report on hypertension in the elderly. *Hypertension* 23, 275–285.
- Izzo J.L., Levy D., Black H.R. (2000). Clinical advisory statement: Importance of systolic blood pressure in older Americans. *Hypertension* 35, 1021–1024.
- Franklin S.S. (1999). Aging and hypertension: The assessment of blood pressure indices in predicting coronary heart disease. *Journal of Hypertension* 17 (Suppl. 5), S29–S36.
- SHEP Cooperative Research Group. (1991). Prevention of stroke by antihypertensive drug treatment in older persons with isolated systolic hypertension. *Journal of the American Medical Association* 265, 3255–3264.
- Smith J.J., Porth C.J.M. (1990). Age and the response to orthostatic stress. In Smith J.J. (Ed.), *Circulatory response to the upright posture* (pp. 121–138). Boca Raton, FL: CRC Press.
- Mathias C.J., Kimber J.R. (1999). Postural hypotension: Causes, clinical features, investigation, and management. *Annual Review of Medicine* 50, 317–336.
- Kochar M.S. (1990). Orthostatic hypotension. In Smith J.J. (Ed.), *Circulatory response to the upright posture* (pp. 170–179). Boca Raton, FL: CRC Press.
- Jansen R.W.M.M., Lipsitz L.A. (1995). Postprandial hypotension: Epidemiology, pathophysiology, and clinical management. *Annals of Internal Medicine* 122, 286–295.
- American Autonomic Society and American Academy of Neurologists. (1996). Consensus statement of the definition of orthostatic hypotension, pure autonomic failure, and multiple system atrophy. *Neurology* 46, 1470.

Alterations in Cardiac Function

Disorders of the Pericardium Pericardial Effusion **Pericarditis Acute Pericarditis Chronic Pericarditis With Effusion Constrictive Pericarditis Coronary Heart Disease Coronary Circulation** Pathogenesis of Coronary Heart Disease **Chronic Ischemic Heart Disease Stable Angina** Variant Angina Silent Myocardial Ischemia **Acute Coronary Syndromes** Unstable Angina/Non-ST-Segment Elevation Myocardial Infarction ST-Segment Elevation Myocardial Infarction Myocardial and Endocardial Diseases **Myocarditis Cardiomyopathies Dilated Cardiomyopathies** Hypertrophic Cardiomyopathies **Restrictive Cardiomyopathies** Arrhythmogenic Right Ventricular Cardiomyopathy Peripartum Cardiomyopathy **Disorders Affecting the Endocardium** Infective Endocarditis **Rheumatic Heart Disease** Valvular Heart Disease Hemodynamic Derangements **Mitral Valve Disorders** Mitral Valve Stenosis Mitral Valve Regurgitation Mitral Valve Prolapse

CHAPTER

Aortic Valve Disorders Aortic Valve Stenosis Aortic Valve Regurgitation Diagnosis and Treatment Heart Disease in Infants and Children Fetal and Perinatal Circulation Congenital Heart Disorders Shunting and Cyanotic Disorders Types of Defects Kawasaki's Disease Clinical Course Diagnosis and Treatment

eart disease remains a leading cause of death and disability in the United States and throughout the world. Currently, it is the number one cause of death in the United States. It is also the leading cause of permanent disability in the U.S. labor force and accounts for 19% of Social Security disability payments.¹ By the year 2020, it is projected that cardiovascular disease will, for the first time in human history, be the most common cause of death worldwide.²

In an attempt to focus on common heart problems that affect persons in all age groups, this chapter is organized into five sections: disorders of the pericardium, coronary heart disease, myocardial and endocardial diseases, valvular heart disease, and heart disease in infants and children.

DISORDERS OF THE PERICARDIUM

The pericardium is a double-layered serous membrane that isolates the heart from other thoracic structures, maintains its position in the thorax, and prevents it from overfilling. The two layers of the pericardium are separated by a thin layer of serous fluid, which prevents frictional forces from developing as the inner visceral layer, or epicardium, comes in contact with the outer parietal layer of the fibrous pericardium.

The pericardium is subject to many of the same pathologic processes (*e.g.*, congenital disorders, infections, trauma, immune mechanisms, and neoplastic disease) that affect other structures of the body. Pericardial disorders frequently are

KEY CONCEPTS

DISORDERS OF THE PERICARDIUM

- The pericardium isolates the heart from other thoracic structures, maintains its position in the thorax, and prevents it from overfilling.
- The two layers of the pericardium are separated by a thin layer of serous fluid, which prevents frictional forces from developing between the visceral and parietal layers of the pericardium.
- Disorders that produce inflammation of the pericardium interfere with the friction-reducing properties of the pericardial fluid and produce pain.
- Disorders that increase the fluid volume of the pericardial sac interfere with cardiac filling and produce a subsequent reduction in cardiac output.

associated with or result from another disease within the heart or in the surrounding structures.

Pericardial Effusion

Pericardial effusion refers to the accumulation of fluid in the pericardial cavity. Normally, there is about 30 to 50 mL of thin, clear, straw-colored fluid in the pericardial sac. The amount of fluid, the rapidity with which it accumulates, and the elasticity of the pericardium determine the effect the effusion has on cardiac function. Small pericardial effusions may produce no symptoms or abnormal clinical findings. Even a large effusion that develops slowly may cause few or no symptoms, provided the pericardium is able to stretch and avoid compressing the heart. However, a sudden accumulation of even 200 mL may raise intracardiac pressure to levels that seriously limit the venous return to the heart. Symptoms of cardiac compression also may occur with relatively small accumulations of fluid when the pericardium has become thickened by scar tissue or neoplastic infiltrations.

Cardiac tamponade represents an increase in intrapericardial pressure caused by an accumulation of fluid or blood in the pericardial sac. It can occur as the result of conditions such as trauma, cardiac surgery, cancer, uremia, or cardiac rupture caused by myocardial infarction. The seriousness of cardiac tamponade results from impairment in diastolic filling and reduction in stroke volume and cardiac output. The severity of the condition depends on the amount of fluid that is present and the rate at which it accumulated. A rapid accumulation of fluid results in an elevation of central venous pressure, jugular venous distention, a decline in venous return to the heart, a decrease in cardiac output despite an increase in heart rate, a fall in systolic blood pressure, pulsus paradoxus, and signs of circulatory shock.

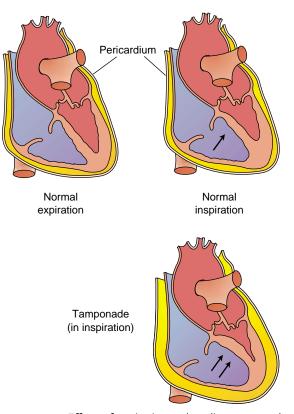
Pulsus paradoxus, which refers to an exaggeration (>10 mm Hg) of the normal (2 to 4 mm Hg) decrease in systolic blood pressure that occurs during inspiration, is a clinical indicator of cardiac tamponade. The decreased intrathoracic pressure that occurs during inspiration normally accelerates venous flow, increasing

right atrial and right ventricular filling. This causes the interventricular septum to bulge to the left, producing a decrease in left ventricular filling, stroke volume output, and systolic blood pressure. In cardiac tamponade, the left ventricle is compressed from within by movement of the interventricular septum and from without by fluid in the pericardium (Fig. 17-1). This produces a marked decrease in left ventricular filling and left ventricular stroke volume output, often within a beat of the beginning of inspiration.

The echocardiogram is a rapid, accurate, and widely used method for evaluating pericardial effusion. Aspiration and laboratory analysis of the pericardial fluid may be used to identify the causative agent. Cardiac catheterization may be used to determine the hemodynamic effects of pericardial effusion and cardiac tamponade. Pericardiocentesis, the removal of fluid from the pericardial sac, may be a lifesaving measure in severe cardiac tamponade.

Pericarditis

Pericarditis represents an inflammatory process of the pericardium.³ It can result from a number of diverse causes. Most forms of pericarditis occur secondary to other systemic or car-



■ FIGURE 17-1 ■ Effects of respiration and cardiac tamponade on ventricular filling and cardiac output. During inspiration venous flow into the right heart increases, causing the interventricular septum to bulge into the left ventricle. This produces a decrease in left ventricular volume, with a subsequent decrease in stroke volume output. In cardiac tamponade, the fluid in the pericardial sac produces further compression of the left ventricle, causing an exaggeration of the normal inspiratory decrease in stroke volume and systolic blood pressure.

diac disease. Primary pericarditis is unusual and most often of viral origin. Most causes of pericarditis evoke an acute pericarditis. Exceptions are tuberculosis and fungal infections, which often produce chronic reactions.⁴

Acute Pericarditis

Acute pericarditis can be classified according to cause (*e.g.*, infections, trauma, rheumatic fever) or the nature of the exudate (*e.g.*, serous, fibrinous, purulent, hemorrhagic). Like other inflammatory conditions, acute pericarditis often is associated with increased capillary permeability. The capillaries that supply the serous pericardium become permeable, allowing plasma proteins, including fibrinogen, to leave the capillaries and enter the pericardial space. This results in an exudate that varies in type and amount according to the causative agent. Acute pericarditis frequently is associated with a fibrinous (fibrin-containing) exudate, which heals by resolution or progresses to deposition of scar tissue and formation of adhesions between the layers of the serous pericardium (Fig. 17-2). Inflammation also may involve the superficial myocardium and the adjacent pleura.

Viral infection (especially infections with the coxsackieviruses and echoviruses, but also influenza, Epstein-Barr, varicella, hepatitis, mumps, and human immunodeficiency viruses) is one of the most common causes of acute pericarditis. Viral pericarditis is seen more frequently in men than in women. It is commonly preceded by a prodromal phase during which fever, malaise, and other flulike symptoms are present. Often it is preceded by an upper respiratory tract infection. Although the acute symptoms of viral pericarditis usually subside in several weeks, easy fatigability often continues for several months.



■ FIGURE 17-2 ■ Fibrinous pericarditis. The heart of a patient who died of uremia displays a shaggy, fibrinous exudate covering the visceral pericardium. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 585]. Philadelphia: Lippincott Williams & Wilkins)

In many cases, the condition is self-limited, resolving in 2 to 6 weeks. Rarely, it may persist and produce recurrent subacute or chronic disease.

Other causes of acute pericarditis are rheumatic fever, postpericardiotomy syndrome, post-traumatic pericarditis, metabolic disorders (*e.g.*, uremia, myxedema), and pericarditis associated with connective tissue diseases (*e.g.*, systemic lupus erythematosus, rheumatoid arthritis). With the increased use of open-heart surgery in the treatment of various heart disorders, postpericardiotomy syndrome has become a commonly recognized form of pericarditis. Pericarditis with effusion is a common complication in persons with renal failure, in those with untreated uremia, and in those being treated with hemodialysis. Irradiation may initiate a subacute pericarditis, with an onset usually within the first year of therapy. It is most commonly associated with high doses of radiation delivered to areas near the heart.

The manifestations of acute pericarditis include a triad of chest pain, pericardial friction rub, and serial electrocardiographic (ECG) changes. The clinical findings and other manifestations may vary according to the causative agent. Nearly all persons with acute pericarditis have chest pain. The pain usually is abrupt in onset, occurs in the precordial area, and is described as sharp. It may radiate to the neck, back, abdomen, or side. It typically is worse with deep breathing, coughing, swallowing, and positional changes because of changes in venous return and cardiac filling. A pericardial friction rub results from the rubbing and friction between the inflamed pericardial surfaces. Serial ECGs are useful in differentiating acute pericarditis from myocardial infarction. Treatment of acute pericarditis depends on the cause. When infection is present, antibiotics specific for the causative agent usually are prescribed. Anti-inflammatory drugs such as aspirin and nonsteroidal anti-inflammatory agents (NSAIDs) may be given to minimize the inflammatory response and the accompanying undesirable effects.

Chronic Pericarditis With Effusion

Chronic pericarditis with effusion is characterized by an increase in inflammatory exudate that continues beyond the acute period. In some cases, the exudate persists for several years. The process commonly is associated with other forms of heart disease, such as rheumatic fever, congenital heart lesions, or hypertensive heart disease. Systemic diseases, such as lupus erythematosus, rheumatoid arthritis, and scleroderma also are causes of chronic pericarditis, as are metabolic disturbances associated with acute and chronic renal failure. In most cases of chronic pericarditis, no specific pathogen or cause can be identified. Unlike with acute pericarditis, the signs and symptoms of chronic pericarditis often are minimal, with the condition being detected for the first time on a routine chest x-ray film. As the condition progresses, the fluid may accumulate and compress the adjacent cardiac structures and impair cardiac filling.

Constrictive Pericarditis

In constrictive pericarditis, fibrous scar tissue develops between the visceral and parietal layers of the serous pericardium. In time, the scar tissue contracts and interferes with diastolic filling of the heart, at which point cardiac output and cardiac reserve becomes fixed. Ascites is a prominent early finding and may be accompanied by pedal edema, dyspnea on exertion, and fatigue. The jugular veins also are distended. Surgical removal of the pericardium may be necessary in persons with severe compromise of cardiac function.

In summary, Disorders of the pericardium include pericardial effusion, acute and chronic pericarditis, and constrictive pericarditis. Pericardial effusion, which refers to the presence of an exudate in the pericardial cavity, can increase intracardiac pressure, compress the heart, and interfere with venous return to the heart. Cardiac tamponade is a life-threatening cardiac compression resulting from excess fluid in the pericardial sac. Acute pericarditis is characterized by chest pain, ECG changes, and a friction rub. Among its causes are infections, uremia, rheumatic fever, connective tissue diseases, and myocardial infarction. In constrictive pericarditis, scar tissue develops between the visceral and parietal layers of the serous pericardium. In time, the scar tissue contracts and interferes with cardiac filling.

CORONARY HEART DISEASE

The term *coronary heart disease* (CHD) describes heart disease caused by impaired coronary blood flow. In most cases, CHD is caused by atherosclerosis. Diseases of the coronary arteries can cause angina, myocardial infarction or heart attack, cardiac dysrhythmias, conduction defects, heart failure, and sudden death. Heart attack is the largest killer of American men and women, claiming more than 218,000 lives annually.¹ Each year, 1.5 million Americans have new or recurrent heart attacks, and one third of those die within the first hour, usually as the result of cardiac arrest resulting from ventricular fibrillation.

During the past 50 years, there have been phenomenal advances in understanding the pathogenesis of CHD and in the development of diagnostic techniques and treatment methods for disease. However, declines in morbidity and mortality have failed to keep pace with these scientific advances, probably because many of the outcomes are more dependent on lifestyle factors and age than on scientific advances.

Coronary Circulation

There are two main coronary arteries, the left and the right, which arise from the coronary sinus just above the aortic valve (Fig. 17-3). The left coronary artery extends for approximately 3.5 cm as the left main coronary artery and then divides into the anterior descending and circumflex branches. The left anterior descending artery passes down through the groove between the two ventricles, giving off diagonal branches, which supply the left ventricle, and perforating branches, which supply the anterior portion of the interventricular septum and the anterior papillary muscle of the left ventricle. The circumflex branch of the left coronary artery passes to the left and moves posteriorly in the groove that separates the left atrium and ventricle, giving off branches that supply the left lateral wall of the left ventricle. The *right coronary artery* lies in the right atrioventricular groove, and its branches supply the right ventricle. The sinoatrial node usually is supplied by the right coronary artery. The right coronary artery usually moves to the back of the heart, where it forms the *posterior descending artery*, which supplies the posterior portion of the heart (the interventricular septum, atrioventricular [AV] node, and posterior papillary muscle). In 10% to 20% of persons, the left circumflex, rather than the right coronary artery, moves posteriorly to form the posterior descending artery.

Although there are no connections between the large coronary arteries, there are anastomotic channels that join the small arteries (Fig. 17-4). With gradual occlusion of the larger vessels, the smaller collateral vessels increase in size and provide alternative channels for blood flow.⁵ One of the reasons CHD does

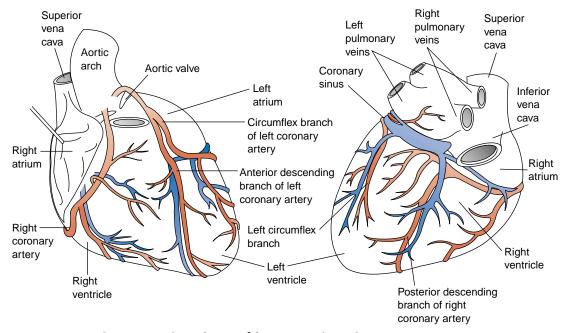


FIGURE 17-3 Coronary arteries and some of the coronary sinus veins.

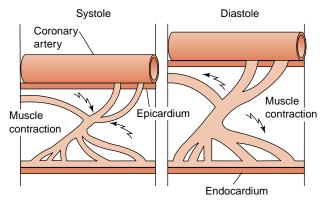


not produce symptoms until it is far advanced is that the collateral channels develop at the same time the atherosclerotic changes are occurring.

The openings for the coronary arteries originate in the root of the aorta just outside the aortic valve; thus, the primary factor responsible for perfusion of the coronary arteries is the aortic blood pressure. Changes in aortic pressure produce parallel changes in coronary blood flow.

In addition to generating the aortic pressure that moves blood through the coronary vessels, the contracting heart muscle influences its own blood supply by compressing the intramyocardial and subendocardial blood vessels.5 The large epicardial coronary arteries lie on the surface of the heart, with the smaller intramyocardial coronary arteries branching off and penetrating the myocardium before merging with a network or plexus of subendocardial vessels that supply the endocardium. During systole, contraction of the cardiac muscle compresses the intramyocardial vessels that feed the subendocardial plexus, and the increased pressure in the ventricle causes further compression of these vessels (Fig. 17-5). As a result, blood flow through the subendocardial vessels occurs mainly during diastole. Thus, there is increased risk of subendocardial ischemia and infarction when diastolic pressure is low, when a rapid heart rate decreases the time spent in diastole, and when an elevation in diastolic intraventricular pressure is sufficient to compress the vessels in the subendocardial plexus.5,6

Heart muscle relies primarily on fatty acids and aerobic metabolism to meet its energy needs. Although the heart can engage in anaerobic metabolism, this process relies on the continuous delivery of glucose and results in the formation of large amounts of lactic acid. Blood flow usually is regulated by the need of the cardiac muscle for oxygen. Even under normal rest-



■ FIGURE 17-5 ■ The compressing effect of the contracting myocardium on intramyocardial blood vessels and subendocardial blood flow during systole and diastole.

ing conditions, the heart extracts and uses 60% to 80% of oxygen in blood flowing through the coronary arteries, compared with the 25% to 30% extracted by skeletal muscle.5 Because there is little oxygen reserve in the blood, myocardial ischemia develops when the coronary arteries are unable to dilate and increase blood flow during periods of increased activity or stress.

Pathogenesis of Coronary Heart Disease

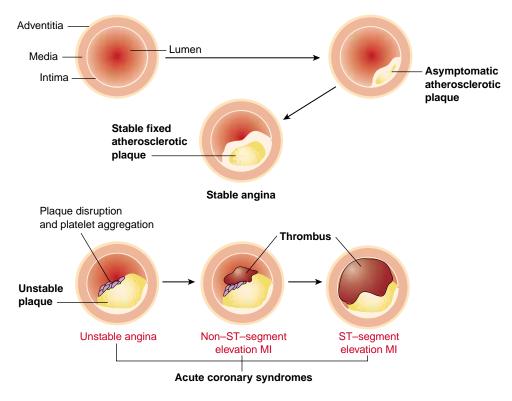
Atherosclerosis (discussed in Chapter 15) is by far the most common cause of CHD, and atherosclerotic plague disruption the most frequent cause of myocardial infarction and sudden death. More than 90% of persons with CHD have coronary atherosclerosis.4 Most, if not all, have one or more lesions causing at least 75% reduction in cross-sectional area, the point at which augmented blood flow provided by compensatory vasodilation no longer is able to keep pace with even moderate increases in metabolic demand.4

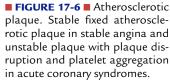
Atherosclerosis can affect one or all three of the major epicardial coronary arteries and their branches (i.e., one-, two-, or three-vessel disease). Clinically significant lesions may be located anywhere in these vessels but tend to predominate in the first several centimeters of the left anterior descending and left circumflex or the entire length of the right coronary artery.⁴ Sometimes the major secondary branches also are involved.

There are two types of atherosclerotic lesions: the *fixed* or stable plaque, which obstructs blood flow, and the unstable or vulnerable plaque, which can rupture and cause platelet adhesion and thrombus formation. The fixed or stable plaque is commonly implicated in chronic ischemic heart disease (stable angina, variant or vasospastic angina, and silent myocardial ischemia) and the unstable plaque in unstable angina and myocardial infarction (Fig. 17-6).

Atherosclerotic plaques are made up of a soft lipid-rich core with a fibrous cap. Plaques with a thin fibrous cap overlying a large lipid core are at greatest risk for rupture. Plaque disruption may occur with or without thrombosis. When the plaque injury is mild, intermittent thrombotic occlusions may occur and cause episodes of anginal pain at rest. More extensive thrombus formation can progress until the coronary artery becomes occluded, leading to myocardial infarction. Platelets play a major role in linking plaque disruption to acute CHD. As a part of the response to plaque disruption, platelets aggregate and release







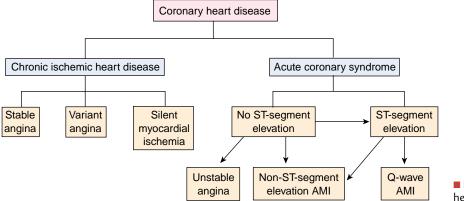
substances that further propagate platelet aggregation, vasoconstriction, and thrombus formation. Because of the role that platelets play in the pathogenesis of CHD, antiplatelet drugs (*e.g.*, low-dose aspirin) are frequently used for preventing heart attack.

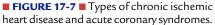
There are two types of thrombi formed as a result of plaque disruption: white platelet-containing thrombi and red fibrincontaining thrombi. The thrombi in unstable angina have been characterized as grayish-white and presumably platelet rich.⁴ Red thrombi, which develop with vessel occlusion in myocardial infarction, are rich in fibrin and red blood cells superimposed on the platelet component and extended by the stasis of blood flow.

Coronary heart disease is commonly divided into two types of disorders: chronic ischemic heart disease and the acute coronary syndromes (Fig. 17-7). There are three types of chronic ischemic heart disease: chronic stable angina, variant or vasospastic angina, and silent myocardial ischemia. The acute coronary syndromes represent the spectrum of ischemic coronary disease ranging from unstable angina through myocardial infarction.

Chronic Ischemic Heart Disease

The term *ischemia* means "to suppress or withhold blood flow." Limitations in coronary blood flow most commonly are the result of atherosclerosis, with vasospasm and thrombosis as contributing factors. The metabolic demands of the heart are increased with everyday activities such as mental stress, exercise, and exposure to cold. In certain disease states such as thyrotoxicosis, the metabolic demands may be so excessive that the blood flow may be inadequate despite normal coronary arteries. In other situations, such as aortic stenosis, the coronary arteries may not be diseased, but the perfusion pressure may be insufficient to provide adequate blood flow.



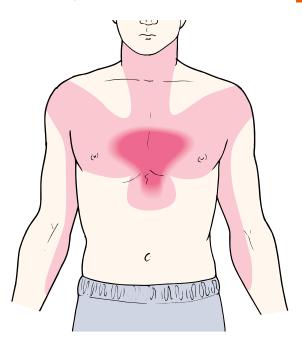


Chapter 17: Alterations in Cardiac Function 297

KEY CONCEPTS

ISCHEMIC HEART DISEASE

- The term ischemic heart disease refers to disorders in coronary blood flow due to stable or unstable atherosclerotic plaques.
- Stable atherosclerotic plaques produce fixed obstruction of coronary blood flow, with myocardial ischemia occurring during periods of increased metabolic need, such as in stable angina.
- Unstable atherosclerotic plaques tend to fissure or rupture, causing platelet aggregation and potential for thrombus formation with production of a spectrum of acute coronary syndromes of increasing severity, ranging from unstable angina, to non-ST-segment elevation myocardial infarction, to ST-segment elevation myocardial infarction.



■ FIGURE 17-8 ■ Areas of pain due to angina.

Stable Angina

The term *angina* is derived from a Latin word meaning "to choke." Angina pectoris is characterized by recurring episodes of chest pain or pressure sensation associated with transient myocardial ischemia. Stable angina is the initial manifestation of ischemic heart disease in approximately half of persons with CHD.⁷ Although most persons with stable angina have atherosclerotic heart disease; angina does not develop in a considerable number of persons with advanced coronary atherosclerosis. This probably is because of their sedentary lifestyle, the development of adequate collateral circulation, or the inability of these persons to perceive pain. In many instances, myocardial infarction occurs without a history of angina.

Stable angina is usually precipitated by situations that increase the work demands of the heart, such as physical exertion, exposure to cold, and emotional stress. The pain typically is described as a constricting, squeezing, or suffocating sensation. It usually is steady, increasing in intensity only at the onset and end of the attack. The pain of angina commonly is located in the precordial or substernal area of the chest; it is similar to myocardial infarction in that it may radiate to the left shoulder, jaw, arm, or other areas of the chest (Fig. 17-8). In some persons, the arm or shoulder pain may be confused with arthritis; in others, epigastric pain is confused with indigestion. Angina commonly is categorized according to whether it occurs with exercise or during rest, is of new onset, or is of increasing severity.

The diagnosis of stable angina is based on a detailed pain history and the presence of risk factors. Typically, chronic stable angina is provoked by exertion or emotional stress and relieved within minutes by rest or the use of nitroglycerin. A delay of more than 5 to 10 minutes before relief is obtained suggests that the symptoms are not caused by ischemia or that they are caused by severe ischemia.⁸ Angina that occurs at rest, is of new onset, or is increasing in intensity or duration denotes an increased risk for myocardial infarction and should be evaluated using the criteria for acute coronary syndromes. Noncoronary causes of chest pain, such as that caused by esophageal or musculoskeletal disorders, are excluded. ECG, echocardiography, exercise stress testing or pharmacologic imaging studies, and coronary angiography may be used to confirm the diagnosis and describe the type of angina (exercise vs. vasospastic).

The treatment goals for stable angina are directed toward prevention of myocardial infarction and symptom reduction.⁷ Both nonpharmacologic and pharmacologic treatment methods are used. Nonpharmacologic methods are aimed at symptom control and lifestyle modifications to lower risk factors for coronary disease. They include smoking cessation in persons who smoke, stress reduction, a regular exercise program, limiting dietary intake of cholesterol and saturated fats, weight reduction if obesity is present, and avoidance of cold or other stresses that produce vasoconstriction. Immediate cessation of activity often is sufficient to abort an anginal attack. Sitting down or standing quietly may be preferable to lying down because these positions decrease preload by producing pooling of blood in the lower extremities.

Pharmacologic methods include the use of antiplatelet drugs, β -adrenergic blocking drugs, calcium channel blockers, and/or long-acting nitrates. Sublingual nitroglycerin or nitroglycerin spray may be prescribed for immediate relief of anginal pain. Coronary artery bypass surgery or percutaneous transluminal coronary angioplasty may be indicated in persons with significant coronary artery occlusion (to be discussed).

Variant Angina

The syndrome of variant angina or *Prinzmetal's angina* was first described by the American cardiologist Myron Prinzmetal and associates in 1959.⁹ Subsequent evidence indicated variant angina is caused by spasms of the coronary arteries, so the condition is often referred to as *vasospastic angina*.¹⁰ In most instances, the spasms occur in the presence of coronary artery stenosis; however, variant angina has occurred in the absence of visible disease.

298 Unit Four: Alterations in the Cardiovascular System

Unlike stable angina that occurs with exertion or stress, variant angina usually occurs during rest or with minimal exercise and frequently occurs during sleep. It commonly follows a cyclic or regular pattern of occurrence (*e.g.*, it happens at the same time each day). Dysrhythmias often occur when the pain is severe, and most persons are aware of their presence during an attack. ECG changes are significant if recorded during an attack. Persons with variant angina who have serious dysrhythmias during spontaneous episodes of pain are at a higher risk of sudden death.

Persons with variant angina usually experience response to treatment with calcium antagonists. These agents, along with short- and long-term nitrates, are the mainstay of treatment of variant angina. Because the two drugs act through different mechanisms, their beneficial effects may be additive.

Silent Myocardial Ischemia

Silent myocardial ischemia occurs in the absence of anginal pain. The factors that cause silent myocardial ischemia appear to be the same as those responsible for angina: impaired blood flow from the effects of coronary atherosclerosis or vasospasm. Silent myocardial ischemia affects three populations: persons who are asymptomatic without other evidence of CHD, persons who have had a myocardial infarct and continue to have episodes of silent ischemia, and persons with angina who also have episodes of silent ischemia.11 The reason for the painless episodes of ischemia is unclear. The episodes may be shorter and involve less myocardial tissue than those producing pain. Another explanation is that persons with silent angina have defects in pain threshold or pain transmission, or autonomic neuropathy with sensory denervation. There is evidence of an increased incidence of silent myocardial ischemia in persons with diabetes mellitus, probably the result of autonomic neuropathy, which is a common complication of diabetes.12

Acute Coronary Syndromes

The term *acute coronary syndromes* (ACS) has recently been accepted to describe the spectrum of acute ischemic heart diseases that include (1) unstable angina, (2) non–ST-segment elevation (non–Q-wave) myocardial infarction, and (3) ST-segment elevation (Q-wave) myocardial infarction^{13,14} (Fig. 17-9).

Persons with an ACS are routinely classified as low risk or high risk based on presenting characteristics, ECG variables, serum cardiac markers, and the timing of presentation. Persons with ST-segment elevation on ECG are usually found to have complete coronary occlusion on angiography, and many ultimately have Q-wave myocardial infarction. Persons without ST-segment elevation are those in which coronary occlusion is subtotal or intermittent.

Unstable Angina/Non–ST-Segment Elevation Myocardial Infarction

Unstable angina is characterized by symptoms at rest (usually prolonged, *i.e.*, >20 minutes); new-onset (<2 months) exertional angina; or recent (<2 months) acceleration in angina severity.¹⁵ It is considered to be a clinical syndrome of myocardial ischemia ranging between stable angina and myocardial infarction. Unlike chronic stable angina, which is caused by a fixed obstruction, unstable angina most frequently results from atherosclerotic plaque disruption and repair. Unstable angina

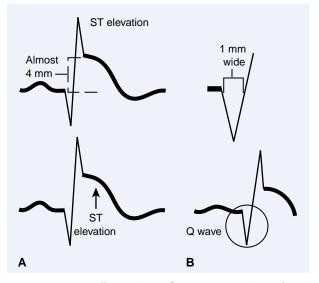


FIGURE 17-9 Illustration of an ECG tracing showing ST-segment elevation (\mathbf{A}) and Q wave in acute coronary syndromes (\mathbf{B}) .

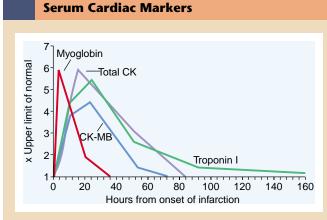
and non–ST-segment elevation myocardial infarction are similar conditions but have different severity.¹⁵ They differ primarily in whether the ischemia is severe enough to cause sufficient myocardial damage to release detectable quantities of serum cardiac markers.

The diagnosis of unstable angina/non–ST-segment elevation myocardial infarction is based on pain severity and presenting symptoms, hemodynamic stability, ECG findings, and serum cardiac markers (see Box 17-1). Persons who have no serum markers for myocardial damage are considered to have unstable angina, whereas those with serum markers are considered to have no–ST-segment elevation myocardial infarction. When chest pain has been unremitting for longer than 20 minutes, the possibility of ST-segment elevation myocardial infarction usually is considered.¹³

ST-Segment Elevation Myocardial Infarction

Acute myocardial infarction (AMI), also known as a heart attack, is characterized by the ischemic death of myocardial tissue associated with atherosclerotic disease of the coronary arteries. Elevation of the ST segment usually indicates acute myocardial injury. When the ST segment is elevated without associated Q waves, it is called a *non–Q-wave infarction*. A non–Q-wave infarction usually represents a small infarct that may evolve into a larger infarct. The area of infarction is determined by the coronary artery that is affected and by its distribution of blood flow. Approximately 30% to 40% of infarcts affect the right coronary artery, 40% to 50% affect the left anterior descending artery, and the remaining 15% to 20% affect the left circumflex artery.⁴

The onset of AMI usually is abrupt, with pain as the significant symptom. The pain typically is severe and crushing, often described as being constricting, suffocating, or like "someone sitting on my chest." It usually is substernal, radiating to the left arm, neck, or jaw, although it may be experienced in other areas of the chest. Unlike that of angina, the pain associated



The relative timing, rate of rise, peak values, and duration of cardiac marker elevation above the upper limit of normal for multiple serum markers following AMI. (Modified from Antman E.M. [1994]. General hospital management. In Julian D.G. & Braunwald E. [Eds.], *Management of acute myocardial infarction* [p. 63]. London: W.B. Saunders Ltd.).

Myoglobin is an oxygen-carrying protein, similar to hemoglobin, that is normally present in cardiac and skeletal muscle. It is a small molecule that is released quickly from infarcted myocardial tissue and becomes elevated within 1 hour after myocardial cell death, with peak levels reached within 4 to 8 hours. Because myoglobin is present in both cardiac and skeletal muscle, it is not cardiac specific.

- Creatine kinase (CK), formerly called creatinine phosphokinase, is an intracellular enzyme found in muscle cells. Muscles, including cardiac muscle, use adenosine triphosphate (ATP) as their energy source. Creatine, which serves as a storage form of energy in muscle, uses CK to convert ADP to ATP. CK exceeds normal range within 4 to 8 hours of myocardial injury and declines to normal within 2 to 3 days. There are three isoenzymes of CK, with the MB isoenzyme (CK-MB) being highly specific for injury to myocardial tissue.
- The troponin complex consists of three subunits (i.e., troponin C, troponin I, and troponin T) that regulate calcium-mediated contractile process in striated muscle. These subunits are released during myocardial infarction. Cardiac muscle forms of both troponin T and troponin I are used in diagnosis of myocardial infarction. Troponin I (and troponin T; not shown) rises more slowly than myoglobin and may be useful for diagnosis of infarction, even up to 3 to 4 days after the event. It is thought that cardiac troponin assays are more capable of detecting episodes of myocardial infarction in which cell damage is below that detected by CK-MB level.

with myocardial infarction is more prolonged and not relieved by rest or nitroglycerin, and narcotics frequently are required. Women often experience atypical ischemic-type chest discomfort, whereas the elderly may report shortness of breath more frequently than chest pain.¹⁶

Gastrointestinal complaints are common. There may be a sensation of epigastric distress; nausea and vomiting may occur. Reports of fatigue and weakness, especially of the arms and legs, are common. Pain and sympathetic stimulation combine to give rise to tachycardia, anxiety, restlessness, and feelings of impending doom. The skin often is pale, cool, and moist. The impaired myocardial function may lead to hypotension and shock.

Sudden death from AMI is death that occurs within 1 hour of symptom onset. It usually is attributed to fatal dysrhythmias, which may occur without evidence of infarction. Approximately 30% to 50% of persons with AMI die of ventricular fibrillation within the first few hours after symptoms begin.¹⁶ Early hospitalization after onset of symptoms greatly improves chances of averting sudden death because appropriate resuscitation facilities are immediately available when potentially fatal ventricular dysrhythmias occur.

Pathologic Changes. The extent of the infarct depends on the location and degree of occlusion, amount of heart tissue supplied by the vessel, duration of the occlusion, metabolic needs of the affected tissue, extent of collateral circulation, and other factors such as heart rate, blood pressure, and cardiac rhythm. A myocardial infarct may involve the endocardium, myocardium, epicardium, or a combination of these. *Transmural infarcts* involve the full thickness of the ventricular wall and most commonly occur when there is obstruction of a single artery.⁴ *Subendocardial infarcts* involve the inner one third to one half of the ventricular wall and occur more frequently in the presence of severely narrowed but still patent arteries. Most infarcts are transmural, involving the free wall of the left ventricle and often the interventricular septum (Fig. 17-10).

The principal biochemical consequence of AMI is the conversion from aerobic to anaerobic metabolism with inadequate production of energy to sustain normal myocardial function. Although gross tissue changes are not apparent for hours after onset of an AMI, the ischemic area ceases to function within a matter of minutes, and irreversible myocardial cell damage occurs after 20 to 40 minutes of severe ischemia.⁴

The term *reperfusion* refers to re-establishment of blood flow through use of thrombolytic therapy or revascularization procedures. Early reperfusion (within 15 to 20 minutes) after onset of ischemia can prevent necrosis. Reperfusion after a longer interval can salvage some of the myocardial cells that would have died because of longer periods of ischemia. Even though much of the viable myocardium existing at the time of reflow ultimately recovers, critical abnormalities in biochemical function may persist, causing impaired ventricular function. The recovering area of the heart is often referred to as a *stunned myocardium*. Because myocardial function is lost before cell death occurs, a stunned myocardium may not be capable of sustaining life, and persons with large areas of dysfunctional myocardium may require life support until the stunned regions regain their function.⁴



■ FIGURE 17-10 ■ Acute myocardial infarct. A cross-section of the ventricles of a man who died a few days after the onset of severe chest pain shows a transmural infarct in the posterior and septal regions of the left ventricle. The necrotic myocardium is soft, yellowish, and sharply demarcated. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 558]. Philadelphia: Lippincott Williams & Wilkins)

Myocardial cells that undergo necrosis are gradually replaced with scar tissue. An acute inflammatory response develops in the area of necrosis approximately 2 to 3 days after infarction. Thereafter, macrophages begin removing the necrotic tissue; the damaged area is gradually replaced with an ingrowth of highly vascularized granulation tissue, which gradually becomes less vascular and more fibrous.⁴ At approximately 3 to 7 days, the center of the infarcted area is soft and yellow; if rupture of the ventricle, interventricular septum, or valve structures occurs, it usually happens at this time. Replacement of the necrotic myocardial tissue usually is complete by the seventh week. Fibrous scar tissue lacks the contractile, elastic, and conductive properties of normal myocardial cells.

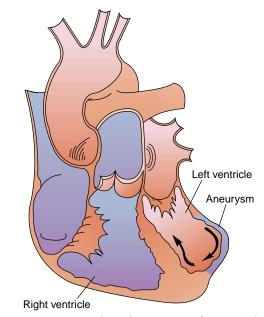
Complications. The size and location of the infarction determine the acute course, clinical complications, and long-term prognosis. Among the early complications of AMI are life-threatening dysrhythmias, sudden death, heart failure, and cardiogenic shock (see Chapter 18). Recurrent infarction may occur in the region of infarction (infarct extension). It may be associated with prolonged or intermittent episodes of chest pain.

Other complications of AMI include pericarditis, development of thromboemboli, rupture of the heart, and formation of ventricular aneurysms. Pericarditis may complicate the course of AMI. It usually appears on the second or third day after infarction. Dressler's syndrome describes the signs and symptoms associated with pericarditis, pleurisy, and pneumonitis: fever, chest pain, dyspnea, and abnormal laboratory test results (*i.e.*, elevated white blood cell count and sedimentation rate) and ECG findings. The symptoms may arise between 1 day and several weeks after infarction and are thought to represent a hypersensitivity response to tissue necrosis. Thromboemboli are a potential complication of AMI, arising as venous thrombi or occasionally as clots from the wall of the ventricle. Immobility and impaired cardiac function contribute to stasis of blood in the venous system. Infrequent, but dreaded, complications of AMI include rupture of the myocardium, the interventricular septum, or a papillary muscle. Complete rupture of the left ventricular wall, which usually occurs 3 to 7 days after AMI when the injured ventricular tissue is soft and weak, usually results in instant death.

An aneurysm is an outpouching of the ventricular wall that develops in 10% to 20% of persons surviving AMI, usually those with anterior Q-wave infarctions.¹⁶ Scar tissue does not have the characteristics of normal myocardial tissue; when a large section of ventricular muscle is replaced by scar tissue, an aneurysm may develop (Fig. 17-11). This section of the myocardium does not contract with the rest of the ventricle during systole. Instead, it diminishes the pumping efficiency of the heart and increases the work of the left ventricle, predisposing the patient to heart failure. Ischemia in the surrounding area predisposes the patient to development of dysrhythmias, and stasis of blood in the aneurysm can lead to thrombus formation. Surgical resection may be performed to improve ventricular function.

Diagnosis and Treatment. Diagnosis of AMI is based on presenting signs and symptoms, ECG charges, and serum cardiac markers. ECG changes may not be present immediately after the onset of symptoms, except as dysrhythmias. The occurrence of dysrhythmias and conduction defects depends on the areas of the heart and conduction pathways that are included in the infarct. Typical ECG changes include ST-segment elevation, prolongation of the Q wave, and inversion of the T wave.

The treatment of ACS depends on the extent of ischemia and/or infarction. Because the specific diagnosis of AMI is often difficult to make at the time of entry into the health care system, the immediate management of all ACS is generally



■ FIGURE 17-11 ■ Paradoxical movement of a ventricular aneurysm during systole.

the same. The American College of Cardiology/American Heart Association (ACC/AHA) Task Force guidelines for management of AMI recommend that the initial emergency department management of myocardial infarction include administration of oxygen by nasal prongs; sublingual nitroglycerin; adequate analgesia; and aspirin (160 to 325 mg).¹⁶ ECG monitoring should be instituted, and a 12-lead electrocardiogram should be performed. The severe pain of AMI gives rise to anxiety and recruitment of autonomic nervous system responses, both of which increase the work demands of the heart. Morphine is often given intravenously for pain relief because it has a rapid onset of action and the intravenous route does not elevate serum enzymes. Aspirin is given for its antiplatelet effects.

Immediate reperfusion therapy, using thrombolytic agents or revascularization procedures, is usually indicated for persons with ECG evidence of infarction. The best results occur if treatment is initiated within 60 to 90 minutes of symptom onset.¹⁷ The magnitude of benefit declines after this period, but it is possible that some benefit can be achieved for as long as 12 hours after the onset of pain.

Thrombolytic agents dissolve blood and platelet clots and are used to reduce mortality and limit infarct size. Revascularization interventions include percutaneous transluminal coronary angioplasty (PTCA), coronary stent implantation, and coronary artery bypass surgery. PTCA involves the recanalization of a stenotic vessel using balloon dilatation (Fig. 17-12). The procedure is done under local anesthesia in the cardiac catheterization laboratory and is similar to cardiac catheterization for coronary angiography. Implantation of coronary stents (fenestrated, stainless-steel tubes) reduces the occurrence of restenosis after PTCA. A new approach to the prevention of coronary restenosis after balloon angioplasty and stent placement is the use of localized intracoronary radiation. The procedure, also known as *brachytherapy*, is credited with inhibiting cell pro-



liferation, vascular lesion formation, and prevention of constrictive arterial remodeling. The radiation source can be impregnated into stents, or the radiation can be delivered by a radiation catheter containing a sealed source of radiation (radioactive seeds, wire, or ribbon) that is inserted into the treatment site and then removed.¹⁸

Coronary artery bypass surgery remains an option for patients who have significant CHD uncontrolled by angioplasty or pharmacologic therapy. In this surgical procedure, revascularization of the myocardium is effected by placing a saphenous vein graft between the aorta and the affected coronary artery distal to the site of occlusion or by using the internal mammary artery as a means of revascularizing the left anterior descending artery or its branches (Fig. 17-13).

Rehabilitation Programs. Rehabilitation programs for persons with ACS incorporate rest, exercise, and risk factor modification. An exercise program is an integral part of a cardiac rehabilitation program. It includes activities such as walking, swimming, and bicycling. These exercises involve changes in muscle length and rhythmic contractions of muscle groups. Most exercise programs are individually designed to meet each person's physical and psychological needs. The goal of the exercise program is to increase the maximal oxygen consumption by the muscle tissues, so that these persons are able to perform more work at a lower heart rate and blood pressure. In addition to

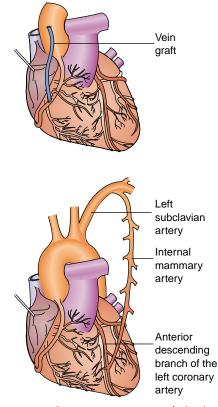


FIGURE 17-13 Coronary artery revascularization. (**Top**) Saphenous vein bypass graft. The vein segment is sutured to the ascending aorta and the right coronary artery at a point distal to the occluding lesion. (**Bottom**) Mammary artery bypass. The mammary artery is anastomosed to the anterior descending left coronary artery, bypassing the obstructing lesion.

exercise, cardiac risk factor modification incorporates strategies for smoking cessation, weight loss, stress reduction, and control of hypertension and diabetes.

In summary, CHD is a disorder of impaired coronary blood flow, usually caused by atherosclerosis. Myocardial ischemia occurs when there is a disparity between coronary blood flow and the metabolic needs of the heart. The *chronic ischemic heart diseases* include stable angina, variant angina, and silent myocardial ischemia. Stable angina is associated with a fixed atherosclerotic obstruction and pain that is precipitated by increased work demands on the heart and relieved by rest. Variant angina results from spasms of the coronary arteries. Silent myocardial ischemia occurs without symptoms. Treatment includes nonpharmacologic methods, such as pacing of activities and avoidance of activities that cause angina, and the use of pharmacologic agents, including aspirin, β -adrenergic blockers, calcium channel blockers, and nitrates.

The acute coronary syndromes result from unstable atherosclerotic plaques with plaque disruption, platelet aggregation, and thrombus formation. They include unstable angina, non–ST-segment elevation AMI, and ST-segment (Q wave) elevation AMI. Unstable angina is an accelerated form of angina in which the pain occurs more frequently, is more severe, and lasts longer than chronic stable angina. AMI refers to the ischemic death of myocardial tissue associated with obstructed blood flow in the coronary arteries. ST-segment elevation and Q wave AMI differ in terms of the extent of myocardial damage. The complications of AMI include potentially fatal dysrhythmias, heart failure and cardiogenic shock, pericarditis, thromboemboli, rupture of cardiac structures, and ventricular aneurysms. Treatment goals focus on the reestablishment of myocardial blood flow through rapid recanalization of the occluded coronary artery, prevention of clot extension through use of aspirin and other antiplatelet and antithrombotic agents, alleviation of pain, and measures such as the administration of oxygen to increase the oxygen saturation of hemoglobin. Thrombolytic agents, PTCA, and coronary artery bypass surgery are measures used to recanalize or bypass the occluded artery.

MYOCARDIAL AND ENDOCARDIAL DISEASE

Myocardial diseases, including myocarditis and the primary cardiomyopathies, are disorders originating in the myocardium, but not from cardiovascular disease. Both myocarditis and the primary cardiomyopathies are causes of sudden death and heart failure.

Myocarditis

The term *myocarditis* is used to describe an inflammation of the heart muscle and conduction system without evidence of myocardial infarction.^{19,20} Viruses are the most important cause of myocarditis in North America and Europe.²⁰ Coxsackieviruses

A and B and other enteroviruses probably account for most of the cases. Myocarditis is a frequent pathologic cardiac finding in persons with acquired immunodeficiency syndrome (AIDS), although it is unclear whether it is caused by the human immunodeficiency virus itself or by a secondary infection. Other causes of myocarditis are radiation therapy, hypersensitivity reactions, or exposure to chemical or physical agents that induce acute myocardial necrosis and secondary inflammatory changes. A drug that is increasingly associated with myocarditis is cocaine, probably because of its vasoconstrictor properties.²⁰

Myocardial injury attributable to infectious agents is thought to result from necrosis caused by direct invasion of the offending organism, toxic effects of exogenous toxins or endotoxins produced by a systemic pathogen, or destruction of cardiac tissue by immunologic mechanisms initiated by the infectious agent. The immunologic response may be directed at foreign antigens of the infectious agent that share molecular characteristics with those of the host cardiac myocardial cells (*i.e.*, molecular mimicry; see Chapter 10), providing a continuous stimulus for the immune response even after the infectious agent has been cleared from the body.

The manifestations of myocarditis vary from an absence of symptoms to profound heart failure or sudden death. When viral myocarditis occurs in children or young adults, it often is asymptomatic. Acute symptomatic myocarditis typically manifests as a flulike syndrome with malaise, low-grade fever, and tachycardia that is more pronounced than would be expected for the level of fever present. There commonly is a history of an upper respiratory tract or gastrointestinal tract infection, followed by a latent period of several days. Cardiac auscultation may reveal an S₃ ventricular gallop rhythm and a transient pericardial or pleurocardial rub. In approximately one half of the cases, myocarditis is transient, and symptoms subside within 1 to 2 months. In other cases, fulminant heart failure and life-threatening dysrhythmias develop, causing sudden death. Still others progress to subacute and chronic disease.

The diagnosis of myocarditis is based on clinical manifestations and ECG changes. Serum creatinine kinase often is elevated. Troponin T or troponin I, or both, may be elevated, providing evidence of myocardial cell damage.²⁰ Confirmation of active myocarditis requires endomyocardial biopsy.

Treatment measures focus on symptom management and prevention of myocardial damage. Bed rest is necessary, and activity restriction must be maintained until fever and cardiac symptoms subside to decrease the myocardial workload. Activity is gradually increased but kept at a sedentary level for 6 months to 1 year. The use of corticosteroids and immunosuppressant drugs remains a matter of controversy. Although treatment of myocarditis is successful in many persons, some experience congestive heart failure and can expect only a limited life span. For these persons, heart transplantation becomes an alternative.

Cardiomyopathies

The cardiomyopathies are a group of disorders that affect the heart muscle. They can develop as primary or secondary disorders. The primary cardiomyopathies, which are discussed in this chapter, are heart muscle diseases of unknown origin. Secondary cardiomyopathies are conditions in which the cardiac abnormality results from another cardiovascular disease, such as myocardial infarction. The onset of the primary cardiomyopathies often is silent, and symptoms do not occur until the disease is well advanced. The diagnosis is suspected when a young, previously healthy, normotensive person experiences cardiomegaly and heart failure.

In 1989, the International Society and Federation of Cardiology and the World Health Organization categorized the primary cardiomyopathies into three groups: dilated, hypertrophic, and restrictive²¹ (Fig. 17-14). This classification was enlarged in 1996 to include arrhythmogenic right ventricular cardiomyopathy.²² Peripartum cardiomyopathy is a disorder of pregnancy.

Dilated Cardiomyopathies

Dilated cardiomyopathies are characterized by progressive cardiac hypertrophy and dilation and impaired pumping ability of one or both ventricles. Although all four chambers of the heart are affected, the ventricles are more dilated than the atria. Because of the wall thinning that accompanies dilation, the thickness of the ventricular wall often is less than would be expected for the amount of hypertrophy present.²³ Mural thrombi are common and may be a source of thromboemboli. The cardiac valves are intrinsically normal. Microscopically, there is evidence of scarring and atrophy of myocardial cells.

Dilated cardiomyopathy may result from a number of different myocardial insults, including infectious myocarditis, alcohol and other toxic agents, metabolic influences, neuromuscular diseases, and immunologic disorders. Genetic influences have been documented in some cases. Often the cause is unknown; these cases are appropriately designated as *idiopathic dilated cardiomyopathy*.



The most common initial manifestations of dilated cardiomyopathy are those related to heart failure. There is a profound reduction in the left ventricular ejection fraction (*i.e.*, ratio of stroke volume to end-diastolic volume) to 40% or less, compared with a normal value of approximately 67%. After symptoms have developed, the course of the disorder is characterized by a worsening of heart failure, development of mural thrombi, and ventricular dysrhythmias. The most striking symptoms of dilated cardiomyopathy are dyspnea on exertion, paroxysmal nocturnal dyspnea, orthopnea, weakness, fatigue, ascites, and peripheral edema. The systolic blood pressure is normal or low, and the peripheral pulses often are of low amplitude. Pulsus alternans, in which the pulse regularly alternates between weaker and stronger volume, may be present. Tachycardia, atrial fibrillation, and complex ventricular dysrhythmias leading to sudden cardiac death are common.

The treatment of dilated cardiomyopathy is directed toward relieving the symptoms of heart failure and reducing the workload of the heart. Digoxin, diuretics, and afterload-reducing drugs are used to improve myocardial contractility and decrease left ventricular filling pressures. Avoiding myocardial depressants, including alcohol, and pacing rest with asymptomatic levels of exercise or activity is imperative. Insertion of an internal cardioverter-defibrillator may be indicated for controlling recurrent ventricular dysrhythmias. In persons with severe heart failure that is refractory to treatment, cardiac transplantation may be considered.

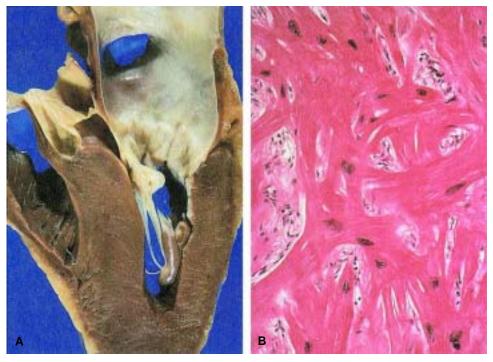
Hypertrophic Cardiomyopathies

Hypertrophic cardiomyopathy is characterized by left, right, or left and right ventricular hypertrophy and abnormal diastolic filling. Although the hypertrophy may be symmetric, the involvement of the ventricular septum often is disproportionate, producing intermittent left ventricular outflow obstruction.²⁴ Synonyms for this disorder include *idiopathic hypertrophic subaortic stenosis* and *asymmetric septal hypertrophy*.

Symptomatic hypertrophic cardiomyopathy commonly is a disease of young adulthood. The cause of the disorder is unknown, although it often is of familial origin, with the disorder being inherited as an autosomal dominant trait. Molecular studies of the genetic alterations responsible for hypertrophic cardiomyopathy suggest that the disease is caused by mutation in one of four genes encoding the proteins of the cardiac muscle fibers.²⁴ More than 50 mutations in these proteins have been identified. The prognosis of persons with different myosin mutations varies greatly; some mutations are relatively benign, whereas others are associated with premature death.

A distinctive microscopic finding in hypertrophic cardiomyopathy is myofibril disarray (Fig. 17-15). Instead of the normal parallel arrangement of myofibrils, the myofibrils branch off at random angles, sometimes at right angles to an adjacent fiber with which they connect. Small bundles of fibers may course haphazardly through normally arranged muscle fibers.^{4,25} These disordered fibers may produce abnormal movements of the ventricles, with uncoordinated contraction and impaired relaxation. Arrhythmias and premature sudden death are common with this disorder. Study results have shown that 36% of young athletes who die suddenly have probable or definite hypertrophic cardiomyopathy.¹

The manifestations of hypertrophic cardiomyopathy are variable; for reasons that are unclear, in some persons, the dis-



■ FIGURE 17-15 ■ Hypertrophic cardiomyopathy. (A) The heart has been opened to show striking asymmetric left ventricular hypertrophy. The interventricular septum is thicker than the free wall of the ventricle and impinges on the outflow tract. (B) A section of the myocardium shows myocardial fiber disarray characterized by oblique and often perpendicular orientation of adjacent hypertrophic myocytes. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 580]. Philadelphia: Lippincott Williams & Wilkins)

order remains stable for many years, with symptoms gradually increasing as the disease progresses, but others experience sudden cardiac death as first evidence of the disease.²⁴ Atrial fibrillation is a common precursor to sudden death in those who die of dysrhythmias. Dyspnea is the most common symptom associated with a gradual elevation in left ventricular diastolic pressure resulting from impaired ventricular filling and increased wall stiffness caused by ventricular hypertrophy. Because of the obstruction to outflow from the left ventricle, increasingly greater levels of ventricular pressure are needed to eject blood into the aorta, limiting cardiac output. Chest pain, fatigue, and syncope are common and worsen during exertion.

The treatment of hypertrophic cardiomyopathy includes medical and surgical management. The goal of medical management is to relieve the symptoms by lessening the pressure difference between the left ventricle and the aorta, thereby improving cardiac output. Drugs that block the β -adrenergic receptors may be used in persons with chest pain, dysrhythmias, or dyspnea.²⁴ These drugs reduce the heart rate and improve myocardial function by allowing more time for ventricular filling and reducing ventricular stiffness. The calcium channelblocking drug verapamil may be used as an alternative to the β -adrenergic blockers. Increased calcium uptake and increased intracellular calcium content are associated with an increased contractile state, a characteristic finding in patients with hypertrophic cardiomyopathy.

Surgical treatment may be used if severe symptoms persist despite medical treatment. It involves incision of the septum with or without the removal of part of the muscle tissue (*i.e.*, myectomy). It is accompanied by all the risks of open heart surgery. Implantable cardioverter-defibrillators may be used to abort lethal arrhythmias.²⁴

Restrictive Cardiomyopathies

Of the three categories of cardiomyopathies, the restrictive type is the least common in Western countries. With this form of cardiomyopathy, ventricular filling is restricted because of excessive rigidity of the ventricular walls, although the contractile properties of the heart remain relatively normal. The condition is endemic in parts of Africa, India, South and Central America, and Asia.²⁶ Outside the tropics, the most common causes of restrictive cardiomyopathy are endocardial infiltrations such as amyloidosis. Amyloid infiltrations of the heart are common in the elderly. The idiopathic form of the disorder may have a familial origin.

Symptoms of restrictive cardiomyopathy include dyspnea, paroxysmal nocturnal dyspnea, orthopnea, peripheral edema, ascites, fatigue, and weakness. The manifestations of restrictive cardiomyopathy resemble those of constrictive pericarditis. In the advanced form of the disease, all the signs of heart failure are present except cardiomegaly.

Arrhythmogenic Right Ventricular Cardiomyopathy

In arrhythmogenic right ventricular cardiomyopathy, the right ventricular myocardium is replaced with a fibrofatty deposit. This condition frequently has a familial predisposition, with an autosomal dominant inheritance pattern. Sudden death caused by arrhythmias is common, particularly in the young.²²

Peripartum Cardiomyopathy

Peripartum cardiomyopathy refers to left ventricular dysfunction developing during the last month before delivery to 5 months after delivery. The condition is relatively rare, with an estimated incidence of 1 per 3000 to 4000 live birth pregnancies.²⁷ Risk factors for peripartum cardiomyopathy include advanced maternal age, African-American race, multifetal pregnancies, preeclampsia, and gestational hypertension.²⁷ The reported mortality rate ranges from 18% to 56%. Survivors may not recover completely and may require heart transplantation.

The cause of peripartum cardiomyopathy is uncertain. A number of causes have been proposed, including myocarditis, an abnormal immune response to pregnancy, maladaptive response to the hemodynamic stresses of pregnancy, or prolonged inhibition of contractions in premature labor.

The signs and symptoms resemble those of dilated cardiomyopathy. Because many women experience dyspnea, fatigue, and pedal edema during the last month of normal pregnancy, the symptoms may be ignored and the diagnosis delayed. The diagnosis is based on echocardiography studies, ECG, and other tests of cardiac function. Treatment methods are similar to those used in dilated cardiomyopathy.

There are two possible outcomes of peripartum cardiomyopathy. In approximately one half of cases, the heart returns to normal within 6 months, and the chances for long-term survival are good. In these women, heart failure returns only during subsequent pregnancies. In the other one half of cases, cardiomegaly persists, the prognosis is poor, and death is probable if another pregnancy occurs. In women with cardiomyopathy associated with documented viral myocarditis, the likelihood of recurrence is low.

Disorders Affecting the Endocardium

Infective Endocarditis

Infective endocarditis is a relatively uncommon, life-threatening infection of the endocardial surface of the heart, including the heart valves. Because bacteria are the most common infecting organisms, the condition may be referred to as *bacterial endocarditis*. Despite important advances in antimicrobial therapy and improved ability to diagnose and treat complications, infective endocarditis continues to produce substantial morbidity and mortality.

Two factors contribute to the development of infective endocarditis: a damaged endocardial surface and a portal of entry by which the organism gains access to the circulatory system. The presence of valvular disease, prosthetic heart valves, or congenital heart defects provides an environment conducive to bacterial growth.^{4,28,29} In persons with pre-existing valvular or endocardial defects, simple gum massage or an innocuous oral lesion may afford the pathogenic bacteria access to the bloodstream. Transient bacteremia may emerge in the course of seemingly minor health problems, such as an upper respiratory tract infection, a skin lesion, or a dental procedure.

Although infective endocarditis usually occurs in persons with pre-existing heart lesions, it also can develop in normal hearts of intravenous drug abusers. The mode of infection is a contaminated drug solution or a needle contaminated with skin flora. Intravenous drug abuse is the most common source of right-sided (tricuspid) lesions. Although staphylococcal infections are common, intravenous drug users may be infected with unusual organisms, such as gram-negative bacilli, yeasts, and fungi.

In hospitalized patients, infective endocarditis may arise as a complication of infected intravascular or urinary tract catheters. Infective endocarditis also may complicate prosthetic heart valve replacement. It can develop as an early infection that follows surgery or as a later infection that results from the longterm presence of the prosthesis.

Depending on the duration of the disease, presenting manifestations, and complications, cases of infective endocarditis can be classified as acute, subacute, or chronic.³⁰ Acute infective endocarditis is thought primarily to affect persons with normal hearts and usually is caused by suppurative organisms such as Staphylococcus aureus and Streptococcus pyogenes. Subacute endocarditis is seen most frequently in patients with damaged hearts and usually is caused by less virulent organisms such as Streptococcus viridans and Staphylococcus epidermis. Certain lowvirulence organisms such as Legionella and Brucella may produce a chronic form of the disease.

Pathophysiology. The pathophysiology of infective endocarditis involves the formation of intracardiac vegetative lesions that have local and distant systemic effects. The vegetative lesion that is characteristic of infective endocarditis consists of a collection of infectious organisms and cellular debris enmeshed in the fibrin strands of clotted blood. The infectious loci continuously release bacteria into the bloodstream and are a source of persistent bacteremia. These lesions may be singular or multiple, may grow to be as large as several centimeters, and usually are found loosely attached to the free edges of the valve surface (Fig. 17-16). As the lesions grow, they cause valve destruction, leading to valvular regurgitation, ring abscesses with heart block, and valve perforation. The loose organization of these lesions permits the organisms and fragments of the lesions to form emboli and travel in the bloodstream. The fragments may lodge in small blood vessels, causing small hemorrhages, abscesses, and infarction of tissue. The bacteremia also can initiate immune responses thought to be responsible for the skin manifestations, arthritis, glomerulonephritis, and other immune disorders associated with the condition.



■ FIGURE 17-16 ■ Bacterial endocarditis. The mitral valve shows destructive vegetations, which have eroded through the free margin of the valve leaflet. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 572]. Philadelphia: Lippincott Williams & Wilkins)

306 Unit Four: Alterations in the Cardiovascular System

Clinical Course. The clinical course of infective endocarditis is determined by the extent of heart damage, the type of organism involved, site of infection (*i.e.*, right or left side of the heart), and whether embolization from the site of infection occurs. Destruction of infected heart valves is common with certain forms of organisms, such as *S. aureus*. Peripheral embolization can lead to metastatic infections and abscess formation; these are particularly serious when they affect organs such as the brain and kidneys. In right-sided endocarditis, which usually involves the tricuspid valve, septic emboli travel to the lung, causing infarction and lung abscesses.

The signs and symptoms of infective endocarditis include fever and signs of systemic infection, change in the character of an existing heart murmur, and evidence of embolic distribution of the vegetative lesions. In the acute form, the fever usually is spiking and accompanied by chills. In the subacute form, the fever usually is low grade, of gradual onset, and frequently accompanied by other systemic signs of inflammation, such as anorexia, malaise, and lethargy. Small petechial hemorrhages frequently result when emboli lodge in the small vessels of the skin, nail beds, and mucous membranes. Splinter hemorrhages (*i.e.*, dark red lines) under the nails of the fingers and toes are common. Cough, dyspnea, arthralgia or arthritis, diarrhea, and abdominal or flank pain may occur as the result of systemic emboli.

The blood culture is the most definitive diagnostic procedure and is essential in determining treatment. The optimal time to obtain cultures is during a chill, just before a temperature rise. Usually two separate positive cultures are required for diagnosis. The echocardiogram is useful in detecting underlying valvular disease. Transesophageal echocardiography is rapid and noninvasive, and has proved useful for detecting vegetations. Other signs, such as the presence of embolic disease or evidence of immunologic phenomena such as glomerulonephritis or rheumatoid factor, provide useful diagnostic information.

Treatment of infective endocarditis focuses on identifying and eliminating the causative microorganism, minimizing the residual cardiac effects, and treating the pathologic effects of the emboli. Antibiotic therapy is used to eradicate the pathogen. Surgery may be indicated for moderate to severe heart failure, progressive renal failure, significant emboli, dysrhythmias, or left-sided endocarditis. Infected prosthetic valves may need to be replaced.

Of great importance is the prevention of infective endocarditis in persons with prosthetic heart valves, previous bacterial endocarditis, certain congenital heart defects, and other known risk factors.³⁰ Prevention can be largely accomplished through prophylactic administration of an antibiotic before dental and other procedures that may cause bacteremia.³⁰

Rheumatic Heart Disease

Rheumatic fever is an acute, immune-mediated, multisystem inflammatory disease that follows a group A (β -hemolytic) streptococcal (GAS) throat infection. The most serious aspect of rheumatic fever is the development of chronic valvular disorders that produce permanent cardiac dysfunction and sometimes cause fatal heart failure years later. In the United States and other industrialized countries, the incidence of rheumatic fever and the prevalence of rheumatic heart disease has markedly declined during the past 40 to 50 years. This decline has been attributed to the introduction of antimicrobial agents for improved treatment of GAS pharyngitis, increased access to medical care, and improved economic standards, along with better and less crowded housing. Unfortunately, rheumatic fever and rheumatic heart disease continue to be major health problems in many underdeveloped countries, where inadequate health care, poor nutrition, and crowded living conditions still prevail.

Rheumatic fever is primarily a disease of school-aged children. The incidence of acute rheumatic fever peaks between 5 and 15 years of age.⁴ The disease usually follows an inciting GAS throat infection by 1 to 4 weeks. Rheumatic fever and its cardiac complications can be prevented by antibiotic treatment of the initial GAS throat infection. Evidence of a streptococcal infection is established through the use of throat cultures, antigen tests, and antibodies to products liberated by the streptococci. Throat cultures taken at the time of the acute infection usually are positive for GAS infection. It takes several days to obtain the results of a throat culture. The development of rapid tests for direct detection of GAS antigens has provided at least a partial solution for this problem. Penicillin (or another antibiotic in penicillin-sensitive patients) is the treatment of choice for GAS.

Pathogenesis. The pathogenesis of rheumatic fever is unclear, and why only a small percentage of persons with uncomplicated streptococcal infections contract rheumatic fever remains unknown. The time frame for development of symptoms in relation to the sore throat and the presence of antibodies to the GAS organism strongly suggest an immunologic origin. Like other immunologic phenomena, rheumatic fever requires an initial sensitizing exposure to the offending streptococcal agent, and the risk of recurrence is high after each subsequent exposure.

Rheumatic fever can manifest as an acute, recurrent, or chronic disorder. The acute stage of rheumatic fever includes a history of an initiating streptococcal infection and subsequent involvement of the connective tissue of the heart, blood vessels, joints, and subcutaneous tissues. Common to all is a lesion called the *Aschoff body*, which is a localized area of tissue necrosis surrounded by immune cells.⁴ The *recurrent phase* usually involves extension of the cardiac effects of the disease. The *chronic phase* of rheumatic fever is characterized by permanent deformity of the heart valves and is a common cause of mitral valve stenosis. Chronic rheumatic heart disease usually does not appear until at least 10 years after the initial attack, sometimes decades later.

Clinical Course. Most children with rheumatic fever have a history of sore throat, headache, fever, abdominal pain, nausea, vomiting, swollen glands (usually at the angle of the jaw), and other signs and symptoms of streptococcal infection. Other clinical features associated with an acute episode of rheumatic fever are related to the acute inflammatory process and the structures involved in the disease process. The course of rheumatic fever is characterized by a constellation of disorders that include carditis, migratory polyarthritis of the large joints, erythema marginatum, subcutaneous nodules, and Sydenham's chorea.

Acute rheumatic carditis, which complicates the acute phase of rheumatic fever, may progress to chronic valvular disorders. The carditis can affect the pericardium, myocardium, or endocardium, and all of these layers of the heart usually are involved. Both the pericarditis and myocarditis usually are selflimited manifestations of the acute stage of rheumatic fever.

Chapter 17: Alterations in Cardiac Function 307

The involvement of the endocardium and valvular structures produces the permanent and disabling effects of rheumatic fever. Although any of the four valves can be involved, the mitral and aortic valves are affected most often. During the acute inflammatory stage of the disease, the valvular structures become red and swollen; small vegetative lesions develop on the valve leaflets. The acute inflammatory changes gradually proceed to development of fibrous scar tissue, which tends to contract and cause deformity of the valve leaflets and shortening of the chordae tendineae. In some cases, the edges or commissures of the valve leaflets fuse together as healing occurs.

The manifestations of acute rheumatic carditis include a heart murmur in a child without a previous history of rheumatic fever, change in the character of a murmur in a person with a previous history of the disease, cardiomegaly or enlargement of the heart, friction rub or other signs of pericarditis, and congestive heart failure in a child without discernible cause.

Although not a cause of permanent disability, polyarthritis is the most common finding in rheumatic fever. The arthritis involves the larger joints, particularly the knees, ankles, elbows, and wrists, and almost always is migratory, affecting one joint and then moving to another. In untreated cases, the arthritis lasts approximately 4 weeks. A striking feature of rheumatic arthritis is the dramatic response (usually within 48 hours) to salicylates.

Erythema marginatum lesions are maplike, macular areas most commonly seen on the trunk or inner aspects of the upper arm and thigh. Skin lesions are present only in approximately 10% of patients who have rheumatic fever; they are transitory and disappear during the course of the disease. The subcutaneous nodules are 1 to 4 cm in diameter. They are hard, painless, and freely movable and usually overlie the extensor muscles of the wrist, elbow, ankle, and knee joints. Subcutaneous nodules are rare, but when present, they occur most often in persons with carditis.

Chorea (i.e., Sydenham's chorea), sometimes called St. Vitus' dance, is the major central nervous system manifestation. It is seen most frequently in girls and is the least common of the clinical manifestations. The choreic movements are spontaneous, rapid, purposeless, jerking movements that interfere with voluntary activities. Facial grimaces are common, and even speech may be affected. The chorea is self-limited, usually running its course within a matter of weeks or months.

Diagnosis and Treatment. The diagnosis of rheumatic fever is based on the Jones criteria, which were initially proposed in 1955 and revised in 1984 and 1992 by a committee of the AHA.^{31,32} The criteria group the signs and symptoms of rheumatic fever into major and minor categories. The presence of two major signs (i.e., carditis, polyarthritis, chorea, erythema marginatum, and subcutaneous nodules) or one major and two minor signs (*i.e.*, arthralgia, fever, and prolonged PR interval) accompanied by evidence of a preceding GAS infection indicates a high probability of rheumatic fever. The erythrocyte sedimentation rate, C-reactive protein, and white blood cell count commonly are used to confirm recent infection. Echocardiography/Doppler ultrasound (echo-Doppler) may be used to identify cardiac lesions in persons who do not have typical signs of cardiac involvement during an attack of rheumatic fever.32

Treatment of acute rheumatic fever is designed to control the acute inflammatory process and prevent cardiac complications and recurrence of the disease. During the acute phase, antibiotics, anti-inflammatory drugs, and selective restriction of physical activities are prescribed. Penicillin is usually the antibiotic of choice. Salicylates are used to reduce fever and relieve joint pain and swelling. A short course of corticosteroids may be used when the response to salicylates is ineffective. Because of the high risk for recurrence after subsequent GAS throat infections, treatment during the acute phase of the disease is usually followed by secondary prophylaxis using penicillin or an alternative antibiotic.³³ The duration of prophylaxis depends on whether residual valvular disease is present or absent. Usually, prophylaxis is also instituted during dental or other procedures that might provide the GAS with access to the bloodstream.

In summary, myocardial disorders represent a diverse group of disorders of myocardial muscle cells, not related to coronary artery disease. Myocarditis is an acute inflammation of cardiac muscle cells, most often of viral origin. Myocardial injury from myocarditis is thought to result from necrosis caused by direct invasion of the offending organism, toxic effects of exogenous toxins or endotoxins produced by a systemic pathogen, and destruction of cardiac tissue by immunologic mechanisms initiated by the infectious agent. Although the disease usually is benign and self-limited, it can result in sudden death or chronic heart failure, for which heart transplantation may be considered.

The cardiomyopathies represent disorders of the heart muscle. Cardiomyopathies may manifest as primary or secondary disorders such as myocardial infarction. There are four main types of primary cardiomyopathies: dilated cardiomyopathy, in which fibrosis and atrophy of myocardial cells produces progressive dilation and impaired pumping ability of the heart; hypertrophic cardiomyopathy, characterized by myocardial hypertrophy, abnormal diastolic filling, and in many cases intermittent left ventricular outflow obstruction; restrictive cardiomyopathy, in which there is excessive rigidity of the ventricular wall; and arrhythmogenic right ventricular cardiomyopathy. Peripartum cardiomyopathy occurs during pregnancy. The cause of many of the primary cardiomyopathies is unknown. The disease is suspected when cardiomegaly and heart failure develop in a young, previously healthy person.

Infective endocarditis involves the invasion of the endocardium by pathogens that produce vegetative lesions on the endocardial surface. The loose organization of these lesions permits the organisms and fragments of the lesions to be disseminated throughout the systemic circulation. Two predisposing factors contribute to the development of infective endocarditis: a damaged endocardium and a portal of entry through which the organisms gain access to the bloodstream.

Rheumatic fever, which is associated with an antecedent GAS throat infection, is an important cause of heart disease. Its most serious and disabling effects result from involvement of the heart valves. Because there is no single laboratory test, sign, or symptom that is pathognomonic of acute rheumatic fever, the Jones criteria are used to establish the diagnosis during the acute stage of the disease.

VALVULAR HEART DISEASE

The function of the heart valves is to promote directional flow of blood through the chambers of the heart. Dysfunction of the heart valves can result from a number of disorders, including congenital defects, trauma, ischemic damage, degenerative changes, and inflammation. Although any of the four heart valves can become diseased, the most commonly affected are the mitral and aortic valves. Disorders of the pulmonary and tricuspid valves are uncommon, probably because of the low pressure in the right side of the heart.

Hemodynamic Derangements

The heart valves consist of thin leaflets of tough, flexible, endothelium-covered fibrous tissue firmly attached at the base to the fibrous valve rings (see Chapter 14). Capillaries and smooth muscle are present at the base of the leaflet but do not extend up into the valve. The leaflets of the heart valves may be injured or become the site of an inflammatory process that can deform their line of closure. Healing of the valve leaflets often is associated with increased collagen content and scarring, causing the leaflets to shorten and become stiffer. The edges of the valve leaflets can heal together so that the valve does not open or close properly.

Two types of mechanical disruptions occur with valvular heart disease: narrowing of the valve opening so it does not open properly and distortion of the valve so it does not close properly (Fig. 17-17). *Stenosis* refers to a narrowing of the valve orifice and failure of the valve leaflets to open normally. Blood flow through a normal valve can increase by five to seven times the resting volume; consequently, valvular stenosis must be severe before it causes problems. Significant narrowing of the valve orifice increases the resistance to blood flow through the valve, converting the normally smooth laminar flow to a less efficient turbulent flow. This increases the volume and

KEY CONCEPTS

VALVULAR HEART DISEASE

- The heart valves determine the direction of blood flow through the heart chambers.
- Valvular heart defects exert their effects by obstructing flow of blood (stenotic valve disorder) or allowing backward flow of blood (regurgitant valve disorders).
- Stenotic valvular defects produce distention of the heart chamber that empties blood through the diseased valve and impaired filling of the chamber that receives blood that moves through the valve.
- Regurgitant valves allow blood to move back through the valve when it should be closed. This produces distention and places increased work demands on the chamber ejecting blood through the diseased valve.

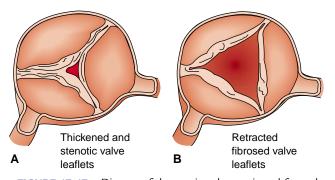


FIGURE 17-17 Disease of the aortic valve as viewed from the aorta. (A) Stenosis of the valve opening. (B) An incompetent or regurgitant valve that is unable to close completely.

work of the chamber emptying through the narrowed valve the left atrium in the case of mitral stenosis and the left ventricle in aortic stenosis. Symptoms usually are noticed first during situations of increased flow, such as exercise. An *incompetent* or *regurgitant valve* permits backward flow to occur when the valve should be closed, with blood flowing back into the left ventricle during diastole when the aortic valve is affected and back into the left atrium during systole when the mitral valve is diseased.

The effect that valvular heart disease has on cardiac function is related to alterations in blood flow across the valve and to the resultant increase in work demands on the heart that the disorder generates. Many valvular heart defects are characterized by heart murmurs resulting from turbulent blood flow through a diseased valve. Disorders in valve flow and heart chamber size for mitral and aortic valve disorders are illustrated in Figure 17-18.

Mitral Valve Disorders

The mitral valve controls the directional flow of blood between the left atrium and the left ventricle. The edges or cusps of the AV valves are thinner than those of the semilunar valves; they are anchored to the papillary muscles by the chordae tendineae. During much of systole, the mitral valve is subjected to the high pressure generated by the left ventricle as it pumps blood into the systemic circulation. During this period of increased pressure, the chordae tendineae prevent the eversion of the valve leaflets into the left atrium.

Mitral Valve Stenosis

Mitral valve stenosis represents the incomplete opening of the mitral valve during diastole with left atrial distention and impaired filling of the left ventricle. Mitral valve stenosis most commonly is the result of rheumatic fever. Less frequently, the defect is congenital and manifests during infancy or early childhood.³⁴ Mitral valve stenosis is a continuous, progressive, lifelong disorder, consisting of a slow, stable course in the early years and progressive acceleration in later years. The 10-year survival rate for persons with untreated mitral stenosis is 50% to 60%, depending on symptoms at the time of presentation.^{34,35}

Mitral valve stenosis is characterized by fibrous replacement of valvular tissue, along with stiffness and fusion of the valve apparatus (Fig. 17-19). Typically, the mitral cusps fuse at the edges, and involvement of the chordae tendineae causes short-

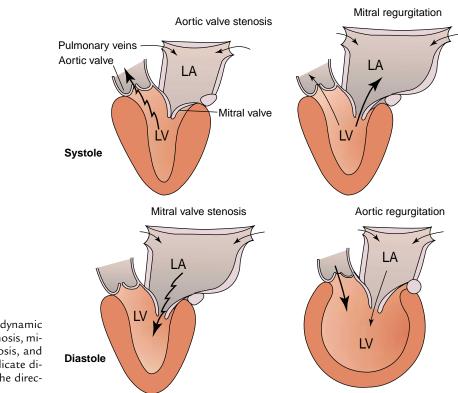


FIGURE 17-18 Alterations in hemodynamic function that accompany aortic valve stenosis, mitral valve regurgitation, mitral valve stenosis, and aortic valve regurgitation. *Thin arrows* indicate direction of normal flow, and *thick arrows* the direction of abnormal flow.

ening, which pulls the valvular structures more deeply into the ventricles. As the resistance to flow through the valve increases, the left atrium becomes dilated and left atrial pressure rises (see Fig. 17-18). The increased left atrial pressure eventually is transmitted to the pulmonary venous system, causing pulmonary congestion.

The rate of flow across the valve depends on the size of the valve orifice, the driving pressure (*i.e.*, atrial minus ventricular pressure), and the time available for flow during diastole. As the condition progresses, symptoms of decreased cardiac out-



■ FIGURE 17-19 ■ Chronic rheumatic valvulitis. A view of the mitral valve from the left atrium shows rigid, thickened, and fused leaflets with a narrow orifice, creating the characteristic "fish mouth" appearance of the rheumatic mitral stenosis. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 570]. Philadelphia: Lippincott Williams & Wilkins)

put occur during extreme exertion or other situations that cause tachycardia and thereby reduce diastolic filling time. In the late stages of the disease, an increase in pulmonary vascular resistance leads to the development of pulmonary hypertension; this increases the pressure against which the right heart must pump and eventually leads to right-sided heart failure.

The signs and symptoms of mitral valve stenosis depend on the severity of the obstruction and are related to the elevation in left atrial pressure and pulmonary congestion, decreased cardiac output caused by impaired left ventricular filling, and left atrial enlargement with development of atrial arrhythmias and mural thrombi. The murmur of mitral valve stenosis is heard during diastole when blood is flowing through the constricted valve orifice; it is characteristically a low-pitched, rumbling murmur, best heard at the apex of the heart.

The symptoms are those of pulmonary congestion, including recurrent nocturnal dyspnea and orthopnea. Palpitations, chest pain, weakness, and fatigue are common complaints. Premature atrial beats, paroxysmal atrial tachycardia, and atrial fibrillation may occur as a result of distention of the left atrium. Atrial fibrillation develops in 30% to 40% of persons with symptomatic mitral stenosis.³⁴ Together, the fibrillation and distention predispose to mural thrombus formation. The risk of arterial embolization, particularly stroke, is significantly increased in persons with atrial fibrillation. Anticoagulation therapy is often used to prevent systemic embolization in persons with atrial fibrillation.

Mitral Valve Regurgitation

Mitral valve regurgitation is characterized by incomplete closure of the mitral valve, with the left ventricular stroke volume being divided between the forward stroke volume that moves

310 Unit Four: Alterations in the Cardiovascular System

into the aorta and the regurgitant stroke volume that moves back into the left atrium during systole (see Fig. 17-18). Mitral valve regurgitation can result from many processes. Rheumatic heart disease is associated with a rigid and thickened valve that does not open or close completely. In addition to rheumatic disease, mitral regurgitation can occur as the result of papillary muscle dysfunction, stretching of the valve structures caused by dilatation of the left ventricle or valve orifice, or mitral valve prolapse. Acute mitral regurgitation can result from rupture of the chordae tendineae or papillary muscles, most commonly the result of myocardial infarction.

The hemodynamic changes that occur with mitral valve regurgitation occur more slowly, allowing for recruitment of compensatory mechanisms. An increase in left ventricular enddiastolic volume permits an increase in total stroke volume, with restoration of forward flow into the aorta. Augmented preload and reduced or normal afterload (provided by unloading the left ventricle into the left atrium) facilitates left ventricular ejection. At the same time, a gradual increase in left atrial size allows for accommodation of the regurgitant volume at a lower filling pressure.

The increased volume work associated with mitral regurgitation is relatively well tolerated, and many persons with the disorder remain asymptomatic for 10 to 20 years despite severe regurgitation.^{34,35} The degree of left ventricular enlargement reflects the severity of regurgitation. As the disorder progresses, left ventricular function becomes impaired, the forward (aortic) stroke volume decreases, and the left atrial pressure increases, with the subsequent development of pulmonary congestion. Mitral regurgitation, like mitral stenosis, predisposes to atrial fibrillation.

Mitral Valve Prolapse

Sometimes referred to as the *floppy mitral valve syndrome*, mitral valve prolapse occurs in 2% to 6% of the population.³⁶ The disorder is seen more frequently in women than in men and may have a familial basis. Although the cause of the disorder usually is unknown, it has been associated with Marfan's syndrome, osteogenesis imperfecta, and other connective tissue disorders and with cardiac, hematologic, neuroendocrine, metabolic, and psychological disorders.

Pathologic findings in persons with mitral valve prolapse include a myxedematous (mucinous) degeneration of mitral valve leaflets that causes them to become enlarged and floppy so that they prolapse or balloon back into the left atrium during systole (Fig. 17-20).²⁵ Secondary fibrotic changes reflect the stresses and injury that the ballooning movements impose on the valve. Certain forms of mitral valve prolapse may arise from disorders of the myocardium that result in abnormal movement of the ventricular wall or papillary muscle; this places undue stress on the mitral valve.

Most persons with mitral valve prolapse are asymptomatic, and the disorder is discovered during a routine physical examination. A minority of persons have chest pain mimicking angina, dyspnea, fatigue, anxiety, palpitations, and lightheadedness. Unlike angina, the chest pain often is prolonged, ill defined, and not associated with exercise or exertion. The pain has been attributed to ischemia resulting from traction of the prolapsing valve leaflets. The anxiety, palpitations, and dysrhythmias may result from abnormal autonomic nervous



■ FIGURE 17-20 ■ Mitral valve prolapse. A view of the mitral valve from the left atrium shows redundant and deformed leaflets that billow into the left atrial cavity. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 574]. Philadelphia: Lippincott Williams & Wilkins)

system function that commonly accompanies the disorder. Rare cases of sudden death have been reported for persons with mitral valve prolapse, mainly those with a family history of similar occurrences.

The treatment of mitral valve prolapse focuses on the relief of symptoms and the prevention of complications. Persons with palpitations and mild tachyarrhythmias or increased adrenergic symptoms and those with chest discomfort, anxiety, and fatigue often have response to therapy with the β adrenergic–blocking drugs. In many cases, the cessation of stimulants such as caffeine, alcohol, and cigarettes may be sufficient to control symptoms. Infective endocarditis is an uncommon complication in persons with a murmur; antibiotic prophylaxis usually is recommended before dental or surgical procedures associated with bacteremia. Persons with severe valve dysfunction may require valve surgery.

Aortic Valve Disorders

The aortic valve is located between the aorta and left ventricle. The aortic valve has three cusps and sometimes is referred to as the *aortic semilunar valve* because its leaflets are crescent or moon shaped (see Chapter 14, Fig. 14-10). The aortic valve has no chordae tendineae. Although their structures are similar, the cusps of the aortic valve are thicker than those of the mitral valve. The middle layer of the aortic valve is thickened near the middle, where the three leaflets meet, ensuring a tight seal. Between the thickened tissue and their free margins, the leaflets are more thin and flimsy.

An important aspect of the aortic valve is the location of the orifices for the two main coronary arteries, which are located behind the valve and at right angles to the direction of blood flow. It is the lateral pressure in the aorta that propels blood into the coronary arteries. During the ejection phase of the cardiac cycle, the lateral pressure is diminished by conversion of potential energy to kinetic energy as blood moves forward into the aorta. This process is grossly exaggerated in aortic stenosis because of the high flow velocities.

Aortic Valve Stenosis

Aortic stenosis is characterized by increased resistance to ejection of blood from the left ventricle into the aorta (see Fig. 17-18). Because of the increased resistance, the work demands on the left ventricle are increased, and the volume of blood ejected into the systemic circulation is decreased. The most common causes of aortic stenosis are rheumatic fever and congenital valve malformations. Congenital malformations may result in unicuspid, bicuspid, or misshaped valve leaflets. In elderly persons, stenosis may be related to degenerative atherosclerotic changes of the valve leaflets. Approximately 25% of persons older than 65 years and 35% of those older than 70 years of age have echocardiographic evidence of sclerosis, with 2% to 3% having evidence of aortic stenosis.³⁵

The progression of aortic stenosis varies widely among individuals. The progression may be more rapid in persons with degenerative calcific disease than in those with congenital or rheumatic disease.³⁵ The aortic valve must be reduced to approximately one fourth its normal size before critical changes in cardiac function occur.³⁴ Significant obstruction to aortic outflow causes a decrease in stroke volume, along with a reduction in systolic blood pressure and pulse pressure. Because of the narrowed valve opening, it takes longer for the heart to eject blood; the heart rate often is slow, and the pulse is of low amplitude. There is a soft, absent, or paradoxically split S₂ sound and a harsh systolic ejection murmur that is heard best along the left sternal border.

Persons with aortic stenosis tend to be asymptomatic for many years despite severe obstruction. Eventually, symptoms of angina, syncope, and heart failure develop. Angina occurs in approximately two thirds of persons with advanced aortic stenosis and is similar to that observed in CHD. Syncope (fainting) is most commonly caused by the reduced cerebral circulation that occurs during exertion when the arterial pressure declines consequent to vasodilation in the presence of a fixed cardiac output. Exertional hypotension may cause "graying out" spells or dizziness on exercise.³⁶ Dyspnea, marked fatigability, peripheral cyanosis, and other signs of low-output heart failure usually are not prominent until late in the course of the disease.

Aortic Valve Regurgitation

Aortic regurgitation is the result of an incompetent aortic valve that allows blood to flow back to the left ventricle during diastole (see Fig. 17-18). As a result, the left ventricle must increase its stroke volume to include blood entering from the lungs as well as that leaking back through the regurgitant valve. This defect may result from conditions that cause scarring of the valve leaflets or from enlargement of the valve orifice to the extent that the valve leaflets no longer meet. Rheumatic fever ranks first on the list of causes of aortic regurgitation; failure of a prosthetic valve is another cause.

Chronic aortic regurgitation, which usually has a gradual onset, represents a condition of combined left ventricular volume and pressure overload. As the valve deformity increases, regurgitant flow into the left ventricle increases, diastolic blood pressure falls, and the left ventricle progressively enlarges. Hemodynamically, the increase in left ventricular volume results in the ejection of a large stroke volume that usually is adequate to maintain the cardiac output until late in the course of the disease. Most persons remain asymptomatic during this compensated phase, which may last decades. The only sign for many years may be soft systolic aortic murmur.

As the disease progresses, signs and symptoms of left ventricular failure begin to appear. These include exertional dyspnea, orthopnea, and paroxysmal nocturnal dyspnea. In aortic regurgitation, failure of aortic valve closure during diastole causes an abnormal drop in diastolic pressure. Because coronary blood flow is greatest during diastole, the drop in diastolic pressure produces a decrease in coronary perfusion. Although angina is rare, it may occur when the heart rate and diastolic pressure fall to low levels. Persons with severe aortic regurgitation often report an uncomfortable awareness of heartbeat, particularly when lying down, and chest discomfort caused by pounding of the heart against the chest wall. Tachycardia, occurring with emotional stress or exertion, may produce palpitations, head pounding, and premature ventricular contractions.

The major physical findings relate to the widening of the arterial pulse pressure. The pulse has a rapid rise and fall, with an elevated systolic pressure and low diastolic pressure caused by the large stroke volume and rapid diastolic runoff of blood back into the left ventricle. Korotkoff sounds may persist to zero, even though intra-arterial pressure rarely falls below 30 mm Hg.³⁴ The large stroke volume and wide pulse pressure may result in prominent carotid pulsations in the neck, throbbing peripheral pulses, and a left ventricular impulse that causes the chest to move with each beat. The turbulence of flow across the aortic valve during diastole produces a high-pitched or blowing sound.

Diagnosis and Treatment

Valvular defects usually are detected through cardiac auscultation (*i.e.*, heart sounds). Diagnosis is aided by echocardiography and cardiac catheterization. Echocardiography uses ultrasound signals in the range of 2 to 5 million Hz to create an image of the internal structures of the heart because the chest wall, blood, and different heart structures all reflect ultrasound differently. The echocardiogram is useful for determining ventricular dimensions and valve movements, obtaining data on the movement of the left ventricular wall and septum, estimating diastolic and systolic volumes, and viewing the motion of individual segments of the left ventricular wall during systole and diastole. Transesophageal echocardiography is particularly useful in assessing valve function.

The treatment of valvular defects consists of medical management of heart failure and associated problems and surgical intervention to repair or replace the defective valve. Surgical valve repair or replacement depends on the valve that is involved and the extent of deformity. Valvular replacement with a prosthetic device usually is reserved for severe disease. Percutaneous balloon valvuloplasty involves the opening of a stenotic valve by guiding an inflated balloon through the valve orifice. The procedure is done in the cardiac catheterization laboratory and involves the insertion of a balloon catheter into the heart by way of a peripheral blood vessel. In summary, dysfunction of the heart valves can result from a number of disorders, including congenital defects, trauma, ischemic heart disease, degenerative changes, and inflammation. Rheumatic endocarditis is a common cause. Valvular heart disease produces its effects through disturbances of blood flow. A stenotic valvular defect is one that causes a decrease in blood flow through a valve, resulting in impaired emptying and increased work demands on the heart chamber that empties blood across the diseased valve. A regurgitant valvular defect permits the blood flow to continue when the valve is closed. Valvular heart disorders produce blood flow turbulence and often are detected through cardiac auscultation.

HEART DISEASE IN INFANTS AND CHILDREN

Heart disease in infants and children encompasses both congenital and acquired disorders. About 40,000 infants are born each year with a congenital heart defect, and 25% of these have defects that are severe enough to cause death within the first year if not corrected.¹ Premature infants have a higher incidence of congenital heart defects, most commonly patent ductus arteriosus and atrial septal defects. Advances in diagnostic methods and surgical treatment have greatly increased the long-term survival and outcomes for children born with congenital heart defects. This section of the chapter provides a discussion of the fetal and perinatal circulations; congenital heart disorders, and Kawasaki's disease, an acquired heart disorder of young children.

The major development of the fetal heart occurs between the fourth and seventh weeks of gestation, and during this time, most congenital heart defects arise. The development of the heart may be altered by environmental, genetic, and chromosomal influences. Most congenital heart defects are thought to be multifactorial in origin, resulting from an interaction between a genetic predisposition to develop a heart defect and environmental influences. Infants born to parents with congenital heart defects or with siblings who have congenital heart defects are at higher risk. A number of chromosomal abnormalities are associated with congenital heart disease, most prominently Down syndrome and Turner's syndrome (see Chapter 4). Other intrauterine factors such as maternal diabetes, congenital rubella, maternal alcohol ingestion, and treatment with anticonvulsant drugs are also associated with congenital heart disorders.37

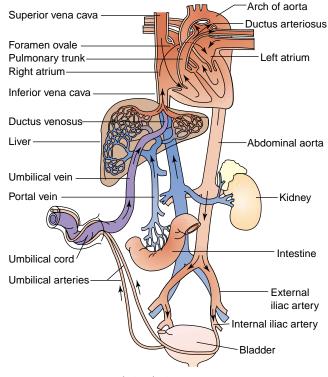
Fetal and Perinatal Circulation

The birth process produces dramatic changes in the circulation. It produces an increased risk of disorders such as patent ductus arteriosus in infants who are born prematurely, and it challenges the circulatory function in infants with congenital heart defects.

The fetal circulation is different anatomically and physiologically from the postnatal circulation. Before birth, oxygenation of blood occurs by way of the placenta, and after birth, it occurs by way of the lungs. The fetus is maintained in a lowoxygen state (PO₂ to 30 to 35 mm Hg and 60% to 70% hemoglobin saturation).³⁸ To compensate, fetal cardiac output is higher than at any other time (400 to 500 mL/kg/minute).

In the fetus, blood enters the circulation through the umbilical vein and returns to the placenta by way of the two umbilical arteries (Fig. 17-21). A vessel called the ductus venosum allows blood from the umbilical vein to bypass the hepatic circulation and pass directly into the inferior vena cava. From the inferior vena cava, blood flows into the right atrium and then is directed through the foramen ovale into the left atrium. Blood then passes into the left ventricle and is ejected into the ascending aorta to perfuse the head and upper extremities. In this way, the best-oxygenated blood from the placenta is used to perfuse the brain. At the same time, venous blood from the head and upper extremities returns to the right side of the heart by way of the superior vena cava, moves into the right ventricle, and is ejected into the pulmonary artery. The pulmonary vascular resistance is very high because the lungs are fluid filled, and the resultant alveolar hypoxia contributes to intense vasoconstriction. Because of the high pulmonary vascular resistance, the blood that is ejected into the pulmonary artery is diverted through the ductus arteriosus into the descending aorta. This blood perfuses the lower extremities and is returned to the placenta by way of the umbilical arteries.

At birth, the infant takes its first breath and switches from placental to pulmonary oxygenation of the blood. The most dramatic alterations in the circulation after birth are the elimination of the low-resistance placental vascular bed and the marked pulmonary vasodilation that is produced by initiation of ventilation. The pressure in the pulmonary circulation and the right side of the heart fall as fetal lung fluid is replaced by



■ FIGURE 17-21 ■ Fetal circulation.

air and as lung expansion decreases the pressure transmitted to the pulmonary blood vessels. With lung inflation, the alveolar oxygen tension increases, causing reversal of the hypoxemiainduced pulmonary vasoconstriction of the fetal circulation. Cord clamping and removal of the low-resistance placental circulation produce an increase in peripheral vascular resistance and a resultant increase in left ventricular pressure. The accompanying increase in left atrial pressure as compared to right atrial pressure promotes closure of the foramen ovale. Reversal of the fetal hypoxemic state also produces constriction of ductal smooth muscle, contributing to closure of the ductus arteriosus. Closure of the foramen ovale and the ductus arteriosus normally occur within the first day of life, effectively separating the pulmonary and systemic circulations.

The birth process also initiates a sequence of maturational changes in the pulmonary blood vessels and pulmonary vascular resistance. After the initial precipitous fall in pulmonary vascular resistance, a more gradual decrease in pulmonary vascular resistance is related to regression of the medial smooth muscle layer in the pulmonary arteries. During the first 2 to 9 weeks of life, gradual thinning of the medial smooth muscle layer of pulmonary arteries results in further decreases in pulmonary vascular resistance. By the time a healthy, term infant is several weeks old, the pulmonary vascular resistance has fallen to adult levels.

Several factors, including prematurity, alveolar hypoxia, lung disease, and congenital heart defects, may affect postnatal pulmonary vascular development.37 If an infant is born prematurely, the smooth muscle layers of the pulmonary vasculature may develop incompletely or regress in a shorter period. Much of the development of the smooth muscle layer in the pulmonary arterioles occurs during the latter part of gestation; as a result, infants who are born prematurely have less medial smooth muscle. These infants follow the same pattern of smooth muscle regression, but because less muscle exists, the muscle layer may regress in a shorter period. The pulmonary vascular smooth muscle in premature infants also may be less responsive to hypoxia. For these reasons, a premature infant may demonstrate a larger decrease in pulmonary vascular resistance and a resultant shunting of blood from the aorta through the ductus arteriosus to the pulmonary artery within hours of birth.

Hypoxia during the first days of life may delay or prevent the normal decrease in pulmonary vascular resistance. During this period, the pulmonary arteries remain reactive and can constrict in response to hypoxia, acidosis, hyperinflation of the alveoli, and hypothermia. Alveolar hypoxia is one of the most potent stimuli of pulmonary vasoconstriction and pulmonary hypertension in the neonate.

Congenital Heart Disorders

Congenital heart diseases are commonly classified according to their anatomic defects (atrial septal or ventricular septal defects), the hemodynamic alterations caused by the anatomic defects (left-to-right or right-to-left shunts), and their effect on tissue oxygenation (cyanotic or noncyanotic defects).^{39,40}

Shunting and Cyanotic Disorders

Shunting of blood refers to the diverting of blood flow from one system to the other—from the arterial to the venous system (*i.e.*, left-to-right shunt) or from the venous to the arterial system (*i.e.*, right-to-left shunt). The shunting of blood in congenital heart defects is determined by the presence of an abnormal opening between the right and left circulations and the degree of resistance to flow through the opening. The shunting of blood can affect both the oxygen content of the blood and the volume of blood being delivered to the vessels in the pulmonary circulation.

A *right-to-left* shunt results in unoxygenated blood moving from the right side of the heart into the left side of the heart and then being ejected into the systemic circulation. Cyanosis develops when sufficient unoxygenated blood mixes with oxygenated blood in the left side of the heart. Children with rightto-left shunts are considered to have a cyanotic heart defect, whether they have recognizable cyanosis or not. In a *left-toright* shunt, blood intended for ejection into the systemic circulation is recirculated through the right side of the heart and back through the lungs; this increased volume distends the right side of the heart and pulmonary circulation and increases the workload placed on the right ventricle. Children with leftto-right shunts are considered to have a noncyanotic heart defect, even though they are cyanotic for other reasons, such as low cardiac output.

Congenital heart defects manifest with numerous signs and symptoms. Some defects, such as patent ductus arteriosus and small ventricular septal defects, close spontaneously, and in other, less severe defects, there are no signs and symptoms. The disorder typically is discovered during a routine health examination. Pulmonary congestion, heart failure, and decreased peripheral perfusion are the chief concerns in children with more severe defects. Such defects often cause problems shortly after birth or early in infancy. The child may exhibit cyanosis, respiratory difficulty, and fatigability and is likely to have difficulty with feeding and failure to thrive. A generalized cyanosis that persists longer than 3 hours after birth suggests congenital heart disease.

One technique for evaluating the cyanosis consists of administering 100% oxygen for 10 minutes. If the infant "pinks up," the cyanosis probably was caused by a respiratory problem and not a heart deffect. Because infant cyanosis may appear as a duskiness, it is important to assess the color of the mucous membranes, fingernails, toenails, tongue, and lips.

The manifestations and treatment of heart failure in the infant and young child are similar to those in the adult, but the infant's small size and limited physical reserve makes the manifestations more serious and treatment more difficult (see Chapter 18). The treatment plan usually includes supportive therapy designed to help the infant compensate for the limitations in cardiac reserve and to prevent complications. Surgical intervention often is required for severe defects; it may be done in the early weeks of life or, conditions permitting, delayed until the child is older.

Most children with structural congenital heart disease and those who have had corrective surgery are at risk for the development of infectious endocarditis. These children should receive prophylactic antibiotic therapy during periods of increased risk of bacteremia.

Types of Defects

Congenital heart defects can affect almost any of the cardiac structures or central blood vessels. Defects include communication between heart chambers, interrupted development of the heart chambers or valve structures, malposition of heart chambers and great vessels, and altered closure of fetal communication channels. The particular defect reflects the embryo's stage of development at the time it occurred. Some congenital heart disorders, such as tetralogy of Fallot, involve several defects. At least 35 types of defects have been identified, the most common being patent ductus arteriosus (6% to 11%), atrial septal defects (8% to 13%), and ventricular septal defects (20% to 25%).¹

Patent Ductus Arteriosus. Patent ductus arteriosus results from persistence of the fetal ductus beyond the prenatal period. In fetal life, the ductus arteriosus is the vital link by which blood from the right side of the heart bypasses the lungs and enters the systemic circulation (Fig. 17-22G). After birth, this passage no longer is needed, and it usually closes during the first 24 to 72 hours. The physiologic stimulus and mechanisms associated with permanent closure of the ductus are not entirely known, but the fact that infant hypoxia predisposes to a delayed closure suggests that the increase in arterial oxygen levels that occurs immediately after birth plays a role. Additional factors that contribute to closure are a fall in endogenous levels of prostaglandins and adenosine and the release of

other vasoactive substances. After constriction, the lumen of the ductus becomes permanently sealed with fibrous tissue within 2 to 3 weeks. Ductal closure may be delayed or prevented in very premature infants, probably as a result of a combination of factors, including decreased medial muscle in the ductus wall, decreased constrictive response to oxygen, and increased circulating levels of vasodilating prostaglandins. Hemodynamically significant patent ductus arteriosus is observed in approximately one half of infants with birth weights of less than 1000 g.³⁸ Ductal closure also may be delayed in infants with congenital heart defects that produce a decrease in oxygen tension.

As is true of other heart and circulatory defects, patency of the ductus arteriosus may vary; the size of the opening may be small, medium, or large. After the infant's pulmonary vascular resistance falls, the patent ductus arteriosus provides for a continuous runoff of aortic blood into the pulmonary artery, causing a decrease in aortic diastolic and mean arterial pressure and a widening of the pulse pressure. With a large patent ductus, the runoff is continuous, resulting in increased pulmonary blood flow, pulmonary congestion, and increased resistance against which the right side of the heart must pump. Increased pul-

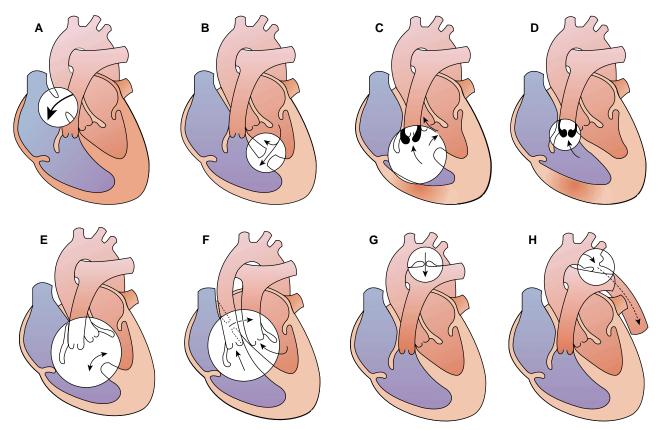


FIGURE 17-22 Congenital heart defects. (A) Atrial septal defect. Blood is shunted from left to right. (B) Ventricular septal defect. Blood is usually shunted from left to right. (C) Tetralogy of Fallot. This involves a ventricular septal defect, dextroposition of the aorta, right ventricular outflow obstruction, and right ventricular hypertrophy. Blood is shunted from right to left. (D) Pulmonary stenosis, with decreased pulmonary blood flow and right ventricular hypertrophy. (E) Endocardial cushion defects. Blood flows between the chambers of the heart. (F) Transposition of the great vessels. The pulmonary artery is attached to the left side of the heart and the aorta to the right side. (G) Patent ductus arteriosus. The high-pressure blood of the aorta is shunted back to the pulmonary artery. (H) Postductal coarctation of the aorta.

monary venous return and increased work demands may lead to left ventricular failure.

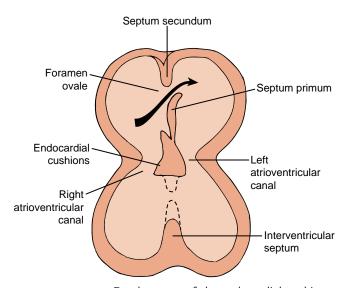
Patent ductus arteriosus can be treated either pharmacologically or surgically. Drugs that inhibit prostaglandin synthesis (*e.g.*, indomethacin), may be used to induce closure of a patent ductus arteriosus.

Atrial Septal Defects. An atrial septal defect is an abnormal opening in the atrial septum that allows communication between the left and right atrium (see Fig. 17-22A). It differs from a patent foramen ovale, which does not usually permit flow unless right atrial pressures are elevated.

Partitioning of the atria takes place during the 4th and 5th weeks of development and occurs in two stages, beginning with the formation of a thin, crescent-shaped membrane called the *septum primum* followed by the development of a second membrane called the *septum secundum*. As the septum secundum develops, it gradually overlaps an opening in the upper part of septum primum, forming an oval opening with a flap-type valve called the foramen ovale (see Fig. 17-23). The foramen ovale, which closes shortly after birth, allows blood from the umbilical vein to pass directly into the left heart, bypassing the lungs.

Atrial septal defects may be single or multiple and vary from a small, asymptomatic opening to a large, symptomatic opening. Most atrial septal defects are small and discovered inadvertently during a routine physical examination.⁴¹ In the case of an isolated septal defect that is large enough to allow shunting, the flow of blood usually is from the left to the right side of the heart because of the more compliant right ventricle and because the pulmonary vascular resistance is lower than the systemic vascular resistance. This produces right ventricular volume overload and increased pulmonary blood flow.

Young children with atrial septal defects are usually asymptomatic but experience symptoms later in life, usually during



■ FIGURE 17-23 ■ Development of the endocardial cushions, right and left atrioventricular canals, the interventricular septum, and septum primum and septum secundum of the foramen ovale. Note that blood from the right atrium flows through the foramen ovale to the left atrium. *Dotted line* indicates site for development of the membranous intraventricular septum.

adolescence when the changes in pulmonary vasculature may reverse the direction flow through the defect and create a rightto-left shunt with development of cyanosis. Adolescents and young adults may experience atrial fibrillation or atrial flutter and palpations because of atrial dilation. Symptomatic defects are usually treated surgically.

Ventricular Septal Defects. A ventricular septal defect is an opening in the ventricular septum that results from an imperfect separation of the ventricles during early fetal development (see Fig. 17-22B). Ventricular septal defects are the most common form of congenital heart defect, accounting for 20% to 25% of congenital heart disorders.³⁸ Ventricular septal defects may be the only cardiac defect, or they may be one of multiple cardiac anomalies.

The ventricular septum originates from two sources: the interventricular groove of the folded tubular heart that gives rise to the muscular part of the septum, and the endocardial cushions that extend to form the membranous portion of the septum (Fig. 17-23). The upper membranous portion of the septum is the last area to close, and it is here that most defects occur.

Depending on the size of the opening, the signs and symptoms of a ventricular septal defect may range from an asymptomatic murmur to congestive heart failure. If the defect is small, it allows a small shunt and small increases in pulmonary blood flow. These defects produce few symptoms, and approximately one third close spontaneously.³⁷ With mediumsize defects, a larger shunt occurs, producing a larger increase in pulmonary blood flow (*i.e.*, twice as much blood may pass through the pulmonary circulation as through the systemic circulation). The increased pulmonary flow most often occurs under relatively low pressure. Most of the children with such defects are asymptomatic and have a low risk for development of pulmonary vascular disease. Children with large defects often have a severe left-to-right shunt, often complicated by pulmonary hypertension and congestive heart failure.

Most infants with a ventricular septal defect are asymptomatic during early infancy because the higher pulmonary vascular resistance prevents shunting. The infant with a large, uncomplicated ventricular septal defect usually is asymptomatic until pulmonary vascular resistance begins to fall at approximately 4 to 25 weeks of age. After a large shunt develops, the infant breathes rapidly, feeds poorly, and is diaphoretic (*i.e.*, signs of congestive heart failure). Right-to-left shunting produces cyanosis.

The treatment of a ventricular septal defect depends on the size of the defect and accompanying hemodynamic derangements. Children with small or medium-size defects are followed up closely in the hope that the defect will close spontaneously. Prophylactic antibiotic therapy is given during periods of increased risk for bacteremia. Congestive heart failure is treated medically. Surgical intervention is required for infants who do not have a response to medical management.

Endocardial Cushion Defects. The endocardial cushions form the AV canals, the upper part of the ventricular septum, and the lower part of the atrial septum (Fig. 17-23). Endocardial cushion defects are responsible for approximately 5% of all congenital heart defects. As many as 50% of children with Down's syndrome have endocardial cushion defects.³⁷

316 Unit Four: Alterations in the Cardiovascular System

Because the endocardial cushions contribute to multiple aspects of heart development, several variations with this type of defect are possible. The terms most commonly used to categorize endocardial cushion defects are *partial* and *complete AV canal defects*.³⁷ In partial AV canal defects, the two AV valve rings are complete and separate. In complete canal defect, there is a common AV valve orifice along with defects in both the atrial and ventricular septal tissue. Many variations of these two forms of endocardial cushion defect are possible (see Fig. 17-22E).

The direction and magnitude of a shunt in a child with endocardial cushion defects are determined by the combination of defects and the child's pulmonary and systemic vascular resistance. With complete AV canal defects, congestive heart failure and intercurrent pulmonary infections appear early in infancy. There is left-to-right shunting and transatrial and transventricular mixing of blood. Pulmonary hypertension and increased pulmonary vascular resistance are common. Cyanosis develops with progressive shunting.

The treatment for endocardial cushion defects is determined by the severity of the defect. With an atrial septal defect, surgical repair usually is planned on an elective basis before the child enters school. Palliative or corrective surgery is required in infants with complete AV canal defects who have congestive heart failure and do not experience a response to medical treatment.

Pulmonary Valve Stenosis. Pulmonary valve stenosis may occur as an isolated valvular lesion or in conjunction with more complex defects, such as tetralogy of Fallot. In isolated valvular defects, the pulmonary cusps may be absent or malformed, or they may remain fused at their commissural edges; all three abnormalities often coexist.

Pulmonary valvular defects usually cause some impairment of pulmonary blood flow and increase the workload imposed on the right side of the heart (see Fig. 17-22D). Most children with pulmonic valve stenosis have mild to moderate stenosis that does not increase in severity. These children are largely asymptomatic. Severe defects are manifested by marked impairment of pulmonary blood flow that begins during infancy and is likely to become more severe as the child grows. Cyanosis develops in approximately one third of children younger than 2 years of age.³⁷ The ductus arteriosus may provide the vital accessory route for perfusing the lungs in infants with severe stenosis. When pulmonary stenosis is extreme, increased pressures in the right side of the heart may delay closure of the foramen ovale.

Treatment measures designed to maintain the patency of the ductus arteriosus may be used as a temporary measure to maintain or increase pulmonary blood flow in infants with severe pulmonary stenosis. Pulmonary valvotomy often is the treatment of choice. Transcatheter balloon valvuloplasty may be used in some infants with moderate degrees of obstruction.³⁸

Tetralogy of Fallot. As the name implies, tetralogy of Fallot consists of four associated congenital heart defects: (1) a ventricular septal defect involving the membranous septum and the anterior portion of the muscular septum; (2) dextroposition or shifting to the right of the aorta, so that it overrides the right ventricle and is in communication with the septal defect; (3) obstruction or narrowing of the pulmonary outflow channel, including pulmonic valve stenosis, a decrease in the size of the pulmonary trunk, or both; and (4) hypertrophy of the

right ventricle because of the increased work required to pump blood through the obstructed pulmonary channels³⁸ (see Fig. 17-22C).

Most children with tetralogy of Fallot display some degree of cyanosis, thus the term blue babies. The cyanosis develops as the result of decreased pulmonary blood flow and because the right-to-left shunt causes mixing of unoxygenated blood with the oxygenated blood, which is ejected into the peripheral circulation. Hypercyanotic attacks ("tet spells") may occur during the first months of life. These spells typically occur in the morning during crying, feeding, or defecating. These activities increase the infant's oxygen requirements. Crying and defecating may further increase pulmonary vascular resistance, thereby increasing right-to-left shunting and decreasing pulmonary blood flow. With the hypercyanotic spell, the infant becomes acutely cyanotic, hyperpneic, irritable, and diaphoretic. Later in the spell, the infant becomes limp and may lose consciousness. Placing the infant in the knee-chest position increases systemic vascular resistance, which increases pulmonary blood flow and decreases right-to-left shunting. During a hypercyanotic spell, toddlers and older children may spontaneously assume the squatting position, which functions like the knee-chest position to relieve the spell.37

Because of the hypoxemia that occurs in these children, palliative surgery designed to increase pulmonary blood flow often is needed during early infancy, with corrective surgery performed at a later age. Palliative surgery involves the creation of a surgical shunt to increase pulmonary blood flow. The most popular procedures use the subclavian artery or prosthetic material to create a shunt between the aorta and pulmonary artery.

Transposition of the Great Vessels. In complete transposition of the great vessels, the aorta originates in the right ventricle, and the pulmonary artery originates in the left ventricle (see Fig. 17-22F). The defect is more common in infants whose mothers have diabetes and in boys. In infants born with this defect, survival depends on communication between the right and left sides of the heart in the form of a patent ductus arteriosus or septal defect. Prostaglandin E1 may be administered in an effort to maintain the patency of the ductus arteriosus. Balloon atrial septostomy may be done to increase the blood flow between the two sides of the heart. In this procedure, a balloon-tipped catheter is inserted into the heart through the vena cava and then passed through the foramen ovale into the left atrium. The balloon is inflated and brought back through the foramen ovale, enlarging the opening as it goes. Corrective surgery is essential for long-term survival.38

Coarctation of the Aorta. Coarctation of the aorta is a localized narrowing of the aorta, proximal (preductal or coarctation of infancy) or distal (postductal) to the ductus. Approximately 98% of coarctations are postductal (see Fig. 17-22H). The anomaly occurs twice as often in males as in females. Coarctation of the aorta may be a feature of Turner's syndrome (see Chapter 4).

The classic sign of coarctation of the aorta is a disparity in pulsations and blood pressures in the arms and legs. The femoral, popliteal, and dorsalis pedis pulsations are weak or delayed compared with the bounding pulses of the arms and carotid vessels. The systolic blood pressure in the legs obtained by the cuff method normally is 10 to 20 mm Hg higher than in the arms.³⁸ In coarctation, the pressure is lower and may be difficult to obtain. The differential in blood pressure is common in children older than 1 year of age, approximately 90% of whom have hypertension in the upper extremities greater than the 95th percentile for age (see Chapter 16).

Children with significant coarctation should be treated surgically; the optimal age for surgery is 2 to 4 years. If untreated, most persons with coarctation of the aorta die between 20 and 40 years of age. The common serious complications are related to the hypertensive state. In some centers, balloon valvoplasty has been used for treatment of unoperated coarctation. This method is still being developed, and ongoing clinical trials are needed to determine its long-term effectiveness and possible complications.³⁸

Kawasaki's Disease

Kawasaki's disease is an acute febrile disease of young children. First described in Japan in 1967 by Dr. Tomisaku Kawasaki, the disease affects the skin, brain, eyes, joints, liver, lymph nodes, and heart. The disease can produce aneurysmal disease of the coronary arteries and is the most common cause of acquired heart disease in young children. Although first reported in Japanese children, the disease affects children of many races, occurs worldwide, and is increasing in frequency.

Kawasaki's disease is characterized by a vasculitis (*i.e.*, inflammation of the blood vessels) that begins in the small vessels (*i.e.*, arterioles, venules, and capillaries) and progresses to involve some of the larger arteries, such as the coronaries.⁴² The cause of Kawasaki's disease is unknown, but it is thought to be of immunologic origin. It has been hypothesized that some unknown antigen, possibly a common infectious agent, triggers the immune response in a genetically predisposed child.

Clinical Course

The course of the disease is triphasic and includes an acute febrile phase that lasts approximately 7 to 14 days; a subacute phase that follows the acute phase and lasts from days 10 through 24; and a convalescent phase that follows the subacute stage and continues until the signs of the acute-phase inflammatory response have subsided and the signs of the illness have disappeared.⁴²

The *acute phase* begins with an abrupt onset of fever, followed by bilateral conjunctivitis, usually without exudates; erythema of the oral and pharyngeal mucosa with "strawberry tongue" and dry fissured lips; redness and swelling of the hands and feet; rash of various forms; and enlarged cervical lymph nodes. The fever typically is high, reaching 40 °C (104 °F) or more, has an erratic spiking pattern, is unresponsive to antibiotics, and persists for 5 or more days.⁴² The conjunctivitis begins shortly after the onset of fever, persists throughout the febrile course of the disease, and may last as long as 3 to 5 weeks.

The *subacute phase* begins when fever and other acute signs have subsided and lasts until all signs of the disease have disappeared. The subacute phase is associated with peeling of the skin of the fingertips and ends of the toes, thrombocytosis, the development of coronary aneurysms, and the greatest risk of sudden death. The *convalescent stage* persists from the complete resolution of symptoms until all signs of inflammation have disappeared. This usually takes approximately 8 weeks. In addition to the major manifestations that occur during the acute stage of the illness, there are several associated, less specific characteristics of the disease, including arthritis, urethritis and pyuria, gastrointestinal manifestations (*e.g.*, diarrhea, abdominal pain), hepatitis, and hydrops of the gallbladder.⁴² Arthritis or arthralgia occurs in approximately 30% of children with the disease, characterized by symmetric joint swelling that involves large and small joints. Central nervous system involvement occurs in almost all children and is characterized by pronounced irritability and lability of mood.

Cardiac involvement is the most important manifestation of Kawasaki's disease. Coronary vasculitis develops in between 10% and 40% of children within the first 2 weeks of the illness, manifested by dilatation and aneurysm formation in the coronary arteries, as seen on two-dimensional echocardiography. The manifestations of coronary artery involvement include signs and symptoms of myocardial ischemia or, rarely, overt myocardial infarction or rupture of the aneurysm. Pericarditis, myocarditis, endocarditis, heart failure, and dysrhythmias also may develop.

Diagnosis and Treatment

As with rheumatic fever, the diagnosis of Kawasaki's disease is based on clinical findings because no specific laboratory test for the disease exists. Clinical criteria developed by the Japan Kawasaki Disease Research Committee and subsequently by the AHA are used in establishing a diagnosis of Kawasaki's disease.⁴³ Chest radiographs, ECG tests, and two-dimensional echocardiography are used to detect coronary artery involvement and follow its progress. Coronary angiography may be used to determine the extent of coronary artery involvement.

Intravenous gamma globulin and aspirin are considered the best therapy for prevention of coronary artery abnormalities in children with Kawasaki's disease.⁴² During the acute phase of the illness, aspirin usually is given in larger doses and for its anti-inflammatory and antipyretic effects.⁴² After the fever is controlled, the aspirin dose is lowered, and the drug is given for its anti-platelet-aggregating effects.

In summary, congenital heart defects arise during fetal heart development and reflect the stage of development at the time the causative event occurred. Several factors contribute to the development of congenital heart defects, including genetic and chromosomal influences, viruses, and environmental agents such as drugs and radiation. The cause of the defect often is unknown. The defect may produce no effects, or it may markedly affect cardiac function. Congenital heart defects commonly produce shunting of blood from the right to the left side of the heart or from the left to the right side of the heart. Left-to-right shunts typically increase the volume of the right side of the heart and pulmonary circulation, and right-to-left shunts transfer unoxygenated blood from the right side of the heart to the left side, diluting the oxygen content of blood that is being ejected into the systemic circulation and causing cyanosis. The direction and degree of shunt depend on the size of the defect that connects the two sides of the heart and the difference in resistance between the two sides of the circulation. Congenital heart defects often are classified as defects that produce cyanosis and those that produce little or no cyanosis. Depending on the

severity of the defect, congenital heart defects may be treated medically or surgically. Medical and surgical treatment often is indicated in children with severe defects.

Kawasaki's disease is an acute febrile disease of young children that affects the skin, brain, eyes, joints, liver, lymph nodes, and heart. The disease can produce aneurysmal disease of the coronary arteries and is the most common cause of acquired heart disease in young children.

REVIEW QUESTIONS

Explain the decrease in cardiac output and pulsus paradoxus that occur with cardiac tamponade.

Describe blood flow in the coronary circulation and relate it to anginal pain that occurs during increased activity and tachycardia.

Characterize the pathogenesis of atherosclerosis in terms of fixed atherosclerotic lesions, unstable plaque, and thrombosis with obstruction. Explain the role of low-dose aspirin in prevention of CHD.

Define the term acute coronary syndromes and distinguish among chronic stable angina, unstable angina, non–ST-segment elevation myocardial infarction, and ST-segment elevation myocardial infarction in terms of pathology, symptomatology, ECG changes, and serum cardiac markers.

Compare the heart changes that occur with dilated, hypertrophic, constrictive cardiomyopathies, and arrhythmogenic right ventricular cardiomyopathy.

Describe the relation between the infective vegetations associated with infective endocarditis and explain their relationship to the extracardiac manifestations of the disease.

State the function of the heart valves and relate to alterations in hemodynamic changes (obstruction or regurgitation of blood flow), changes in stroke volume output, and nature of the heart murmur in terms of its occurrence during systole or diastole.

Describe the anatomic defects and altered patterns of blood flow in children with atrial septal defects, ventricular septal defects, endocardial cushion defects, pulmonary stenosis, tetralogy of Fallot, patent ductus arteriosus, transposition of the great vessels, and coarctation of the aorta.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- 1. American Heart Association (AHA). (2000). 1999 Heart and stroke facts. Dallas: Author.
- Hoyert D.L., Kochanek K.D., Murphy S.L. (1999). Deaths: Final data for 1997. National Vital Statistics Report 47 (19), 1–104.

- Lorell B.H. (1997). Pericardial diseases. In Braunwald E. (Ed.), *Heart disease: A textbook of cardiovascular medicine* (5th ed., Vol. 2, pp. 1478–1505). Philadelphia: W.B. Saunders.
- 4. Cotran R.S., Kumar V., Collins T. (Eds.). (1999). *Robbins pathologic basis of disease* (6th ed., pp. 528, 566–656, 570–576). Philadelphia: W.B. Saunders.
- Guyton A., Hall J.E. (2000). Textbook of medical physiology (10th ed., pp. 226–229). Philadelphia: W.B. Saunders.
- Ganz P., Braunwald E. (1997). Coronary blood flow and myocardial ischemia. In Braunwald E. (Ed.), *Heart disease: A textbook of cardiovascular medicine* (5th ed., Vol. 2, pp. 1164–1168). Philadelphia: W.B. Saunders.
- Gibbons R.J., Chatterjee K., Daley J., et al. Committee Members. (1999). ACC/ACP-ASIM guidelines for the management of patients with chronic stable angina: Executive summary and recommendations. *Circulation* 99, 2829–2848.
- Gersh B.J., Braunwald E., Rutherford J.D. (1997). Chronic coronary artery disease. In Braunwald E. (Ed.), *Heart disease: A textbook of cardiovascular medicine* (5th ed., Vol. 2, pp. 1289–1316). Philadelphia: W.B. Saunders.
- Prinzmetal M., Kennamer R., Merliss R., et al. (1959). A variant form of angina pectoris. *American Journal of Medicine* 27, 375–388.
- Pepine C.J., El-Tamimi H., Lambert C.R. (1992). Prinzmetal's angina (variant angina). *Heart Disease and Stroke* 1, 281–286.
- Cohn P.F. (1994). Silent myocardial ischemia: To treat or not to treat. Hospital Practice 29 (6), 107–116.
- Chiariello M., Indolfi C. (1996). Silent myocardial ischemia in patients with diabetes mellitus. *Circulation* 93, 2089–2091.
- Braunwald E., Antman E.M., Beasley J.W., et al., Committee Members. (2000). ACC/AHA guidelines for the management of patients with unstable angina and non–ST-segment elevation myocardial infarction: Executive summary and recommendations. *Circulation* 102, 1193–1209.
- Fullwood J., Butler G., Smith T., et al. (2000). New strategies in management of acute coronary syndromes. *Nursing Clinics of North America* 35, 877–896.
- Unstable Angina Guideline Panel. (1994). Unstable angina: Diagnosis and management. AHCPR publication no. 94B0602. Rockville, MD: U.S. Department of Health and Social Services.
- Ryan T.J., Antman E.M., Brooks N.H., et al., Committee Members. (1999). 1999 update: ACC/AHA guidelines for the management of acute myocardial infarction: Executive summary and recommendation. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Management of Acute Myocardial Infarction). *Journal of the American College of Cardiology* 28, 1328–1428.
- 17. Bittl J.A. (1997). Advances in coronary angioplasty. *New England Journal of Medicine* 337, 1290–1302.
- Morris N.B. (1999). Brachytherapy. Critical Care Clinics of North America 11, 333–343.
- Brown C.A., O'Connell J.B. (1995). Myocarditis and idiopathic dilated cardiomyopathy. *American Journal of Medicine* 99, 309–314.
- 20. Feldman A.M., McNamara D. (2000). Myocarditis. New England Journal of Medicine 343, 1388–1398.
- Bradenburg R.O., Chazo J.E., Cherian G., et al., Committee Members. (1982). Report of WHO/ISF Cardiology Task Force on the Definition and Classification of Cardiomyopathies. *British Heart Journal* 44, 672–673.
- Richardson P., Rapporteur W., McKenna W., et al. (1996). Report of the 1995 World Health Organization/International Society and Federation of Cardiology Task Force on Definition and Classification of Cardiomyopathies. *Circulation* 93, 841–842.
- Wyne J., Braunwald E. (1997). The cardiomyopathies and myocarditides. In Braunwald E. (Ed.), *Heart disease: A textbook of cardiovascular medicine* (5th ed., Vol. 2, pp. 1404–1451). Philadelphia: W.B. Saunders.
- Spirito P., Seidman C.E., McKenna W.J., et al. (1997). The management of hypertrophic cardiomyopathy. *New England Journal of Medicine* 336, 775–783.

319 **Chapter 17: Alterations in Cardiac Function**

- 25. Jennings R.B., Steenbergen C. Jr. (1999). The heart (ed. 3, pp. 577-583). In Rubin E., Farber J.L. (Eds.), Pathology. Philadelphia: Lippincott Williams & Wilkins
- 26. Kushwaha S.S., Fallon J.T., Fuster V. (1997). Restrictive cardiomyopathy. New England Journal of Medicine 336, 267-274.
- 27. Pearson G.D., Veille J., Rahimtoola S., et al. (2000). Peripartum cardiomyopathy: National Heart, Lung, and Blood Institute and Office of Rare Diseases (National Institutes of Health) Workshop Recommendations and Review. Journal of the American Medical Association 283, 83-88.
- 28. Karchmer A.W. (1997). Infective endocarditis. In Braunwald E. (Ed.), Heart disease: A textbook of cardiovascular medicine (5th ed., Vol. 2, pp. 1077-1099). Philadelphia: W.B. Saunders.
- 29. American Heart Association Advisory and Coordinating Committee. (1998). Diagnosis and management of infective endocarditis and its complications. Circulation 98, 2936-2948.
- 30. American Heart Association Advisory and Coordinating Committee. (1998). Diagnosis and management of infective endocarditis and its complications. Circulation 98, 2936-2948.
- 31. Ad Hoc Committee to Revise Jones Criteria (Modified) of the Council on Rheumatic Fever and Congenital Heart Disease of the American Heart Association. (1984). Jones criteria (revised) for guidance in the diagnosis of rheumatic fever. Circulation 69, 203A-208A.
- 32. Committee on Rheumatic Fever, Endocarditis, and Kawasaki Disease of the Council on Cardiovascular Disease in the Young of the American Heart Association. (1993). Guidelines for the diagnosis of rheumatic fever. Journal of the American Medical Association 268, 2069-2073.
- 33. Committee on Rheumatic Fever, Endocarditis, and Kawasaki Disease of the Council on Cardiovascular Disease in the Young of the American Heart Association. (1995). Treatment of acute streptococcal pharyngitis and prevention of rheumatic fever. Dallas: American Heart Association.

- 34. Braunwald E. (1997). Valvular heart disease. In Braunwald E. (Ed.), Heart disease: A textbook of cardiovascular medicine (5th ed., Vol. 2, pp. 1007-1076). Philadelphia: W.B. Saunders.
- 35. Bonow R.O., Carabello B., deLeon A.D. Jr., et. al., Committee on Management of Patients with Valvular Heart Disease. (1998). Guideline for the management of patients with valvular heart disease: Executive summary. A report of the American College of Cardiology/ American Heart Association Task Force on Guidelines. Circulation 98, 1949-1984
- 36. Massie B.M., Amidon T.M. (2001). Heart. In Tierney L.M., McPhee S.J., Papadakis M.A. (Eds.), Current medical diagnosis and treatment (40th ed., pp. 361-369). New York: Lange Medical Books/McGraw-Hill.
- 37. Hazinski M.F. (1992). Nursing care of the critically ill child (2nd ed., pp. 112-131, 271-361). St. Louis: Mosby-Year Book.
- 38. Bernstein D. (2000). The cardiovascular system. In Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.), Nelson textbook of pediatrics (16th ed., pp. 1337-1413). Philadelphia: W.B. Saunders.
- 39. Friedman W.F. (1997). Congenital heart disease in infancy and childhood. In Braunwald E. (Ed.), Heart disease: A textbook of cardiovascular medicine (5th ed., Vol. 2, pp. 877-962). Philadelphia: W.B. Saunders
- 40. Nouri S. (1997). Congenital heart defects: Cyanotic and acyanotic. Pediatric Annals 26, 92–98. Philadelphia: W.B. Saunders.
- 41. Driscoll D.J. (1999). Left-to-right shunt lesions. Pediatric Clinics of North America 46, 355-368.
- 42. Rowley A.H., Shulman S.T. (2000). Kawasaki disease. In Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.), Nelson textbook of pediatrics (16th ed., pp. 725-727). Philadelphia: W.B. Saunders.
- 43. American Heart Association. (1999). Diagnostic guidelines for Kawasaki disease. [On-line]. Available: http://www.american heart.org/catalog/ Kawasaki, html.

CHAPTER 18

Heart Failure and Circulatory Shock

Heart Failure

Physiology of Heart Failure Cardiac Output Compensatory Mechanisms Congestive Heart Failure Types of Heart Failure **Manifestations of Heart Failure Diagnosis and Treatment** Acute Pulmonary Edema Manifestations Treatment **Cardiogenic Shock Manifestations** Treatment Mechanical Support of the Failing Heart and Heart Transplantation **Circulatory Failure (Shock)** Types of Shock Hypovolemic Shock **Obstructive Shock Distributive Shock** Sepsis and Septic Shock **Complications of Shock** Acute Respiratory Distress Syndrome **Acute Renal Failure Gastrointestinal Complications Disseminated Intravascular Coagulation** Multiple Organ Dysfunction Syndrome Heart Failure in Children and the Elderly Heart Failure in Infants and Children Manifestations **Diagnosis and Treatment** Heart Failure in the Elderly Manifestations **Diagnosis and Treatment**

dequate perfusion of body tissues depends on the pumping ability of the heart, a vascular system that transports blood to the cells and back to the heart, sufficient blood to fill the circulatory system, and tissues that are able to extract and use the oxygen and nutrients from the blood. Heart failure and circulatory shock are separate conditions that reflect failure of the circulatory system. Both conditions exhibit common compensatory mechanisms even though they differ in terms of pathogenesis and causes.

HEART FAILURE

Heart failure affects an estimated 4.8 million Americans.¹ Although morbidity and mortality rates from other cardiovascular diseases have decreased during the past several decades, the incidence of heart failure is increasing at an alarming rate. This change undoubtedly reflects treatment improvements and survival from other forms of heart disease. Despite advances in treatment, the 5-year survival rate for heart failure is only about 50%.

Physiology of Heart Failure

The term *heart failure* denotes the failure of the heart as a pump. The heart has the amazing capacity to adjust its pumping ability to meet the varying needs of the body. During sleep, its output declines, and during exercise, it increases markedly. The ability to increase cardiac output during increased activity is called the *cardiac reserve*. For example, competitive swimmers and long-distance runners have large cardiac reserves. During exercise, the cardiac output of these athletes rapidly increases to as much as five to six times their resting level. In sharp contrast with healthy athletes, persons with heart failure often use their cardiac reserve at rest. For them, just climbing a flight of stairs may cause shortness of breath because they have exceeded their cardiac reserve.

The pathophysiology of heart failure involves an interaction between two factors: (1) a decrease in cardiac output with a consequent decrease in blood flow to the kidneys and other body organs and tissues; (2) the recruitment of compensatory mechanisms designed to maintain tissue perfusion.

KEY CONCEPTS

HEART FAILURE

- The function of the heart is to move deoxygenated blood from the venous system through the right heart into the pulmonary circulation and to move the oxygenated blood from the pulmonary circulation through the left heart into the arterial system.
- To function effectively, the right and left hearts must maintain an equal output.
- Right heart failure represents failure of the right heart to pump blood forward into the pulmonary circulation; blood backs up in the systemic circulation, causing peripheral edema and congestion of the abdominal organs.
- Left heart failure represents failure of the left heart to move blood from the pulmonary circulation into the systemic circulation; blood backs up in the pulmonary circulation.

Cardiac Output

The cardiac output is the amount of blood that the heart pumps each minute. It reflects how often the heart beats each minute (heart rate) and how much blood the heart pumps with each beat (stroke volume) and can be expressed as the product of the heart rate and stroke volume (cardiac output = heart rate × stroke volume). Heart rate is a function of sympathetic nervous system reflexes, which accelerate heart rate and parasympathetic nervous system reflexes, which slows it down. Stroke volume is a function of preload, afterload, and cardiac contractility.

Preload and Afterload. The work that the heart performs consists mainly of ejecting blood into the pulmonary or systemic circulations. It is determined largely by the loading conditions or what is called the preload and afterload.

Preload reflects the loading condition of the heart at the end of diastole just before the onset of systole. It is the volume of blood stretching the resting heart muscle and is determined mainly by the venous return to the heart. Afterload represents the force that the contracting heart must generate to eject blood from the filled heart. The main components of afterload are ventricular wall tension and the peripheral vascular resistance. The greater the peripheral vascular resistance, the greater the ventricular wall tension and intraventricular pressure required to open the aortic valve and pump blood into the peripheral circulation.

Cardiac Contractility. Cardiac contractility refers to the mechanical performance of the heart: the ability of the contractile elements (actin and myosin filaments) of the heart muscle to interact and shorten against a load. Contractility increases cardiac output independent of preload filling and muscle stretch.

An inotropic influence is one that increases cardiac contractility. Sympathetic stimulation increases the strength of cardiac contraction (*i.e.*, positive inotropic action), and hypoxia and ischemia decrease contractility (*i.e.*, negative inotropic effect). The inotropic drug digitalis, which is used in the treatment of heart failure, increases cardiac contractility such that the heart is able to eject more blood at any level of preload filling.

Compensatory Mechanisms

Heart failure is characterized by a decrease in cardiac output with a consequent decline in blood flow to the kidneys as well as other body organs and tissues. With a decrease in cardiac performance, tissue and organ perfusion is largely maintained through compensatory mechanisms such as the Frank-Starling mechanism, activation of the sympathetic nervous system and the renin-angiotensin-aldosterone mechanism, and myocardial hypertrophy (Fig. 18-1). In the failing heart, early decreases in cardiac function often go unnoticed because these compensatory mechanisms are used to maintain the cardiac output. This state is called compensated heart failure. Unfortunately, these mechanisms were not intended for long-term use. In severe and prolonged heart failure, the compensatory mechanisms are no longer effective and may themselves worsen the failure, causing what is termed *decompensated failure*.

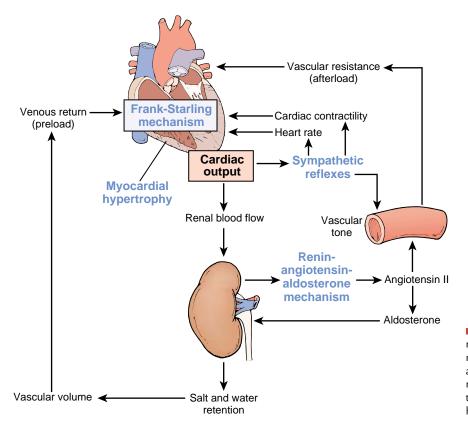
Frank-Starling Mechanism. The Frank-Starling mechanism relies on an increase in venous return and a resultant increase in diastolic filling of the ventricles. Known as the *end-diastolic vol*ume, this volume causes the tension in the wall of the ventricles and the pressure in the ventricles to rise. With increased ventricular end-diastolic volume, there is increased stretching of the myocardial fibers, more optimal approximation of the actin and myosin filaments, and a resultant increase in the stroke volume in accord with the Frank-Starling mechanism (see Chapter 14, Fig. 14-20).

In heart failure a decrease in cardiac output and renal blood flow leads to increased salt and water retention, a resultant increase in vascular volume and venous return to the heart, and an increase in ventricular end-diastolic volume. Within limits, as preload and ventricular end-diastolic volume increase, there is a resultant increase in cardiac output. Thus, cardiac output may be normal at rest in persons with heart failure. However, as myocardial function deteriorates, the heart becomes overfilled, the muscle fibers become overstretched, and the ventricular function curve flattens (Fig. 18-2). The maximal increase in cardiac output that can be achieved may severely limit activity, while producing an elevation in left ventricular and pulmonary capillary pressure and development of dyspnea and pulmonary congestion.

An important determinant of myocardial energy consumption is ventricular wall tension. Overfilling of the ventricle produces a decrease in wall thickness and an increase in wall tension. Because increased wall tension increases myocardial oxygen requirements, it can produce ischemia and further impairment of cardiac function. The use of diuretics in the treatment of heart failure helps to reduce vascular volume and ventricular filling, thereby unloading the heart and reducing ventricular wall tension.

Increased Sympathetic Nervous System Activity. Stimulation of the sympathetic nervous system plays an important role in the compensatory response to decreased cardiac output and the pathogenesis of heart failure.^{2,3} Both cardiac sympathetic tone

322 Unit Four: Alterations in the Cardiovascular System



■ FIGURE 18-1 ■ Compensatory mechanisms in heart failure. The Frank-Starling mechanism, sympathetic reflexes, reninangiotensin-aldosterone mechanism, and myocardial hypertrophy function in maintaining the cardiac output for the failing heart.

and circulating catecholamine (epinephrine and norepinephrine) levels are elevated during the late stages of most forms of heart failure. By direct stimulation of heart rate and cardiac contractility and by regulation of vascular tone, the sympathetic nervous system helps to maintain perfusion of the various organs, particularly the heart and brain.

The negative aspects of increased sympathetic activity include an increase in peripheral vascular resistance and the after-



load against which the heart must pump. Excessive sympathetic stimulation also may result in decreased blood flow to skin, muscle, kidney, and abdominal organs. The catecholamines also may contribute to the high rate of sudden death by promoting dysrhythmias.⁴

Renin-Angiotensin Mechanism. One of the most important effects of a lowered cardiac output in heart failure is a reduction in renal blood flow and glomerular filtration rate, which leads to salt and water retention. Normally, the kidneys receive approximately 25% of the cardiac output, but this may be decreased to as low as 8% to 10% in persons with heart failure. With decreased renal blood flow, there is a progressive increase in renin secretion by the kidneys along with parallel increases in circulating levels of angiotensin II. The increased concentration of angiotensin II contributes to a generalized vasoconstriction and serves as a stimulus for aldosterone production by the adrenal cortex (see Chapter 16). Aldosterone, in turn, increases tubular reabsorption of sodium, with an accompanying increase in water retention. Because aldosterone is metabolized in the liver, its levels are further increased when heart failure causes liver congestion.

Recent evidence suggests that angiotensin is also a growth factor for cardiac muscle cells and fibroblasts and, as such, may play a central role in modifying the structure and function of the myocardium in persons with heart failure.² Angiotensin-converting enzyme (ACE) inhibitor drugs, which block the conversion of angiotensin I to angiotensin II, are often used in the treatment of heart failure.^{2–4}

Myocardial Hypertrophy. Myocardial hypertrophy is a longterm compensatory mechanism. Cardiac muscle, like skeletal muscle, responds to an increase in work demands by undergoing hypertrophy. Hypertrophy increases the number of contractile elements in myocardial cells as a means of increasing their contractile performance.

Myocardial hypertrophy occurs early in the course of heart failure and is an important risk factor for subsequent morbidity and mortality. Although hypertrophy increases the systolic function of the heart, it also eventually can lead to diastolic dysfunction and myocardial ischemia. Some forms of hypertrophy may lead to abnormal remodeling of the ventricular wall with a reduction in chamber size, reduced diastolic filling, and increased ventricular wall tension. For example, untreated hypertension causes hypertrophy that may preserve systolic function for a time, but eventually the work performed by the ventricle exceeds the augmented muscle mass and the heart dilates.² The increased muscle mass of the hypertrophied heart increases the need for oxygen delivery. When the oxygen requirements of the increased muscle mass exceed the ability of the coronary vessels to bring blood to the area, myocardial hypertrophy is no longer beneficial and may result in ischemia with decreased contractility. In addition, hypertrophy of cardiac muscle cells may be accompanied by the growth of nonmyocardial tissue (e.g., fibrous tissue) that produces stiffness of the ventricle and further impairment of ventricular function.

Congestive Heart Failure

Heart failure occurs when the pumping ability of the heart becomes impaired. The term congestive heart failure (CHF) refers to heart failure that is accompanied by congestion of body tissues.

Heart failure may be caused by a variety of conditions, including acute myocardial infarction, hypertension, valvular heart disease, or degenerative conditions of the heart muscle known collectively as cardiomyopathies (see Chapter 17). Heart failure also may occur because of excessive work demands, such as occurs with hypermetabolic states, or with volume overload, such as occurs with renal failure. Either of these states may exceed the work capacity of even a healthy heart. In persons with asymptomatic heart disease, heart failure may be precipitated by an unrelated illness or stress. Table 18-1 lists major causes of heart failure.

Types of Heart Failure

Heart failure may be described as high-output or low-output failure, systolic or diastolic failure, and right-sided or left-sided failure.

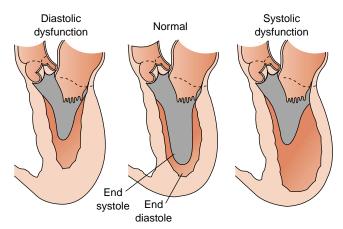
High- and Low-Output Failure. High- and low-output failure are described in terms of cardiac output. High-output failure is an uncommon type of heart failure that is caused by an excessive need for cardiac output. With high-output failure, the function of the heart may be supernormal but inadequate because of excessive metabolic needs. Causes of high-output failure include severe anemia, thyrotoxicosis, and conditions that cause arteriovenous shunting. High-output failure tends to be specifically treatable. Low-output failure is caused by disorders that im-

TABLE 18-1 Causes	of Heart Failure
Impaired Cardiac Function	Excess Work Demands
Myocardial Disease Cardiomyopathies Myocarditis Coronary insufficiency Myocardial infarction	Increased Pressure Work Systemic hypertension Pulmonary hypertension Coarctation of the aorta
Valvular Heart Disease Stenotic valvular disease Regurgitant valvular disease	Increased Volume Work Arteriovenous shunt Excessive administration of intravenous fluids
Congenital Heart Defects	Increased Perfusion Work Thyrotoxicosis Anemia
Constrictive Pericarditis	

pair the pumping ability of the heart, such as ischemic heart disease and cardiomyopathy.

Systolic Versus Diastolic Failure. A recent classification separates the pathophysiology of CHF into two categories—systolic dysfunction and diastolic dysfunction. Systolic dysfunction is characterized by impaired ejection of blood from the heart during systole and diastolic dysfunction by impaired filling of the ventricles during diastole (Fig. 18-3). Many persons with heart failure fall into an intermediate category, with combined elements of both systolic and diastolic failure.

Systolic dysfunction involves a decrease in cardiac contractility and ejection fraction. It commonly results from conditions that impair the contractile performance of the heart (e.g., ischemic heart disease and cardiomyopathy), produce a volume



■ FIGURE 18-3 ■ Congestive heart failure due to systolic and diastolic dysfunction. The ejection fraction represents the difference between the end-diastolic and end-systolic volumes. Normal systolic and diastolic function with normal ejection fraction (middle); diastolic dysfunction with decreased ejection fraction due to decreased diastolic filling (left); systolic dysfunction with decreased ejection fraction due to impaired systolic function (right).

overload (*e.g.*, valvular insufficiency and anemia), or generate a pressure overload (*e.g.*, hypertension and valvular stenosis) on the heart.

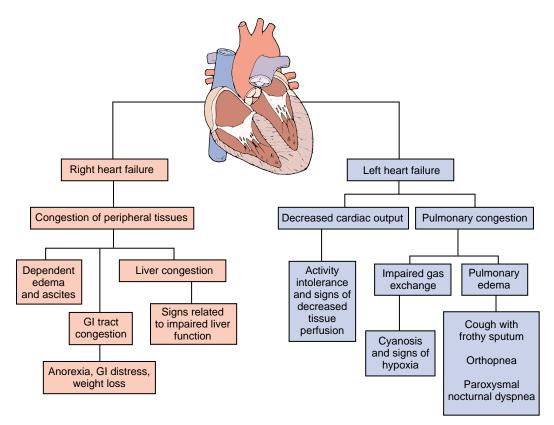
A normal heart ejects approximately 65% of the blood that is present in the ventricle at the end of diastole when it contracts. This is called the *ejection fraction*. In systolic heart failure, the ejection fraction declines progressively with increasing degrees of myocardial dysfunction. In very severe forms of heart failure, the ejection fraction may drop to a single-digit percentage. With a decrease in ejection fraction, there is a resultant increase in diastolic volume, ventricular dilation, ventricular wall tension, and ventricular end-diastolic pressure. The symptoms of persons with systolic dysfunction result mainly from reductions in ejection fraction and cardiac output.

Diastolic dysfunction, which reportedly accounts for approximately 40% of all cases of CHF, is characterized by a smaller ventricular chamber size, ventricular hypertrophy, and poor ventricular compliance (*i.e.*, ability to stretch during filling).⁵ Because of impaired filling, congestive symptoms tend to predominate in diastolic dysfunction. Among the conditions that cause diastolic dysfunction are those that restrict diastolic filling (*e.g.*, mitral stenosis), those that increase ventricular wall thickness and reduce chamber size (*e.g.*, myocardial hypertrophy caused by lung disease and hypertrophic cardiomyopathy), and those that delay diastolic relaxation (*e.g.*, aging, ischemic heart disease). Aging often is accompanied by a delay in relaxation of the heart during diastole; diastolic filling begins while the ventricle is still stiff and resistant to stretching to accept an increase in volume.⁶ A similar delay occurs with myocardial ischemia, resulting from a lack of energy to break the rigor bonds that form between the actin and myosin filaments of the contracting cardiac muscle.⁷ Because tachycardia produces a decrease in diastolic filling time, persons with diastolic dysfunction often become symptomatic during activities and situations that increase heart rate.

Right-Sided Versus Left-Sided Heart Failure. Heart failure also can be classified according to the side of the heart (right or left) that is affected. An important feature of the circulatory system is that the right and left ventricles act as two pumps that are connected in series. To function effectively, the right and left ventricles must maintain an equal output. Although the initial event that leads to heart failure may be primarily right sided or left sided in origin, long-term heart failure usually involves both sides.

Right-Sided Heart Failure. Right-sided heart failure impairs the ability to move deoxygenated blood from the systemic venous circulation into the pulmonary circulation. Consequently, when the right heart fails, there is an accumulation of blood in the systemic venous system (Fig. 18-4). This causes an increase in the right atrial, right ventricular end-diastolic, and systemic venous pressures.

A major effect of right-sided heart failure is the development of peripheral edema. Because of the effects of gravity, the edema is most pronounced in the dependent parts of the body—in the lower extremities when the person is in the upright position and in the area over the sacrum when the person is supine. The



■ FIGURE 18-4 ■ Manifestations of left- and right-sided heart failure.

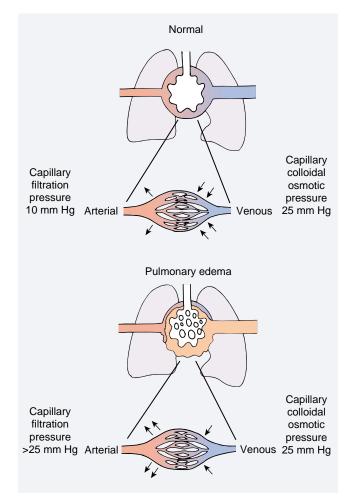
accumulation of edema fluid is evidenced by a gain in weight (*i.e.*, 1 pint of accumulated fluid results in a 1-lb weight gain). Daily measurement of weight can be used as a means of assessing fluid accumulation in a patient with chronic CHF. As a rule, a weight gain of more than 2 lb in 24 hours or 5 lb in 1 week is considered a sign of worsening failure.

Right-sided heart failure also produces congestion of the viscera. As venous distention progresses, blood backs up in the hepatic veins that drain into the inferior vena cava, and the liver becomes engorged. This may cause hepatomegaly and right upper quadrant pain. In severe and prolonged right-sided failure, liver function is impaired and hepatic cells may die. Congestion of the portal circulation also may lead to engorgement of the spleen and the development of ascites. Congestion of the gastrointestinal tract may interfere with digestion and absorption of nutrients, causing anorexia and abdominal discomfort. The jugular veins, which are above the level of the heart, are normally collapsed in the standing position or when sitting with the head at higher than a 30-degree angle. In severe rightsided failure, the external jugular veins become distended and can be visualized when the person is sitting up or standing.

The causes of right-sided heart failure include conditions that restrict blood flow into or through the lungs. Stenosis or regurgitation of the tricuspid or pulmonic valves, right ventricular infarction, cardiomyopathy, and persistent left-sided failure are common causes. *Cor pulmonale* refers to right heart failure resulting from pulmonary disease. *Acute cor pulmonale* results from right heart strain or overload secondary to acute pulmonary hypertension, often caused by massive pulmonary embolism. *Chronic cor pulmonale* occurs secondary to diseases that affect pulmonary tissues (chronic obstructive lung disease) or pulmonary vasculature (primary pulmonary hypertension).

Left-Sided Heart Failure. Left-sided heart failure impairs the pumping of blood from the low-pressure pulmonary circulation into the high-pressure arterial side of the systemic circulation (Fig. 18-4). With impairment of left heart function, there is a decrease in cardiac output, an increase in left atrial and left ventricular end-diastolic pressures, and congestion in the pulmonary circulation. When the pulmonary capillary filtration pressure (normally approximately 10 mm Hg) exceeds the capillary osmotic pressure (normally approximately 25 mm Hg), there is a shift of intravascular fluid into the interstitium of the lung and development of pulmonary edema (Fig 18-5). An episode of pulmonary edema often occurs at night, after a person has been reclining for some time and gravitational forces have been removed from the circulatory system. It is then that edema fluid that has been sequestered in the lower extremities during the day is returned to the vascular compartment and redistributed to the pulmonary circulation.

The most common causes of left-sided heart failure are acute myocardial infarction and cardiomyopathy. Left-sided heart failure and pulmonary congestion can develop very rapidly in persons with acute myocardial infarction. Even when the infarcted area is small, there may be a surrounding area of ischemic tissue. This may result in a large area of nonpumping ventricle and rapid onset of pulmonary edema. Aortic and aortic valve disorders can also cause left-sided heart failure. Pulmonary edema also may develop during rapid infusion of intravenous fluids or blood transfusions in an elderly person or in a person with limited cardiac reserve.



■ FIGURE 18-5 ■ Mechanism of respiratory symptoms in leftsided heart failure. Normal exchange of fluid in the pulmonary capillaries (top). The capillary filtration pressure that moves fluid out of the capillary into the lung is less than the capillary colloidal osmotic pressure that pulls fluid back into the capillary. Development of pulmonary edema (bottom) occurs when the capillary filtration pressure exceeds the capillary colloidal osmotic pressure that pulls fluid back into the capillary.

Manifestations of Heart Failure

The manifestations of heart failure depend on the extent and type of cardiac dysfunction (e.g., systolic versus diastolic) that is present and the rapidity with which it develops. Many persons with heart failure have concomitant cardiovascular and noncardiovascular disorders that may exacerbate the condition.7 The concomitant conditions include cardiovascular disorders such as hypertension, coronary artery disease, diabetes, and cardiac dysrhythmias. Associated noncardiovascular diseases include renal insufficiency, thyroid disease, and pulmonary disease. A person with previously stable compensated heart failure may experience signs of heart failure for the first time when the condition has advanced to a critical point, such as with a progressive increase in pulmonary hypertension in a person with mitral valve regurgitation. Overt heart failure also may be precipitated by conditions such as infection, emotional stress, uncontrolled hypertension, administration of fluid overload, or inappropriate reduction in therapy. Many persons with serious underlying heart disease, regardless of whether they have previously experienced heart failure, may be relatively asymptomatic as long they carefully adhere to their treatment regimen. A dietary excess of sodium is a frequent cause of sudden cardiac decompensation.

The manifestations of heart failure reflect the physiologic effects of the impaired pumping ability of the heart, decreased renal blood flow, and activation of the sympathetic compensatory mechanisms. The signs and symptoms include fluid retention and edema, shortness of breath and other respiratory manifestations, fatigue and limited exercise tolerance, cyanosis, cachexia and malnutrition, and cyanosis. Distention of the jugular veins may be present in right-sided failure. Excessive sympathetic stimulation may produce diaphoresis and tachycardia in persons with severe heart failure.

Fluid Retention and Edema. Many of the manifestations of CHF result from the increased capillary pressures that develop in the peripheral circulation in right-sided heart failure and in the pulmonary circulation in left-sided heart failure. The increased capillary pressure reflects an overfilling of the vascular system because of increased salt and water retention and venous congestion resulting from the impaired pumping ability of the heart.

Nocturia is a nightly increase in urine output that occurs relatively early in the course of CHF. It results from the return to the circulation of edema fluids from the dependent parts of the body when the person assumes the supine position for the night. As a result, the cardiac output, renal blood flow, glomerular filtration, and urine output increase. Oliguria is a late sign related to a severely reduced cardiac output and resultant renal failure.

Respiratory Manifestations. Shortness of breath caused by congestion of the pulmonary circulation is one of the major manifestations of left-sided heart failure. Perceived shortness of breath (i.e., breathlessness) is called dyspnea. Dyspnea related to an increase in activity is called exertional dyspnea. Orthopnea is shortness of breath that occurs when a person is supine. The gravitational forces that cause fluid to become sequestered in the lower legs and feet when the person is standing or sitting are removed when a person with CHF assumes the supine position; fluid from the legs and dependent parts of the body is mobilized and redistributed to an already distended pulmonary circulation. Paroxysmal nocturnal dyspnea is a sudden attack of dyspnea that occurs during sleep. It disrupts sleep, and the person awakens with a feeling of extreme suffocation that resolves when he or she sits up. Initially, the experience may be interpreted as awakening from a bad dream.

A subtle and often overlooked symptom of heart failure is a chronic dry, nonproductive cough, which becomes worse when the person is lying down. Bronchospasm caused by congestion of the bronchial mucosa may cause wheezing and difficulty in breathing. This condition is sometimes referred to as *cardiac asthma*.

Cheyne-Stokes respiration, also known as periodic breathing, is characterized by a slow waxing and waning of respiration. The person breathes deeply for a period when the arterial carbon dioxide pressure (PCO_2) is high and then slightly or not at all when the PCO_2 falls. In persons with left-sided heart failure, the condition is thought to be caused by a prolongation of the

heart-to-brain circulation, particularly in persons with hypertension and associated cerebral vascular disease. Cheyne-Stokes breathing may contribute to daytime sleepiness, and occasionally the person awakens at night with dyspnea precipitated by Cheyne-Stokes breathing.

Fatigue and Limited Exercise Tolerance. Fatigue and limb weakness often accompany diminished output from the left ventricle. Cardiac fatigue is different from general fatigue in that it usually is not present in the morning but appears and progresses as activity increases during the day. In acute or severe left-sided failure, cardiac output may fall to levels that are insufficient for providing the brain with adequate oxygen, and there are indications of mental confusion and disturbed behavior. Confusion, impairment of memory, anxiety, restlessness, and insomnia are common in elderly persons with advanced heart failure, particularly in those with cerebral atherosclerosis. These very symptoms may confuse the diagnosis of heart failure in the elderly because of the myriad other causes associated with aging.

Cachexia and Malnutrition. Cardiac cachexia is a condition of malnutrition and tissue wasting that occurs in persons with end-stage heart failure. A number of factors probably contribute to its development, including the fatigue and depression that interfere with food intake, congestion of the liver and gastrointestinal structures that impairs digestion and absorption and produces feelings of fullness, and the circulating toxins and mediators released from poorly perfused tissues that impair appetite and contribute to tissue wasting.

Cyanosis. Cyanosis is the bluish discoloration of the skin and mucous membranes caused by excess desaturated hemoglobin in the blood; it often is a late sign of heart failure. Cyanosis may be central or peripheral. Central cyanosis is caused by conditions that impair oxygenation of the arterial blood, such as pulmonary edema, left heart failure, or right-to-left shunting. Central cyanosis is best monitored in the lips and mucous membranes because these areas are not subject to conditions such as cold that cause peripheral cyanosis. Peripheral cyanosis is caused by conditions such as low-output failure that cause delivery of poorly oxygenated blood to the peripheral tissues, or by conditions such as peripheral vasoconstriction that cause excessive removal of oxygen from the blood.

Diagnosis and Treatment

Diagnostic methods in heart failure are directed toward establishing the cause of the disorder and determining the extent of the dysfunction. Because heart failure represents the failure of the heart as a pump and can occur in the course of a number of heart diseases or other systemic disorders, the diagnosis of heart failure often is based on signs and symptoms related to the failing heart itself, such as shortness of breath and fatigue. The functional classification of the New York Heart Association is one guide to classifying the extent of dysfunction (Table 18-2).

The diagnostic methods used for determining the presence of heart failure, its cause, and extent of dysfunction include history and physical examination, laboratory studies, chest radiography, electrocardiography (ECG), and echocardiography.⁸ The history should include information related to dyspnea, cough, nocturia, generalized fatigue, and other signs and symptoms of heart failure. A complete physical examination in-

TABLE 18-2	New Yorl	k Heart
Association		Classification
of Patients V	With Heart	Disease

Classification	Characteristics
Class I	Patients with cardiac disease but without the resulting limitations in physical activity. Ordinary activity does not cause undue fatigue, palpitation, dyspnea, or anginal
	pain.
Class II	Patients with heart disease resulting in slight limitations of physical activity. They are
	comfortable at rest. Ordinary physical
	activity results in fatigue, palpitation,
a	dyspnea, or anginal pain.
Class III	Patients with cardiac disease resulting in
	marked limitation of physical activity. They
	are comfortable at rest. Less than ordinary
	physical activity causes fatigue, palpitation,
	dyspnea, or anginal pain.
Class IV	Patients with cardiac disease resulting in
	inability to carry on any physical activity
	without discomfort. The symptoms of
	cardiac insufficiency or of the anginal syn-
	drome may be present even at rest. If any
	physical activity is undertaken, discomfort
	increases.

(From Criteria Committee of the New York Heart Association. [1964]. Diseases of the heart and blood vessels: Nomenclature and criteria for diagnosis [6th ed., pp. 112–113]. Boston: Little, Brown)

cludes assessment of heart rate, heart sounds, blood pressure, jugular veins for venous congestion, lungs for signs of pulmonary congestion, and lower extremities for edema. Laboratory tests are used in the diagnosis of anemia and electrolyte imbalances, and to detect signs of chronic liver congestion.

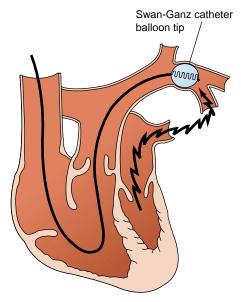
Chest radiographs provide information about the size and shape of the heart and pulmonary vasculature. The cardiac silhouette can be used to detect cardiac hypertrophy and dilatation. X-ray films can indicate the relative severity of the failure by revealing if pulmonary edema is predominantly vascular, interstitial, or advanced to the alveolar and bronchial stages. ECG findings may indicate atrial or ventricular hypertrophy, underlying disorders of cardiac rhythm, or conduction abnormalities such as right or left bundle branch block. Echocardiography plays a key role in assessing the anatomic and functional abnormalities in CHF, which include the size and function of cardiac valves, the motion of both ventricles, and the ventricular ejection fraction.9 Radionuclide angiography and cardiac catheterization are other diagnostic tests used to detect the underlying causes of heart failure, such as heart defects and cardiomyopathy.

Invasive hemodynamic monitoring often is used in the management of acute, life-threatening episodes of heart failure. These monitoring methods include measurement of central venous pressure (CVP), pulmonary capillary wedge pressure (PCWP), thermodilution cardiac output, and intra-arterial blood pressure. Measurements of CVP are obtained from a catheter inserted into the right atrium; they provide information about the pumping ability of the right heart and its ability to move blood into the pulmonary circulation. PCWP is obtained from a balloon-tipped catheter that is advanced into a small pulmonary vessel (Fig. 18-6). When the balloon on pulmonary catheter is inflated, the catheter monitors pulmonary capillary pressures that are in direct communication with those of the left heart. Thus, PCWP provide information about the pumping ability of the left heart.

The goals of treatment for chronic heart failure are directed toward relieving the symptoms and improving the quality of life, with a long-term goal of slowing, halting, or reversing the cardiac dysfunction.^{10,11} Treatment measures include correction of reversible causes such as anemia or thyrotoxicosis, surgical repair of a ventricular defect or an improperly functioning valve, pharmacologic and nonpharmacologic control of afterload stresses such as hypertension, modification of activities and lifestyle to a level consistent with the functional limitations of a reduced cardiac reserve, and the use of medications to improve cardiac function and limit excessive compensatory mechanisms. Restriction of salt intake and diuretic therapy facilitate the excretion of edema fluid.

In severe heart failure, restriction of activity, including bed rest if necessary, often facilitates temporary recompensation of cardiac function. However, there is no convincing evidence that continued bed rest is of benefit. Carefully designed and managed exercise programs for patients with CHF are well tolerated and beneficial to patients with stable New York Heart Association (NYHA) class I to III heart failure.¹²

Pharmacologic Treatment. Once heart failure is moderate to severe, polypharmacy becomes a management standard and often includes diuretics, digoxin, ACE inhibitors, and β -adrenergic–blocking agents.^{13,14} The choice of pharmacologic agents is determined by problems caused by the disorder (*i.e.*, systolic or diastolic dysfunction) and those brought about by activation of compensatory mechanisms (*e.g.*, excess fluid retention, inappropriate activation of sympathetic mechanisms).



■ FIGURE 18-6 ■ Swan-Ganz balloon-tipped catheter positioned in a pulmonary capillary. The pulmonary capillary wedge pressure, which reflects the left ventricular diastolic pressure, is measured with the balloon inflated.

328 Unit Four: Alterations in the Cardiovascular System

Diuretics are among the most frequently prescribed medications for heart failure. They promote the excretion of edema fluid and help to sustain cardiac output and tissue perfusion by reducing preload and allowing the heart to operate at a more optimal part of the Frank-Starling curve. Digitalis drugs are inotropic agents that improve cardiac function by increasing the force and strength of ventricular contraction. They also produce a decrease in sinoatrial node activity and conduction through the atrioventricular node, thus slowing the heart rate and increasing diastolic filling time. Although not a diuretic, digitalis promotes urine output by improving cardiac output and renal blood flow. The ACE inhibitors, which prevent the conversion of angiotensin I to angiotensin II, have been effectively used in the treatment of heart failure. In heart failure, renin activity frequently is elevated because of decreased renal blood flow. The net result is an increase in angiotensin II, which causes vasoconstriction and increased aldosterone production with a subsequent increase in salt and water retention by the kidney. The ACE inhibitors reduce both mechanisms to decrease the workload of the heart.

 β -adrenergic–blocking agents are used to decrease left ventricular dysfunction associated with activation of the sympathetic nervous system.^{13,15} Chronic elevation of norepinephrine levels has been shown to cause cardiac muscle cell death and progressive left ventricular dysfunction, and is associated with poor prognosis in heart failure. The β -adrenergic–blocking agents also decrease the risk of serious cardiac dysrhythmias in the person with heart failure.

Acute Pulmonary Edema

Acute pulmonary edema is the most dramatic symptom of left heart failure. It is a life-threatening condition in which capillary fluid moves into the alveoli. The accumulated fluid in the alveoli and respiratory airways causes lung stiffness, makes lung expansion more difficult, and impairs the gas exchange function of the lung. With the decreased ability of the lungs to oxygenate the blood, the hemoglobin leaves the pulmonary circulation without being fully oxygenated.

Manifestations

Acute pulmonary edema usually is a terrifying experience. The person usually is seen sitting and gasping for air, in obvious apprehension. The pulse is rapid, the skin is moist and cool, and the lips and nail beds are cyanotic. As the lung edema worsens and oxygen supply to the brain drops, confusion and stupor appear. Dyspnea and air hunger are accompanied by a cough productive of frothy and often blood-tinged sputum—the effect of air mixing with serum albumin and red blood cells that have moved into the alveoli. The movement of air through the alveolar fluid produces fine crepitant sounds called *crackles*, which can be heard through a stethoscope placed on the chest. As fluid moves into the larger airways, the breathing becomes louder. The crackles heard earlier become louder and coarser. In the terminal stage the breathing pattern is called the *death rattle*. Persons with severe pulmonary edema literally drown in their own secretions.

Treatment

Treatment of acute pulmonary edema is directed toward reducing the fluid volume in the pulmonary circulation. This can be accomplished by reducing the amount of blood that the right heart delivers to the lungs or by improving the work performance of the left heart. Several measures can decrease the blood volume in the pulmonary circulation; the seriousness of the pulmonary edema determines which are used. One of the simplest measures to relieve orthopnea is assumption of the seated position, in which gravity causes blood to be redistributed to the lower extremities. For many persons, sitting up or standing is almost instinctive and may be sufficient to relieve the symptoms associated with mild accumulation of fluid.

Measures to improve left heart performance focus on decreasing the preload by reducing the filling pressure of the left ventricle and on reducing the afterload against which the left heart must pump. This can be accomplished through the use of diuretics, vasodilator drugs, treatment of arrhythmias that impair cardiac function, and improvement of the contractile properties of the left ventricle with digitalis. Rapid digitalization can be accomplished with intravenous administration of the drug.

Oxygen therapy increases the oxygen content of the blood and helps relieve anxiety. Positive-pressure breathing, which is administered through a specially designed mask, increases the intra-alveolar pressure, opposes the capillary filtration pressure in the pulmonary capillaries, and sometimes is used as a temporary measure to decrease the amount of fluid moving into the alveoli. However, in the most severe cases, endotracheal intubation and mechanical ventilation may be necessary. Although its mechanisms of action are unclear, morphine sulfate usually is the drug of choice in acute pulmonary edema. Morphine relieves anxiety and depresses the pulmonary reflexes that cause spasm of the pulmonary vessels. It also increases venous pooling by vasodilatation.

Cardiogenic Shock

Cardiogenic shock refers to the pronounced failure of the heart as a pump. Cardiogenic shock can occur relatively quickly because of the damage to the heart that occurs during myocardial infarction; ineffective pumping caused by cardiac dysrhythmias; mechanical defects that may occur as a complication of myocardial infarction, such as ventricular septal defect; ventricular aneurysm; acute disruption of valvular function; or problems associated with open heart surgery. Cardiogenic shock also may ensue as an end-stage condition of coronary artery disease or cardiomyopathy.

The most common cause of cardiogenic shock is myocardial infarction. Most patients who die of cardiogenic shock have lost at least 40% of the contracting muscle of the left ventricle because of a recent infarct or a combination of recent and old infarcts.¹⁶ Cardiogenic shock can follow other types of shock associated with inadequate coronary blood flow.

In all cases of cardiogenic shock, there is failure to eject blood from the heart, hypotension, and inadequate cardiac output. Increased systemic vascular resistance often contributes to the deterioration of cardiac function by increasing afterload or the resistance to ventricular systole. The filling pressure, or preload of the heart, also is increased as blood returning to the heart is added to blood that previously was returned but not pumped forward, resulting in an increase in end-systolic ventricular volume. Increased resistance to ventricular systole (*i.e.*, afterload) combined with the decreased myocardial contractility causes the increased end-systolic ventricular volume and increased preload, which further complicate cardiac status.

Manifestations

The signs and symptoms of cardiogenic shock are consistent with those of extreme heart failure. The lips, nail beds, and skin are cyanotic because of stagnation of blood flow and increased extraction of oxygen from the hemoglobin as it passes through the capillary bed. The CVP and PCWP rise as a result of volume overload caused by the pumping failure of the heart.

Treatment

Treatment of cardiogenic shock requires a precarious balance between improving cardiac output, reducing the workload and oxygen needs of the myocardium, and preserving coronary perfusion. Fluid volume must be regulated within a level that maintains the filling pressure (*i.e.*, venous return) of the heart and maximum use of the Frank-Starling mechanism without causing pulmonary congestion.

Pharmacologic treatment includes the use of vasodilator drugs such as nitroprusside and nitroglycerin that cause arterial and venous dilatation. These drugs produce venous and arterial dilation, thus decreasing the venous return to the heart and the arterial resistance against which the left heart must pump. Catecholamines increase cardiac contractility but must be used with caution because they also produce vasoconstriction and increase cardiac workload by increasing the afterload.

The *intra-aortic balloon pump* may be used for persons who do not experience response to medical treatment. It provides a means of increasing aortic diastolic pressure and enhancing coronary and peripheral blood flow without increasing systolic pressure and the afterload, against which the left ventricle must pump.¹⁶ The device, which pumps in synchrony with the heart, consists of a balloon that is inserted through a catheter into the descending aorta (Fig. 18-7). The balloon is filled with helium

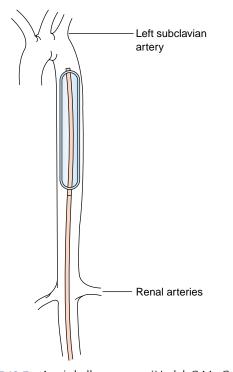


FIGURE 18-7 Aortic balloon pump. (Hudak C.M., Gallo B.M. [1994]. *Critical care nursing* [6th ed.]. Philadelphia: J.B. Lippincott)

and is timed to inflate during ventricular diastole and deflate just before ventricular systole. Diastolic inflation creates a pressure wave in the ascending aorta that increases coronary artery flow and a less intense wave in the lower aorta that enhances organ perfusion. The sudden balloon deflation at the onset of systole lowers the resistance to ejection of blood from the left ventricle, thereby increasing the heart's pumping efficiency and decreasing myocardial oxygen consumption.

Mechanical Support of the Failing Heart and Heart Transplantation

Refractory heart failure reflects deterioration in cardiac function that is unresponsive to medical or surgical interventions. With improved methods of treatment, more people are reaching a point where a cure is unachievable and death is imminent without mechanical support or heart transplantation.

Since the early 1960s, significant progress has been made in improving the efficacy of *ventricular assist devices* (VADs), which are mechanical pumps used to support ventricular function. VADs are used to decrease the workload of the myocardium while maintaining cardiac output and systemic arterial pressure.¹⁷ This decreases the workload on the ventricle and allows it to rest and recover.

Heart transplantation remains the treatment of choice for many persons with end-stage cardiac failure. The number of successful heart transplantations has been steadily climbing, with more than 2800 procedures performed per year. Patients with heart transplants who are treated with triple-immunosuppressant therapy have a 5-year survival rate of 70% to 80%.¹⁸ Despite the overall success of heart transplantation, donor availability and complications associated with infection, rejection, and immunosuppression drug therapy remain problems.

In summary, heart failure occurs when the heart fails to pump sufficient blood to meet the metabolic needs of body tissues. The physiology of heart failure reflects an interplay between a decrease in cardiac output that accompanies impaired function of the failing heart and the compensatory mechanisms designed to preserve the cardiac reserve. Compensatory mechanisms that contribute to maintenance of the cardiac reserve include the Frank-Starling mechanism, sympathetic nervous system responses, the renin-angiotensinaldosterone mechanism, and myocardial hypertrophy. In the failing heart, early decreases in cardiac function may go unnoticed because these compensatory mechanisms maintain the cardiac output. This is called compensated heart failure. Unfortunately, the mechanisms were not intended for longterm use, and in severe and prolonged heart failure, the compensatory mechanisms no longer are effective and further impair cardiac function.

Heart failure may be described as high-output or lowoutput failure, systolic or diastolic failure, and right-sided or left-sided failure. With high-output failure, the function of the heart may be supernormal but inadequate because of excessive metabolic needs, and low-output failure is caused by disorders that impair the pumping ability of the heart. With systolic dysfunction, there is impaired ejection of blood from the heart during systole; with diastolic dysfunction, there is impaired filling of the heart during diastole. Right-sided failure is characterized by congestion in the peripheral circulation, and left-sided failure by congestion in the pulmonary circulation.

The manifestations of heart failure include edema, nocturia, fatigue and impaired exercise tolerance, cyanosis, signs of increased sympathetic nervous system activity, and impaired gastrointestinal function and malnutrition. In rightsided failure, there is dependent edema of the lower parts of the body, engorgement of the liver, and ascites. In left-sided failure, shortness of breath and chronic, nonproductive cough are common.

Acute pulmonary edema is a life-threatening condition in which the accumulation of fluid in the interstitium of the lung and alveoli interferes with lung expansion and gas exchange. It is characterized by extreme breathlessness, crackles, frothy sputum, cyanosis, and signs of hypoxemia. In cardiogenic shock, there is failure to eject blood from the heart, hypotension, inadequate cardiac output, and impaired perfusion of peripheral tissues. Mechanical support devices, including the intra-aortic balloon pump (for acute failure) and the VAD, sustain life in persons with severe heart failure. Heart transplantation remains the treatment of choice for many persons with end-stage heart failure.

CIRCULATORY FAILURE (SHOCK)

The functions of the circulatory system are to perfuse body tissues and supply them with oxygen. Whereas heart failure results from impaired ability of the heart as a pump, circulatory shock results from a failure of the circulatory system to supply the peripheral tissues and organs of the body with an adequate blood supply. As with heart failure, circulatory shock is not a specific disease but can occur in the course of many life-threatening traumatic or disease states. In situations of severe cardiovascular compromise, signs of heart failure and vascular compromise may coexist. Although circulatory shock produces hypotension, it should not be equated with a drop in blood pressure. Hypotension often is a late sign and indicates a failure of compensatory mechanisms.

Adequate perfusion of body tissues depends on the pumping ability of the heart, a vascular system that transports blood to the cells and back to the heart, sufficient blood to fill the vascular system, and tissues that are able to use and extract oxygen and nutrients from the blood. As with heart failure, circulatory shock produces compensatory physiologic responses that eventually decompensate into various shock states if the condition is not properly treated in a timely manner.

Types of Shock

Circulatory shock is used to describe a critical decrease in tissue perfusion caused by a loss or redistribution of intravascular fluid. It can be classified as hypovolemic, obstructive, or distributive. These three main types of shock are summarized in Chart 18-1 and depicted in Figure 18-8. Cardiogenic shock, which results from failure of the heart as a pump, was discussed earlier in the chapter.

KEY CONCEPTS

CIRCULATORY SHOCK

- Circulatory shock represents the inability of the circulation to adequately perfuse the tissues of the body.
- It can result from a loss of fluid from the vascular compartment, an increase in the size of the vascular compartment that interferes with the distribution of blood, or obstruction of flow through the vascular compartment.
- The manifestations of shock reflect both the impaired perfusion of body tissues and the body's attempt to maintain tissue perfusion through conservation of water by the kidney, translocation of fluid from the extracellular to the intravascular compartment, and activation of sympathetic nervous system mechanisms that increase heart rate and divert blood from less to more essential body tissues.

Hypovolemic Shock

Hypovolemic shock is characterized by diminished blood volume such that there is inadequate filling of the vascular compartment (see Fig. 18-9). It occurs when there is an acute loss of 15% to 20% of the circulating blood volume. The decrease may be caused by an external loss of whole blood (*e.g.*, hemorrhage), plasma (*e.g.*, severe burns), or extracellular fluid (*e.g.*, gastrointestinal fluids lost in vomiting or diarrhea). Hypovolemic shock also can result from an internal hemorrhage or from third-space losses, when extracellular fluid is shifted from the vascular compartment to the interstitial space or compartment.

Hypovolemic shock has been the most widely studied type of shock and usually serves as a prototype in discussions of the manifestations of shock. Approximately 10% of the total

CHART 18-1 Classification of Circulatory Shock

Hypovolemic

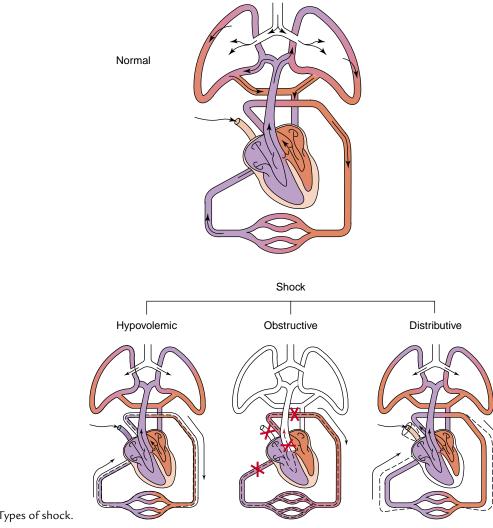
Loss of whole blood Loss of plasma Loss of extracellular fluid

Obstructive

Inability of the heart to fill properly (cardiac tamponade) Obstruction to outflow from the heart (pulmonary embolus, cardiac myxoma, pneumothorax, or dissecting aneurysm)

Distributive

Loss of sympathetic vasomotor tone Presence of vasodilating substances in the blood (anaphylactic shock) Presence of inflammatory mediators (septic shock)



■ **FIGURE 18-8** ■ Types of shock.

blood volume can be removed without changing the cardiac output or arterial pressure.¹⁹ The average blood donor loses a pint of blood without experiencing adverse effects. As increasing amounts of blood (10% to 25%) are removed, the cardiac output falls, but the arterial pressure is maintained because of sympathetic-mediated increases in heart rate and vasoconstriction. Blood pressure is the product of cardiac output and peripheral vascular resistance (BP = CO × PVR); an increase in peripheral vascular resistance maintains blood pressure in the presence of decreased cardiac output for a short period. Cardiac output and tissue perfusion decrease before signs of hypotension occur. The arterial pressure falls to zero when approximately 35% to 45% of the total blood volume has been removed.¹⁹

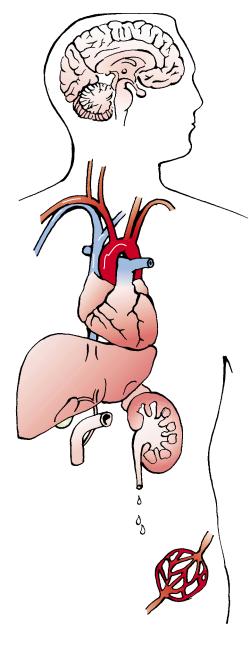
Manifestations. The signs and symptoms of hypovolemic shock depend on shock stage and are closely related to low peripheral blood flow and excessive sympathetic stimulation. They include thirst, an increase in heart rate, cool and clammy skin, a decrease in arterial blood pressure, a decrease in urine output, and changes in mentation (Fig. 18-9). Laboratory tests of hemoglobin and hematocrit provide information regarding

the severity of blood loss or hemoconcentration caused by dehydration. Serum lactate and arterial pH provide information about the severity of acidosis.

An increase in heart rate is often an early sign of shock. As shock progresses, the pulse becomes weak and thready, indicating vasoconstriction and a reduction in filling of the vascular compartment, and the respirations become rapid and deep. Arterial blood pressure is decreased in moderate to severe shock. However, controversy exists regarding the value of blood pressure measurements in the early diagnosis and management of shock. This is because compensatory mechanisms tend to preserve blood pressure until shock is relatively far advanced. Furthermore, an adequate arterial pressure does not ensure adequate perfusion and oxygenation of vital organs at the cellular level. This does not imply that blood pressure should not be closely monitored in patients at risk for development of shock, but it does indicate the need for other assessment measures by which shock may be detected at an earlier stage.

Decreased intravascular volume results in decreased venous return to the heart and a decrease in CVP. When shock becomes severe, the peripheral veins collapse, making it difficult to insert peripheral venous lines. Sympathetic stimulation also leads

332 Unit Four: Alterations in the Cardiovascular System



Brain

- Thirst
- Increased production and release ADH

Cardiovascular system

- Increased heart rate
- Increased force of cardiac contraction
- Increased systemic vascular resistance
- · Decreased blood flow to the
- kidney
- gastrointestinal tract
- skin
- skeletal muscles
- · Constriction of the veins

Adrenal gland

- Increased production and release of aldosterone by the adrenal cortex
- Increased production and release of the catecholamines (epinephrine and norepinephrine) by the adrenal medulla
- Liver
- Constriction of veins and sinusoids with mobilization of blood stored in the liver

Kidney

- Increased retention of sodium and water
- · Decreased urine output

Capillary bed

 Increased reabsorption of water into the capillary from the interstitial spaces due to constriction of the arterioles with a resultant decrease in capillary pressure

■ **FIGURE 18-9** ■ Compensatory mechanisms in hypovolemic shock.

to intense vasoconstriction of the skin vessels and activation of the sweat glands. As a result, the skin is cool and moist. When shock is caused by hemorrhage, the loss of red blood cells leaves the skin and mucous membranes looking pale.

Urine output decreases very quickly in hypovolemic and other forms of shock. Compensatory mechanisms decrease renal blood flow as a means of diverting blood flow to the heart and brain. Oliguria of 20 mL/hour or less indicates severe shock and inadequate renal perfusion. Continuous measurement of urine output is essential for assessing the circulatory status of the person in shock.

Restlessness and apprehension are common behaviors in early shock. As the shock progresses and blood flow to the brain decreases, restlessness is replaced by apathy and stupor. If shock is unchecked, the apathy progresses to coma. Coma caused by blood loss alone and not related to head injury or other factors is an unfavorable sign. *Treatment.* The treatment of hypovolemic shock is directed toward correcting or controlling the underlying cause and improving tissue perfusion. Persons who have sustained blood loss are commonly placed in supine position with the legs elevated to maximize cerebral blood flow. Oxygen is administered to persons with signs of hypoxemia. Because subcutaneous administration is unpredictable, pain medications usually are administered intravenously. Frequent measurements of heart rate and cardiac rhythm, blood pressure, and urine flow are used to assess the severity of circulatory compromise and to monitor treatment.

In hypovolemic shock, the goal of treatment is to restore vascular volume. This can be accomplished through intravenous administration of fluids and blood. The crystalloids (*e.g.*, isotonic saline and glucose solutions) are readily available for emergencies and mass casualties. They often are effective, at least temporarily, when given in adequate doses. Blood or

blood products (packed or frozen red cells) are administered based on hematocrit and hemodynamic findings. Fluids and blood are best administered based on volume indicators such as CVP and PCWP.

Vasoactive drugs (*e.g.*, adrenergic agents) are agents capable of constricting or dilating blood vessels. As a general rule, the adrenergic drugs are not used as a primary form of therapy in hypovolemic shock. Simple blood pressure elevation produced by vasopressor drugs has little effect on the underlying cause of shock and in many cases may be detrimental. These agents are given only when hypotension persists after volume deficits have been corrected.

Dopamine, which induces more favorable actions than many of the other adrenergic drugs, may be used in the treatment of severe and prolonged shock. Dopamine, when given in low doses, is thought to increase blood flow to the kidneys, liver, and other abdominal organs while maintaining vasoconstriction in less vital structures, such as the skin and skeletal muscles. In severe shock, higher doses may be needed to maintain blood pressure. After dopamine administration exceeds this low-dose range, it has vasoconstrictive effects on blood flow to the kidneys and abdominal organs that are similar to those of epinephrine.

Obstructive Shock

The term *obstructive shock* is used to describe circulatory shock that results from mechanical obstruction of the flow of blood through the central circulation (great veins, heart, or lungs; see Fig. 18-8). Obstructive shock may be caused by a number of conditions, including dissecting aortic aneurysm, cardiac tamponade, pneumothorax, atrial myxoma, or evisceration of abdominal contents into the thoracic cavity because of a ruptured hemidiaphragm. The most common cause of obstructive shock is pulmonary embolism.

The primary physiologic results of obstructive shock are elevated right heart pressure and impaired venous return to the heart. The signs of right heart failure are seen, including elevation of central venous pressure and jugular vein distention. Treatment modalities focus on correcting the cause of the disorder, frequently with surgical interventions such as pulmonary embolectomy, pericardiocentesis (*i.e.*, removal of fluid from the pericardial sac) for cardiac tamponade, or the insertion of a chest tube for correction of a tension pneumothorax or hemothorax. In select cases of pulmonary embolus, thrombolytic drugs may be used to dissolve the clots causing the obstruction.

Distributive Shock

Distributive shock is characterized by loss of blood vessel tone, enlargement of the vascular compartment, and displacement of the vascular volume away from the heart and central circulation. With distributive shock, the capacity of the vascular compartment expands to the extent that a normal volume of blood does not fill the circulatory system (see Fig. 18-8). Loss of vessel tone has two main causes: a decrease in the sympathetic control of vasomotor tone and the presence of vasodilator substances in the blood. Venous return is decreased in distributive shock, which leads to a diminished cardiac output but not a decrease in total blood volume; this type of shock is also referred to as *normovolemic shock*. There are two shock states that share the basic circulatory pattern of distributive shock: neurogenic shock and anaphylactic shock. Septic shock shares many of the features of distributive shock.

Neurogenic Shock. Neurogenic shock describes shock caused by decreased sympathetic control of blood vessel tone caused by a defect in the vasomotor center in the brain stem or the sympathetic outflow to the blood vessels. Output from the vasomotor center can be interrupted by brain injury, the depressant action of drugs, general anesthesia, hypoxia, or lack of glucose (e.g., insulin reaction). Fainting attributable to emotional causes is a transient form of neurogenic shock. Spinal anesthesia or spinal cord injury above the midthoracic region can interrupt the transmission of outflow from the vasomotor center. The term spinal shock is used to describe the neurogenic shock that occurs in persons with spinal cord injury. Many general anesthetic agents can cause a neurogenic shock-like reaction, especially during induction, because of interference with sympathetic nervous system function. In contrast to hypovolemic shock, the heart rate in neurogenic shock often is slower than normal, and the skin is dry and warm. This type of distributive shock is rare and usually transitory.

Anaphylactic Shock. Anaphylactic shock is characterized by massive vasodilatation, pooling of blood in the peripheral blood vessels, and increased capillary permeability.²⁰ This type of shock, which is a manifestation of systemic anaphylaxis, is caused by an immune-mediated reaction in which vasodilator substances such as histamine are released into the blood (see Chapter 10). These substances cause dilatation of arterioles and venules along with a marked increase in capillary permeability. The vascular response in anaphylactic shock is often accompanied by bronchospasm, contraction of gastrointestinal and uterine smooth muscle, and urticaria or angioedema (swelling of the face and throat).

Among the most common causes of anaphylactic shock are reactions to drugs, such as penicillin; foods, such as nuts and shellfish; and insect venoms. The most common cause is stings from insects of the order Hymenoptera (*i.e.*, bees, wasps, and fire ants). Latex allergy has caused life-threatening anaphylaxis in a growing segment of the population (see Chapter 10).²¹ Health care workers and others who are exposed to latex are developing latex sensitivities that range from mild urticaria, contact dermatitis, and mild respiratory distress to anaphylactic shock.²⁰ Children with spina bifida also are at extreme risk for this increasingly serious allergy.

The onset of anaphylaxis depends on the sensitivity of the person and the rate and quantity of antigen exposure. Anaphylactic shock often develops suddenly; death can occur within a matter of minutes unless appropriate medical intervention is promptly instituted. Signs and symptoms of impending anaphylactic shock include abdominal cramps; apprehension; burning and warm sensation of the skin, itching, urticaria (*i.e.*, hives); coughing, choking, wheezing, chest tightness, and difficulty in breathing. After blood begins to pool peripherally, there is a precipitous drop in blood pressure, and the pulse becomes so weak that it is difficult to detect. Life-threatening airway obstruction may ensue as a result of laryngeal edema or bronchial spasm.

Treatment includes immediate discontinuance of the inciting agent or institution of measures to decrease its absorption (*e.g.*, application of ice); close monitoring of cardiovascular and respiratory function; and maintenance of adequate respiratory gas exchange, cardiac output, and tissue perfusion. Epinephrine constricts the blood vessels and relaxes the smooth muscle in the bronchioles; it usually is the first drug to be given to a patient believed to be experiencing an anaphylactic reaction. Other treatment measures include the administration of oxygen, antihistaminic drugs, and corticosteroids. Resuscitation measures may be required.

The prevention of anaphylactic shock is preferable to treatment. Once a person has been sensitized to an antigen, the risk of repeated anaphylactic reactions with subsequent exposure is high. Persons with known hypersensitivities should carry some form of medical identification to alert medical personnel if they become unconscious or unable to relate this information. Persons who are risk for anaphylaxis should be provided with emergency medications (*e.g.*, epinephrine autoinjector) and instructed in procedures to follow in case they are inadvertently exposed to the offending antigen.

Sepsis and Septic Shock

Septic shock is associated with severe infection and the systemic response to infection. It is associated most frequently with gram-negative bacteremia, although it can be caused by gram-positive bacilli and other microorganisms such as fungi, which carry an even greater risk of mortality.²² Unlike other types of shock, septic shock commonly is associated with pathologic complications, such as acute respiratory distress syndrome, disseminated intravascular coagulation, and multiple organ dysfunction syndrome.

Septic shock has become the most common type of distributive shock. There are approximately 400,000 to 500,000 septic episodes each year in the United States. The growing incidence has been attributed to an increased awareness of the diagnosis, increased numbers of immunocompromised patients, increased use of invasive procedures, increased number of resistant organisms, and an increased number of elderly patients with critical illnesses.²³ Despite advances in treatment methods, the mortality rate is approximately 40%.²⁴

Septic shock has been described in the context of the *systemic inflammatory response*.²⁵ Although usually associated with infection, the systemic inflammatory response syndrome can be initiated by noninfectious disorders such as acute trauma and pancreatitis.

Mechanisms. The mechanisms of septic shock are thought to be related to mediators of the inflammatory response.^{26,27} Although the immune system and the inflammatory response are designed to overcome infection, the dysregulated release of inflammatory mediators such as tumor necrosis factor (TNF) and the interleukins (see Chapter 10) may initiate the potentially fatal sepsis syndrome. These inflammatory mediators, which are released in the presence of bacterial toxins, promote endothelial cell–leukocyte adhesion, release of cell-damaging proteases and prostaglandins, and activation of the blood coagulation cascade. The prostaglandins participate in the generation of fever, tachycardia, ventilation-perfusion abnormalities, and lactic acidosis that occur with septic shock.²⁸

In addition to inducing the release of inflammatory mediators, endotoxins may themselves induce tissue damage by directly activating pathways such as the coagulation cascade, the complement cascade, vessel injury, or release of vasodilating prostaglandins.²⁷ Thus, the sepsis syndrome represents the complex consequences of microbial products that produce profound dysregulation of the inflammatory response.

Manifestations. Septic shock typically manifests with fever, vasodilatation, and a warm, flushed skin. Mild hyperventilation, respiratory alkalosis, and abrupt changes in personality and behavior caused by reduction in cerebral blood flow may be the earliest signs and symptoms of septic shock. These manifestations, which are thought to be a primary response to the bacteremia, commonly precede the usual signs and symptoms of sepsis by several hours or days.

Unlike other forms of shock (*i.e.*, cardiogenic, hypovolemic, and obstructive) that are characterized by a compensatory increase in systemic vascular resistance, septic shock often presents with hypovolemia because of arterial and venous dilatation and leakage of plasma into the interstitial spaces. Aggressive treatment of the hypovolemia in septic shock leads to a decrease in systemic vascular resistance and increased cardiac output and tachycardia. With the development of refined resuscitation methods and better hemodynamic monitoring systems, approximately 90% of patients in septic shock convert to this hyperdynamic response with high cardiac output and low systemic vascular resistance.²⁷ However, as the condition progresses, cardiac function is depressed; the heart becomes dilated, and the ejection fraction decreases.

Treatment. The treatment of septic shock focuses on the causative agent and support of the circulation. The administration of antibiotics specific to the infectious agent is essential. The cardiovascular status of the patient must be supported to maintain oxygen delivery to the cells. Swift and aggressive fluid administration is needed to compensate for third spacing, and equally aggressive use of vasopressor agents is needed to counteract the vasodilation caused by endotoxins.

Complications of Shock

Wiggers, a noted circulatory physiologist, stated, "Shock not only stops the machine, but it wrecks the machinery."²⁹ Many body systems are wrecked by severe shock. Some of the major complications of severe shock are acute respiratory distress syndrome, acute renal failure, gastrointestinal ulceration, disseminated intravascular coagulation, and multiple organ dysfunction syndrome.

Acute Respiratory Distress Syndrome

Acute respiratory distress syndrome (ARDS) is a potentially lethal form of respiratory failure that can follow severe shock (see Chapter 21). The symptoms of ARDS usually do not develop until 24 to 48 hours after the initial trauma; in some instances, they occur much later. The respiratory rate and effort of breathing increase, and arterial blood gas analysis demonstrates the presence of profound hypoxemia with hypercapnia.

The exact cause of ARDS is unknown. Neutrophils are thought to play a key role in the pathogenesis of ARDS. A cytokine-mediated activation and accumulation of neutrophils in the pulmonary vasculature and subsequent endothelial injury is thought to cause leaking of fluid and plasma proteins into the interstitium and alveolar spaces.^{30,31} The fluid leakage impairs gas exchange and makes the lung stiffer and more difficult to inflate. Abnormalities in the production, composition, and function of surfactant may contribute to alveolar collapse and gas exchange abnormalities.³¹

Interventions for ARDS focus on increasing the oxygen concentration in the inspired air and supporting ventilation mechanically to optimize gas exchange while avoiding oxygen toxicity and preventing further lung injury.^{30,31} Despite the delivery of high levels of oxygen using high-pressure mechanical ventilatory support and positive end-expiratory pressure, many persons with ARDS remain hypoxic, often with a fatal outcome.

Acute Renal Failure

The renal tubules are particularly vulnerable to ischemia, and acute renal failure is one important late cause of death in severe shock. Sepsis and trauma account for most cases of acute renal failure (see Chapter 24). The endotoxins implicated in septic shock are powerful vasoconstrictors that are capable of activating the sympathetic nervous system and causing intravascular clotting. They have been shown to trigger all the separate physiologic mechanisms that contribute to the onset of acute renal failure. The degree of renal damage is related to the severity and duration of shock. The normal kidney is able to tolerate severe ischemia for 15 to 20 minutes. The renal lesion most frequently seen after severe shock is acute tubular necrosis. Acute tubular necrosis usually is reversible, although return to normal renal function may require weeks or months. Continuous monitoring of urine output during shock provides a means of assessing renal blood flow. Frequent monitoring of serum creatinine and blood urea nitrogen levels also provides valuable information regarding renal status.

Gastrointestinal Complications

The gastrointestinal tract is particularly vulnerable to ischemia because of the changes in distribution of blood flow to its mucosal surface. In shock, there is widespread constriction of blood vessels that supply the gastrointestinal tract, causing a redistribution of blood flow that severely diminishes mucosal perfusion. Superficial mucosal lesions of the stomach and duodenum can develop within hours of severe trauma, sepsis, or burn.

Bleeding is a common symptom of gastrointestinal ulceration caused by shock. Hemorrhage has its onset usually within 2 to 10 days after the original insult and often begins without warning. Poor perfusion in the gastrointestinal tract has been credited with allowing intestinal bacteria to enter the bloodstream, thereby contributing to the development of sepsis and shock.

Histamine type 2 receptor antagonists, proton pump inhibitors, or mucosal protective agents may be given prophylactically to prevent gastrointestinal ulcerations caused by shock. Nasogastric tubes, when attached to intermittent suction, also help to diminish the accumulation of hydrogen ions in the stomach.

Disseminated Intravascular Coagulation

Disseminated intravascular coagulation (DIC) is characterized by widespread activation of the coagulation system with resultant formation of fibrin clots and thrombotic occlusion of small and midsize vessels (see Chapter 12). The systemic formation of fibrin results from increased generation of thrombin, the simultaneous suppression of physiologic anticoagulation mechanisms, and the delayed removal of fibrin as a consequence of impaired fibrinolysis. Clinically overt DIC is reported to occur in as many as 30% to 50% of persons with sepsis and septic shock.³² As with other systemic inflammatory responses, the derangement of coagulation and fibrinolysis is thought to be mediated by inflammatory mediators.

The contribution of DIC to morbidity and mortality in sepsis depends on the underlying clinical condition and the intensity of the coagulation disorder. Depletion of the platelets and coagulation factors increases the risk of bleeding. Deposition of fibrin in the vasculature of organs contributes to ischemic damage and organ failure.

The management of sepsis-induced DIC focuses on treatment of the underlying disorder and measures to interrupt the coagulation process. Anticoagulation therapy and administration of platelets and plasma may be used.

Multiple Organ Dysfunction Syndrome

Multiple organ dysfunction syndrome (MODS) represents the presence of altered organ function in an acutely ill patient such that homeostasis cannot be maintained without intervention. As the name implies, MODS commonly affects multiple organ systems, including the kidneys, lungs, liver, brain, and heart. MODS is a particularly life-threatening complication of shock, especially septic shock. It has been reported as the most frequent cause of death in the noncoronary intensive care unit. Mortality rates vary from 30% to 100%, depending on the number of organs involved.33 Mortality rates increase with an increased number of organs failing. A high mortality rate is associated with failure of the brain, liver, kidney, and lung. The pathogenesis of MODS is not clearly understood, so current management is primarily supportive. Major risk factors for the development of MODS are sepsis, shock, prolonged periods of hypotension, hepatic dysfunction, trauma, infarcted bowel, advanced age, and alcohol abuse.³³ Interventions for multiple organ failure are focused on support of the affected organs.

In summary, circulatory shock is an acute emergency in which body tissues are deprived of oxygen and cellular nutrients or are unable to use these materials in their metabolic processes. Circulatory shock may develop because there is not enough blood in the circulatory system (i.e., hypovolemic shock), blood flow or venous return is obstructed (i.e., obstructive shock), or there is excessive vasodilation and pooling of blood (i.e., distributive shock). Two types of shock share the circulatory features of distributive shock: neurogenic shock and anaphylactic shock. Septic shock, which shares many of the features of distribution shock, is associated with a severe, overwhelming infection and has a mortality rate of approximately 40%.

The manifestations of circulatory shock are related to low peripheral blood flow and excessive sympathetic stimulation. The low peripheral blood flow produces thirst, changes in skin temperature, a decrease in blood pressure, an increase in heart rate, decreased venous pressure, decreased urine output, and changes in the sensorium. Signs and symptoms, such as changes in skin temperature (i.e., increased in septic shock and decreased in hypovolemic and other forms of shock), may differ with the type of shock. The intense vasoconstriction that serves to maintain blood flow to the heart and brain causes a decrease in tissue perfusion, impaired cellular metabolism, liberation of lactic acid, and, eventually, cell death. Whether the shock will be irreversible or the patient will survive is determined largely by changes that occur at the cellular level.

The complications of shock result from the deprivation of blood flow to vital organs or systems, such as the lungs, kidneys, gastrointestinal tract, and blood coagulation system. ARDS is characterized by changes in the permeability of the alveolar-capillary membrane with the development of interstitial edema and severe hypoxia that does not respond to oxygen therapy. The renal tubules are particularly vulnerable to ischemia, and acute renal failure is an important complication of shock. Gastrointestinal ischemia may lead to gastrointestinal bleeding and increased permeability to the intestinal bacteria, which cause further sepsis and shock. DIC is characterized by formation of small clots in the circulation. MODS, perhaps the most ominous complication of shock, rapidly depletes the body's ability to compensate and recover from a shock state.

HEART FAILURE IN CHILDREN AND THE ELDERLY

Heart Failure in Infants and Children

As in adults, heart failure in infants and children results from the inability of the heart to maintain the cardiac output required to sustain metabolic demands.^{34,35} Congenital heart defects are the most common cause of CHF during childhood. Surgical correction of congenital heart defects may cause heart failure as a result of intraoperative manipulation of the heart and resection of heart tissue, with subsequent alterations in pressure, flow, and resistance relations. Usually, the heart failure that results is acute and resolves after the effects of the surgical procedure have subsided. Chronic congestive failure occasionally is observed in children with severe chronic anemia, inflammatory heart disease, end-stage congenital heart disease, or cardiomyopathy. Inflammatory heart disorders (e.g., myocarditis, rheumatic fever, bacterial endocarditis, Kawasaki's disease), cardiomyopathy, and congenital heart disorders are discussed in Chapter 17.

Manifestations

Many of the signs and symptoms of heart failure in infants and children are similar to those in adults. They include fatigue, effort intolerance, cough, anorexia, and abdominal pain. A subtle sign of cardiorespiratory distress in infants and children is a change in disposition or responsiveness, including irritability or lethargy. Sympathetic stimulation produces peripheral vasoconstriction and diaphoresis. Decreased renal blood flow often results in a urine output of less than 0.5 to 1.0 mL/kg/hour, despite adequate fluid intake. When right ventricular function is impaired, systemic venous congestion develops. Hepatomegaly caused by liver congestion often is one of the first signs of systemic venous congestion in infants and children. However, dependent edema or ascites rarely is seen unless the CVP is extremely high. Because of their short, fat necks, jugular venous distention is difficult to detect in infants; it is not a reliable sign until the child is of school age or older.

Most commonly, children experience interstitial edema, rather than alveolar pulmonary edema. This reduces lung compliance and increases the work of breathing, causing tachypnea and increased respiratory effort. Older children display use of accessory muscles (*i.e.*, scapular and sternocleidomastoid). Head bobbing and nasal flaring may be observed in infants. Signs of respiratory distress often are the first and most noticeable indication of heart failure in infants and young children. Pulmonary congestion may be mistaken for bronchiolitis or lower respiratory tract infection. The infant or young child with respiratory distress often grunts with expiration. This grunting effort (essentially, exhaling against a closed glottis) is an instinctive effort to increase end-expiratory pressures and prevent collapse of small airways and the development of atelectasis. Respiratory crackles are uncommon in infants and usually suggest development of a respiratory tract infection. Wheezes may be heard, particularly if there is a large left-to-right shunt.

Infants with heart failure often have increased respiratory problems during feeding.³⁴ The history is one of prolonged feeding with excessive respiratory effort and fatigue. Weight gain is slow because of high energy requirements and low calorie intake. Other frequently occurring manifestations of heart failure in infants are excessive sweating (caused by increased sympathetic tone), particularly over the head and neck, and repeated lower respiratory tract infections. Peripheral perfusion usually is poor, with cool extremities; tachycardia is common (resting heart rate >150 beats per minute); and respiratory rate is increased (resting rate >50 breaths per minute).

Diagnosis and Treatment

Diagnosis of congestive failure in infants and children is based on symptomatology, chest radiographic films, ECG findings, echocardiographic techniques to assess cardiac structures and ventricular function (*i.e.*, end-systolic and end-diastolic diameters), arterial blood gases to determine intracardiac shunting and ventilation-perfusion inequalities, and other laboratory studies to determine anemia and electrolyte imbalances.

Treatment of congestive failure in infants and children includes measures aimed at improving cardiac function and eliminating excess intravascular fluid. Oxygen delivery must be supported and oxygen demands controlled or minimized. When possible, the cause of the disorder is corrected (*e.g.*, medical treatment of sepsis and anemia, surgical correction of congenital heart defects). With congenital anomalies that are amenable to surgery, medical treatment often is needed for a time before surgery and usually is continued during the immediate postoperative period. For many children, only medical management can be provided.

Medical management of heart failure in infants and children is similar to that in the adult, although it is tailored to the special developmental needs of the child. Inotropic agents such as digitalis often are used to increase cardiac contractility. Diuretics may be given to reduce preload and vasodilating drugs used to manipulate the afterload. Drug doses must be carefully tailored to control for the child's weight and conditions such as reduced renal function. Daily weighing and accurate measurement of intake and output are imperative during acute episodes of failure.

Most children feel better in the semiupright position. An infant seat is useful for infants with chronic CHF. Activity restrictions usually are designed to allow children to be as active as possible within the limitations of their heart disease. Infants with congestive failure often have problems feeding. Small, frequent feedings usually are more successful than larger, less frequent feedings. Severely ill infants may lack sufficient strength to suck and may need to be tube fed.

Heart Failure in the Elderly

Heart failure is one of the most common causes of disability in the elderly and is the most common hospital discharge diagnosis for the elderly. More than 75% of patients with heart failure are older than 65 years.³⁶ Among the factors that have contributed to the increased numbers of older people with heart failure are the improved therapies for ischemic and hypertensive heart disease. Thus, persons who would have died of acute myocardial disease 20 years ago are now surviving, but with residual left ventricular dysfunction. Similarly, improved blood pressure control has led to a 60% decline in stroke mortality rates, yet these same people remain at risk for CHF as a complication of hypertension. In addition, advances in treatment of other diseases have contributed indirectly to the rising prevalence of heart failure in the older population.

Coronary heart disease, hypertension, and valvular heart disease (particularly aortic stenosis and mitral regurgitation) are common causes of heart failure in older adults.^{36,37} Although the pathophysiology of heart failure is similar in younger and older persons, elderly persons tend to experience cardiac failure when confronted with stresses that would not produce failure in younger persons. There are four principal changes associated with cardiovascular aging that impair the ability to respond to stress.³⁶ First, reduced responsiveness to β-adrenergic stimulation limits the heart's capacity maximally to increase heart rate and contractility. A second major effect of aging is increased vascular stiffness, which results in an increased resistance to left ventricular ejection (afterload) and contributes to the development of systolic hypertension in the elderly. Third, in addition to increased vascular stiffness, the heart itself becomes stiffer and less compliant with age. The changes in diastolic stiffness result in important alterations in diastolic filling and atrial function. A reduction in ventricular filling not only affects cardiac output, but also produces an elevation in diastolic pressure that is transmitted back to the left atrium, where it stretches the muscle wall and predisposes to atrial ectopic beats and atrial fibrillation. The fourth major effect of cardiovascular aging is altered myocardial metabolism at the level of the mitochondria. Although older mitochondria may be able to generate sufficient adenosine triphosphate to meet the normal energy needs of the heart, they may not be able to respond under stress.

Manifestations

The manifestations of heart failure in the elderly often are masked by other disease conditions. Nocturia is an early symptom but may be caused by other conditions such as prostatic hypertrophy. Dyspnea on exertion may result from lung disease, lack of exercise, and deconditioning. Lower extremity edema commonly is caused by venous insufficiency.

Among the acute manifestations of heart failure in the elderly are increasing lethargy and confusion, probably the result of impaired cerebral perfusion. Activity intolerance is common. Instead of dyspnea, the prominent sign may be restlessness. Impaired perfusion of the gastrointestinal tract is a common cause of anorexia and profound loss of lean body mass. Loss of lean body mass may be masked by edema. The elderly also maintain a precarious balance between the managed symptom state and acute symptom exacerbation. During the managed symptom state, they are relatively symptom free while adhering to their treatment regimen. Acute symptom exacerbation, often requiring emergency medical treatment, can be precipitated by seemingly minor conditions such as poor compliance with sodium restriction, infection, or stress. Failure promptly to seek medical care is a common cause of progressive acceleration of symptoms.

Diagnosis and Treatment

The diagnosis of heart failure in the elderly is based on the history, physical examination, chest radiograph, and ECG findings. However, the presenting symptoms of heart failure often are difficult to evaluate.

Treatment of heart failure in the elderly involves many of the same methods as in younger persons. Activities are restricted to a level that is commensurate with the cardiac reserve. Seldom is bed rest recommended or advised. Bed rest causes rapid deconditioning of skeletal muscles and increases the risk of complications, such as orthostatic hypotension and thromboemboli. Instead, carefully prescribed exercise programs can help to maintain activity tolerance. Even walking around a room usually is preferable to continuous bed rest. Sodium restriction usually is indicated.

Age- and disease-related changes increase the likelihood of adverse drug reactions and drug interactions. Drug dosages and the number of drugs that are prescribed should be kept to a minimum. Compliance with drug regimens often is difficult; the simpler the regimen, the more likely it is that the older person will comply. In general, the treatment plan for elderly persons with heart failure must be put in the context of the person's overall needs. An improvement in the quality of life may take precedence over increasing the length of survival.

In summary, the mechanisms of heart failure in children and the elderly are similar to those in adults. However, the causes and manifestations may differ because of age. In children, CHF is seen most commonly during infancy and immediately after heart surgery. It can be caused by congenital and acquired heart defects and is characterized by fatigue, effort intolerance, cough, anorexia, abdominal pain, and impaired growth. Treatment of CHF in children includes correction of the underlying cause when possible. For congenital anomalies that are amenable to surgery, medical treatment often is needed for a time before surgery and usually is continued in the immediate postoperative period. For many children, only medical management can be provided.

In the elderly, age-related changes in cardiovascular functioning contribute to CHF but are not in themselves sufficient to cause heart failure. The manifestations of congestive failure often are different and superimposed on other disease conditions; therefore, CHF often is more difficult to diagnose in the elderly than in younger persons. Because the elderly are more susceptible to adverse drug reactions and have more problems with compliance, the number of drugs that are prescribed is kept to a minimum, and the drug regimen is kept as simple as possible.

REVIEW QUESTIONS

Explain the effect of decreased cardiac reserve on symptom development in heart failure.

Explain how increased sympathetic activity, fluid retention, the Frank-Starling mechanism, and myocardial hypertrophy function as compensatory mechanisms in heart failure.

Differentiate high-output versus low-output heart failure, systolic versus diastolic heart failure, and right-sided versus left-sided heart failure.

Relate the effect of left ventricular failure to the development of and manifestations of pulmonary edema.

Describe the pathophysiology of cardiogenic shock.

Describe the compensatory mechanisms that are activated in circulatory shock.

List the chief characteristics of hypovolemic shock, cardiogenic shock, obstructive shock, and distributive shock.

Characterize changes in thirst, skin blood flow, pulse rate, urine output, and sensorium that are indicative of shock.

Describe the complications of shock as they relate to the lung, kidney, gastrointestinal tract, and blood clotting.

Define multiple organ dysfunction syndrome and cite its significance in shock.

Describe the manifestations of heart failure in infants and children.

Explain how the aging process affects heart failure in the elderly.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- 1. American Heart Association. (2002). American Heart Association 2002 statistical update. [On-line]. Available: http://www. americanheart.org/statistics/othercvd.html.
- Colucci W.C., Braunwald E. (1997). Pathophysiology of heart failure. In Braunwald E. (Ed.), *Heart disease* (5th ed., pp. 394–420). Philadelphia: W.B. Saunders.
- Schrier R.W., Abraham W.T. (1999). Hormones and hemodynamics in heart failure. *New England Journal of Medicine* 341, 577–584.
- Schlant R.C., Sonnenblick E.H., Katz A.M. (1997). Physiology of heart failure. In Alexander R.W., Schlant R.C., Fuster V., et al. (Eds.), *Hurst's* the heart (9th ed., pp. 681–726). New York: McGraw-Hill.
- 5. Weinberger H.D. (1999). Diagnosis and treatment of diastolic heart failure. *Hospital Practice* 34 (3), 115–126.
- Tresch D.D., McGough M.F. (1995). Heart failure with normal systolic function: A common disorder in older people. *Journal of the American Geriatric Society* 49(9), 1035–1042.
- American College of Cardiology/American Heart Association Task Force on Practice Guides. (2001). ACC/AHA guidelines for evaluation and management of chronic heart failure in the adult: Executive Summary. *Circulation* 104, 2996–3007.
- Shamsham F., Mitchell J. (2000). Essentials of the diagnosis of heart failure. American Family Physician 61, 1319–1328.
- Vitarelli A., Gheorghiade M. (2000). Transthoracic and transesophageal echocardiogram in the hemodynamic assessment of patients with CHF. American Journal of Cardiology 86, 366–406.

- Parker W.R., Anderson A.S. (2001). Slowing the progression of CHF. Postgraduate Medicine 109 (3), 36–45.
- 11. Hoyt R.E., Bowling L.S. (2001). Reducing readmission for congestive heart failure. *American Family Physician* 63, 1593–1600.
- Wielenga R.P., Huisveld A., Bol E., et al. (1999). Safety and effects of physical training in chronic heart failure: Results of the chronic heart failure and graded exercise study. *European Heart Journal* 20(12), 872–879.
- Stanek B. (2000). Optimizing management of patients with advanced heart failure: The importance of preventing progression. *Drugs and Aging* 16, 87–106.
- 14. Katz A.M., Silverman D.I. (2000). Treatment of heart failure. *Hospital Practice* 35 (12B), 19–26.
- 15. Ramahi T.M. (2000). Beta blocker therapy for chronic heart failure. *American Family Physician* 62, 2267–2274.
- Califf R.M., Bengton J.R. (1994). Cardiogenic shock. New England Journal of Medicine 330, 1724–1730.
- 17. Mussivand T. (1999). Mechanical circulatory devices for the treatment of heart failure. *Journal of Cardiac Surgery* 14, 218–228.
- The Registry of the International Society of Heart and Lung Transplant. (1992). Ninth official report. *Journal of Heart and Lung Transplantation* 11, 599–606.
- Guyton A.C., Hall J.E. (2000). *Textbook of medical physiology* (10th ed., pp. 253–262). Philadelphia: W.B. Saunders.
- Bochner B.S., Lichtenstein L.M. (1991). Anaphylaxis. New England Journal of Medicine 324, 1785–1790.
- 21. Stankiewicz J., Ruta W., Gorski P. (1995). Latex allergy. International Journal of Occupational Medicine and Environmental Health 8, 139–148.
- 22. Leibovici L., Smara Z., Konigsberger H., et al. (1995). Long-term survival following bacteremia or fungemia. *Journal of the American Medical Association* 274, 897–912.
- 23. Balk R.A. (2000). Severe sepsis and septic shock. *Critical Care Clinics* 16, 179–191.
- 24. Carcillo J.A., Cunnin R.E. (1997) Septic shock. Critical Care Clinics 13, 553–574.
- 25. Members of the American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference Committee. (1992). American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference: Definitions of sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. *Critical Care Medicine* 20, 864–874.
- Glauser M.P. (2000). Pathologic basis of sepsis: Considerations for future strategies of intervention. *Critical Care Medicine* 28 (9 Suppl.), S4–S8.
- 27. Wheeler A.P., Bernard G.R. (1999). Treating patients with severe sepsis. *New England Journal of Medicine* 340, 207–214.
- Parrillo J.E. (1995). Pathogenetic mechanisms of septic shock. New England Journal of Medicine 328, 1471–1477.
- 29. Smith J.J., Kampine J.P. (1980). *Circulatory physiology* (p. 298). Baltimore: Williams & Wilkins.
- Fein A.M., Calalang-Colucci M.G. (2000). Acute lung injury and acute respiratory distress syndrome in sepsis and septic shock. *Critical Care Clinics* 4, 289–317.
- 31. Ware L.B., Mattay M.A. (2000). The acute respiratory distress syndrome. *New England Journal of Medicine* 342, 1334–1349.
- Levi M., Ten Cate H.T. (1999). Disseminated intravascular coagulation. New England Journal of Medicine 341, 586–592.
- Balk R.A. (2000). Pathogenesis and management of multiple organ dysfunction or failure in severe sepsis and septic shock. *Critical Care Clinics* 16, 337–352.
- Bernstein D. (2000). Heart failure. In Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.), *Nelson textbook of pediatrics* (16th ed., pp. 1440– 1444). Philadelphia: W.B. Saunders.
- 35. O'Laughlin M.P. (1999). Congestive heart failure in children. *Pediatric Clinics of North America* 46, 263–273.
- Rich M.W. (1997). Epidemiology, pathophysiology, and etiology of congestive heart failure in older adults. *Journal of the American Geriatrics Society* 45, 968–974.
- Duncan A.K., Vittone J., Fleming K.C., et al. (1996). Cardiovascular disease in elderly patients. *Mayo Clinic Proceedings* 71, 184–196.

Five

CHAPTER

Alterations in the Respiratory System

Structure and Function of the Respiratory System

Structural Organization of the Respiratory System The Conducting Airways Nasopharyngeal Airways Laryngotracheal Airways **Tracheobronchial Tree** The Lungs and Respiratory Airways **Respiratory Lobules** Lung Circulation Pleura Exchange of Gases Between the Atmosphere and the Lungs **Basic Properties of Gases** Ventilation and the Mechanics of Breathing **Respiratory Pressures** The Chest Cage and Respiratory Muscles Lung Compliance **Airway Resistance** Lung Volumes **Pulmonary Function Studies** Efficiency and the Work of Breathing Exchange and Transport of Gases Ventilation Distribution of Ventilation

Perfusion **Distribution of Blood Flow** Effects of Hypoxia Diffusion Matching of Ventilation and Perfusion **Dead Air Space** Shunt Mismatching of Ventilation and Perfusion **Gas Transport** Oxygen Transport **Carbon Dioxide Transport Control of Breathing Respiratory Center Chemoreceptors** Lung Receptors **Cough Reflex**

Respiration provides the body with a means of gas exchange. It is the process whereby oxygen from the air is transferred to the blood and carbon dioxide is eliminated from the body. The nervous system controls the movement of the respiratory muscles and adjusts the rate of breathing so that it matches the needs of the body during various levels of activity. The content in this chapter focuses on the structure and function of the respiratory system as it relates to these aspects of respiration. The function of the red blood cell in the transport of oxygen is discussed in Chapter 13.

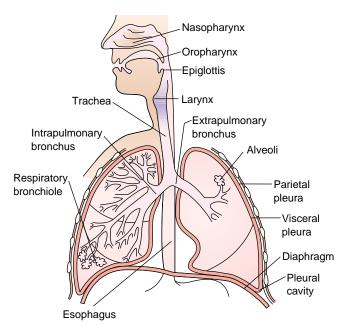
STRUCTURAL ORGANIZATION OF THE RESPIRATORY SYSTEM

The respiratory system consists of the air passages and the lungs. Functionally, the respiratory system can be divided into two parts: the *conducting airways*, through which air moves as it passes between the atmosphere and the lungs, and the *respiratory tissues* of the lungs, where gas exchange takes place.

The Conducting Airways

The conducting airways consist of the nasal passages, mouth and pharynx, larynx, trachea, bronchi, and bronchioles (Fig. 19-1). The air we breathe is warmed, filtered, and moistened as it moves through these structures. Heat is transferred to the air from the blood flowing through the walls of the respiratory passages; the mucociliary blanket removes foreign materials; and water from the mucous membranes is used to moisten the air.

The conducting airways are lined with a pseudostratified columnar epithelium that contains a mosaic of mucus-secreting goblet cells and cells that contain hairlike projections called cilia (Fig. 19-2). The larger bronchi contain additional mucussecreting glands that connect to the epithelium through long epithelial ducts. The epithelial layer gradually becomes thinner as it moves from the pseudostratified epithelium of the bronchi to cuboidal epithelium of the bronchioles and then to squamous epithelium of the alveoli. The mucus produced by the goblet cells in the conducting airways forms a layer, called the *mucocil*iary blanket, that protects the respiratory system by entrapping dust and other foreign particles that enter the airways. The cilia, which constantly are in motion, move the mucociliary blanket with its entrapped particles in an escalator-like fashion toward the oropharynx, from which it is expectorated or swallowed. The function of the mucociliary blanket in clearing the lower airways and alveoli is optimal at normal oxygen levels and is impaired



■ FIGURE 19-1 ■ Structures of the respiratory system.

in situations of low and high oxygen levels. It is impaired by drying, such as breathing heated but unhumidified indoor air during winter. Cigarette smoking slows down or paralyzes the motility of the cilia. This slowing allows the residue from tobacco smoke, dust, and other particles to accumulate in the lungs, decreasing the efficiency of this pulmonary defense system.

The airways are kept moist by water contained in the mucous layer. Moisture is added to the air as it moves through the conducting airways. The capacity of the air to contain moisture or water vapor without condensation increases as the temperature rises. Thus, the air in the alveoli, which is maintained at body temperature, usually contains considerably more water vapor than the atmospheric-temperature air that we breathe. The difference between the water vapor contained in the air we breathe and that found in the alveoli is drawn from the moist surface of the mucous membranes that line the conducting airways and is a source of insensible water loss (see Chapter 6). Under normal conditions, approximately 1 pint of water per day is lost in humidifying the air breathed. During fever, the water vapor in the lungs increases, causing more water to be lost from the respiratory mucosa. In addition, fever usually is accompanied by an increase in respiratory rate so that more air passes through the airways, withdrawing moisture from its mucosal surface. As a result, respiratory secretions thicken, preventing free movement of the cilia and impairing the protective function of the mucociliary defense system. This is particularly true in persons whose water intake is inadequate.

Nasopharyngeal Airways

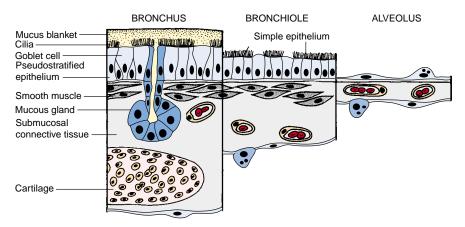
The nose is the preferred route for the entrance of air into the respiratory tract during normal breathing. As air passes through the nasal passages, it is filtered, warmed, and humidified. The outer nasal passages are lined with coarse hairs, which filter and trap dust and other large particles from the air. The upper portion of the nasal cavity is lined with mucous membrane that contains a rich network of small blood vessels; this portion of the nasal cavity supplies warmth and moisture to the air we breathe.

The mouth serves as an alternative airway when the nasal passages are plugged or when there is a need for the exchange of large amounts of air, as occurs during exercise. The oropharvnx extends posteriorly from the soft palate to the epiglottis. The oropharynx is the only opening between the nose and mouth and the lungs. Both swallowed food on its way to the esophagus and air on its way to the larynx pass through it. Obstruction of the oropharynx leads to immediate cessation of ventilation. Neural control of the tongue and pharyngeal muscles may be impaired in coma and certain types of neurologic disease. In these conditions, the tongue falls back into the pharynx and obstructs the airway, particularly if the person is lying on his or her back. Swelling of the pharyngeal structures caused by injury, infection, or severe allergic reaction also predisposes a person to airway obstruction, as does the presence of a foreign body.

Laryngotracheal Airways

The larynx connects the oropharynx with the trachea. The walls of the larynx are supported by firm cartilaginous structures that prevent collapse during inspiration. The functions of the larynx can be divided into two categories: those associated with speech and those associated with protecting the lungs from substances other than air. The larynx is located in a strate-

■ FIGURE 19-2 ■ Airway wall structure: bronchus, bronchiole, and alveolus. The bronchial wall contains pseudostratified epithelium, smooth muscle cells, mucous glands, connective tissue, and cartilage. In smaller bronchioles, a simple epithelium is found, cartilage is absent, and the wall is thinner. The alveolar wall is designed for gas exchange, rather than structural support. (From Weibel E.R., Taylor R.C. [1988]. Design and structure of the human lung. In Fishman A.P. [ed.]. *Pulmonary diseases and disorders*. Vol. 1. [p. 14]. New York: McGraw-Hill. Reproduced with permission of the McGraw-Hill Companies)



gic position between the upper airways and the lungs and sometimes is referred to as the "watchdog of the lungs."

The cavity of the larynx is divided into two pairs of two-bytwo folds of mucous membrane stretching from front to back with an opening in the midline (Fig. 19-3). The upper pair of folds, called the *vestibular folds*, has a protective function. The lower pair of folds has cordlike margins; they are termed the *vocal folds* because their vibrations are required for making vocal sounds. The true vocal folds and the elongated opening between them are called the *glottis*. A complex set of muscles controls the opening and closing of the glottis. Speech involves the intermittent release of expired air and opening and closing of the glottis. The epiglottis, which is located above the vocal folds, is a large, leaf-shaped piece of cartilage that is covered with epithelium. During swallowing, the free edges of the epiglottis move downward to cover the larynx, thus routing liquids and foods into the esophagus.

In addition to opening and closing the glottis for speech, the vocal folds of the larynx can perform a sphincter function in

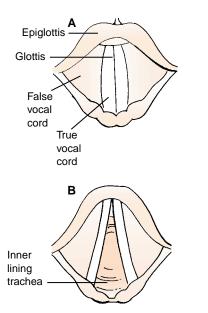


FIGURE 19-3 Epiglottis and vocal cords viewed above with (A) glottis closed and (B) glottis open.

closing off the airways. When confronted with substances other than air, the laryngeal muscles contract and close off the airway. At the same time, the cough reflex is initiated as a means of removing a foreign substance from the airway. If the swallowing mechanism is partially or totally paralyzed, food and fluids can enter the airways instead of the esophagus when a person attempts to swallow. These substances are not easily removed; and when they are pulled into the lungs, they can cause a serious inflammatory condition called *aspiration pneumonia*.

Tracheobronchial Tree

The tracheobronchial tree, which consists of the trachea, bronchi, and bronchioles, can be viewed as a system of branching tubes (Fig. 19-5). It is similar to a tree whose branches become smaller and more numerous as they divide. There are approximately 23 levels of branching, beginning with the conducting airways and ending with the respiratory airways, where gas exchange takes place (Fig. 19-5).

The trachea, or windpipe, is a continuous tube that connects the larynx and the major bronchi of the lungs (see Fig. 19-4). The walls of the trachea are supported by horseshoe-shaped cartilages, which prevent it from collapsing when the pressure in the thorax becomes negative.

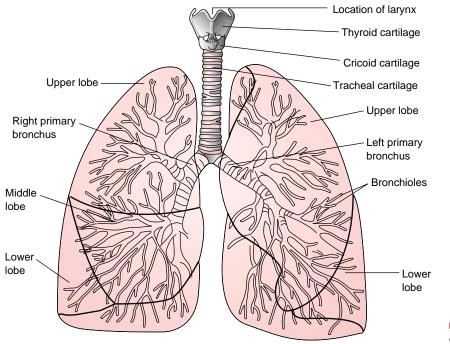
The trachea divides to form the right and the left primary bronchi. Each bronchus enters the lung through a slit called the hilus. The structure of the primary bronchi is similar to that of the trachea in that these airways are supported by cartilaginous rings.

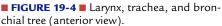
Each primary bronchus divides into *secondary*, or *lobular*, *bronchi*, which supply each of the lobes of the lungs—three in the right lung and two in the left. The right middle lobe bronchus is of relatively small diameter and length and sometimes bends sharply near its bifurcation. It is surrounded by a collar of lymph nodes that drain the middle and the lower lobe and is particularly subject to obstruction. The secondary bronchi divide to form the segmental bronchi, which supply the bronchopulmonary segments of the lung. These segments are identified according to their location in the lung (*e.g.*, the apical segment of the right upper lobe) and are the smallest named units in the lung. Lung lesions such as atelectasis and pneumonia often are localized to a particular bronchopulmonary segment.

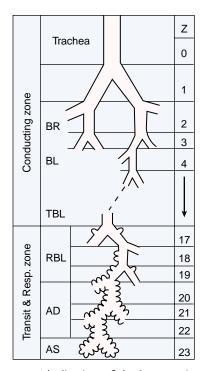
The bronchi continue to branch, forming smaller bronchi, until they become the terminal bronchioles, the smallest of

Dr Murtadha Alshareifi e-Library

342 Unit Five: Alterations in the Respiratory System







■ FIGURE 19-5 ■ Idealization of the human airways. The first 16 generations of branching (Z) make up the conducting airways, and the last seven constitute the respiratory zone (or transitional and respiratory zone). BR, bronchus; BL, bronchiole; TBL, terminal bronchiole; RBL, respiratory bronchiole; AD, alveolar ducts; AS, alveolar sacs. (Weibel E.R. [1962]. *Morphometry of the human lung* [p. 111]. Berlin: Springer-Verlag)

the conducting airways. As these bronchi branch and become smaller, this cartilaginous support becomes thinner and then disappears at the level of the respiratory bronchioles. Between the cartilaginous support and the mucosal surface are two crisscrossing layers of smooth muscle that wind in opposite directions (Fig. 19-6). Bronchospasm, or contraction of these muscles, causes narrowing of the bronchioles and impairs air flow. The epithelial lining of conducting airways gradually becomes reduced from pseudostratified epithelium in the bronchi to a thin layer of tightly joined epithelial cells in the alveoli.

The Lungs and Respiratory Airways

The lungs are soft, spongy, cone-shaped organs located side by side in the chest cavity (see Fig. 19-1). They are separated from each other by the *mediastinum* (*i.e.*, the space between the lungs) and its contents—the heart, blood vessels, lymph nodes, nerve fibers, thymus gland, and esophagus. The upper part of the lung, which lies against the top of the thoracic cavity, is called the *apex*, and the lower part, which lies against the diaphragm, is called the *base*. The lungs are divided into lobes: three in the right lung and two in the left (see Fig. 19-4).

The lungs are the functional structures of the respiratory system. In addition to their gas exchange function, they inactivate vasoactive substances such as bradykinin; they convert angiotensin I to angiotensin II; and they serve as a reservoir for blood storage. Heparin-producing cells are particularly abundant in the capillaries of the lung, where small clots may be trapped.

Respiratory Lobules

The gas exchange function of the lung takes place in the lobules of the lungs. Each lobule, which is the smallest functional unit of the lung, is supplied by a branch of a terminal bron-

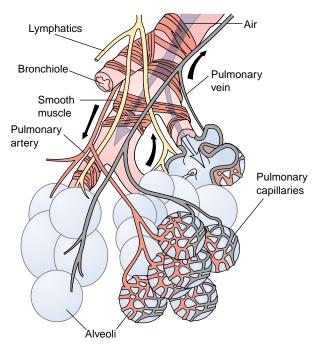


FIGURE 19-6 Lobule of the lung, showing the bronchial smooth muscle fibers, pulmonary blood vessels, and lymphatics.

chiole, an arteriole, the pulmonary capillaries, and a venule (see Fig. 19-6). Gas exchange takes place in the terminal respiratory bronchioles and the alveolar ducts and sacs. Blood enters the lobules through a pulmonary artery and exits through a pulmonary vein. Lymphatic structures surround the lobule and aid in the removal of plasma proteins and other particles from the interstitial spaces.

Unlike larger bronchi, the respiratory bronchioles are lined with simple epithelium, rather than ciliated pseudostratified epithelium. The respiratory bronchioles also lack the cartilaginous support of the larger airways. Instead, they are attached to the elastic spongework of tissue that contains the alveolar air spaces. When the air spaces become stretched during inspiration, the bronchioles are pulled open by expansion of the surrounding tissue.

The alveolar sacs are cup-shaped, thin-walled structures that are separated from each other by thin alveolar septa. Most of the septa are occupied by a single network of capillaries so that blood is exposed to air on both sides. There are approximately 300 million alveoli in the adult lung, with a total surface area of approximately 50 to 100 m². Unlike the bronchioles, which are tubes with their own separate walls, the alveoli are interconnecting spaces that have no separate walls (Fig. 19-7). As a result of this arrangement, there is a continual mixing of air in the alveolar structures.

The alveolar structures are composed of two types of cells: type I alveolar cells and type II alveolar cells (Fig. 19-8). The type I alveolar cells are flat squamous epithelial cells across which gas exchange takes place. The type II alveolar cells produce surfactant, a lipoprotein substance that decreases the surface tension in the alveoli. The alveoli also contain alveolar macrophages, which are responsible for the removal of offending substances from the alveolar epithelium.

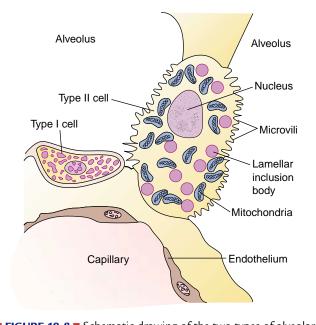


■ FIGURE 19-7 ■ Close-up of a cross-section of a small bronchus and surrounding alveoli. (Courtesy of Janice A. Nowell, University of California, Santa Cruz)

Lung Circulation

The lungs are provided with a dual blood supply: the pulmonary and bronchial circulations. The pulmonary circulation arises from the pulmonary artery and provides for the gas exchange function of the lungs. Deoxygenated blood leaves the right heart through the pulmonary artery, which divides into a left pulmonary artery that enters the left lung and a right pulmonary artery that enters the right lung. Return of oxygenated blood to the heart occurs by way of the pulmonary veins, which empty into the left atrium.

The bronchial circulation distributes blood to the conducting airways and supporting structures of the lung. The bronchial circulation has a secondary function of warming and humidifying incoming air as it moves through the conducting airways. The bronchial arteries arise from the thoracic aorta and enter the lungs with the major bronchi, dividing and subdividing along with the bronchi as they move out into the lung, supplying them and other lung structures with oxygen. The capillaries of the bronchial circulation drain into the bronchial veins, the larger of which empties into the vena cava. The smaller of the bronchial veins empties into the pulmonary veins. This blood is unoxygenated because the bronchial circulation does



■ FIGURE 19-8 Schematic drawing of the two types of alveolar cells and their relation to alveoli and capillaries. Alveolar type I cells comprise most of the alveolar surface. Alveolar type II cells are located in the corner between two adjacent alveoli. Also shown are endothelial cells that line the pulmonary capillaries. (Rhoades R.A., Tanner G.A. [1996]. *Medical physiology* [p. 362]. Boston: Little, Brown).

not participate in gas exchange. As a result, this blood dilutes the oxygenated blood returning to the left side of the heart.

Pleura

A thin, transparent, double-layered serous membrane, called the *pleura*, lines the thoracic cavity and encases the lungs. The outer parietal layer lies adjacent to the chest wall, and the inner visceral layer adheres to the outer surface of the lung (see Fig. 19-1). The parietal pleura forms part of the mediastinum and lines the inner wall of the thoracic or chest cavity. A thin film of serous fluid separates the two pleural layers, and this allows the two layers to glide over each other and yet hold together, so there is no separation between the lungs and the chest wall. The pleural cavity is a potential space in which serous fluid or inflammatory exudate can accumulate. The term *pleural effusion* is used to describe an abnormal collection of fluid or exudate in the pleural cavity.

In summary, the respiratory system consists of the air passages and the lungs, where gas exchange takes place. Functionally, the air passages of the respiratory system can be structurally divided into two parts: the conducting airways, through which air moves as it passes into and out of the lungs, and the respiratory tissues, where gas exchange actually takes place. The conducting airways include the nasal passages, mouth and nasopharynx, larynx, and tracheobronchial tree. Air is warmed, filtered, and humidified as it passes through these structures. The lungs are the functional structures of the respiratory system. In addition to their gas exchange function, they inactivate vasoactive substances such as bradykinin; they convert angiotensin I to angiotensin II; and they serve as a reservoir for blood. The lobules, which are the functional units of the lung, consist of the respiratory bronchioles, alveoli, and pulmonary capillaries. It is here that gas exchange takes place. Oxygen from the alveoli diffuses across the alveolar capillary membrane into the blood, and carbon dioxide from the blood diffuses into the alveoli.

The lungs are provided with a dual blood supply: the pulmonary circulation provides for the gas exchange function of the lungs, and the bronchial circulation distributes blood to the conducting airways and supporting structures of the lung. The lungs are encased in a thin, transparent, doublelayered serous membrane called the *pleura*. The pressure in the pleural space, which is negative in relation to alveolar pressure, prevents the lungs from collapsing.

EXCHANGE OF GASES BETWEEN THE ATMOSPHERE AND THE LUNGS

Basic Properties of Gases

The air we breathe is made up of a mixture of gases, mainly nitrogen and oxygen. These gases exert a combined pressure called the *atmospheric pressure*. The pressure at sea level is defined as 1 atmosphere, which is equal to 760 millimeters of mercury (mm Hg) or 14.7 pounds per square inch (PSI). When measuring respiratory pressures, atmospheric pressure is assigned a value of 0. A respiratory pressure of +15 mm Hg means that the pressure is 15 mm Hg above atmospheric pressure, and a respiratory pressure of -15 mm Hg is 15 mm Hg less than atmospheric pressure. Respiratory pressures often are expressed in centimeters of water (cm H₂O) because of the small pressures involved (1 mm Hg = 1.35 cm H₂O pressure).

The pressure exerted by a single gas in a mixture is called the *partial pressure*. The capital letter "P" followed by the chemical symbol of the gas (PO₂) is used to denote its partial pressure. The law of partial pressures states that the total pressure of a mixture of gases, as in the atmosphere, is equal to the sum of the partial pressures of the different gases in the mixture. If the concentration of oxygen at 760 mm Hg (1 atmosphere) is 20%, its partial pressure is 152 mm Hg (760 × 0.20).

Water vapor is different from other types of gases; its partial pressure is affected by temperature but not atmospheric pressure. The relative humidity refers to the percentage of moisture in the air compared with the amount that the air can hold without causing condensation (100% saturation). Warm air holds more moisture than cold air. This is the reason that precipitation in the form of rain or snow commonly occurs when the relative humidity is high and there is a sudden drop in atmospheric temperature. The air in the alveoli, which is 100% saturated at normal body temperature, has a water vapor pressure of 47 mm Hg. The water vapor pressure must be included in the sum of the total pressure of the gases in the alveoli (*i.e.*, the total pressure of the other gases in the alveoli is 760 – 47 = 713 mm Hg). Air moves between the atmosphere and the lungs because of a pressure difference. According to the laws of physics, the pressure of a gas varies inversely with the volume of its container, provided the temperature remains constant. If equal amounts of a gas are placed in two different-size containers, the pressure of the gas in the smaller container is greater than the pressure in the larger container. The movement of gases is always from the container with the greater pressure to the one with the lesser pressure. The chest cavity can be viewed as a volume container. During inspiration, the size of the chest cavity increases and air moves into the lungs; during expiration, air moves out as the size of the chest cavity decreases.

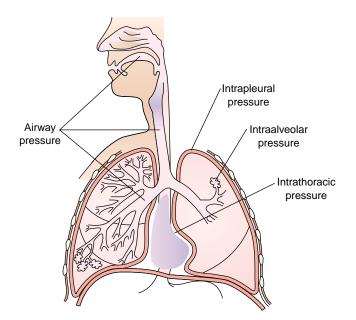
Ventilation and the Mechanics of Breathing

Ventilation is concerned with the movement of gases into and out of the lungs. It relies on a system of open airways and the respiratory pressures created as the movements of the respiratory muscles change the size of the chest cage. The degree to which the lungs inflate and deflate depends on the respiratory pressures inflating the lung, compliance of the lungs, and airway resistance.

Respiratory Pressures

The pressure inside the airways and alveoli of the lungs is called the *intrapulmonary pressure* or *alveolar pressure*. The gases in this area of the lungs are in communication with atmospheric pressure (Fig. 19-9). When the glottis is open and air is not moving into or out of the lungs, as occurs just before inspiration or expiration, the intrapulmonary pressure is zero or equal to atmospheric pressure.

The pressure in the pleural cavity is called the *intrapleural pressure*. The intrapleural pressure is always negative in relation to alveolar pressure, approximately –4 mm Hg between breaths when the glottis is open and the alveolar spaces are open to the atmosphere. The lungs and the chest wall have elastic proper-



■ FIGURE 19-9 ■ Partitioning of respiratory pressures.

ties, each pulling in the opposite direction. If removed from the chest, the lungs would contract to a smaller size, and the chest wall, if freed from the lungs, would expand. The opposing forces of the chest wall and lungs create a pull against the visceral and parietal layers of the pleura, causing the pressure in the pleural cavity to become negative. During inspiration, the elastic recoil of the lungs increases, causing intrapleural pressure to become more negative than during expiration. Without the negative intrapleural pressure holding the lungs against the chest wall, their elastic recoil properties would cause them to collapse. Although the intrapleural pressure of the inflated lung is always negative in relation to alveolar pressure, it may become positive in relation to atmospheric pressure (*e.g.*, during forced expiration and coughing).

The *intrathoracic pressure* is the pressure in the thoracic cavity. It is essentially equal to intrapleural pressure and is the pressure to which the lungs, heart, and great vessels are exposed. Forced expiration against a closed glottis (Valsalva maneuver) compresses the air in the thoracic cavity and produces marked increases in intrathoracic and intrapleural pressures.

The Chest Cage and Respiratory Muscles

The lungs and major airways share the chest cavity with the heart, great vessels, and esophagus. The chest cavity is a closed compartment bounded on the top by the neck muscles and at the bottom by the diaphragm. The outer walls of the chest cavity are formed by 12 pairs of ribs, the sternum, the thoracic vertebrae, and the intercostal muscles that lie between the ribs. Mechanically, the act of breathing depends on the fact that the chest cavity is a closed compartment whose only opening to the exterior is the trachea.

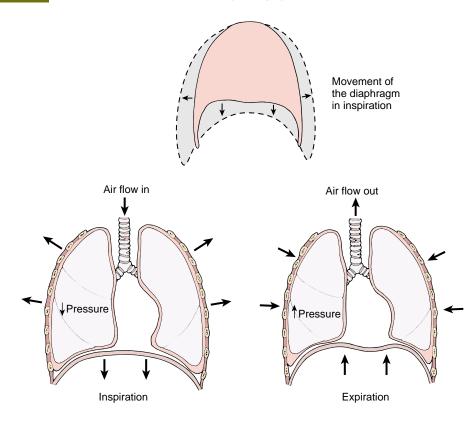
Ventilation consists of inspiration and expiration. During *inspiration*, the size of the chest cavity increases, the intrathoracic pressure becomes more negative, and air is drawn into the lungs. *Expiration* occurs as the elastic components of the chest wall and lung structures that were stretched during inspiration recoil, causing the size of the chest cavity to decrease and the pressure in the chest cavity to increase (Fig. 19-10).

Inspiration. The diaphragm is the principal muscle of inspiration. When the diaphragm contracts, the abdominal contents are forced downward and the chest expands from top to bottom (Fig. 19-10). The diaphragm is innervated by the phrenic nerve roots, which arise from the cervical level of the spinal cord, mainly from C4 but also from C3 and C5.

The external intercostal muscles, which also aid in inspiration, connect to the adjacent ribs and slope downward and forward (Fig. 19-11). When they contract, they raise the ribs and rotate them slightly so that the sternum is pushed forward; this enlarges the chest from side to side and from front to back. The intercostal muscles receive their innervation from nerves that exit the central nervous system at the thoracic level of the spinal cord. Paralysis of these muscles usually does not have a serious effect on respiration because of the effectiveness of the diaphragm.

The accessory muscles of inspiration include the scalene muscles and the sternocleidomastoid muscles. The scalene muscles elevate the first two ribs, and the sternocleidomastoid muscles raise the sternum to increase the size of the chest cavity. These muscles contribute little to quiet breathing but

346 Unit Five: Alterations in the Respiratory System



■ FIGURE 19-10 ■ Movement of the diaphragm and changes in chest volume and pressure during inspiration and expiration. (A) Movement of diaphragm and expansion of the chest cavity during inspiration. (B) During inspiration, contraction of the diaphragm and expansion of the chest cavity produce a decrease in intrathoracic pressure, causing air to move into the lungs. (C) During expiration, relaxation of the diaphragm and chest cavity produce an increase in intrathoracic pressure, causing air to move out of the lungs.

contract vigorously during exercise. For the accessory muscles to assist in ventilation, they must be stabilized in some way. For example, persons with bronchial asthma often brace their arms against a firm object during an attack as a means of stabilizing their shoulders so that the attached accessory muscles can exert their full effect on ventilation. The head commonly is bent backward so that the scalene and sternocleidomastoid muscles



can elevate the ribs more effectively. Other muscles that play a minor role in inspiration are the alae nasi, which produce flaring of the nostrils during obstructed breathing.

Expiration. Expiration is largely passive. It occurs as the elastic components of the chest wall and lung structures that were stretched during inspiration recoil, causing air to leave the lungs as the intrathoracic pressure increases. When needed, the abdominal and the internal intercostal muscles can be used to increase expiratory effort (see Fig. 19-11). The increase in intraabdominal pressure that accompanies the forceful contraction of the abdominal muscles pushes the diaphragm upward and results in an increase in intrathoracic pressure. The internal intercostal muscles move inward, which pulls the chest downward, increasing expiratory effort.

Lung Compliance

Lung compliance refers to the ease with which the lungs can be inflated. It is determined by the elastin and collagen fibers of the lung, its water content, and surface tension. Compliance can be appreciated by comparing the ease of blowing up a new balloon that is stiff and noncompliant with one that has been previously blown up and stretched.

Specifically, lung compliance describes the change in lung volume that can be accomplished with a given change in respiratory pressure. The normal compliance of both lungs in the average adult is approximately 200 mL/cm H₂O. This means that every time the transpulmonary pressure increases by 1 cm/H₂O, the lung volume expands by 200 mL. It would take more pressure to move the same amount of air into a non-compliant lung.

Changes in Elastin/Collagen Composition of Lung Tissue. Lung tissue is made up of elastin and collagen fibers. The elastin fibers are easily stretched and increase the ease of lung inflation, whereas the collagen fibers resist stretching and make lung inflation more difficult. In lung diseases such as interstitial lung disease and pulmonary fibrosis, the lungs become stiff and noncompliant as the elastin fibers are replaced with scar tissue. Pulmonary congestion and edema produce a reversible decrease in pulmonary compliance.

Elastic recoil describes the ability of the elastic components of the lung to recoil to their original position after having been stretched. Overstretching the airways, as occurs with emphysema, causes the elastic components of the lung to lose their recoil, making the lung easier to inflate but more difficult to deflate because of its inability to recoil.

Surface Tension. An important factor in lung compliance is the *surface tension* in the alveoli. The alveoli are lined with a thin film of liquid, and it is at the interface between this liquid film and the alveolar air that surface tension develops. This is because the forces that hold the liquid film molecules together are stronger than those that hold the air molecules together. In the alveoli, excess surface tension causes the liquid film to contract, making lung inflation more difficult.

The pressure in the alveoli (which are modeled as spheres with open airways projecting from them) can be predicted using Laplace's law (pressure = $2 \times \text{surface tension/radius}$). If the surface tension were equal throughout the lungs, the alveoli with the smallest radii would have the greatest pressure, and this would cause them to empty into the larger alveoli (Fig. 19-12). The reason this does not occur is because of special surface tension-lowering molecules, called *surfactant*, that line the inner surface of the alveoli.

Surfactant is a complex mixture of lipoproteins (largely phospholipids) and small amounts of carbohydrates that is synthesized in the type II alveolar cells. The surfactant molecule has two ends: a hydrophobic (water-insoluble) tail and a hydrophilic (water-soluble) head (Fig. 19-13). The hydrophilic head of the surfactant molecule attaches to the liquid mole-

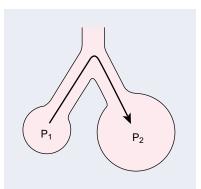


FIGURE 19-12 Law of Laplace (P = 2 T/r, P = pressure, T = tension, r = radius). The effect of the radius on the pressure and movement of gases in the alveolar structures is depicted. Air moves from P₁ with a small radius and higher pressure to P₂ with its larger radius and lower pressure.

cules and the hydrophobic tail to the gas molecules, interrupting the intermolecular forces that are responsible for creating the surface tension.

Surfactant exerts four important effects on lung inflation: (1) it lowers the surface tension; (2) it increases lung compliance and ease of inflation; (3) it provides for stability and more even inflation of the alveoli; and (4) it assists in preventing pulmonary edema by keeping the alveoli dry. Without surfactant, lung inflation would be extremely difficult, requiring an intrapleural pressure of -20 to -30 mm Hg, compared with the -3 to -5 mm Hg pressure that normally is needed. The surfactant molecules are more densely packed in the small alveoli

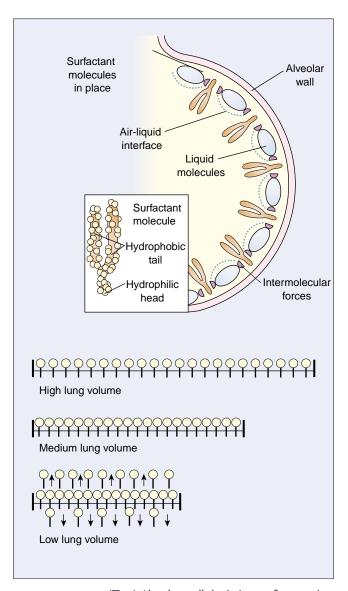


FIGURE 19-13 (**Top**) Alveolar wall depicting surface tension resulting from the intramolecular forces in the air-liquid film interface; the surfactant molecule with its hydrophobic tail and hydrophilic head; and its function in reducing surface tension by disrupting the intermolecular forces. (**Bottom**) The surface concentration of surfactant molecules at high, medium, and low lung volumes.

348 Unit Five: Alterations in the Respiratory System

than in larger alveoli, where the density of the molecules is less. Therefore, surfactant reduces the surface tension more effectively in the small alveoli, which have the greatest tendency to collapse, providing for stability and more even distribution of ventilation. Surfactant also helps to keep the alveoli dry and prevent pulmonary edema. This is because water is pulled out of the pulmonary capillaries into the alveoli when increased surface tension causes the alveoli to contract.

The type II alveolar cells that produce surfactant do not begin to mature until the 26th to 28th week of gestation; consequently, many premature infants have difficulty producing sufficient amounts of surfactant. This can lead to alveolar collapse and severe respiratory distress. This condition, called *infant respiratory distress syndrome*, is the single most common cause of respiratory disease in premature infants. Surfactant dysfunction also is possible in the adult. This usually occurs as the result of severe injury or infection and can contribute to the development of a condition called *acute respiratory distress syndrome* (see Chapter 21).

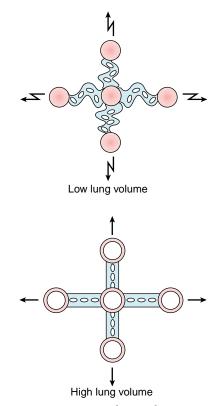
Airway Resistance

The volume of air that moves into and out of the air exchange portion of the lungs is directly related to the pressure difference between the lungs and the atmosphere and inversely related to the resistance that the air encounters as it moves through the airways. The effects of airway resistance on airflow can be illustrated using *Poiseuille's law*. According to Poiseuille's law, the resistance to flow is inversely related to the fourth power of the radius ($R = 1/r^4$). If the radius is reduced by one half, the resistance increases 16-fold ($2 \times 2 \times 2 \times 2 = 16$). Because the resistance of the airways is inversely proportional to the fourth power of the radius, small changes in airway caliber, such as those caused by pulmonary secretions or bronchospasm, can produce a marked increase in airway resistance.

Airway resistance is also affected by lung volumes, being less during inspiration than during expiration. This is because elastic-type fibers connect the outside of the airways to the surrounding lung tissues. As a result, these airways are pulled open as the lungs expand during inspiration, and they become narrower as the lungs deflate during expiration (Fig. 19-14). This is one of the reasons that persons with conditions that increase airway resistance, such as bronchial asthma, usually have less difficulty during inspiration than during expiration.

Airway Compression. Airflow through the collapsible airways in the lungs depends on the distending airway (intrapulmonary) pressures that hold the airways open and the external (intrapleural or intrathoracic) pressures that surround and compress the airways. The difference between these two pressures (airway pressure minus intrathoracic pressure) is called the *transpulmonary pressure*. For airflow to occur, the distending pressure inside the airways must be greater than the compressing pressure outside the airways (Fig. 19-15).

During forced expiration, the transpulmonary pressure is decreased because of a disproportionate increase in the intrathoracic pressure compared with airway pressure. The resistance that air encounters as it moves out of the lungs causes a further drop in airway pressure. If this drop in airway pressure is sufficiently great, the surrounding intrathoracic pressure will compress the collapsible airways (*i.e.*, those that lack cartilaginous support), causing airflow to be interrupted and air to be

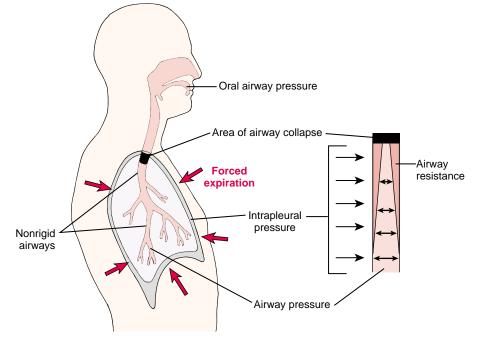


■ FIGURE 19-14 ■ Interaction of tissue forces on airways during low and high lung volumes. At low lung volumes, the tissue forces tend to fold and place less tension on the airways and they become smaller; during high lung volumes, the tissue forces are stretched and pull the airways open.

trapped in the alveoli (Fig. 19-15). Although this type of airway compression usually is seen only during forced expiration in persons with normal respiratory function, it may occur during normal breathing in persons with lung disease. For example, in conditions that increase airway resistance, such as chronic obstructive airway disease (COPD), the pressure drop along the smaller airways is magnified, and an increase in intra-airway pressure is needed to maintain airway patency (see Chapter 20). Measures such as pursed-lip breathing increase airway pressure and improve expiratory flow rates in persons with COPD.

Lung Volumes

Lung volumes, or the amount of air exchanged during ventilation, can be subdivided into three components: (1) the tidal volume (TV), (2) the inspiratory reserve volume (IRV), and (3) the expiratory reserve volume (ERV). The TV, usually about 500 mL, is the amount of air that moves into and out of the lungs during a normal breath (Fig. 19-6). The IRV is the maximum amount of air that can be inspired in excess of the normal TV, and the ERV is the maximum amount that can be exhaled in excess of the normal TV. Approximately 1200 mL of air always remains in the lungs after forced expiration; this air is the *residual volume* (RV). The RV increases with age because there is more trapping of air in the lungs at the end of expiration. These volumes can be measured using an instrument called a *spirometer*. ■ FIGURE 19-15 ■ Mechanism that limits maximal expiratory flow rate. (A) Airway patency and airflow in the nonrigid airways of the lungs rely on a transpulmonary pressure gradient in which airway pressure is greater than intrapleural pressure. (B) Airway resistance normally produces a drop in airway pressure as air moves out of the lungs. The increased intrapleural pressure that occurs with forced expiration produces airway collapse in the nonrigid airways at the point where intrapleural pressure exceeds airway pressure.



Lung capacities include two or more lung volumes. The *vital capacity* (VC) equals the IRV plus the TV plus the ERV and is the amount of air that can be exhaled from the point of maximal inspiration. The *inspiratory capacity* (IC) equals the TV plus the IRV. It is the amount of air a person can breathe in beginning at the normal expiratory level and distending the lungs to the maximal amount. The *functional residual capacity* (FRC) is the sum of the RV and ERV; it is the volume of air that remains in the lungs at the end of normal expiration. The *total lung capacity* (TLC) is the sum of all the volumes in the lungs. The RV cannot be measured with the spirometer because this air cannot be expressed from the lungs. It is measured by indirect methods, such as the helium dilution methods, the nitrogen washout methods, or body plethysmography. Lung volumes and capacities are summarized in Table 19-1.

Pulmonary Function Studies

The previously described lung volumes and capacities are anatomic or static measures, determined by lung volumes and measured without relation to time. The spirometer also is used to measure dynamic lung function (*i.e.*, ventilation with respect to time); these tests often are used in assessing pulmonary function (Table 19-2). The *maximum voluntary ventilation* measures the volume of air that a person can move into and out of the lungs during maximum effort lasting for a specific period of time. This measurement usually is converted to liters per minute. Two other useful tests are the forced vital capacity (FVC) and the *forced expiratory volume* (FEV). The FVC involves full inspiration to total lung capacity followed by forceful maximal expiration. Obstruction of airways produces a FVC that is lower than that observed with more slowly performed vital

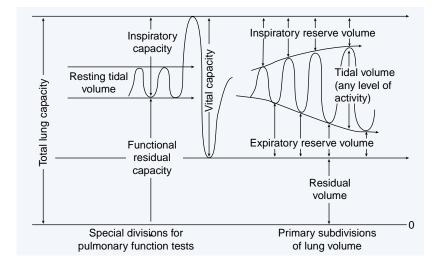


FIGURE 19-16 A tracing of respiratory volumes and capacities made with the use of a spirometer.

TABLE 19-1	LE 19-1 Lung Volumes and Capacities				
Volume		Symbol	Measurement		
Tidal volume (about 500 mL at rest) Inspiratory reserve volume (about 3000 mL)		TV IRV	Amount of air that moves into and out of the lungs with each breath Maximum amount of air that can be inhaled from the point of maximal expiration		
Expiratory reserve volume (about 1100 mL)		ERV	Maximum volume of air that can be exhaled from the resting end- expiratory level		
Residual volume (about 1200 mL)		RV	Volume of air remaining in the lungs after maximal expiration. This volur cannot be measured with the spirometer; it is measured indirectly using methods such as the helium dilution method, the nitrogen washout technique, or body plethysmography.		
Functional residual capacity (about 2300 mL)		FRC	Volume of air remaining in the lungs at end-expiration (sum of RV and ERV)		
Inspiratory capacit	y (about 3500 mL)	IC	Sum of IRV and TV		
Vital capacity (abo	ut 4600 mL)	VC	Maximal amount of air that can be exhaled from the point of maximal inspiration		
Total lung capacity	y (about 5800 mL)	TLC	Total amount of air that the lungs can hold; it is the sum of all the volume components after maximal inspiration. This value is about 20% to 25% less in females than in males.		

capacity measurements. The forced expiratory volume (FEV) is the expiratory volume achieved in a given time period. The $FEV_{1.0}$ is the forced expiratory volume that can be exhaled in 1 second. The $FEV_{1.0}$ frequently is expressed as a percentage of the FVC. The $FEV_{1.0}$ and FVC are used in the diagnosis of obstructive lung disorders. contrast, persons with obstructive airway disease usually find it less difficult to inflate their lungs but expend more energy in moving air through the airways. As a result, these persons take deeper breaths and breathe at a slower rate (*e.g.*, $600 \times 10 = 6000$ mL) to achieve their oxygen needs.

Efficiency and the Work of Breathing

The *minute volume*, or total ventilation, is the amount of air that is exchanged in 1 minute. It is determined by the metabolic needs of the body. The minute volume is equal to the TV multiplied by the respiratory rate. During normal activity it is about 6000 mL (500 mL TV \times respiratory rate of 12 breaths per minute). The efficiency of breathing is determined by matching the TV and respiratory rate in a manner that provides an optimal minute volume while minimizing the work of breathing.

The work of breathing is determined by the amount of effort required to move air through the conducting airways and by the ease of lung expansion or compliance. Expansion of the lungs is difficult for persons with stiff and noncompliant lungs; they usually find it easier to breathe if they keep their TV low and breathe at a more rapid rate (*e.g.*, $300 \times 20 = 6000$ mL) to achieve their minute volume and meet their oxygen needs. In

In summary, the movement of air between the atmosphere and the lungs follows the laws of physics as they relate to gases. The air in the alveoli contains a mixture of gases, including nitrogen, oxygen, carbon dioxide, and water vapor. With the exception of water vapor, each gas exerts a pressure that is determined by the atmospheric pressure and the concentration of the gas in the mixture. Water vapor pressure is affected by temperature but not atmospheric pressure. Air moves into the lungs along a pressure gradient. The pressure inside the airways and alveoli of the lungs is called *intrapulmonary* (or *alveolar*) *pressure*; the pressure in the pleural cavity is called *pleural pressure*; and the pressure in the thoracic cavity is called *intrathoracic pressure*.

Breathing is the movement of gases between the atmosphere and the lungs. It requires a system of open airways and pressure changes resulting from the action of the respiratory

TABLE 19-2	Pulmonary Function Tests				
Test		Symbol	Measurement*		
Maximal voluntar Forced vital capac	,	MVV FVC	Maximum amount of air that can be breathed in a given time Maximum amount of air that can be rapidly and forcefully exhaled from the lungs after full inspiration. The expired volume is plotted against time.		
Forced expiratory volume achieved in 1 second		FEV _{1.0}	Volume of air expired in the first second of FVC		
Percentage of forced vital capacity FEV		FEV _{1.0} /FVC%	Volume of air expired in the first second, expressed as a percentage of FVC		

*By convention, all the lung volumes and rates of flow are expressed in terms of body temperature and pressure and saturated with water vapor (BTPS), which allows for a comparison of the pulmonary function data from laboratories with different ambient temperatures and altitudes. muscles in changing the volume of the chest cage. The diaphragm is the principal muscle of inspiration, assisted by the external intercostal muscles. The scalene and sternocleidomastoid muscles elevate the ribs and act as accessory muscles for inspiration. Expiration is largely passive, aided by the elastic recoil of the respiratory muscles that were stretched during inspiration. When needed, the abdominal and internal intercostal muscles can be used to increase expiratory effort.

Lung compliance describes the ease with which the lungs can be inflated. It reflects the elasticity of the lung tissue and the surface tension in the alveoli. Surfactant molecules, produced by type II alveolar cells, reduce the surface tension in the lungs, thereby increasing lung compliance. Airway resistance refers to the impediment to flow that the air encounters as it moves through the airways. The minute volume, which is determined by the metabolic needs of the body, is the amount of air that is exchanged in 1 minute (*i.e.*, respiratory rate and TV). The efficiency and work of breathing are determined by factors such as impaired lung compliance and airway diseases that increase the work involved in maintaining the minute volume. Lung volumes and lung capacities can be measured using a spirometer. Pulmonary function studies are used to assess ventilation with respect to time.

EXCHANGE AND TRANSPORT OF GASES

The primary functions of the lungs are oxygenation of the blood and removal of carbon dioxide. Pulmonary gas exchange is conventionally divided into three processes: (1) ventilation or the flow of gases into and out of the alveoli of the lungs, (2) perfusion or flow of blood in the adjacent pulmonary capillaries, and (3) diffusion or transfer of gases between the alveoli and the pulmonary capillaries. The efficiency of gas exchange requires that alveolar ventilation occur adjacent to perfused pulmonary capillaries.

Ventilation

Ventilation refers to the exchange of gases in the respiratory system. There are two types of ventilation: pulmonary and alveolar. *Pulmonary ventilation* refers to the total exchange of gases between the atmosphere and the lungs. *Alveolar ventilation* is the exchange of gases within the gas exchange portion of the lungs. Ventilation requires a system of open airways and a pressure difference that moves air into and out of the lungs. It is affected by body position and lung volume as well as by disease conditions that affect the heart and respiratory system.

Distribution of Ventilation

The distribution of ventilation between the base (bottom) and apex (top) of the lung varies with body position and the effects of gravity on intrapleural pressure. Compliance reflects the change in volume that occurs with a change in intrapleural pressure. It is less in fully expanded alveoli, which have difficulty accommodating more air, and greater in alveoli that are less inflated. In the seated or standing position, gravity exerts a downward pull on the lung, causing intrapleural pressure at the apex of the lung to become more negative. As a result, the alveoli at the apex of the lung are more fully expanded and less compliant than those at the base of the lung. The same holds true for lung expansion in the dependent portions of the lung in the supine or lateral position. In the supine position, ventilation in the lowermost (posterior) parts of the lung exceeds that in the uppermost (anterior) parts. In the lateral position (*i.e.*, lying on the side), the dependent lung is better ventilated.

The distribution of ventilation also is affected by lung volumes. During full inspiration (high lung volumes) in the seated or standing position, the airways are pulled open and air moves into the more compliant portions of the lower lung. At low lung volumes, the opposite occurs. At functional residual capacity (see Fig. 19-16), the pleural pressure at the base of the lung exceeds airway pressure, compressing the airways so that ventilation is greatly reduced. In contrast, the airways in the apex of the lung remain open, and this area of the lung is well ventilated.

Perfusion

The term *perfusion* is used to describe the flow of blood through the pulmonary capillary bed. The primary functions of the pulmonary circulation are to perfuse or provide blood flow to the gas exchange portion of the lung and to facilitate gas exchange. The pulmonary circulation serves several important functions in addition to gas exchange. It filters all the blood that moves from the right to the left side of the circulation; it removes most of the thromboemboli that might form; and it serves as a reservoir of blood for the left side of the heart.

The gas exchange function of the lungs requires a continuous flow of blood through the respiratory portion of the lungs. Deoxygenated blood enters the lung through the pulmonary artery, which has its origin in the right side of the heart and enters the lung at the hilus, along with the primary bronchus. The pulmonary arteries branch in a manner similar to that of the airways. The small pulmonary arteries accompany the bronchi as they move down the lobules and branch to supply the capillary network that surrounds the alveoli (see Fig. 19-6). The oxygenated capillary blood is collected in the small pulmonary veins of the lobules, and then it moves to the larger veins to be collected in the four large pulmonary veins that empty into the left atrium.

Distribution of Blood Flow

As with ventilation, the distribution of pulmonary blood flow is affected by body position and gravity. In the upright position, the distance of the upper apices of the lung above the level of the heart may exceed the perfusion capabilities of the mean pulmonary arterial pressure (approximately 12 mm Hg); therefore, blood flow in the upper part of the lungs is less than that in the base or bottom part of the lungs. In the supine position, the lungs and the heart are at the same level, and blood flow to the apices and base of the lungs becomes more uniform. In this position, blood flow to the posterior or dependent portions (*e.g.*, bottom of the lung when lying on the side) exceeds flow in the anterior or nondependent portions of the lungs. In persons with left-sided heart failure, congestion develops in the dependent portions of the lungs exposed to increased blood flow.

Effects of Hypoxia

The blood vessels in the pulmonary circulation undergo marked vasoconstriction when they are exposed to hypoxia. The precise mechanism for this response is unclear. When alveolar oxygen levels drop below 60 mm Hg, marked vasoconstriction may occur, and at very low oxygen levels, the local flow may be almost abolished. In regional hypoxia, as occurs with a localized airway obstruction (*e.g.*, atelectasis), vasoconstriction is localized to a specific region of the lung. Vasoconstriction has the effect of directing blood flow away from the hypoxic regions of the lungs. When alveolar hypoxia no longer exists, blood flow is restored.

Generalized hypoxia causes vasoconstriction throughout the lung. Generalized vasoconstriction occurs when the partial pressure of oxygen is decreased at high altitudes, or it can occur in persons with chronic hypoxia caused by lung disease. Prolonged hypoxia can lead to pulmonary hypertension and increased workload on the right heart. A low blood pH also produces vasoconstriction, especially when alveolar hypoxia is present (*e.g.*, during circulatory shock).

Diffusion

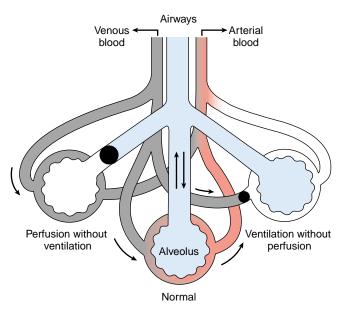
Diffusion refers to the movement of gases in the alveoli and across the alveolar capillary membrane. Diffusion of gases in the lung is affected by the difference in the pressure of gas across the membrane, the surface area that is available for diffusion, and the thickness of the alveolar capillary membrane through which the gas must pass. Administration of high concentrations of oxygen increases the pressure difference between the two sides of the membrane and increases the diffusion of the gas. Diseases that destroy lung tissue and the surface area for diffusion and those that increase the thickness of the alveolar-capillary membrane adversely influence the diffusing capacity of the lungs. For example, the removal of one lung reduces the diffusing capacity by one half. The thickness of the alveolar-capillary membrane and the distance for diffusion are increased in persons with pulmonary edema or pneumonia. The characteristics of the gas and its molecular weight and solubility constitute the diffusion coefficient and determine how rapidly the gas diffuses through the respiratory membranes. For example, carbon dioxide diffuses 20 times more rapidly than oxygen because of its greater solubility in the respiratory membranes.

Matching of Ventilation and Perfusion

The gas exchange properties of the lung depend on matching ventilation and perfusion, ensuring that equal amounts of air and blood are entering the respiratory portion of the lungs (Fig. 19-17). There are two factors that may interfere with the matching of ventilation and perfusion: (1) dead air space and (2) shunt.

Dead Air Space

Dead space refers to the air that must be moved with each breath but does not participate in gas exchange. The movement of air through dead space contributes to the work of breathing but not to gas exchange. There are two types of dead space: that contained in the conducting airways, called the *anatomic dead space*, and that contained in the respiratory portion of the lung,



■ **FIGURE 19-17** ■ Matching of ventilation and perfusion. (**Center**) Normal matching of ventilation and perfusion; (**left**) perfusion without ventilation (*i.e.*, shunt); (**right**) ventilation without perfusion (*i.e.*, dead air space).

called the *alveolar dead space*. The volume of anatomic airway dead space is fixed at approximately 150 to 200 mL, depending on body size. It constitutes air contained in the nose, pharynx, trachea, and bronchi. The creation of an opening in the trachea to facilitate ventilation (tracheostomy) decreases anatomic dead space ventilation because air does not have to move through the nasal and oral airways. Alveolar dead space, normally about 5 to 10 mL, constitutes alveolar air that does not participate in gas exchange. When alveoli are ventilated but deprived of blood

KEY CONCEPTS

MATCHING OF VENTILATION AND PERFUSION

- Exchange of gases between the air in the alveoli and the blood in pulmonary capillaries requires a matching of ventilation and perfusion.
- Dead air space refers to air that is moved with each breath but is not ventilated. Anatomic dead space is that contained in the conducting airways that normally do not participate in gas exchange. Alveolar dead space results from alveoli that are ventilated but not perfused.
- Shunt refers to blood that moves from the right to the left side of the circulation without being oxygenated. With an anatomic shunt, blood moves from the venous to the arterial side of the circulation without going through the lungs. Physiologic shunting results from blood moving through unventilated parts of the lung.

flow, they do not contribute to gas exchange and thereby constitute alveolar dead space.

The *physiologic dead space* includes the anatomic dead space plus alveolar dead space. In persons with normal respiratory function, physiologic dead space is about the same as anatomic dead space. Only in lung disease does physiologic dead space increase.

Shunt

Shunt refers to blood that moves from the right to the left side of the circulation without being oxygenated. There are two types of shunts: physiologic and anatomic. In a physiologic shunt, there is mismatching of ventilation and perfusion, resulting in insufficient ventilation to provide the oxygen needed to oxygenate the blood flowing through the alveolar capillaries. Physiologic shunting of blood usually results from destructive lung disease that impairs ventilation or from heart failure that interferes with movement of blood through sections of the lungs. In an anatomic shunt, blood moves from the venous to the arterial side of the circulation without moving through the lungs. Anatomic intracardiac shunting of blood caused by congenital heart defects is discussed in Chapter 17.

Mismatching of Ventilation and Perfusion

Mismatching of ventilation and perfusion occurs when there is perfusion without ventilation or ventilation without perfusion (see Figure 19-17). Perfusion without ventilation (shunt) results in a low ventilation-perfusion ratio. This is the type of situation that occurs when there is incomplete expansion of the lung, such as in atelectasis (Chap-ter 21). Ventilation without perfusion (dead air space) results in a high ventilation-perfusion ratio. An example of this type of situation is pulmonary embolism, when a blood clot obstructs flow (Chapter 21). The PO₂ in the arterial blood leaving the pulmonary circulation reflects mixing of blood from areas of shunt and dead air space.

Gas Transport

The lungs enable inhaled air to come in contact with blood flowing through the pulmonary capillaries so that exchange of gases between the external environment and the internal environment of the body can occur. The lungs restore the oxygen content of the arterial blood and remove carbon dioxide from the venous blood.

The blood carries oxygen and carbon dioxide in the dissolved state and in combination with hemoglobin. Carbon dioxide also is converted to bicarbonate and transported in that form. The amount of a gas that can dissolve in plasma is determined by two factors: the solubility of the gas in the plasma and the partial pressure of the gas in the alveoli. Oxygen and carbon dioxide dissolve in plasma. The gases that are dissolved in plasma are similar to the carbon dioxide that is dissolved in a capped bottle of a carbonated drink. In the case of the carbonated drink, increased pressure is used to increase the dissolved carbon dioxide to the drink. When the bottle cap is removed and the pressure reduced, tiny bubbles can be seen as the gas moves from the dissolved to the gaseous state.

In the clinical setting, blood gas measurements are used to determine the level of the partial pressure of oxygen (PO_2) and carbon dioxide (PCO₂) in the blood. Arterial blood commonly is used for measuring blood gases. Venous blood is not used because venous levels of oxygen and carbon dioxide reflect the metabolic demands of the tissues, rather than the gas exchange function of the lungs. The PO₂ of arterial blood normally is greater than 80 mm Hg, and the PCO_2 is in the range of 35 to 45 mm Hg. Normally, the arterial blood gases are the same or nearly the same as the partial pressure of the gases in the alveoli. The arterial PO_2 often is written PaO_{2} , and the alveolar PO_2 as PAO_{2} , with the same types of designations being used for PCO_2 . This text uses PO_2 and PCO_2 to designate both arterial and alveolar levels of the gases.

Oxygen Transport

Oxygen is transported in two forms: (1) in chemical combination with hemoglobin and (2) in the dissolved state. The hemoglobin in red blood cells serves as a transport vehicle for oxygen. It binds oxygen in the pulmonary capillaries and releases it in the tissue capillaries. As oxygen moves into or out of the red blood cells, it dissolves in the plasma. It is the dissolved form of oxygen that leaves the capillary, crosses cell membranes, and participates in cell metabolism. Only approximately 1% of the oxygen in the blood is carried in the dissolved state; the remainder is carried in combination with hemoglobin. The oxygen content of the blood (measured in milliliters per 100 milliliters of blood) includes the oxygen carried by hemoglobin and dissolved oxygen.

Hemoglobin Transport. Hemoglobin is a highly efficient carrier of oxygen, and approximately 98% to 99% of the oxygen used by body tissues is carried in this manner. Hemoglobin with bound oxygen is called *oxyhemoglobin*, and when oxygen is removed, it is called deoxygenated or reduced hemoglobin. Each gram of hemoglobin carries approximately 1.34 mL of oxygen when it is fully saturated. This means that a person with a hemoglobin of 14 g/100 mL carries 18.8 mL of oxygen per 100 mL of blood. In the lungs, oxygen moves across the alveolar-capillary membrane, through the plasma, and into

KEY CONCEPTS

OXYGEN TRANSPORT

- Oxygen is transported in chemical combination with hemoglobin and as a gas dissolved in the plasma.
- Hemoglobin, which is the main transporter for oxygen, binds oxygen as it passes through the lungs and releases it as it moves through the tissues.
- The amount of oxygen that is carried as a dissolved gas is determined by the partial pressure of the gas in the lungs.
- The oxygen content of the blood, or the amount of oxygen that is available to the tissues, represents the total amount of oxygen carried by the hemoglobin plus the amount of oxygen that is carried in the dissolved state.

the red blood cell, where it forms a loose and reversible bond with the hemoglobin molecule. In normal lungs, this process is rapid, so that even with a fast heart rate, the hemoglobin is almost completely saturated with oxygen during the short time it spends in the pulmonary capillaries.

The oxygenated hemoglobin is transported in the arterial blood to the peripheral capillaries, where the oxygen is released and made available to the tissues for use in cell metabolism. As the oxygen moves out of the capillaries in response to the needs of the tissues, the hemoglobin saturation, which usually is approximately 95% to 97% as the blood leaves the left side of the heart, drops to approximately 75% as the mixed venous blood returns to the right side of the heart.

Dissolved Oxygen. The partial pressure of oxygen (PO₂) represents the level of dissolved oxygen in plasma. The amount of gas that can be dissolved in a liquid depends on the solubility of the gas and its pressure. The solubility of oxygen in plasma is fixed and very small. For every 1 mm Hg of PO₂ present in the alveoli, 0.003 mL of oxygen becomes dissolved in 100 mL of plasma. This means that at a normal alveolar PO₂ of 100 mm Hg, the blood carries only 0.3 mL of dissolved oxygen in each 100 mL of plasma. This amount is very small compared with the amount that can be carried in an equal amount of blood when oxygen is attached to hemoglobin.

Although the amount of oxygen carried in plasma under normal conditions is small, it can become a life-saving mode of transport in carbon monoxide poisoning, when most of the hemoglobin sites are occupied by carbon monoxide and are unavailable for transport of oxygen. The use of a hyperbaric chamber, in which 100% oxygen can be administered at high atmospheric pressures, increases the amount of oxygen that can be carried in the dissolved state.

Oxygen-Hemoglobin Dissociation. Oxygen that remains bound to hemoglobin cannot participate in tissue metabolism. The efficiency of the oxygen dissociation transport system depends on the ability of the hemoglobin molecule to bind oxygen in the lungs and release it, as it is needed in the tissues. The affinity of hemoglobin refers to its capacity to bind oxygen. Hemoglobin binds oxygen more readily when its affinity is increased and releases it more readily when its affinity is decreased. As described in Chapter 13, the hemoglobin molecule is composed of four polypeptide chains bound to an iron-containing heme group. Because oxygen binds to the iron atom, each hemoglobin molecule can bind four molecules of oxygen. When oxygen is bound to all four of the heme groups, the hemoglobin molecule is said to be fully saturated. Hemoglobin is partially saturated when it contains only one, two, or three molecules of oxygen.

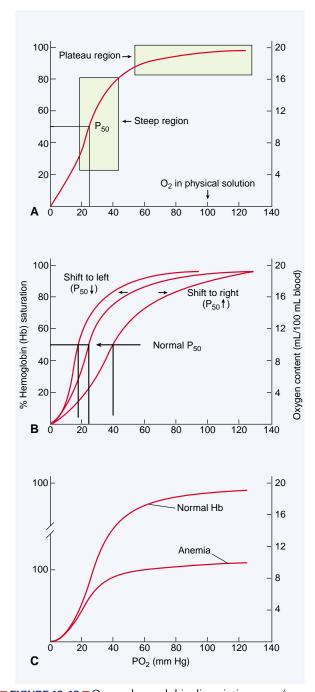
Hemoglobin's affinity for oxygen is influenced by pH, carbon dioxide concentration, and temperature. Hemoglobin binds oxygen more strongly under conditions of increased pH (alkalosis), decreased carbon dioxide concentration, and decreased body temperature, and releases it more readily under conditions of decreased pH (acidosis), increased carbon dioxide concentration, and fever. Conditions that decrease affinity and favor unloading of oxygen reflect the level of tissue metabolism and need for oxygen. For example, increased tissue metabolism generates carbon dioxide and metabolic acids, thereby decreasing the affinity of hemoglobin for oxygen. Heat also is a byproduct of tissue metabolism, explaining the effect of fever on oxygen binding. Red blood cells contain a metabolic intermediate called 2,3-diphosphoglycerate (2,3-DPG) that also affects the affinity of hemoglobin for oxygen. An increase in 2,3-DPG enhances unloading of oxygen from hemoglobin at the tissue level. An increase in 2,3-DPG occurs with exercise and the hypoxia that occurs with high altitude and chronic lung disease.

The relation between the oxygen carried in combination with hemoglobin and the PO₂ of the blood is described by the oxygen-hemoglobin dissociation curve, which is shown in Figure 19-18A. The x axis depicts the PO_2 ; the left y axis, hemoglobin saturation, and the right y axis, the oxygen content or total amount of oxygen in the blood, including the dissolved oxygen and that carried by the hemoglobin. There are three important things to observe about the relationships among PO_{2} , hemoglobin saturation, and oxygen content. First, PO_{2} is the dissolved oxygen. It reflects the partial pressure of the gas in the lung (*i.e.*, the PO_2 is approximately 100 mm Hg when room air is being breathed but can rise to 200 mm Hg or higher when oxygen-enriched air is breathed). Second, hemoglobin saturation reflects the amount of oxygen that is carried by the hemoglobin. Third, it is the oxygen content of the blood, rather than the PO₂, or even hemoglobin saturation, that determines the amount of oxygen that is delivered to the tissues. An anemic person may have a normal PO₂ and hemoglobin saturation level but decreased oxygen content because of the lower amount of hemoglobin for binding oxygen (Fig. 19-18C).

The oxygen-hemoglobin dissociation curve has a flat top portion representing binding of oxygen by the hemoglobin in the lungs and a steep portion representing its release into the tissue capillaries. The S shape of the curve reflects the effect that oxygen saturation has on the affinity of hemoglobin for oxygen. At approximately 100 mm Hg PO₂, a plateau occurs, at which point the hemoglobin is approximately 98% saturated. Increasing the alveolar PO₂ above this level does not increase the hemoglobin saturation. Even at high altitudes, when the partial pressure of oxygen is considerably decreased, the hemoglobin remains relatively well saturated. For example, at 60 mm Hg PO₂, the hemoglobin is still approximately 89% saturated.

The steep portion of the dissociation curve—between 60 and 40 mm Hg—represents the removal of oxygen from the hemoglobin as it moves through the tissue capillaries. This portion of the curve shows that there is considerable transfer of oxygen from hemoglobin to the tissues with only a small drop in PO_2 , thereby ensuring a gradient for oxygen to move into body cells. The tissues normally remove approximately 5 mL of oxygen per 100 mL of blood, and the hemoglobin of mixed venous blood is approximately 75% saturated as it returns to the right side of the heart.

Hemoglobin can be regarded as an oxygen buffer system that regulates oxygen pressure in the tissues. Hemoglobin affinity for oxygen must change with the metabolic needs of the tissues. This change is represented by a shift to the right or left in the dissociation curve, as shown in Figure 19-18B. The P_{50} represents the PO₂ at a hemoglobin of 50%. It can be used



■ FIGURE 19-18 ■ Oxygen-hemoglobin dissociation curve (oxygen content at hemoglobin of 14 gm/dL). (A) Left boxed area represents the steep portion of the curve where oxygen is released from hemoglobin (Hb) to the tissues, and the top boxed area the plateau of the curve where oxygen is loaded onto hemoglobin in the lung. P₅₀ partial pressure of oxygen required to saturate 50% of hemoglobin with oxygen. (B) The effect of body temperature, arterial PCO₂, and pH on hemoglobin affinity for oxygen as indicated by a shift in the curve and position of the P₅₀. A shift of the curve to the right due to an increase in temperature, PCO₂, or decreased pH favors release of oxygen to the tissues. A decrease in temperature, PCO₂, or increase in pH shifts the curve to the left. (C) Effect of anemia on the oxygen-carrying capacity of blood. The hemoglobin can be completely saturated, but the oxygen content of the blood is reduced. (Adapted from Rhoades R.A., Tanner G.A. [1996]. Medical physiology. Boston: Little, Brown)

to determine whether the dissociation curve has shifted to the right or to the left. A *shift to the right* indicates that the tissue PO_2 is greater for any given level of hemoglobin saturation and represents reduced affinity of the hemoglobin for oxygen at any given PO_2 . It usually is caused by conditions such as fever or acidosis or by an increase in PCO_2 , which reflects increased tissue metabolism. High altitude and conditions such as pulmonary insufficiency, heart failure, and severe anemia also cause the oxygen dissociation curve to shift to the right.

A *shift to the left* in the oxygen dissociation curve represents enhanced affinity of hemoglobin for oxygen and occurs in situations associated with a decrease in tissue metabolism, such as alkalosis, decreased body temperature, and decreased carbon dioxide levels.

Carbon Dioxide Transport

Carbon dioxide is transported in the blood in three forms: as dissolved carbon dioxide (10%), attached to hemoglobin (30%), and as bicarbonate (60%). Acid-base balance is influenced by the amount of dissolved carbon dioxide and the bicarbonate level in the blood (see Chapter 6).

As carbon dioxide is formed during the metabolic process, it diffuses out of cells into the tissue spaces and then into the capillaries. The amount of dissolved carbon dioxide that can be carried in plasma is determined by the partial pressure of the gas and its solubility coefficient (0.03 mL/100 mL/1 mm Hg PCO₂). Carbon dioxide is 20 times more soluble in plasma than oxygen. Thus, the dissolved state plays a greater role in transport of carbon dioxide compared with oxygen.

Most of the carbon dioxide diffuses into the red blood cells, where it forms carbonic acid or combines with hemoglobin. Carbonic acid (H_2CO_3) is formed when carbon dioxide combines with water ($CO_2 + H_2O = H^+ + HCO_3^-$). The process is catalyzed by an enzyme called *carbonic anhydrase*, which greatly increases the rate of the reaction. Carbonic acid readily ionizes to form a bicarbonate (HCO_3^-) and a hydrogen (H^+) ion. The hydrogen ion that is generated combines with the hemoglobin, which is a powerful acid-base buffer, and the bicarbonate ion diffuses into plasma in exchange for a chloride ion.

In addition to the carbonic anhydrase-mediated reaction with water, carbon dioxide reacts directly with hemoglobin to form carbaminohemoglobin. The combination of carbon dioxide with hemoglobin is a reversible reaction that involves a loose bond, which allows transport of carbon dioxide from tissues to the lungs, where it is released into the alveoli for exchange with the external environment. The release of oxygen from hemoglobin in the tissues enhances the binding of carbon dioxide to hemoglobin; in the lungs, the combining of oxygen with hemoglobin displaces carbon dioxide. The binding of carbon dioxide to hemoglobin is determined by the acidic nature of hemoglobin. Binding with carbon dioxide causes the hemoglobin to become a stronger acid. In the lungs, the highly acidic hemoglobin has a lesser tendency to form carbaminohemoglobin, and carbon dioxide is released from hemoglobin into the alveoli. In the tissues, the release of oxygen from hemoglobin causes hemoglobin to become less acid, thereby increasing its ability to combine with carbon dioxide and form carbaminohemoglobin.

356 Unit Five: Alterations in the Respiratory System

In summary, the primary functions of the lungs are oxygenation of the blood and removal of carbon dioxide. Pulmonary gas exchange is conventionally divided into three processes: ventilation, or the flow of gases into the alveoli of the lungs; perfusion, or movement of blood through the adjacent pulmonary capillaries; and diffusion, or transfer of gases between the alveoli and the pulmonary capillaries.

Ventilation is the movement of air between the atmosphere and the lungs. Pulmonary ventilation refers to the total exchange of gases between the atmosphere and the lungs, and alveolar ventilation refers to ventilation in the gas exchange portion of the lungs. The distribution of alveolar ventilation and pulmonary capillary blood flow varies with lung volume and body position. In the upright position and at high lung volumes, ventilation is greatest in the lower parts of the lungs. The upright position also produces a decrease in blood flow to the upper parts of the lung, resulting from the distance above the level of the heart and the low mean arterial pressure in the pulmonary circulation.

The diffusion of gases in the lungs is influenced by four factors: the surface area available for diffusion; the thickness of the alveolar-capillary membrane, through which the gases diffuse; the differences in the partial pressure of the gas on either side of the membrane; and the characteristics of the gas. The efficiency of gas exchange requires matching of ventilation and perfusion so that equal amounts of air and blood enter the respiratory portion of the lungs. Two factors—dead air space and shunt—interfere with matching of ventilation and perfusion and do not contribute to gas exchange. Dead air space occurs when areas of the lungs are ventilated but not perfused. Shunt is the condition under which areas of the lungs are perfused but not ventilated.

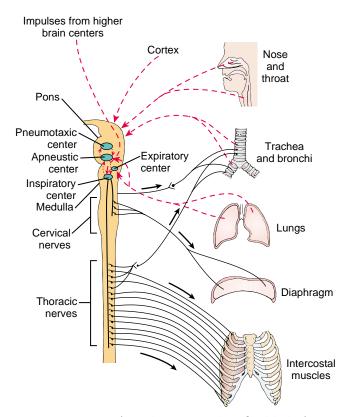
The blood transports oxygen to the cells and returns carbon dioxide to the lungs. Oxygen is transported in two forms: in chemical combination with hemoglobin and physically dissolved in plasma (PO_2). Hemoglobin is an efficient carrier of oxygen, and approximately 98% to 99% of oxygen is transported in this manner. Carbon dioxide is carried in three forms: carbaminohemoglobin (30%), dissolved carbon dioxide (10%), and bicarbonate (60%).

CONTROL OF BREATHING

Unlike the heart, which has inherent rhythmic properties and can beat independently of the nervous system, the muscles that control respiration require continuous input from the nervous system. Movement of the diaphragm, intercostal muscles, sternocleidomastoid, and other accessory muscles that control ventilation is integrated by neurons located in the pons and medulla. These neurons are collectively referred to as the *respiratory center* (Fig. 19-19).

Respiratory Center

The medullary respiratory center consists of two groups of neurons involved in initiating inspiration and expiration and incorporating afferent impulses into the motor responses of the respiratory muscles. The first, or inspiratory area, which is



■ FIGURE 19-19 ■ Schematic representation of activity in the respiratory center. Impulses traveling over afferent neurons (*red dashed lines*) communicate with central neurons, which activate efferent neurons that supply the muscles of respiration. Respiratory movements can be altered by a variety of stimuli.

located in the dorsal region of the medulla, controls the activity of the phrenic nerves that innervate the diaphragm and another group of neurons, which innervate the spinal motor neurons of the intercostals and abdominal muscles. The second group of neurons is concerned with the expiratory phase of respiration.

The pacemaker properties of the respiratory center result from the cycling of the two groups of respiratory neurons in the pons: the *pneumotaxic center* in the upper pons and the *apneustic center* in the lower pons. These two groups of neurons contribute to the function of the respiratory center in the medulla. The apneustic center has an excitatory effect on inspiration, tending to prolong inspiration. The pneumotaxic center switches inspiration off, assisting in the control of respiratory rate and inspiratory volume. Brain injury that damages the connection between the pneumotaxic and apneustic centers results in an irregular breathing pattern that consists of prolonged inspiratory gasps interrupted by expiratory efforts.

Axons from the neurons in the respiratory center cross in the midline and descend in the ventrolateral columns of the spinal cord. The tracts that control expiration and inspiration are spatially separated in the cord, as are the tracts that transmit specialized reflexes (*i.e.*, coughing and hiccupping) and voluntary control of ventilation. Only at the level of the spinal cord are the respiratory impulses integrated to produce a reflex response.

The control of breathing has automatic and voluntary components. The automatic regulation of ventilation is controlled by input from two types of sensors or receptors: chemoreceptors and lung receptors. Chemoreceptors monitor blood levels of oxygen, carbon dioxide, and pH and adjust ventilation to meet the changing metabolic needs of the body. Lung receptors monitor breathing patterns and lung function. Voluntary regulation of ventilation integrates breathing with voluntary acts, such as speaking, blowing, and singing. These voluntary acts, initiated by the motor and premotor cortex, cause temporary suspension of automatic breathing. The automatic and voluntary components of respiration are regulated by afferent impulses that come to the respiratory center from a number of sources. Afferent input from higher brain centers is evidenced by the fact that a person can consciously alter the depth and rate of respiration. Fever, pain, and emotion exert their influence through lower brain centers. Vagal afferents from sensory receptors in the lungs and airways are integrated in the dorsal area of the respiratory center.

Chemoreceptors

Tissue needs for oxygen and the removal of carbon dioxide are regulated by chemoreceptors that monitor blood levels of these gases. Input from these sensors is transmitted to the respiratory center, and ventilation is adjusted to maintain the arterial blood gases within a normal range.

There are two types of chemoreceptors: central and peripheral. The most important chemoreceptors for sensing changes in blood carbon dioxide content are the *central chemoreceptors*. These receptors are located in chemosensitive regions near the respiratory center in the medulla and are bathed in cerebrospinal fluid. Although the central chemoreceptors monitor carbon dioxide levels, the actual stimulus for these receptors is provided by hydrogen ions in the cerebrospinal fluid (CSF). The CSF is separated from the blood by the blood-brain barrier, which permits free diffusion of carbon dioxide but not bicarbonate or hydrogen ions. The carbon dioxide combines rapidly with water to form carbonic acid, which dissociates into hydrogen and bicarbonate ions. The carbon dioxide content in the blood regulates ventilation through its effect on the pH of the extracellular fluid of the brain. The central chemoreceptors are extremely sensitive to short-term changes in carbon dioxide. An increase in plasma carbon dioxide levels increases ventilation, reaching its peak within a minute or so and then declining if the carbon dioxide level remains elevated. Thus, persons with chronically elevated levels of carbon dioxide no longer have a response to this stimulus for increased ventilation but rely on the stimulus provided by a decrease in blood oxygen levels.

The *peripheral chemoreceptors* are located in the carotid and aortic bodies, which are found at the bifurcation of the common carotid arteries and in the arch of the aorta, respectively. These chemoreceptors monitor arterial blood oxygen levels. Although the peripheral chemoreceptors also monitor carbon dioxide, they play a much more important role in monitoring oxygen levels. These receptors exert little control over ventilation until the PO₂ has dropped below 60 mm Hg. Thus, hypoxia is the main stimulus for ventilation in persons with chronically elevated levels of carbon dioxide. If these patients are given oxygen therapy at a level sufficient to increase the PO₂

above that needed to stimulate the peripheral chemoreceptors, their ventilation may be seriously depressed.

Lung Receptors

Lung and chest wall receptors provide information on the status of breathing in terms of airway resistance and lung expansion. There are three types of lung receptors: stretch, irritant, and juxtacapillary receptors.

Stretch receptors are located in the smooth muscle layers of the conducting airways. They respond to changes in pressure in the walls of the airways. When the lungs are inflated, these receptors inhibit inspiration and promote expiration. They are important in establishing breathing patterns and minimizing the work of breathing by adjusting respiratory rate and TV to accommodate changes in lung compliance and airway resistance.

The *irritant receptors* are located between the airway epithelial cells. They are stimulated by noxious gases, cigarette smoke, inhaled dust, and cold air. Stimulation of the irritant receptors leads to airway constriction and a pattern of rapid, shallow breathing. This pattern of breathing probably protects respiratory tissues from the damaging effects of toxic inhalants. It also is thought that the mechanical stimulation of these receptors may ensure more uniform lung expansion by initiating periodic sighing and yawning. It is possible that these receptors are involved in the bronchoconstriction response that occurs in some persons with bronchial asthma.

The *juxtacapillary* or *J receptors* are located in the alveolar wall, close to the pulmonary capillaries. It is thought that these receptors sense lung congestion. These receptors may be responsible for the rapid, shallow breathing that occurs with pulmonary edema, pulmonary embolism, and pneumonia.

Cough Reflex

Coughing is a neurally mediated reflex that protects the lungs from the accumulation of secretions and from entry of irritating and destructive substances. It is one of the primary defense mechanisms of the respiratory tract. The cough reflex is initiated by receptors located in the tracheobronchial wall; these receptors are extremely sensitive to irritating substances and to the presence of excess secretions. Afferent impulses from these receptors are transmitted through the vagus to the medullary center, which integrates the cough response.

Coughing itself requires the rapid inspiration of a large volume of air (usually about 2.5 L), followed by rapid closure of the glottis and forceful contraction of the abdominal and expiratory muscles. As these muscles contract, intrathoracic pressures are elevated to levels of 100 mm Hg or more. The rapid opening of the glottis at this point leads to an explosive expulsion of air.

Many conditions can interfere with the cough reflex and its protective function. The reflex is impaired in persons whose abdominal or respiratory muscles are weak. This problem can be caused by disease conditions that lead to muscle weakness or paralysis, by prolonged inactivity, or as an outcome of surgery involving these muscles. Bed rest interferes with expansion of the chest and limits the amount of air that can be taken into the lungs in preparation for coughing, making the cough weak and ineffective. Disease conditions that prevent effective closure of

358 Unit Five: Alterations in the Respiratory System

the glottis and laryngeal muscles interfere with production of the marked increase in intrathoracic pressure that is needed for effective coughing. For example, the presence of a nasogastric tube may prevent closure of the upper airway structures and may fatigue the receptors for the cough reflex that are located in the area. The cough reflex also is impaired when there is depressed function of the medullary centers in the brain that integrate the cough reflex. Interruption of the central integration aspect of the cough reflex can arise as the result of disease of this part of the brain or the action of drugs that depress the cough center.

Although the cough reflex is a protective mechanism, frequent and prolonged coughing can be exhausting and painful and can exert undesirable effects on the cardiovascular and respiratory systems and on the elastic tissues of the lungs. This is particularly true in young children and elderly persons.

In summary, the respiratory system requires continuous input from the nervous system. Movement of the diaphragm, intercostal muscles, and other respiratory muscles is controlled by neurons of the respiratory center located in the pons and medulla. The control of breathing has automatic and voluntary components. Voluntary respiratory control is needed for integrating breathing and actions such as speaking, blowing, and singing. These acts, which are initiated by the motor and premotor cortex, cause temporary suspension of automatic breathing.

The automatic regulation of ventilation is controlled by two types of receptors: lung receptors, which protect respiratory structures, and chemoreceptors, which monitor the gas exchange function of the lungs by sensing changes in blood levels of carbon dioxide, oxygen, and pH. There are three types of lung receptors: stretch receptors, which monitor lung inflation; irritant receptors, which protect against the damaging effects of toxic inhalants; and J receptors, which are thought to sense lung congestion. There are two groups of chemoreceptors: central and peripheral. The central chemoreceptors are the most important in sensing changes in carbon dioxide levels, and the peripheral chemoreceptors function in sensing arterial blood oxygen levels.

The cough reflex protects the lungs from the accumulation of secretions and from the entry of irritating and destructive substances; it is one of the primary defense mechanisms of the respiratory tract.

REVIEW QUESTIONS

State the difference between the conducting and the respiratory airways.

Describe the function of the mucociliary blanket.

Define the term water vapor pressure and cite the source of water for humidification of air as it moves through the airways.

State the function of the two types of alveolar cells and use Laplace's law to explain the need for surfactant in maintaining the inflation of small alveoli.

State the definition of intrathoracic, intrapleural, and intraalveolar pressures, and state how each of these pressures changes in relation to atmospheric pressure during inspiration and expiration.

Explain why ventilation and perfusion must be matched and describe their relation to dead air space and shunt.

Explain the difference between PO_2 and hemoglobin-bound oxygen and O_2 saturation and content and explain the effect of administering 40% oxygen on O_2 saturation and PO_2 .

Explain the significance of a shift to the right and a shift to the left in the oxygen-hemoglobin dissociation curve.

Describe the function of the chemoreceptors and lung receptors in the regulation of ventilation.

Trace the integration of the cough reflex from stimulus to explosive expulsion of air that constitutes the cough.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

BIBLIOGRAPHY

- Berne R.M., Levy M.N. (2000). Principles of physiology (3rd ed., pp. 302– 352). St. Louis: C.V. Mosby.
- Fishman A.P. (1980). Assessment of pulmonary function. New York: McGraw-Hill.
- Guyton A., Hall J.E. (2000). Textbook of medical physiology (10th ed., pp. 432–482). Philadelphia: W.B. Saunders.
- Rhoades R.A., Tanner G.A. (1996). *Medical physiology* (pp. 341–414). Boston: Little, Brown.
- West J.B. (2000). Respiratory physiology: The essentials. Philadelphia: Lippincott Williams & Wilkins.
- West J.B. (2001). Pulmonary physiology and pathophysiology: An integrated, case-based approach. Philadelphia: Lippincott Williams & Wilkins.

CHAPTER 20

Alterations in Respiratory Function: Infectious Disorders and Neoplasia

Respiratory Tract Infections Influenza **Pneumonias** Acute Bacterial (Typical) Pneumonias **Primary Atypical Pneumonias Tuberculosis Primary Tuberculosis** Secondary Tuberculosis **Diagnosis and Treatment Fungal Infections** Cancer of the Lung **Bronchogenic Carcinoma Manifestations Diagnosis and Treatment Respiratory Disorders in Infants and Children** Lung Development Breathing in the Fetus and Neonate Lung Volumes and Gas Exchange **Control of Ventilation Respiratory Disorders in the Neonate Respiratory Distress Syndrome Bronchopulmonary Dysplasia Respiratory Infections in Children Upper Airway Infections** Lower Airway Infections Signs of Impending Respiratory Failure

Respiratory illnesses represent one of the more common reasons for visits to the physician, admission to the hospital, and forced inactivity among all age groups. Pneumonia is the sixth leading cause of death in the United States, particularly among the elderly and those with compromised immune function. Tuberculosis remains one of the deadliest diseases in the world. It has been estimated that between 19% and 43% of the world population is infected with tuberculosis. Of all neoplasms, lung cancer remains the leading cause of death in the United States.

RESPIRATORY TRACT INFECTIONS

Respiratory tract infections can involve the upper respiratory tract (*i.e.*, nose, oropharynx, and larynx), the lower respiratory tract (*i.e.*, lower airways and lungs), or the upper and lower airways. The discussion in this section of the chapter focuses on influenza, pneumonia, tuberculosis, and fungal infections of the lung. Acute respiratory infections in children are discussed in the last section of the chapter.

Influenza

Influenza is a viral infection that can affect the upper and lower respiratory tracts. It usually occurs in epidemics or pandemics. Until the advent of acquired immunodeficiency syndrome (AIDS), it was the last uncontrolled pandemic killer of humans. More persons died in the 1918 and 1919 influenza pandemic than in World War I. In the United States, approximately 20,000 persons die each year of influenza-related illness during nonpandemic years.¹ Most deaths are caused by pneumonia or exacerbation of cardiopulmonary or other conditions; 80% to 90% of those who die are 65 years of age or older.

There are two types of influenza viruses that cause epidemics in humans: types A and B. Infection with type A is most common and causes the most severe disease. Influenza A is further divided into subtypes based on two surface antigens: hemagglutinin (H) and neuraminidase (N). Influenza B has not been categorized into subtypes. Host antibodies to the surface antigens, which provide entrance in host cells, prevent future infection with influenza virus. New variants result from frequent mutations or antigenic shifts in the surface antigens.¹ Influenza B undergoes less frequent shifts than influenza A. The incubation period for influenza is 1 to 4 days, with 2 days being the average. Persons become infectious starting 1 day before their symptoms begin and remain infectious through approximately 5 days after illness onset. Children can be infectious for a longer time.

The influenza viruses can cause three types of infections: an uncomplicated rhinotracheitis, a respiratory viral infection followed by a bacterial infection, and viral pneumonia. In the early stages, the symptoms of influenza often are indistinguishable from other viral infections. There is an abrupt onset of fever and chills, malaise, muscle aching, headache, profuse, watery nasal discharge, nonproductive cough, and sore throat. One distinguishing feature of an influenza viral infection is the rapid onset, sometimes in as little as 1 to 2 minutes, of profound malaise.² The infection causes necrosis and shedding of the serous and ciliated cells that line the respiratory tract, leaving gaping holes between the underlying basal cells and allowing extracellular fluid to escape. This is the reason for the "runny nose" that is characteristic of this phase of the infection. During recovery, the serous cells are replaced more rapidly than the ciliated cells. Mucus is produced, but the ciliated cells are unable to move it adequately; people recovering from influenza continue to blow their nose to clear the nasopharynx and cough to clear the trachea.

The symptoms of uncomplicated rhinotracheitis usually peak by days 3 to 5 and disappear by days 7 to 10. Persons who have secondary complications usually report that they were beginning to feel better when they experienced a return of symptoms. Complications typically include sinusitis, otitis media, bronchitis, and bacterial pneumonia. The clinical course of influenza pneumonia progresses rapidly. It can cause hypoxemia and death within a few days of onset. The rapid onset is thought to be related to the mode of spread and the absence of an initial rhinotracheitis. If the virus is spread by fingers or largedroplet spray, as from sneezing or coughing, only the upper respiratory tract is involved. Infection of the upper respiratory tract is thought to give the immune system enough time to build the defenses needed to protect against viral pneumonia. When the virus is contained in small droplets, it can bypass the upper respiratory tract and travel directly into the lungs to establish infection.2

The goals of treatment for influenza are designed to limit the infection to the upper respiratory tract. The symptomatic approach, which uses rest, keeping warm, and drinking large amounts of liquids, helps to accomplish this. Rest decreases the oxygen requirements of the body and reduces the respiratory rate and the chance of moving the virus from the upper to lower respiratory tract. Keeping warm helps maintain the respiratory epithelium at a core body temperature of 37°C (or higher if fever is present), thereby inhibiting viral replication, which is optimal at 35°C. Drinking large amounts of liquids ensures that the function of the epithelial lining of the respiratory tract is not further compromised by dehydration.

The appropriate pharmacologic treatment of people with influenza depends on accurate and timely diagnosis. Rapid antigen detection tests are available to confirm a diagnosis of influenza A or B infection. These tests allow health care providers to monitor influenza type and its prevalence in their community, to diagnose influenza more accurately, and to consider treatment options more carefully.³

Four antiviral drugs are available for the treatment of influenza.^{3,4} The first-generation antiviral drugs amantadine and rimantadine are similarly effective against influenza A but not influenza B. These agents inhibit the uncoating of viral RNA in the host cells and prevent its replication. Both drugs are effective in prevention of influenza A in high-risk groups and in the treatment of persons who acquire the disease. The secondgeneration antiviral drugs zanamivir and oseltamivir are inhibitors of neuraminidase, a viral glycoprotein that is necessary for viral replication and release. These drugs, which have been approved for treatment of acute uncomplicated influenza infection, are effective against both influenza A and B viruses. To be effective, the antiviral drugs should be initiated within 30 hours after onset of symptoms.

Vaccines are available to protect against influenza infections. The formulation of the vaccine must be changed yearly in response to changes in the influenza virus. Immunization is recommended for high-risk groups who, because of their age or underlying health problems, are unable to cope well with the infection and often require medical attention, including hospitalization. Immunization is also recommended for persons who can transmit the infection to high-risk groups (*e.g.*, health care worker and care givers).

Pneumonias

The term *pneumonia* describes inflammation of parenchymal structures of the lung, such as the alveoli and the bronchioles.⁵ Pneumonia is the sixth leading cause of death in the United States and the most common cause of death from infectious disease, particularly among the elderly and persons with debilitating diseases. Etiologic agents include infectious and non-infectious agents. Although much less common than infectious pneumonia, inhalation of irritating fumes or aspiration of gastric contents can result in severe pneumonia.

Pneumonias can be classified as typical (*i.e.*, bacterial) or atypical (*i.e.*, viral or mycoplasmal) pneumonias (Fig. 20-1). Typical pneumonia results from infection by bacteria that multiply extracellularly in the alveoli and cause inflammation and exudation of fluid into the air-filled spaces of the alveoli. Acute bacterial pneumonias are also classified according to two anatomic and radiologic patterns: *lobar pneumonia* (affecting part or all of a lung lobe) and *bronchopneumonia* (a patchy distribution involving more than one lobe). Atypical pneumonias produce patchy inflammatory changes that are confined to the alveolar septum and the interstitium of the lung. They usually produce less striking symptoms and physical findings than bacterial pneumonia; there is a lack of alveolar infiltration and purulent sputum, leukocytosis, and lobar consolidation on the radiograph.

Because of the overlap in symptomatology and changing spectrum of infectious organisms involved, pneumonias are increasingly being classified as community-acquired and hospitalacquired pneumonias. Persons with compromised immune function constitute a special concern in both categories.

Community-acquired pneumonia results from organisms found in the community, rather than in the hospital or nursing home.^{6,7} Common causes of community-acquired pneumonia include *Streptococcus pneumoniae* (single most common cause), *Haemophilus influenzae*, *Staphylococcus aureus*, *Klebsiella pneumo*-

Dr Murtadha Alshareifi e-Library

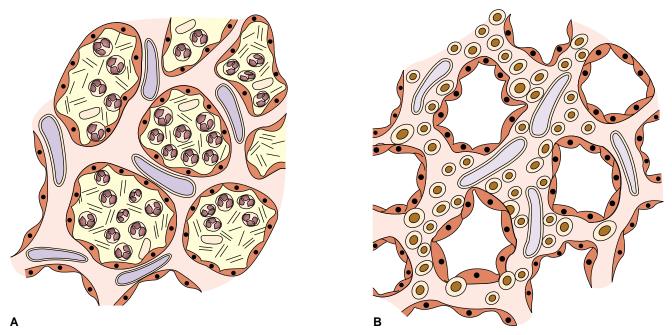


FIGURE 20-1 Location of inflammatory processes in (A) typical and (B) atypical forms of pneumonia.

niae and other gram-negative bacilli, *Legionella pneumophila*, and the influenza and respiratory syncytial viruses.

Hospital-acquired, or nosocomial, pneumonia is defined as a lower respiratory tract infection that was not present or incubating on admission to the hospital. Usually, infections occurring 48 hours or more after admission are considered hospital acquired.^{6,8} Most hospital-acquired pneumonias are bacterial. The organisms are those present in the hospital environment and include *Pseudomonas aeruginosa*, *S. aureus, Enterobacter* species, *Klebsiella* species, *Escherichia coli*, and *Serratia*. The organisms that are responsible for hospital-acquired pneumonias are different from those responsible for community-acquired pneumonia, and many of them have acquired antibiotic resistance and are more difficult to treat. Persons requiring mechanical ventilation are particularly at risk, as are those with compromised immune function, chronic lung disease, or airway instrumentation, such as endotracheal intubation or tracheotomy.

Although almost all types of microorganisms can cause pulmonary infection in immunocompromised persons, certain types of immunologic defects tend to favor certain types of infection. Defects in humoral immunity predispose to bacterial infections against which antibodies play an important role; defects in cellular immunity predispose to infections with viruses, fungi, mycobacteria, and protozoa. Neutropenia and impaired granulocyte function, as occurs in persons with leukemia, chemotherapy, and bone marrow metaplasia, predispose to infections caused by *S. aureus, Aspergillus,* gram-negative bacilli, and *Candida*. Pneumocystis pneumonia is an opportunistic, often fatal form of lung infection that occurs in persons with impaired cell-mediated immunity, particularly those with AIDS (see Chapter 10).

Acute Bacterial (Typical) Pneumonias

Most of the bacteria that cause pneumonia are normal inhabitants of the nasopharynx or oropharynx and are aspirated into the lung.^{5,9} Most persons unknowingly aspirate small amounts of organisms that have colonized their upper airways, particularly during sleep. Normally, these organisms do not cause infection because of the small numbers that are aspirated and because of the respiratory tract's defense mechanisms (Table 20-1). Loss of the cough reflex, damage to the ciliated endothelium that lines the respiratory tract, or impaired immune defenses predispose to colonization and infection of the lower respiratory system. Immune defenses include the bronchial-associated lymphoid tissue, phagocytic cells (*i.e.*, polymorphonuclear cells and macrophages), immunoglobulins (*i.e.*, IgA and IgG), and T-cell–mediated cellular immunity

KEY CONCEPTS

PNEUMONIAS

- Pneumonias are respiratory disorders involving inflammation of the lung structures, such as the alveoli and bronchioles.
- Pneumonia can be caused by infectious agents, such as bacteria and viruses, and noninfectious agents, such as gastric secretions that are aspirated into the lungs.
- The development of pneumonia is facilitated by an exceedingly virulent organism, large inoculum, and impaired host defenses.
- Pneumonias caused by infectious agents commonly are classified according to the source of infection (community- vs. hospital-acquired) and according to the immune status of the host (pneumonia in the immunocompromised person).

Their Effectiveness						
Defense Mechanism	Function	Factors That Impair Effectiveness				
Nasopharyngeal defenses	Remove particles from the air; contact with surface lysosomes and immunoglobulins (IgA) protects against infection	IgA deficiency state, hay fever, common cold, trauma to the nose, others				
Glottic and cough reflexes	Protect against aspiration into tracheo- bronchial tree	Loss of cough reflex due to stroke or neural lesion, neuromuscular disease, abdominal or chest sur- gery, depression of the cough reflex due to seda- tion or anesthesia, presence of a nasogastric tube (tends to cause adaptation of afferent receptors)				
Mucociliary blanket	Removes secretions, microorganisms, and particles from the respiratory tract	Smoking, viral diseases, chilling, inhalation of irritat- ing gases				
Pulmonary macrophages	Remove microorganisms and foreign parti- cles from the lung	Chilling, alcohol intoxication, smoking, anoxia				

TABLE 20-1

espiratory Defense Mechanisms and Conditions That Impair

(see Chapter 8). Bacterial adherence also plays a role in colonization of the lower airways. The epithelial cells of critically and chronically ill persons are more receptive to binding microorganisms that cause pneumonia. Other clinical risk factors favoring colonization of the tracheobronchial tree include antibiotic therapy that alters the normal bacterial flora, diabetes, smoking, chronic bronchitis, and viral infection.

Pneumococcal Pneumonia. Streptococcus pneumoniae (pneumococcus) remains the most common cause of bacterial pneumonia. *S. pneumoniae* colonizes the upper respiratory system and, in addition to pneumonia and lower respiratory tract infections, the organism is an important cause of upper respiratory infections, including sinusitis and otitis media, and of disseminated invasive infections such as bacteremia and meningitis.¹⁰ It is among the leading causes of illness and death of young children, persons with other health problems, and the elderly worldwide.

S. pneumoniae are gram-positive diplococci, possessing a capsule of polysaccharide. There are 90 serologically distinct types of *S. pneumoniae* based on the antigenic properties of their capsular polysaccharides. The virulence of *S. pneumoniae* is a function of its capsule, which prevents or delays digestion by phagocytes. The polysaccharide is an antigen that primarily elicits a B-cell response with antibody production. In the absence of antibody, clearance of the pneumococci from the body relies on the reticuloendothelial system, with the macrophages in the spleen playing a major role in elimination of the organism.¹¹ This, along with the spleen's role in antibody production, increases the risk of pneumococcal bacteremia in persons who are anatomically or functionally asplenic, such as children with sickle cell disease.

The initial step in the pathogenesis of pneumococcal infection is the attachment and colonization of the organism to the mucus and cells of nasopharynx. Colonization does not equate with signs of infection. Perfectly healthy people can be colonized and carry the organism without evidence of infection. The spread of particular strains of pneumococci, particularly antibiotic-resistant strains, is largely by healthy colonized individuals. The factors that permit the pneumococci to spread beyond the nasopharynx vary depending on the virulence of organism, impaired host defense mechanisms, and the existence of preceding viral infection.

The signs and symptoms of pneumococcal pneumonia vary widely, depending on the age and health status of the infected person. In previously healthy persons, the onset usually is sudden and is characterized by malaise; a severe, shaking chill; and fever. The temperature may go as high as 106°F. During the initial or congestive stage, the sputum is watery and breath sounds are limited, with fine crackles. As the disease progresses, the character of the sputum changes; it may be blood tinged or rust colored to purulent. Pleuritic pain, a sharp pain that is more severe with respiratory movements, is common. Elderly persons are less likely to experience marked elevations in temperature; in these persons, the only sign of pneumonia may be a loss of appetite and deterioration in mental status.

Treatment includes the use of antibiotics that are effective against *S. pneumoniae*. In the past, *S. pneumoniae* was uniformly susceptible to penicillin. However, penicillin-resistant and multidrug-resistant strains have been emerging in the United States and other countries.¹²

Pneumococcal pneumonia can be prevented through immunization. A 23-valent pneumococcal vaccine, composed of antigens from 23 types of *S. pneumoniae* capsular polysaccharides, is used. The vaccine is recommended for persons 65 years of age or older and persons aged 2 to 65 years with chronic illnesses, immunocompromised persons 2 years of age or older, and for residents in special environments or social settings in which the risk for invasive pneumococcal disease is increased, and for residents of nursing homes and long-term care facilities. Because their immune system is immature, the antibody response to most pneumococcal capsular polysaccharides usually is poor or inconsistent in children younger than 2 years.¹⁰ A 7-valent pneumococcal polysaccharide-protein conjugate vaccine (Prevnar) is now available for use among infants and children.¹³

Legionnaire's Disease. Legionnaire's disease is a bacterial pneumonia caused by a gram-negative rod, *Legionella pneumophila*. It ranks among the three or four most common causes of community-acquired pneumonia.⁶ Although more than 14 serotypes of *L. pneumophila* have been identified, serotype 1

accounts for more than 80% of reported cases of legionellosis.¹⁴ The organism frequently is found in water, particularly in warm, standing water. The disease was first recognized and received its name after an epidemic of severe and, for some, fatal pneumonia that developed among delegates to the 1976 American Legion convention held in a Philadelphia hotel. The spread of infection was traced to a water-cooled air-conditioning system. Although healthy persons can contract the infection, the risk is greatest among smokers, persons with chronic diseases, and those with impaired cell-mediated immunity.^{6,15}

Symptoms of the disease typically begin approximately 2 to 10 days after infection, with malaise, weakness, lethargy, fever, and dry cough. Other manifestations include disturbances of central nervous system function, gastrointestinal tract involvement, arthralgias, and elevation in body temperature, sometimes to more than 104°F. The presence of pneumonia along with diarrhea, hyponatremia, and confusion is characteristic of Legionella pneumonia. The disease causes consolidation of lung tissues and impairs gas exchange.

Diagnosis is based on clinical manifestations, radiologic studies, and specialized laboratory tests to detect the presence of the organism. Of these, the Legionella urinary antigen test is a relatively inexpensive, rapid test that detects antigens of *L. pneumophila* in the urine.^{14,15} The urine test usually is easier to obtain because people with legionellosis often have a non-productive cough and the results remain positive for weeks despite antibiotic therapy.

Treatment consists of administration of antibiotics that are known to be effective against *L. pneumophila*. Delay in instituting antibiotic therapy significantly increases mortality rates, so antibiotics known to be effective against *L. pneumophila* usually are included in the treatment regimen for severe community-acquired pneumonia.¹⁴

Primary Atypical Pneumonias

The primary atypical pneumonias are characterized by patchy involvement of the lung. They are usually preceded by pharyngitis and systemic flulike symptoms that evolve into laryngitis and finally tracheobronchitis and pneumonia. The most common pathogens are *Mycoplasma pneumoniae*, viruses, and *Chlamydia pneumoniae*.

Mycoplasma and Viral Pneumonias. The mycoplasmas are the smallest free-living agents of disease, having characteristics of viruses and bacteria. The influenza virus is the most common cause of viral pneumonia. Less common offenders are para-influenza and respiratory syncytial viruses. Other viruses sometimes are implicated, including the measles and chickenpox viruses.

The clinical course among persons with mycoplasmal and viral pneumonias varies widely from a mild infection (*e.g.*, influenza types A and B, adenovirus) that masquerades as a chest cold to a more serious and even fatal outcome (*e.g.*, chickenpox pneumonia). The symptoms may remain confined to fever, headache, and muscle aches and pains. Cough, when present, is characteristically dry, hacking, and nonproductive. Viruses impair the respiratory tract defenses and predispose to bronchopneumonia. Some viruses such as herpes simplex, varicella, and adenovirus may be associated with necrosis of the alveolar epithelium and acute inflammation.

Tuberculosis

Tuberculosis is the world's foremost cause of death from a single infectious agent, causing 25% of avoidable deaths in developing countries. With the introduction of antibiotics in the 1950s, the United States and other Western countries enjoyed a long decline in the number of infections until the mid-1980s. Since that time, the rate of infection has increased, particularly among people with the human immunodeficiency virus (HIV). In the United States, the biggest increase in new cases was from 1985 to 1993, after which the rate of cases reported yearly has again declined.¹⁶ Tuberculosis is more common among foreign-born persons from countries with a high incidence of tuberculosis and among residents of high-risk congregate settings, such as correctional facilities, drug treatment facilities, and homeless shelters. Outbreaks of a drug-resistant form of tuberculosis are being reported, complicating the selection of drugs and affecting the duration of treatment.

Tuberculosis is an infectious disease caused by the mycobacterium *M. tuberculosis*. The mycobacteria are slender, rodshaped, aerobic bacteria that do not form spores. They are similar to other bacterial organisms except for an outer waxy capsule that makes them more resistant to destruction; the organism can persist in old necrotic and calcified lesions and remain capable of reinitiating growth. The waxy coat also causes the organism to retain red dye when treated with acid in acidfast staining.¹⁷ Thus, the mycobacteria are often referred to as *acid-fast bacilli*. Although *M. tuberculosis* can infect practically any organ of the body, the lungs are most frequently involved. The tubercle bacilli are strict aerobes that thrive in an oxygenrich environment. This explains their tendency to cause disease in the upper lobe or upper parts of the lower lobe of the lung, where ventilation is greatest.

Tuberculosis is an airborne infection spread by minute, invisible particles, called *droplet nuclei*, that are harbored in the respiratory secretions of persons with active tuberculosis.^{17,18} Coughing, sneezing, and talking all create respiratory droplets; these droplets evaporate, leaving the organisms (droplet nuclei), which remain suspended in the air and are circulated by air currents. Living in crowded and confined conditions increases the risk for spread of the disease.

The tubercle bacillus incites a distinctive chronic inflammatory response referred to as *granulomatous inflammation*. The destructiveness of the disease results from the hypersensitivity response that the bacillus evokes, rather than its inherent destructive capabilities. Cell-mediated immunity and hypersensitivity reactions contribute to the evolution of the disease. Tuberculosis can manifest as a primary or reactivated infection.

Primary Tuberculosis

Primary tuberculosis occurs in a person lacking previous contact with the tubercle bacillus. It typically is initiated as a result of inhaling droplet nuclei that contain the tubercle bacillus (Fig. 20-2).¹⁸ Inhaled droplet nuclei pass down the bronchial tree without settling on the epithelium and implant in a respiratory bronchiole or alveolus beyond the mucociliary system. Soon after entering the lung, the bacilli are surrounded and engulfed by macrophages. *M. tuberculosis* has no known endotoxins or exotoxins; therefore, there is no early immunoglobulin response to infection.

KEY CONCEPTS

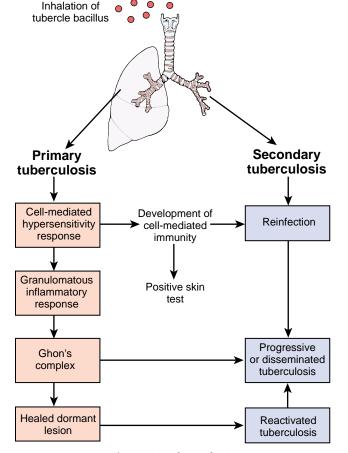
TUBERCULOSIS

- Tuberculosis is an infectious disease caused by Mycobacterium tuberculosis, a rod-shaped aerobic bacteria that is resistant to destruction and can persist in necrotic and calcified lesions for prolonged periods and remain capable of reinstating growth.
- The organism is spread by inhaling the mycobacterium-containing droplet nuclei that circulate in the air.
- The cell-mediated response plays a dominant role in walling off the tubercle bacilli and preventing the development of active tuberculosis. People with impaired cell-mediated immunity are more likely to experience active tuberculosis when infected.
- A positive tuberculin skin test results from a cellmediated immune response and implies that a person has been infected with *M. tuberculosis* and has mounted a cell-mediated immune response. It does not mean that the person has active tuberculosis.

The tubercle bacillus grows slowly, dividing every 25 to 32 hours in the macrophage.¹⁸ As the bacilli multiply, the macrophages degrade some mycobacteria and present antigen to the T lymphocytes for development of a cell-mediated immune response. The organisms grow for 2 to 12 weeks until they reach sufficient numbers to elicit a cellular immune response. In persons with intact cell-mediated immunity, this action is followed by the development of a single, gray-white, circumscribed granulomatous lesion, called a Ghon's focus, that contains the tubercle bacilli, modified macrophages, and other immune cells. Within 2 to 3 weeks, the central portion of the Ghon's focus undergoes soft, caseous (cheeselike) necrosis. This occurs at approximately the time that the tuberculin test result becomes positive, suggesting that the necrosis is caused by the cell-mediated hypersensitivity immune response (see Chapter 10). During this same period, tubercle bacilli, free or inside macrophages, drain along the lymph channels to the tracheobronchial lymph nodes of the affected lung and there evoke the formation of caseous granulomas. The combination of the primary lung lesion and lymph node granulomas is called Ghon's complex (Fig. 20-3).

The cell-mediated hypersensitivity response plays a dominant role in limiting further replication of the bacilli. The immune response also provides protection against additional tubercle bacilli that may be inhaled at a later time. People with HIV infection and others with disorders of cell-mediated immunity are more likely to acquire active tuberculosis if they become infected.

When the number of organisms inhaled is small and the body's resistance is adequate, scar tissue forms and encapsu-



■ FIGURE 20-2 ■ Pathogenesis of TB Infection



■ FIGURE 20-3 ■ Primary tuberculosis. A healed Ghon's complex is represented by a subpleural nodule and involved hilar lymph nodes. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 606]. Philadelphia: Lippincott Williams & Wilkins)

lates the primary lesion. In time, most of these lesions become calcified and are visible on a chest radiograph.

Primary tuberculosis usually is asymptomatic, with the only evidence of the disease being a positive tuberculin skin test result and calcified lesions seen on the chest radiograph. Occasionally, primary tuberculosis may progress, causing more extensive destruction of lung tissue and spreading through the airways and lymphatics to multiple sites within the lung. As the disease spreads, the organism gains access to the sputum, allowing the person to infect others.

In rare instances, tuberculosis may erode into a blood vessel, giving rise to hematogenic dissemination. Miliary tuberculosis describes minute lesions, resembling millet seeds, which can involve almost any organ, resulting from this type of dissemination.

Secondary Tuberculosis

Secondary tuberculosis represents either reinfection from inhaled droplet nuclei or reactivation of a previously healed primary lesion (Fig. 20-2). It often occurs in situations of impaired body defense mechanisms. The partial immunity that follows primary tuberculosis affords protection against reinfection and to some extent aids in localizing the disease should reactivation occur. In secondary tuberculosis, the hypersensitivity reaction can be an aggravating factor, as evidenced by the frequency of cavitation and bronchial dissemination. The cavities may coalesce to a size as large as 10 to 15 cm in diameter (Fig. 20-4).6 Pleural effusion and tuberculous empyema are common as the disease progresses.

Persons with secondary tuberculosis commonly present with low-grade fevers, night sweats, easy fatigability, anorexia, and weight loss. A cough initially is dry but later becomes productive with purulent and sometimes blood-tinged sputum. Dyspnea and orthopnea develop as the disease advances.

Diagnosis and Treatment

The most frequently used screening methods for pulmonary tuberculosis are the tuberculin skin tests and chest radiographic studies. The tuberculin skin test measures delayed hypersensitivity (*i.e.*, cell-mediated, type IV) that follows exposure



FIGURE 20-4 Cavitary tuberculosis in the apex of the left upper lobe of the lung. (Rubin E., Farber J.L. [1999]. Pathology [3rd ed., p. 607]. Philadelphia: Lippincott Williams & Wilkins)

to the tubercle bacillus. Persons who become tuberculin positive usually remain so for the remainder of their lives. A positive reaction to the skin test does not mean that a person has active tuberculosis, only that there has been exposure to the bacillus and that cell-mediated immunity to the organism has developed. False-positive and false-negative skin test reactions can occur. False-positive reactions often result from crossreactions with nontuberculous mycobacteria, such as M. aviumintracellulare complex. Because the hypersensitivity response to the tuberculin test depends on cell-mediated immunity, a falsenegative test result can occur because of immunodeficiency states that result from HIV infection, immunosuppressive therapy, lymphoreticular malignancies, or aging. This is called anergy. In the immunocompromised person, a negative tuberculin test result can mean that the person has a true lack of exposure to tuberculosis or is unable to mount an immune response to the test. Because of the problem with anergy in persons with HIV infection and other immunocompromised states, the use of control tests is recommended. Three antigens that can be used for control testing are *Candida*, mumps virus, and tetanus toxoid. Most healthy persons in the population have been exposed to these antigens and will display a positive response to these control tests.

Diagnosis of active pulmonary tuberculosis requires identification of the organism in respiratory tract secretions. Bacteriologic studies (*i.e.*, acid-fast stain and cultures) of early-sputum specimens, gastric aspirations, or bronchial washings obtained during fiber-optic bronchoscopy may be used. Genotyping can be done to identify different strains of M. tuberculosis. It is useful in investigating outbreaks of tuberculosis, tracing the sources of infection, and determining whether new episodes of the disease are caused by reinfection or reactivation. In addition, genotyping is useful in determining sites and patterns of M. tuberculosis transmission in communities.

The primary drugs used in the treatment of tuberculosis are isoniazid (INH), rifampin, pyrazinamide, ethambutol, and streptomycin. Two groups meet the criteria established for the use of antimycobacterial therapy for tuberculosis: (1) persons with an active form of the disease and (2) those who have had contact with cases of active tuberculosis and who are at risk for development of active tuberculosis.

The tubercle bacillus is an aerobic organism that multiplies slowly and remains relatively dormant in oxygen-poor caseous material. It undergoes a high rate of mutation and tends to acquire resistance to any one drug. For this reason, multidrug regimens are used for treating persons with active tuberculosis. Tuberculosis is an unusual disease in that chemotherapy is required for a relatively long period of time. Short-course programs of therapy (usually for 6 to 12 months) have replaced the earlier 18- to 24-month multidrug regimens. Treatment may need to be prolonged in persons with HIV infection and in those with drug-resistant strains of M. tuberculosis. Drug susceptibility tests are used to guide treatment in drug-resistant forms of the disease.

Prophylactic treatment is used for persons who are infected with M. tuberculosis but do not have active disease (e.g., persons with a positive skin test or those who have had close contact with an active case of tuberculosis).¹⁹

Success of chemotherapy for prophylaxis and treatment of tuberculosis depends on strict adherence to a lengthy drug regimen. This often is a problem, particularly for asymptomatic

persons with tuberculosis infections and for poorly motivated groups, such as intravenous drug abusers. Directly observed therapy, which requires that a health care worker observe while the person takes the antituberculosis drug, is recommended for some persons and for certain types of treatment protocols.

Fungal Infections

Fungal infections are commonly classified as superificial, subcutaneous, deep-seated, and opportunistic pathogenic fungi. The superficial and subcutaneous fungi almost always limit their infections to the skin and subcutaneous tissues. Opportunistic fungi are organisms of low virulence that cause localized or systemic infections in people who are immunocompromised, such as those with AIDS (see Chapter 10). Examples of opportunistic fungi include molds (*e.g., Aspergillus* species) as well as yeast-like species (*e.g., Candida* species).

The deep-seated fungal infections are caused by highly virulent dimorphic fungi, with the ability to invade deeply into tissues and cause systemic disease. They include *Histoplasma capsulatum* (histoplasmosis), *Coccidioides immitis* (coccidioidomycosis), and *Blastomycoses capsulatum* (blastomycosis). Isolated, self-limited pulmonary involvement is commonly seen in people with normal immune function, whereas immunocompromised people often present with disseminated disease. In HIV-infected persons in endemic areas, coccidioidomycosis is now a common opportunistic infection.

Each of the dimorphic fungi has a typical geographic distribution. *H. capsulatum* is endemic along the major river valleys of the Midwest—the Ohio, the Mississippi, and the Missouri.²⁰ The organism grows in soil and other areas that have been enriched with bird excreta: old chicken houses, pigeon lofts, barns, and trees where birds roost. The infection is acquired by inhaling the fungal spores that are released when the dirt or dust from the infected areas is disturbed. C. immitis is most prevalent in the southwestern United States, principally in California, Arizona, and Texas.²¹ Because of its prevalence in the San Joaquin Valley, the disease is sometimes referred to as San Joaquin fever or valley fever. The C. immitis organism lives in soil and can establish new sites in the soil. Events such as dust storms and digging for construction have been associated with increased incidence of the disease. B. capsulatum is most commonly found in the southern and north central United States, especially in areas bordering the Mississippi and Ohio River basins and the Great Lakes.²²

The signs and symptoms of the fungal infections commonly resemble those of tuberculosis. Depending on the host's resistance and immunocompetence, the diseases usually take one of three forms: (1) an acute primary disease, (2) a chronic (cavitary) pulmonary disease, or (3) a disseminated infection. The primary pulmonary lesions consist of nodules containing aggregates of macrophages with engulfed microorganisms. Similar nodules develop in the regional lymph nodes. There is a striking similarity to the primary lesions of tuberculosis. The clinical manifestations consist of a mild, self-limited flulike syndrome.

In the vulnerable host, chronic cavitary lesions develop, with a predilection for the upper lobe, resembling the secondary form of tuberculosis. The most common manifestations are productive cough, fever, night sweats, and weight loss. Disseminated disease most often develops as an acute and fulminating infection in the very old or the very young or in persons with compromised immune function. Although the macrophages of the reticuloendothelial system can remove the fungi from the bloodstream, they are unable to destroy them. Characteristically, this form of the disease presents with a high fever, generalized lymph node enlargement, hepatosplenomegaly, muscle wasting, anemia, leukopenia, and thrombocytopenia. There may be hoarseness, ulcerations of the mouth and tongue, nausea, vomiting, diarrhea, and abdominal pain. Often, meningitis becomes a dominant feature of the disease. Persons with blastomycosis may experience cutaneous infections that induce pseudoepitheliomatous hyperplasia, which may be mistaken for squamous cell carcinoma.

Skin tests similar to the tuberculin test can be used to detect exposure to *Histoplasma* and *Coccidioides*. There is no reliable skin test for *Blastomyces*. The diagnosis of acute infection is usually made by direct visualization of the organism in tissue sections or sputum culture. Serologic tests, detecting antibodies against the specific fungi are available, but lack sensitivity and specificity.

Treatment depends on the severity of infection. Persons without associated risk factors such as HIV infection or without specific evidence of progressive disease usually can be treated without antifungal therapy. The oral or intravenous antifungal drugs are used in the treatment of persons with progressive disease.

In summary, respiratory infections are the most common cause of respiratory illness. The influenza virus causes three syndromes: an uncomplicated rhinotracheitis, a respiratory viral infection followed by a bacterial infection, and viral pneumonia.

Pneumonia describes an infection of the parenchymal tissues of the lung. Loss of the cough reflex, damage to the ciliated endothelium that lines the respiratory tract, or impaired immune defenses predispose to pneumonia. Pneumonia is being increasingly classified as community acquired or hospital acquired. Community-acquired pneumonia involves infections from organisms that are present more often in the community than in the hospital or nursing home. Hospital-acquired, or nosocomial, pneumonia is defined as a lower respiratory tract infection occurring 72 hours or more after hospital admission. Persons with compromised immune function constitute a special concern in both categories. Typical or bacterial pneumonias result from infection by bacteria such as Streptococcus pneumoniae that multiply extracellularly in the alveoli and cause inflammation and exudation of fluid into the air-filled spaces of the alveoli. Atypical pneumonias, such as those caused by Mycoplasma pneumoniae and respiratory viruses, involve the interstitium of the lung and often masquerade as chest colds.

Tuberculosis is a chronic respiratory infection caused by *M. tuberculosis*, which is spread by minute, invisible particles called *droplet nuclei*. Tuberculosis is a particular threat among persons with HIV infection, foreign-born persons from countries with a high incidence of tuberculosis, and residents of high-risk congregate settings, such as correctional facilities, drug treatment facilities, and homeless shelters. The tubercle bacillus incites a distinctive chronic inflammatory response re-

ferred to as *granulomatous inflammation*. The destructiveness of the disease results from the hypersensitivity response that the bacillus evokes, rather than its inherent destructive capabilities. Cell-mediated immunity and hypersensitivity reactions contribute to the evolution of the disease. The treatment of tuberculosis has been complicated by outbreaks of drugresistant forms of the disease.

Infections caused by the fungi *H. capsulatum* (histoplasmosis), *C. immitis* (coccidioidomycosis), and *B. dermatitidis* (blastomycosis) produce pulmonary manifestations that resemble tuberculosis. These infections are common but seldom serious unless they produce progressive destruction of lung tissue or the infection disseminates to organs and tissues outside the lungs.

CANCER OF THE LUNG

Lung cancer is the leading cause of cancer deaths among men and women in the United States, accounting for 25% of all cancer deaths. The increases in lung cancer incidence and deaths during the past 50 years have coincided closely with the increase in cigarette smoking during the same period. Industrial hazards also contribute to the incidence of lung cancer. A commonly recognized hazard is exposure to asbestos, with the mean risk of lung cancer being significantly greater in asbestos workers than in the general population. Tobacco smoke contributes heavily to the development of lung cancer in persons exposed to asbestos; the risk in this population group is estimated to be 50 to 90 times greater than that for nonsmokers.9 Because cancer of the lung usually is far advanced before it is discovered, the prognosis in general is poor. The overall 5-year survival rate is 13% to 15%, a dismal statistic that has not changed since the late 1960s.23

Bronchogenic Carcinoma

Bronchogenic carcinoma, which has its origin in the bronchial or bronchiolar epithelium, constitutes 90% to 95% of all lung cancers. Bronchogenic carcinomas are aggressive, locally invasive, and widely metastatic tumors that arise from the epithelial lining of the major bronchi. These tumors begin as small mucosal lesions that may follow one of several patterns of growth. They may form masses within the lumen of the bronchi that invade the mucosal layer and surrounding connective tissue layer, or they may form large, bulky masses that extend into the adjacent lung tissue. Some large tumors undergo central necrosis and acquire local areas of hemorrhage, and some invade the pleural cavity and chest wall and spread to adjacent intrathoracic structures.⁹

Bronchogenic carcinomas can be subdivided into four major categories: squamous cell lung carcinoma (25% to 40%), adenocarcinoma (20% to 40%), small cell carcinoma (20% to 25%), and large cell carcinoma (10% to 15%).⁹ *Squamous cell carcinoma* is found most commonly in men and is closely correlated with a smoking history. Squamous cell carcinoma tends to originate in the central bronchi as an intraluminal growth and thus is more amenable to early detection through cytologic examination of the sputum than other forms of lung cancer



■ FIGURE 20-5 ■ Squamous cell carcinoma of the lung. The tumor grows within the lumen of a bronchus and invades the adjacent intrapulmonary lymph node. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 656]. Philadelphia: Lippincott Williams & Wilkins)

(Fig. 20-5). It tends to spread centrally into major bronchi and hilar lymph nodes and disseminates outside the thorax later than other types of bronchogenic cancers.

Adenocarcinoma is the most common type of lung cancer in women and nonsmokers. Its association with cigarette smoking is weaker than for squamous cell carcinoma. Adenocarcinomas can have their origin in either the bronchiolar or alveolar tissues of the lung. These tumors tend to be located more peripherally than squamous cell sarcomas and sometimes are associated with areas of scarring (Fig. 20-6). The scars may be attributable to old infarcts, metallic foreign bodies, wounds, or



■ FIGURE 20-6 ■ Adenocarcinoma of the lung. A peripheral tumor in the upper right lobe of the lung. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 657]. Philadelphia: Lippincott Williams & Wilkins)

granulomatous infections such as tuberculosis. In general, these tumors grow more slowly than squamous cell carcinomas.

The *small cell carcinomas* are more common in men than women and are strongly associated with smoking. They are characterized by a distinctive cell type—small, dark, round-tooval shape, lymphocyte-like cells that have a scant cytoplasm and highly colored nuclei. Because of their cellular appearance, these cancers are sometimes referred to as "oat cell" tumors.⁹ The small cell carcinomas are highly malignant, tend to infiltrate widely, disseminate early in their course, and rarely are resectable. These tumors are sensitive to chemotherapy and irradiation, and newer protocols have improved the outlook somewhat.

Large cell carcinomas have large polygonal cells. They constitute a group of neoplasms that are highly anaplastic and difficult to categorize as squamous or adenocarcinoma. They are associated with a poor prognosis because of their tendency to spread to distant sites early in their course.

In general, adenocarcinoma and squamous cell carcinoma tend to remain localized longer and have a better prognosis than do other, less differentiated cancers, which usually are far advanced at the time of diagnosis. All varieties of bronchogenic carcinomas, especially small cell lung carcinoma, have the capacity to synthesize bioactive products and produce paraneoplastic syndromes, including adrenocorticotropic hormone (ACTH), antidiuretic hormone (ADH), parathyroid-like hormone, gonadotropins, and gastrin-releasing peptide.

Manifestations

Cancer of the lung develops insidiously, often giving little or no warning of its presence. Because its symptoms are similar to those associated with smoking and chronic bronchitis, they often are disregarded.

The manifestations of lung cancer can be divided into three categories: those caused by involvement of the lung and adjacent structures, the effects of local spread and metastasis, and the nonmetastatic paraneoplastic manifestations involving endocrine, neurologic, and connective tissue function. As with other cancers, lung cancer causes nonspecific symptoms such as anorexia and weight loss.

Many of the manifestations of lung cancers result from local irritation and obstruction of the airways and from invasion of the mediastinum and pleural space. The earliest symptoms usually are chronic cough, shortness of breath, and wheezing because of airway irritation and obstruction. Hemoptysis (*i.e.*, blood in the sputum) occurs when the lesion erodes blood vessels. Pain receptors in the chest are limited to the parietal pleura, mediastinum, larger blood vessels, and peribronchial afferent vagal fibers. Dull, intermittent, poorly localized retrosternal pain is common in tumors that involve the mediastinum. Pain becomes persistent, localized, and more severe when the disease invades the pleura.

Tumors that invade the mediastinum may cause hoarseness because of the involvement of the recurrent laryngeal nerve and cause difficulty in swallowing because of compression of the esophagus. An uncommon complication called the *superior vena cava syndrome* can occur in some persons with mediastinal involvement. Interruption of blood flow in this vessel usually results from compression by the tumor or involved lymph nodes. The disorder can interfere with venous drainage from the head, neck, and chest wall. The outcome is determined by the speed with which the disorder develops and the adequacy of the collateral circulation.

Tumors adjacent to the visceral pleura often insidiously produce pleural effusion. This effusion can compress the lung and cause atelectasis and dyspnea. It is less likely to cause fever, pleural friction rub, or pain than pleural effusion resulting from other causes.

Metastatic spread occurs by way of lymph channels and the vascular system. Metastases already exist in 50% of patients presenting with evidence of lung cancer and develop eventually in 90% of patients. The most common sites of these metastases are the brain, bone, and liver.

Paraneoplastic disorders are those that are unrelated to metastasis. These include hypercalcemia from secretion of parathyroid-like peptide, Cushing's syndrome from ACTH secretion, inappropriate secretion of ADH, neuromuscular syndromes (e.g., myasthenic syndromes, peripheral neuropathy, polymyositis), and hematologic disorders (e.g., migratory thrombophlebitis, nonbacterial endocarditis, disseminated intravascular coagulation). Neurologic or muscular symptoms can develop 6 months to 4 years before the lung tumor is detected. One of the more common of these problems is weakness and wasting of the proximal muscles of the pelvic and shoulder girdles, with decreased deep tendon reflexes but without sensory changes. Hypercalcemia is seen most often in persons with squamous cell carcinoma, hematologic syndromes in persons with adenocarcinomas, and the remaining syndromes in persons with small cell neoplasms.9 Manifestations of the paraneoplastic syndrome may precede the onset of other signs of lung cancer and may lead to discovery of an occult tumor.

Diagnosis and Treatment

The diagnosis of lung cancer is based on a careful history and physical examination and other tests such as chest radiography, bronchoscopy, cytologic studies (Papanicolaou's test) of the sputum or bronchial washings, percutaneous needle biopsy of lung tissue, and scalene lymph node biopsy. Computed tomography scans, magnetic resonance imaging studies, and ultrasonography are used to locate lesions and evaluate the extent of the disease. The carcinoembryonic antigen (CEA) is produced by undifferentiated lung tumor cells; high CEA titers usually correlate with extensive disease. This test often is used to follow the progress of the disease and its response to treatment.

Like other types of cancer, lung cancers are classified according to cell type (*i.e.*, squamous cell carcinoma, adenocarcinoma, and large cell carcinoma) and staged according to the TNM system. These classifications are used for treatment planning. Small cell carcinoma is not evaluated by the TNM system but staged as limited (confined to the one hemithorax and hilar, mediastinal, and supraclavicular nodes) and extensive (spread to more distant sites).^{6,9}

Treatment methods for lung cancer include surgery, radiation therapy, and chemotherapy.⁶ These treatments may be used singly or in combination. Surgery is used for the removal of small, localized tumors. It can involve a lobectomy, pneumonectomy, or segmental resection of the lung. Radiation therapy can be used as a definitive or main treatment modality, as part of a combined treatment plan, or for palliation of symptoms. Because of the frequency of metastases, chemotherapy often is used in treating lung cancer. Combination chemotherapy, which uses a regimen of several drugs, usually is used. Chemotherapy is the treatment of choice for small cell carcinoma. Advances in the use of combination chemotherapy have improved the outlook for persons with small cell carcinoma.

In summary, cancer of the lung is a leading cause of death among men and women between the ages of 50 and 75 years, and the death rate is increasing among women. In the United States, the increased death rate has coincided with an increase in cigarette smoking. Industrial hazards, such as exposure to asbestos, increase the risk for development of lung cancer. Of all forms of lung cancer, bronchogenic carcinoma is the most common, accounting for 90% to 95% of cases. Because lung cancer develops insidiously, it often is far advanced before it is diagnosed, a fact that is used to explain the poor 5-year survival rate.

The manifestations of lung cancer can be attributed to the involvement of the lung and adjacent structures, the effects of local spread and metastasis, and the nonmetastatic paraneoplastic manifestations involving endocrine, neurologic, and connective tissue function. As with other cancers, lung cancer causes nonspecific symptoms such as anorexia and weight loss. Treatment methods for lung cancer include surgery, irradiation, and chemotherapy.

RESPIRATORY DISORDERS IN INFANTS AND CHILDREN

Acute respiratory disease is the most common cause of illness in infancy and childhood, accounting for 50% of illness in children younger than 5 years.²⁴ This section focuses on respiratory disorders in the neonate and respiratory infections in children. Discussions of bronchial asthma in children and cystic fibrosis are included in Chapter 21.

Lung Development

Although other body systems are physiologically ready for extrauterine life as early as 25 weeks of gestation, the lungs require much longer. Immaturity of the respiratory system is a major cause of morbidity and mortality in infants born prematurely. Even at birth, the lungs are not fully mature, and additional growth and maturation continue well into childhood.

Lung development may be divided into five stages: the embryonic, glandular, canicular, saccular, and alveolar periods.²⁵ The first three phases are devoted to development of the conducting airways, and the last two phases are devoted to development of the gas exchange portion of the lung. By the 25th to 28th weeks of gestation, sufficient terminal air sacs are present to permit survival. It is also during this period that the type II alveolar cells, which produce surfactant, begin to function. Lung development is incomplete at birth; an infant is born with only one eighth to one sixth the adult number of alveoli. Alveoli continue to be formed during early childhood, reaching the adult number of 300 million alveoli by 5 to 6 years of age.25

Breathing in the Fetus and Neonate

Effective ventilation in infants, older children, and adults requires coordinated interaction between the muscles of the upper airways, including those of the pharynx and larynx, the diaphragm, and the intercostal muscles of the chest wall. In the infant, the diaphragm inserts more horizontally than in the adult. As a result, contraction of the diaphragm tends to draw the lower ribs inward, especially if the infant is placed in the horizontal position. The intercostal muscles, which normally lift the ribs during inspiration, are not fully developed in the infant. Instead, they function largely to stabilize the chest. Under circumstances such as crying, the intercostal muscles of the neonate function together with the diaphragm to splint the chest wall and prevent its collapse.

The chest wall of the neonate is highly compliant. A striking characteristic of neonatal breathing is the paradoxical inward movement of the upper chest during inspiration, especially during active sleep. Normally, the infant's lungs also are compliant, which is advantageous to the infant with its compliant chest cage because it takes only small changes in inspiratory pressure to inflate a compliant lung. However, with respiratory disorders that decrease lung compliance, the diaphragm must generate more negative pressure; as a result, the compliant chest wall structures are sucked inward. Retractions are abnormal inward movements of the chest wall during inspiration; they may occur intercostally (between the ribs), in the substernal or epigastric area, and in the supraclavicular spaces.

Airway Resistance. Normal lung inflation requires uninterrupted movement of air through the extrathoracic airways (i.e., nose, pharynx, larynx, and upper trachea) and intrathoracic airways (*i.e.*, bronchi and bronchioles). The neonate (0 to 4 weeks of age) breathes predominantly through the nose and does not adapt well to mouth breathing. Any obstruction of the nose or nasopharynx may increase upper airway resistance and increase the work of breathing.

The airways of the infant and small child are much smaller than those of the adult. Because the resistance to airflow is directly related to the fourth power of the radius, relatively small amounts of mucus secretion, edema, or airway constriction can produce marked changes in airway resistance and airflow. Nasal flaring is a method that infants use to take in more air. This method of breathing increases the size of the nares and decreases the resistance of the small airways.

The airways of infants and small children are also less rigid that those of older children and adults. Cartilaginous support of the extrathoracic (e.g., larynx, trachea) airways is poorly developed in infants and small children. These structures are soft and tend to collapse when the airway is obstructed and the child cries, causing the inspiratory pressures to become more negative.

Lung Volumes and Gas Exchange

In infants, the functional residual capacity (FRC), which is the air left in the lungs at the end of normal expiration, plays an important role in gas exchange (see Chapter 19). The FRC occurs at a higher lung volume in the infant than in the older child or adult.²⁶ This higher end-expiratory volume results from a more rapid respiratory rate, which leaves less time for expiration. The increased end-expiratory volume is important to the neonate because it holds the airways open throughout all phases of respiration to maintain a more uniform lung expansion and gas exchange. During sleep the tone of the upper airway muscles in the neonate is reduced, the intercostal activity that stabilizes the chest wall is less, and the time spent in expiration is increased. As a result, end-expiratory volume decreases and there is less optimal gas exchange during active sleep.²⁶

Control of Ventilation

Fetal blood oxygen (PO₂) levels normally range from 25 to 30 mm Hg, and carbon dioxide (PCO₂) levels range from 45 to 50 mm Hg, independent of any respiratory movements. Any decrease in oxygen levels induces quiet sleep in the fetus, with subsequent cessation of breathing movements, both of which lead to a decrease in oxygen consumption. Switching to oxygen derived from the aerated lung at birth causes an immediate increase in PO₂ to approximately 50 mm Hg; within a few hours, it increases to approximately 70 mm Hg.27 These levels, which greatly exceed fetal levels, cause the chemoreceptors (see Chapter 19) to become silent for several days. Although the infant's PO₂ may fluctuate during this critical time, the chemoreceptors do not respond appropriately. It is not until several days after birth that the chemoreceptors "reset" their PO₂ threshold; only then do they become the major controller of breathing. However, the response seems to be biphasic, with an initial hyperventilation followed by a decreased respiratory rate and even apnea. In neonates, particularly in preterm infants, breathing patterns and respiratory reflexes depend on the arousal state. Periodic breathing and apnea are characteristic of premature infants and reflect patterns of fetal breathing. The fact that they occur with sleep and disappear during wakefulness underscores the importance of arousal.

Alterations in Breathing Patterns. Most lung diseases in infants and small children produce a decrease in lung compliance with manifestations of restrictive lung disease or airway obstruction. Children with restrictive lung disease breathe at faster rates, and their respiratory excursions are shallow. Grunting is an audible noise that occurs as the child tries to raise the FRC and improve gas exchange by closing the glottis at the end of expiration. Airway obstruction produces turbulence. When it occurs in the extrathoracic airways (larynx and trachea) it produces an increase in inspiratory effort and a crowing sound called stridor. With intrathoracic (e.g., bronchi and bronchioles) obstruction, as in bronchiolitis and bronchial asthma, the intrapleural pressure becomes more positive during expiration because of air trapping; this causes collapse of intrathoracic airways. Expiration is prolonged, and the child uses the accessory muscles to aid in expiration. Often an audible wheezing or whistling sound is heard during expiration.

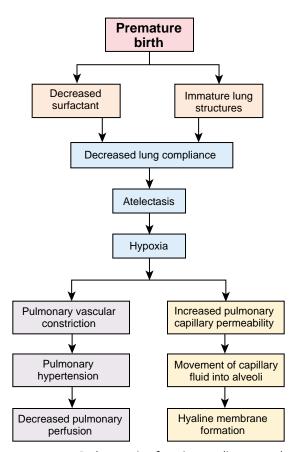
Respiratory Disorders in the Neonate

The neonatal period is one of transition from placental dependency to air breathing. This transition requires functioning of the surfactant system, conditioning of the respiratory muscles, and establishment of parallel pulmonary and systemic circulations. Respiratory disorders develop in infants who are born prematurely or who have other problems that impair this transition. Among the respiratory disorders of the neonate are the respiratory distress syndrome (RDS) and bronchopulmonary dysplasia (BPD).

Respiratory Distress Syndrome

Respiratory distress syndrome, also known as *hyaline membrane disease*, is one of the most common causes of respiratory disease in premature infants. In these infants, pulmonary immaturity, together with surfactant deficiency, leads to alveolar collapse (Fig. 20-7). The type II alveolar cells that produce surfactant do not begin to mature until approximately the 25th to 28th weeks of gestation, and consequently, many premature infants are born with poorly functioning type II alveolar cells and have difficulty producing sufficient amounts of surfactant.²⁵ The incidence of RDS is higher among preterm male infants, white infants, infants of mothers with diabetes, and those who experience asphyxia, cold stress, precipitous deliveries, and delivery by cesarean section (when performed before the 38th week of gestation).

Surfactant synthesis is influenced by several hormones, including insulin and cortisol. Insulin tends to inhibit surfactant production; this explains why infants of mothers with insulindependent diabetes are at increased risk for the development of RDS. Cortisol can accelerate maturation of type II cells and formation of surfactant. The reason premature infants born by cesarean section presumably are at greater risk for the development of RDS is because they are not subjected to the stress of vaginal delivery, which is thought to increase the infants'



■ **FIGURE 20-7** ■ Pathogenesis of respiratory distress syndrome (RSD) in the infant.

cortisol levels. These observations have led to administration of corticosteroid drugs before delivery to mothers with infants at high risk for the development of RDS.²⁷

Surfactant reduces the surface tension in the alveoli, thereby equalizing the retractive forces in the large and small alveoli and reducing the amount of pressure needed to inflate and hold the alveoli open (see Chapter 19). At birth, the first breath requires high inspiratory pressures to expand the lungs. With normal levels of surfactant, the lungs retain as much as 40% of the residual volume after the first breath, and subsequent breaths require far lower inspiratory pressures.9 With a surfactant deficiency, the lungs collapse between breaths, making the infant work as hard with each successive breath as with the first breath. The airless portions of the lungs become stiff and noncompliant. The pulmonary capillary membranes become more permeable, allowing fibrin-rich fluids to be pulled into the alveolar spaces and form a hyaline membrane. The hyaline membrane constitutes a barrier to gas exchange, leading to hypoxemia and carbon dioxide retention, a condition that further impairs surfactant production.

Infants who have RDS present with multiple signs of respiratory distress, usually within the first 24 hours of birth. Central cyanosis is a prominent sign. Breathing becomes more difficult, and retractions occur as the infant's soft chest wall is pulled in as the diaphragm descends. Grunting sounds occur during expiration. As the tidal volume drops because of airway obstruction and alveolar collapse, the respiration rate increases (usually to 60 to 120 breaths/minute) in an effort to maintain normal minute ventilation. Fatigue may develop rapidly because of the increased work of breathing. The stiff lung of infants with RDS also increases the resistance to blood flow in the pulmonary circulation, leading to the development of pulmonary hypertension and decreased pulmonary perfusion.

Infants with suspected RDS require continuous cardiorespiratory monitoring. Oxygen levels can be assessed through an arterial line (umbilical) or by a transcutaneous oxygen sensor. Treatment includes administration of supplemental oxygen, continuous positive airway pressure through nasal prongs, and often assisted mechanical ventilation. A neutral thermal environment and prevention of hypoglycemia are recommended.

Surfactant therapy is used to prevent and treat RDS. There are two types of surfactants available in the United States: surfactants prepared from animal sources and synthetic surfactants.²⁸ The surfactants are suspended in saline and administered into the airways, usually through an endotracheal tube. The treatment often is initiated soon after birth in infants who are at high risk for RDS.

Bronchopulmonary Dysplasia

Bronchopulmonary dysplasia (BPD) is a chronic lung disease that develops in premature infants who were treated with mechanical ventilation, mainly for RDS. The condition is considered to be present if the neonate is oxygen dependent at 36 weeks after gestation. The disorder is thought to be a response of the premature lung to early injury. High concentrations of inspired oxygen and injury from positive-pressure ventilation (*i.e.*, barotrauma) have been implicated. Newer therapies, such as administration of surfactants, high-frequency ventilation, and prenatal or postnatal administration of corticosteroids, may have altered the severity of BPD, but the condition remains a major health problem.^{29,30} BPD is characterized by chronic respiratory distress, persistent hypoxemia when breathing room air, reduced lung compliance, increased airway resistance, and severe expiratory flow limitation. There is a mismatching of ventilation and perfusion with development of hypoxemia and hypercapnia. An increase in pulmonary vascular resistance may lead to the development of pulmonary hypertension and cor pulmonale (*i.e.*, right heart failure associated with lung disease). The infant with BPD may have tachycardia, shallow breathing, chest retractions, cough, barrel chest, and poor weight gain. Clubbing of the fingers occurs in children with severe disease. In infants with right heart failure, tachycardia, tachypnea, hepatomegaly, and periorbital edema develop.

The treatment is mechanical ventilation and administration of adequate oxygenation. Weaning from ventilation is accomplished gradually, and some infants may require ventilation at home. Rapid lung growth occurs during the first year of life, and lung function usually improves. Adequate nutrition is essential for the recovery of infants with BPD.

Most adolescents and young adults who had severe BPD during infancy have some degree of pulmonary dysfunction, consisting of airway obstruction, airway hyperreactivity, or hyperinflation.

Respiratory Infections in Children

In children, respiratory tract infections are common, and although they are troublesome, they usually are not serious. Frequent infections occur because the immune system of infants and small children has not been exposed to many common pathogens; consequently, they tend to contract infections with each new exposure.

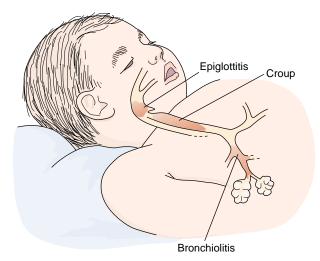
Upper Airway Infections

Acute inflammation of the upper airway is of particular importance in infants and small children because the airway is smaller, predisposing young children to a relatively greater narrowing than is produced by the same degree of inflammation in an older child. Two acute upper respiratory tract infections are relatively common during early childhood—croup and epiglottitis. Croup, the most common form of acute respiratory obstruction in children, is usually relatively benign and selflimited. Epiglottitis is a rapidly progressive and life-threatening condition. The site of involvement is illustrated in Figure 20-8.

Obstruction of the upper airways because of infection tends to exert its greatest effect during the inspiratory phase of respiration. Movement of air through an obstructed upper airway, particularly the vocal cords in the larynx, causes stridor.^{31,32} Impairment of the expiratory phase of respiration also can occur, causing wheezing. With mild to moderate obstruction, inspiratory stridor is more prominent than expiratory wheezing because the airways tend to dilate with expiration. When the swelling and obstruction become severe, the airways no longer can dilate during expiration, and both stridor and wheezing occur.

Viral Croup. Viral croup, more appropriately called *acute laryngotracheobronchitis*, is a viral infection that affects the larynx, trachea, and bronchi. It is characterized by a brassy or "croupy" cough, which may or may not be accompanied by inspiratory stridor, hoarseness, and signs of respiratory distress

372 Unit Five: Alterations in the Respiratory System



■ FIGURE 20-8 ■ Location of airway obstruction in epiglottitis, acute laryngotracheobronchitis (croup), and bronchiolitis. (Courtesy of Carole Russell Hilmer, C.M.I.)

caused by various degrees of laryngeal swelling. The parainfluenza viruses account for approximately 75% all cases; the remaining 25% are caused by adenoviruses, respiratory syncytial virus, influenza A and B viruses, and measles virus.³²

Viral croup usually is seen in children 3 months to 5 years of age. The condition may affect the entire laryngotracheal tree, but because the subglottic area is the narrowest part of the respiratory tree in this age group, the obstruction usually is greatest in this area. Although the respiratory manifestations of croup often appear suddenly, they usually are preceded by upper respiratory infections that cause rhinorrhea (*i.e.*, runny nose), coryza (*i.e.*, common cold), hoarseness, and a low-grade fever. In most children, the manifestation of croup advances only to stridor and slight dyspnea before they begin to recover.

Airway obstruction may progress in some children. As obstruction increases, the stridor becomes continuous and is associated with nasal flaring with substernal and intercostal retractions. Agitation and crying aggravate the signs and symptoms, and the child prefers to sit up or be held upright. In the cyanotic, pale, or obstructed child, any manipulation of the pharynx, including use of a tongue depressor, can cause cardiorespiratory arrest and should be done only in a medical setting that has the facilities for emergency airway management.

Viral croup does not respond to antibiotics. The child should be disturbed as little as possible and carefully monitored for signs of respiratory distress. The use of steam from a shower or bath in a closed bathroom often brings prompt and dramatic relief of symptoms. Exposure to cold air also seems to relieve airway spasm; often, the severe symptoms are relieved simply because the child is exposed to cold air on the way to the hospital emergency room. Other treatment methods may be required when a humidifier or cold mist is ineffective. Administration of a racemic mixture of epinephrine (L-epinephrine and D-epinephrine) by positive-pressure breathing through a face mask often results in transient relief of symptoms.³⁴ Inhaled or systemically administered (oral or intramuscular) corticosteroid therapy may also be used. Children with progressive stridor or signs of respiratory distress may require hospitalization. Establishment of an artificial airway may become necessary in severe airway obstruction.

Spasmodic Croup. Spasmodic croup manifests with symptoms similar to those of acute viral croup. Because the child is afebrile and lacks other manifestations of the viral prodrome, it is thought that it may have an allergic origin. Spasmodic croup characteristically occurs at night and tends to recur with respiratory tract infections. The episode usually lasts several hours and may recur several nights in a row.

Most children with spasmodic croup can be effectively treated at home. An environment of high humidification (*i.e.*, cold-water room humidifier or taking the child into a bathroom with a warm, running shower) lessens irritation and prevents drying of secretions.

Epiglottitis. Acute epiglottitis is a dramatic, potentially fatal condition most often caused by the *H. influenzae* type B bacterium. It is usually seen in children 2 to 7 years of age, with a peak incidence at approximately 3.5 years.³² Epiglottitis is seen less commonly since the widespread use of immunization against *H. influenzae* type B.

Epiglottitis is characterized by inflammatory edema of the supraglottic area, including the epiglottis and pharyngeal structures, that comes on suddenly, bringing danger of airway obstruction and asphyxia. Within a matter of hours, epiglottitis may progress to complete obstruction of the airway and death unless adequate treatment is instituted.

The child appears pale, toxic, and lethargic and assumes a distinctive position—sitting up with the mouth open and the chin thrust forward. The child has difficulty in swallowing, a muffled voice, drooling, fever, and extreme anxiety. Moderate to severe respiratory distress is evident. There is inspiratory and sometimes expiratory stridor, flaring of the nares, and inspiratory retractions of the suprasternal notch and supraclavicular and intercostal spaces. Usually, no other family members are ill with acute respiratory disease.

The child with epiglottitis requires immediate hospitalization. Establishment of an airway by endotracheal tube or tracheotomy usually is needed. If epiglottitis is suspected, the child should never be forced to lie down because this causes the epiglottis to fall backward and may lead to complete airway obstruction. Examination of the throat with a tongue blade or other instrument may cause cardiopulmonary arrest and should be done only by medical personnel experienced in intubation of small children. It also is unwise to attempt any procedure, such as drawing blood, which would heighten the child's anxiety, because this also could precipitate airway spasm and cause death. Recovery from epiglottitis usually is rapid and uneventful after an adequate airway has been established and appropriate antibiotic therapy has been initiated.

Lower Airway Infections

Lower airway infections produce air trapping with prolonged expiration. The child presents with increased expiratory effort, increased respiratory rate, and wheezing. If the infection is severe, there also are marked intercostal retractions and signs of impending respiratory failure.

Acute Bronchiolitis. Acute bronchiolitis is a viral infection of the lower airways, most commonly caused by the respiratory

syncytial virus (RSV).³³ Other viruses, such as parainfluenza 3 virus and some adenoviruses, as well as mycoplasma, also are causative. The infection produces inflammatory obstruction of the small airways and necrosis of the cells lining the lower airways. Because the resistance to airflow in a tube is related to the fourth power of the radius, even minor swelling of bronchioles in an infant can produce profound changes in airflow.

Acute bronchiolitis usually occurs during the first 2 years of life, with a peak incidence between 3 to 6 months of age. The source of infection usually is a family member with a minor respiratory illness. Older children and adults tolerate bronchiolar edema much better than do infants and do not manifest the clinical picture of bronchiolitis.

Most affected infants in whom bronchiolitis develops have a history of a mild upper respiratory tract infection. These symptoms usually last several days and may be accompanied by fever and diminished appetite. There is then a gradual development of respiratory distress, characterized by a wheezy cough, dyspnea, and irritability. The infant usually is able to take in sufficient air but has trouble exhaling it. Air becomes trapped in the lung distal to the site of obstruction and interferes with gas exchange. Hypoxemia and, in severe cases, hypercapnia may develop. Airway obstruction may produce air trapping and hyperinflation of the lungs or collapse of the alveoli.

Infants with acute bronchiolitis have a typical appearance, marked by breathlessness with rapid respirations, a distressing cough, and retractions of the lower ribs and sternum (see Table 20-2). Crying and feeding exaggerate these signs. Wheezing and rales may or may not be present, depending on the degree of airway obstruction. In infants with severe airway obstruction, wheezing decreases as the airflow diminishes.

Usually, the most critical phase of the disease is the first 48 to 72 hours. Cyanosis, pallor, listlessness, and sudden diminution or absence of breath sounds indicate impending respiratory failure. Hospitalization is often indicated during this period. Treatment is largely supportive and includes administration of humidified oxygen to relieve hypoxia. Elevation of the head facilitates respiratory movements and avoids airway compression. Handling is kept at a minimum to avoid tiring. Because the infection is viral, antibiotics are not effective and are given only for a secondary bacterial infection. Dehydration may occur as the result of increased insensible water losses because of the rapid respiratory rate and feeding difficulties, and measures to ensure adequate hydration are needed. Recovery usually begins after the first 48 to 72 hours and usually is rapid and complete.

There is a reported increase in hyperactive airways during later childhood in infants who have had bronchiolitis.³⁴ The reason this occurs is not understood, and further investigation is needed.

Signs of Impending Respiratory Failure

Respiratory problems of infants and small children often are of sudden origin, and recovery usually is rapid and complete. Infants and children are at risk for the development of airway obstruction and respiratory failure resulting from obstructive disorders or lung infection. The child with epiglottitis is at risk for respiratory failure resulting from impaired gas exchange. Children with impending respiratory failure caused by airway or lung disease have rapid breathing, exaggerated use of the accessory muscles, retractions (which are more pronounced in the child than in the adult because of more compliant chest), nasal flaring, and grunting during expiration.³⁵ The signs and symptoms of impending respiratory failure are listed in Chart 20-1.

In summary, acute respiratory disease is the most common cause of illness in infancy and childhood. Although other body systems are physiologically ready for extrauterine life as early as 25 weeks of gestation, the lungs take longer. Immaturity of the respiratory system is a major cause of morbidity and mortality in premature infants. RDS is one of the most common causes of respiratory disease in premature infants. In these infants, pulmonary immaturity, together with surfactant deficiency, leads to alveolar collapse. BPD is a chronic

TABLE 20-2 Characteristics of Epiglottitis, Croup, and Bronchiolitis in Small Children

Characteristics	Epiglottitis	Croup	Bronchiolitis			
Common causative agent	Haemophilus influenzae type B bacterium	Mainly parainfluenza virus	Respiratory syncytial virus			
Most commonly affected age group	2–7 years (peak 3–5 years)	3 months to 5 years	Less than 2 years (most severe in infants younger than 6 months)			
Onset and preceding history	Sudden onset	Usually follows symptoms of a cold	Preceded by stuffy nose and other signs			
Prominent features	Child appears very sick and toxic Sits with mouth open and chin thrust forward Low-pitched stridor, difficulty swallow- ing, fever, drooling, anxiety Danger of airway obstruction and asphyxia	Stridor and a wet, barking cough Usually occurs at night Relieved by exposure to cold or moist air	Breathlessness, rapid, shallow breathing, wheezing, cough, and retractions of lower ribs and sternum during inspiration			
Usual treatment	Hospitalization Intubation or tracheotomy Treatment with appropriate antibiotic	Mist tent or vaporizer Administration of oxygen	Supportive treatment, administra- tion of oxygen and hydration			

CHART 20-1 Signs of Respiratory Distress and Impending Respiratory Failure in the Infant and Small Child

Severe increase in respiratory effort, including severe retractions or grunting, decreased chest movement

Cyanosis that is not relieved by administration of oxygen (40%)

Heart rate of 150 per minute or greater and increasing Bradycardia

Very rapid breathing (rate 60 per minute in the newborn to 6 months or above 30 per minute in children 6 months to 2 years)

Very depressed breathing (rate 20 per minute or below)

Retractions of the supraclavicular area, sternum, epigastrium, and intercostal spaces

Extreme anxiety and agitation

Fatigue

Decreased level of consciousness

pulmonary disease that develops in premature infants who were treated with mechanical ventilation.

Because of the smallness of the airway of infants and children, respiratory tract infections in these groups often are more serious. Infections that may cause only a sore throat and hoarseness in the adult may produce serious obstruction in the child. Obstruction of the extrathoracic airways often produces turbulence of airflow and an audible inspiratory crowing sound called *stridor*, and obstruction of the intrathoracic airways produces an audible expiratory wheezing or whistling sound. Among the respiratory tract infections that produce airway obstruction in small children are croup, epiglottitis, and bronchiolitis. Epiglottitis is a life-threatening supraglottic infection that may cause airway obstruction and asphyxia. Bronchiolitis is a viral infection of the lower airways; infants and small children with bronchiolitis are at risk for respiratory failure resulting from impaired gas exchange.

Infants and children with impending respiratory failure caused by airway or lung disease have rapid and shallow breathing, exaggerated use of the accessory muscles, retractions (which are more pronounced in the child than in the adult because of more compliant chest), and grunting during expiration.

REVIEW QUESTIONS

Differentiate among community-acquired pneumonia, hospital-acquired pneumonia, and pneumonia in immunocompromised persons in terms of pathogens, manifestations, and prognosis.

Explain the rationale for using the 7-valent, rather than the 23-valent, pneumococcus vaccine for children younger than 2 years.

Explain the difference between primary tuberculosis and reactivated tuberculosis on the basis of their pathophysiology and symptoms. Describe the manifestations of lung cancer and list two symptoms of lung cancer that are related to the invasion of the mediastinum.

Define the term *paraneoplastic* and cite three paraneoplastic manifestations of lung cancer.

Characterize the 5-year survival rate for lung cancer.

Cite the function of surfactant in lung function in the neonate and relate it to the development of respiratory distress syndrome.

Cite the possible causes and manifestations of bronchopulmonary dysplasia.

Describe the physiologic basis for sternal and chest wall retractions and grunting, stridor, and wheezing as signs of respiratory distress in infants and small children.

Compare croup, epiglottitis, and bronchiolitis in terms of incidence by age, site of infection, and signs and symptoms.

List the signs of impending respiratory failure in small children.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- Advisory Committee on Immunization Practices. (2002). Prevention and control of influenza: Recommendations of the Advisory Committee on Immunization Practices (ACIP). Morbidity and Mortality Weekly Report 51 (RR-03), 1–31.
- Small P.A. (1990). Influenza: Pathogenesis and host defenses. *Hospital Practice* 25 (11A), 51–62.
- 3. Montalto N.J., Gum K.D., Ashley J.V. (2000). Updated treatment for influenza A and B. American Family Physician 62, 2467–2476.
- 4. Couch R.B. (2000). Drug therapy: Prevention and treatment of influenza. *New England Journal of Medicine* 343, 1778–1787.
- Travis W.D., Farber J.L., Rubin E. (1999). The respiratory system. In Rubin E., Farber J.L. (Eds.), *Pathology* (3rd ed., pp. 601–605, 606–608). Philadelphia: Lippincott Williams & Wilkins.
- Chestnut M.S., Prendergast T.J. (2001). Lung. In Tierney L.M., McPhee S.J., Papadakis M.A. (Eds.), *Current medical diagnosis and treatment* (40th ed., pp. 291–313). New York: Lange Medical Books/ McGraw-Hill.
- Marrie T.J. (1998). Community-acquired pneumonia: Etiology, treatment. Infectious Disease Clinics of North America 12, 723–739.
- McEachen R., Campbell G.D. (1998). Hospital-acquired pneumonia: Epidemiology, etiology, and treatment. *Infectious Disease Clinics of North America* 12 (3), 761–779.
- Kobzik L. (1999). The Lung. In Cotran R.S., Kumar V., Collins T. (Eds.), Robbins' pathologic basis of disease (6th ed., pp. 347–348, 471–473, 718–722, 741–753). Philadelphia: W.B. Saunders.
- Centers for Disease Control and Prevention. (1996). Prevention of pneumococcal disease: Recommendations of the Advisory Committee on Immunization Practices (ACIP). *Morbidity and Mortality Weekly Report* 46 (RR-8), 1–24.
- 11. Catterall J.R. (1999). Streptococcus pneumoniae. Thorax 54, 929-937.
- Centers for Disease Control and Prevention. (1996). Defining the public health impact of drug-resistant Streptococcus pneumoniae. Morbidity and Mortality Weekly Report 45 (RR-1), 1–20.

- Centers for Disease Control and Prevention. (2000). Preventing pneumococcal disease among infants and small children: Recommendations of the Advisory Committee on Immunization Practices (ACIP). *Morbidity and Mortality Weekly Report* 46 (RR-9), 1–38.
- Stout J.E., Yu V.C. (1997). Legionellosis. New England Journal of Medicine 337, 682–688.
- Chambers H.F. (2001). Infectious diseases: Bacterial and chlamydial. In Tierney L.M., McPhee S.J., Papadakis M.A. (Eds.), *Current medical diagnosis and treatment* (40th ed., pp. 1368–1369). New York: Lange Medical Books/McGraw-Hill.
- 16. American Lung Association. (2000). Trends in tuberculosis morbidity and mortality. [On-line]. Available: http://www.lungusa.org.
- Samuelson J. (1999). Infectious diseases. In Cotran R.S., Kumar V., Collins T. (Eds.), *Robbins' pathologic basis of disease* (6th ed., pp. 347– 353). Philadelphia: W.B. Saunders.
- American Thoracic Society and Centers for Disease Control and Prevention. (2000). Diagnostic standards and classification of tuberculosis in adults and children. *American Review of Respiratory Disease* 161, 1376–1395.
- American Thoracic Society and Centers for Disease Control and Prevention. (1994). Treatment of tuberculosis and tuberculosis infection in adults and children. *American Journal of Critical Care Medicine* 149, 1359–1374.
- Galgiani J.N. (1999). Coccidioidomycosis: A regional disease of national importance. Annals of Internal Medicine 130, 293–300.
- Dismukes W.E. (1996). Blastomycosis. In Bennett J.C., Plum F. (Eds.), Cecil textbook of medicine (20th ed., pp. 1821–1822). Philadelphia: W.B. Saunders.
- 22. American Cancer Society. (2000). Lung cancer: Overview. [On-line]. Available: http://www3.cancer.org/cancerinfo/.
- 23. Petty T.L. (1997). Lung cancer. Postgraduate Medicine 101 (3), 121-122.

- Zander J., Hazinski M.F. (1992). Pulmonary disorders. In Hazinski M.F. (Ed.), *Nursing care of the critically ill child* (2nd ed., pp. 395–407). Philadelphia: W.B. Saunders.
- Moore K., Persaud T.V.N. (1998). The developing human (6th ed., pp. 262–269). Philadelphia: W.B. Saunders.
- Oski F.A. (Ed.). (1994). Principles and practices of pediatrics (2nd ed., pp. 336–339, 365–370). Philadelphia: J.B. Lippincott.
- Stoll B.J., Kliegman R.M. (2000). The fetus and neonatal infant. In Behrman R.E., Kliegman R.M., Jenson H.L. (Eds.), *Nelson textbook of pediatrics* (16th ed., pp. 498–505). Philadelphia: W.B. Saunders.
- Drug Facts and Comparisons. (2001). Lung surfactants (55th ed., pp. 704–710). St. Louis: Wolters Kluwer.
- McColley S.A. (1998). Bronchopulmonary dysplasia. Pediatric Clinics of North America 45, 573–585.
- Alexander K.C., Leung M.B.B.C., Cho H. (1999). Diagnosis of stridor in children. American Family Physician 60, 2289–2296.
- Orenstein D.M. (2000). Acute inflammatory upper airway obstruction. In Behrman R.E., Kliegman R.M., Jenson H.L. (Eds.), *Nelson textbook of pediatrics* (16th ed., pp. 1274–1278). Philadelphia: W.B. Saunders.
- Wright R.B., Pomerantz W.J., Luria J.W. (1999). New approaches to respiratory tract infections in children: Bronchiolitis and croup. *Emergency Medical Clinics of North America* 20, 93–114.
- Orenstein D.M. (2000). Bronchitis. In Behrman R.E., Kliegman R.M., Jenson H.L. (Eds.), *Nelson textbook of pediatrics* (16th ed., pp. 1284– 1287). Philadelphia: W.B. Saunders.
- Kattan M. (1999). Epidemiologic evidence of increased airway reactivity in children with a history of bronchiolitis. *Journal of Pediatrics* 135 (part 2) 3–13.
- Derish M.T., Frankel L.R. (2000). Respiratory distress and failure. In Behrman R.E., Kliegman R.M., Jenson H.L. (Eds.), *Nelson textbook of pediatrics* (16th ed., pp. 266–268). Philadelphia: W.B. Saunders.

снартер 21

Alterations in Respiratory Function: Disorders of Gas Exchange

Disorders of Lung Inflation

Disorders of the Pleura Pleuritis and Pleural Pain Pleural Effusion Pneumothorax Atelectasis Obstructive Airway Disorders Physiology of Airway Disease Bronchial Asthma Pathogenesis Clinical Features Status Asthmaticus and Fatal Asthma Bronchial Asthma in Children Chronic Obstructive Pulmonary Disease **Emphysema Chronic Bronchitis Clinical Features Bronchiectasis Cystic Fibrosis** Interstitial Lung Diseases **Pulmonary Vascular Disorders Pulmonary Embolism Pulmonary Hypertension** Secondary Pulmonary Hypertension Primary Pulmonary Hypertension Cor Pulmonale Acute Respiratory Distress Syndrome

Respiratory Failure Alterations in Blood Gases Mechanisms of Altered Gas Exchange Hypoxemia Hypercapnia Treatment of Respiratory Failure

he major function of the lungs is to oxygenate and remove carbon dioxide from the blood as a means of supporting the metabolic functions of body cells. The gas exchange function of the lungs depends on a system of open airways, expansion of the lungs, an adequate area for gas diffusion, and blood flow that carries the gases to the rest of the body. This chapter focuses on diseases that disrupt ventilation and gas exchange and on respiratory failure and hyperventilation.

DISORDERS OF LUNG INFLATION

Air entering through the airways inflates the lung, and the negative pressure in the pleural cavity keeps the lung from collapsing. Disorders of lung inflation are caused by conditions that produce lung compression or lung collapse. There can be compression of the lung by an accumulation of fluid in the intrapleural space, complete collapse of an entire lung as in pneumothorax, or collapse of a segment of the lung as in atelectasis.

Disorders of the Pleura

The pleura is a thin, double-layered membrane that encases the lungs. The two layers of the pleurae are separated by a thin layer of serous fluid (Fig. 21-1). The right and left pleural cavities are separated by the mediastinum, which contains the heart and other thoracic structures. Both the chest wall and the lungs have elastic properties. The pressure in the pleural cavity, which is negative in relation to atmospheric pressure, holds the lungs

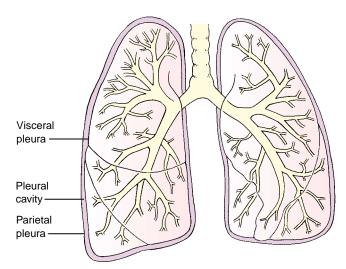


FIGURE 21-1 The parietal and visceral pleura and site of fluid accumulation in pleural effusions.

against the chest wall and keeps them from collapsing (see Chapter 19). Disorders of the pleura include pleural pain, pleural effusion, and pneumothorax.

Pleuritis and Pleural Pain

Pain is a common symptom of pleuritis, or inflammation of the pleura. Pleuritis is common in infectious processes such as viral respiratory infections or pneumonia that extend to involve the pleura. Most commonly the pain is abrupt in onset, such that the person experiencing it can cite almost to the minute when the pain started. It usually is unilateral and tends to be localized to the lower and lateral part of the chest. When the central part of the diaphragm is irritated, the pain may be referred to the shoulder. The pain is usually made worse by chest movements, such as deep breathing and coughing, that exaggerate pressure changes in the pleural cavity and increase movement of the inflamed or injured pleural surfaces. Because deep breathing is painful, tidal volumes usually are kept small, and breathing becomes more rapid. Reflex splinting of the chest muscles may occur, causing a lesser respiratory excursion on the affected side.

It is important to differentiate pleural pain from pain produced by other conditions, such as musculoskeletal strain of chest muscles, bronchial irritation, and myocardial disease. Musculoskeletal pain may occur as the result of frequent, forceful coughing. This type of pain usually is bilateral and located in the inferior portions of the rib cage, where the abdominal muscles insert into the anterior rib cage. It is made worse by movements associated with contraction of the abdominal muscles. The pain associated with irritation of the bronchi usually is substernal and dull, rather than sharp, in character. It is made worse with coughing but is not affected by deep breathing. Myocardial pain, which is discussed in Chapter 17, usually is located in the substernal area and is not affected by respiratory movements.

Pleural Effusion

Pleural effusion refers to an abnormal collection of fluid in the pleural cavity (see Fig. 21-1). The fluid may be a transudate, exudate, purulent drainage (empyema), chyle, or blood. Normally, only a thin layer (<10 to 20 mL) of serous fluid separates the visceral and parietal layers of the pleural cavity. Like fluid developing in other transcellular spaces in the body, pleural effusion occurs when the rate of fluid formation exceeds the rate of its removal (see Chapter 6). Five mechanisms have been linked to the abnormal collection of fluid in the pleural cavity: (1) increased capillary pressure, as in congestive heart failure; (2) increased capillary permeability, which occurs with inflammatory conditions; (3) decreased colloidal osmotic pressure, such as the hypoalbuminemia occurring with liver disease and nephrosis; (4) increased negative intrapleural pressure, which develops with atelectasis; and (5) impaired lymphatic drainage of the pleural space, which results from obstructive processes such as mediastinal carcinoma.

The accumulation of a serous transudate (clear fluid) in the pleural cavity often is referred to as *hydrothorax*. The condition may be unilateral or bilateral. The most common cause of hydrothorax is congestive heart failure.¹ Other causes are renal failure, nephrosis, liver failure, and malignancy. An *exudate* is a pleural fluid that has a specific gravity greater than 1.020 and, often, inflammatory cells. Conditions that produce exudative pleural effusions are infections, pulmonary infarction, malignancies, rheumatoid arthritis, and lupus erythematosus.

Empyema refers to pus in the pleural cavity. It is caused by direct infection of the pleural space from an adjacent bacterial pneumonia, rupture of a lung abscess into the pleural space, invasion from a subdiaphragmatic infection, or infection associated with trauma.

Chylothorax is the effusion of lymph in the thoracic cavity.² Chyle, a milky fluid containing chylomicrons, is found in the lymph fluid originating in the gastrointestinal tract. The thoracic duct transports chyle to the central circulation. Chylothorax also results from trauma, inflammation, or malignant infiltration obstructing chyle transport from the thoracic duct into the central circulation. It also can occur as a complication of intrathoracic surgical procedures and use of the great veins for total parenteral nutrition and hemodynamic monitoring.

Hemothorax is the presence of blood in the pleural cavity. Bleeding may arise from chest injury, a complication of chest surgery, malignancies, or rupture of a great vessel such as an aortic aneurysm. It is usually diagnosed by the presence of blood in the pleural fluid. Hemorrhagic pleural fluid is a mixture of blood and pleural fluid. Hemothorax usually requires drainage, and if the bleeding continues, surgery to control the bleeding may be required.

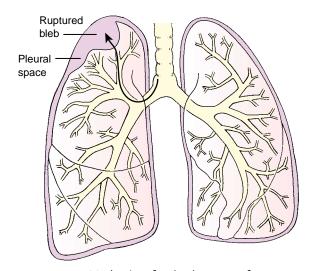
The manifestations of pleural effusion vary with the cause. Hemothorax may be accompanied by signs of blood loss and empyema by fever and other signs of inflammation. Fluid in the pleural cavity acts as a space-occupying mass; it causes a decrease in lung expansion on the affected side that is proportional to the amount of fluid that is present. The effusion may cause a shift in the mediastinal structures toward the opposite side of the chest with a decrease in lung volume on that side as well as the side with the pneumothorax. Characteristic signs of pleural effusion are dullness or flatness to percussion and diminished breath sounds. Dyspnea, the most common symptom, occurs when fluid compresses the lung, resulting in decreased ventilation. Pleuritic pain usually occurs only when inflammation is present, although constant discomfort may be felt with large effusions. Mild hypoxemia may occur and usually is corrected with supplemental oxygen.

Diagnosis of pleural effusion is based on chest radiographs, chest ultrasound, and computed tomography (CT). Thoracentesis is the aspiration of fluid from the pleural space. It can be used to obtain a sample of pleural fluid for diagnosis, or it can be used for therapeutic purposes. The treatment of pleural effusion is directed at the cause of the disorder. With large effusions, thoracentesis may be used to remove fluid from the intrapleural space and allow for re-expansion of the lung. A palliative method used for treatment of pleural effusions caused by a malignancy is the injection of a sclerosing agent into the pleural cavity. This method of treatment causes obliteration of the pleural space and prevents the reaccumulation of fluid. Open surgical drainage may be necessary in cases of continued effusion.

Pneumothorax

Normally, the pleural cavity is free of air and contains only a thin layer of fluid. When air enters the pleural cavity, it is called *pneumothorax*. Pneumothorax causes partial or complete collapse of the affected lung. Pneumothorax can occur without an obvious cause or injury (*i.e.*, spontaneous pneumothorax) or as a result of direct injury to the chest or major airways (*i.e.*, traumatic pneumothorax). Tension pneumothorax describes a life-threatening condition of excessive pressure in the pleural cavity.

Spontaneous Pneumothorax. Spontaneous pneumothorax occurs when an air-filled bleb, or blister, on the lung surface ruptures. Rupture of these blebs allows atmospheric air from the airways to enter the pleural cavity (Fig. 21-2). Because alveolar pressure normally is greater than pleural pressure, air flows from the alveoli into the pleural space, causing the involved portion of the lung to collapse as a result of its own recoil. Air continues to flow into the pleural space until a pressure gradient no longer exists or until the decline in lung size causes the leak to seal. Spontaneous pneumothoraces.³ Primary sponta-



■ FIGURE 21-2 ■ Mechanism for development of spontaneous pneumothorax.

neous pneumothorax occurs in otherwise healthy persons. Secondary spontaneous pneumothorax occurs in persons with underlying lung disease.

In primary spontaneous pneumothorax, the air-filled bleb that ruptures is usually on the top of the lung. The condition is seen most often in tall boys and young men between 10 and 30 years of age.³ It has been suggested that the difference in pleural pressure from the top to the bottom of the lung is greater in tall persons and that this difference in pressure may contribute to the development of blebs. Another factor that has been associated with primary spontaneous pneumothorax is smoking. Disease of the small airways related to smoking probably contributes to the condition.

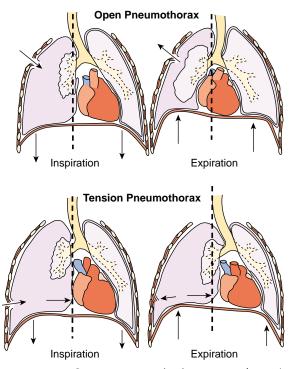
Secondary spontaneous pneumothoraces usually are more serious because they occur in persons with lung disease. They are associated with many different types of lung conditions that cause trapping of gases and destruction of lung tissue, including asthma, tuberculosis, cystic fibrosis, sarcoidosis, bronchogenic carcinoma, and metastatic pleural diseases. The most common cause of secondary spontaneous pneumothorax is emphysema.

Traumatic Pneumothorax. Traumatic pneumothorax may be caused by penetrating or nonpenetrating chest injuries. Fractured or dislocated ribs that penetrate the pleura are the most common cause of pneumothorax from nonpenetrating chest injuries. Hemothorax often accompanies these injuries. Pneumothorax also may accompany fracture of the trachea or major bronchus or rupture of the esophagus. Persons with pneumothorax caused by chest trauma frequently have other complications and may require chest surgery. Medical procedures such as transthoracic needle aspirations, intubation, and positive-pressure ventilation occasionally may cause pneumothorax. Traumatic pneumothorax also can occur as a complication of cardiopulmonary resuscitation.

Tension Pneumothorax. Tension pneumothorax occurs when the intrapleural pressure exceeds atmospheric pressure. It is a life-threatening condition and occurs when injury to the chest or respiratory structures permits air to enter but not leave the pleural space (Fig. 21-3). This results in a rapid increase in pressure in the chest with a compression atelectasis of the unaffected lung, a shift in the mediastinum to the opposite side of the chest, and compression of the vena cava with impairment of venous return to the heart.⁴ Although tension pneumothorax can develop in persons with spontaneous pneumothoraces, it is seen most often in persons with traumatic pneumothoraces.

With tension pneumothorax, the structures in the mediastinal space shift toward the opposite side of the chest (see Fig. 21-3). When this occurs, the position of the trachea, normally located in the midline of the neck, deviates with the mediastinum. There may be distention of the neck veins and subcutaneous emphysema (*i.e.*, air bubbles in the subcutaneous tissues of the chest and neck) and clinical signs of shock.

Clinical Features. The manifestations of pneumothorax depend on its size and the integrity of the underlying lung. In spontaneous pneumothorax, manifestations of the disorder include development of ipsilateral (same side) chest pain in an otherwise healthy person. There is an almost immediate in-



■ FIGURE 21-3 ■ Open or communicating pneumothorax (top) and tension pneumothorax (bottom). In an open pneumothorax, air enters the chest during inspiration and exits during expiration. There may be slight inflation of the affected lung due to a decrease in pressure as air moves out of the chest. In tension pneumothorax, air can enter but not leave the chest. As the pressure in the chest increases, the heart and great vessels are compressed and the mediastinal structures are shifted toward the opposite side of the chest. The trachea is pushed from its normal midline position toward the opposite side of the chest, and the unaffected lung is compressed.

crease in respiratory rate, often accompanied by dyspnea that occurs as a result of the activation of receptors that monitor lung volume. Heart rate is increased. Asymmetry of chest movement may occur because of the air trapped in the pleural cavity on the affected side. Percussion of the chest produces a more hyperresonant sound, and breath sounds are decreased or absent over the area of the pneumothorax.

Hypoxemia usually develops immediately after a large pneumothorax, followed by vasoconstriction of the blood vessels in the affected lung, causing the blood flow to shift to the unaffected lung. In persons with primary spontaneous pneumothorax, this mechanism usually returns oxygen saturation to normal within 24 hours. Hypoxemia usually is more serious in persons with underlying lung disease in whom secondary spontaneous pneumothorax develops. In these persons, the hypoxemia caused by the partial or total loss of lung function can be life threatening.

Diagnosis of pneumothorax can be confirmed by chest radiograph or CT scan. Blood gas analysis may be done to determine the effect of the condition on blood oxygen levels. Treatment varies with the cause and extent of the disorder. Even without treatment, air in the pleural space usually reabsorbs after the pleural leak seals. In small spontaneous pneumothoraces, the air usually reabsorbs, and observation and follow-up chest radiographs are all that is required. Supplemental oxygen may be used to increase the rate at which the air is reabsorbed. In larger pneumothoraces, the air is removed by needle aspiration or a closed drainage system used with or without an aspiration pump. This type of drainage system uses a one-way valve or a tube submerged in water to allow air to exit the pleural space and prevent it from re-entering the chest. In secondary pneumothorax, surgical closure of the chest wall defect, ruptured airway, or perforated esophagus may be required.

Emergency treatment of tension pneumothorax involves the prompt insertion of a large-bore needle or chest tube into the affected side of the chest along with one-way valve drainage or continuous chest suction to aid in lung expansion. Sucking chest wounds, which allow air to pass in and out of the chest cavity, should be treated by promptly covering the area with an airtight covering. Chest tubes are inserted as soon as possible.

Atelectasis

Atelectasis refers to the incomplete expansion of a lung or portion of a lung. It can be caused by airway obstruction, lung compression such as occurs in pneumothorax or pleural effusion, or the increased recoil of the lung caused by inadequate pulmonary surfactant (see Chapter 19).

Atelectasis is caused most commonly by airway obstruction (Fig. 21-4). Obstruction can be caused by a mucus plug in the airway or by external compression by fluid, tumor mass, exudate, or other matter in the area surrounding the airway. A small segment of lung or an entire lung lobe may be involved in obstructive atelectasis. Complete obstruction of an airway is followed by the absorption of air from the dependent alveoli and collapse of that portion of the lung. The danger of obstructive atelectasis increases after surgery. Anesthesia, pain, administration of narcotics, and immobility tend to promote retention of viscid bronchial secretions and thus airway obstruction.

Another cause of atelectasis is compression of lung tissue. It occurs when the pleural cavity is partially or completely filled with fluid, exudate, blood, a tumor mass, or air. It is observed most commonly in persons with pleural effusion from

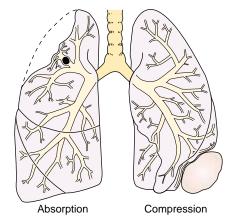


FIGURE 21-4 Atelectasis caused by airway obstruction and absorption of air from the involved lung area on the *left* and by compression of lung tissue on the right.

congestive heart failure or cancer. In compression atelectasis, the mediastinum shifts away from the affected lung.

The clinical manifestations of atelectasis include tachypnea, tachycardia, dyspnea, cyanosis, signs of hypoxemia, diminished chest expansion, absence of breath sounds, and intercostal retractions. Fever and other signs of infection may develop. Both chest expansion and breath sounds are decreased on the affected side. There may be intercostal retraction (pulling in of the intercostal spaces) over the involved area during inspiration. If the collapsed area is large, the mediastinum and trachea shift to the affected side. Signs of respiratory distress are proportional to the extent of lung collapse.

The diagnosis of atelectasis is based on signs and symptoms. Chest radiographs are used to confirm the diagnosis. CT scans may be used to show the exact location of the obstruction. Treatment depends on the cause and extent of lung involvement. It is directed at reducing the airway obstruction or lung compression and at reinflating the collapsed area of the lung. Ambulation and body positions that favor increased lung expansion are used when appropriate. Administration of oxygen may be needed to treat the hypoxemia. Bronchoscopy may be used as a diagnostic and treatment method.

In summary, lung inflation depends on a negative intrapleural pressure and unobstructed intrapulmonary airways. Disorders of the pleura include pleuritis and pain, pleural effusion, and pneumothorax. Pain is commonly associated with conditions that produce inflammation of the pleura. Characteristically, it is unilateral, abrupt in onset, and exaggerated by respiratory movements. Pleural effusion refers to the abnormal accumulation of fluid in the pleural cavity. The fluid may be a transudate (i.e., hydrothorax), exudate (*i.e.*, empyema), blood (*i.e.*, hemothorax), or chyle (i.e., chylothorax). Pneumothorax refers to an accumulation of air in the pleural cavity with the partial or complete collapse of the lung. It can result from rupture of an air-filled bleb on the lung surface or from penetrating or nonpenetrating injuries. A tension pneumothorax is a life-threatening event in which air progressively accumulates in the thorax, collapsing the lung on the injured side and progressively shifting the mediastinum to the opposite side of the thorax, producing severe cardiorespiratory impairment.

Atelectasis refers to an incomplete expansion of the lung. In adults, atelectasis usually results from airway obstruction caused by mucus plug or because of external compression by fluid, tumor mass, exudate, or other matter in the area surrounding the airway.

OBSTRUCTIVE AIRWAY DISORDERS

Obstructive airway diseases are caused by disorders that limit expiratory airflow. Bronchial asthma represents a reversible form of airway disease caused by narrowing of airways due to bronchospasm, inflammation, and increased airway secretions. Chronic obstructive airway disease can be caused by a variety of airway diseases, including chronic bronchitis, emphysema, bronchiectasis, and cystic fibrosis.

Physiology of Airway Disease

Air moves through the upper airways (*i.e.*, trachea and major bronchi) into the lower or pulmonary airways (*i.e.*, bronchi and alveoli), which are located in the lung. In the pulmonary airways, the cartilaginous layer that provides support for the trachea and major bronchi gradually disappears and is replaced with crisscrossing strips of smooth muscle (see Chapter 19). The contraction and relaxation of the smooth muscle layer, which is innervated by the autonomic nervous system (ANS), controls the diameter of the airways and consequent resistance to airflow. Parasympathetic stimulation, through the vagus nerve and cholinergic receptors, produces bronchoconstriction, and sympathetic stimulation, through β_2 -adrenergic receptors, increases bronchodilation. Normally, a slight vagal-mediated bronchoconstrictor tone predominates. When there is need for increased airflow, as during exercise, the vagal-mediated bronchoconstrictor tone is inhibited, and the bronchodilator effects of the sympathetic nervous system are increased.

Bronchial smooth muscle also responds to inflammatory mediators, such as histamine, that act directly on smooth muscle cells to produce bronchoconstriction. During an antigenantibody response, inflammatory mediators are released by a special type of cell, called the *mast cell*, which is present in the airways. The binding of immunoglobulin E (IgE) antibodies to receptors on mast cells prepares them for an allergic response when antigen appears (see Chapter 10).

Bronchial Asthma

Bronchial asthma is a chronic inflammatory airway disease. According to 1998 data, an estimated 26 million Americans have received diagnoses of asthma, and 10.6 million have had

KEY CONCEPTS

- Airway disorders involve the movement of gases into and out of the lung. They involve bronchial smooth muscle tone, mucosal injury, and obstruction due to secretions.
- The tone of the bronchial smooth muscles surrounding the airways determines airway radius, and the presence or absence of airway secretions influences airway patency.
- Bronchial smooth muscle is innervated by the autonomic nervous system—the parasympathetic nervous system, via the vagus nerve, produces bronchoconstriction and the sympathetic nervous system produces bronchodilation.
- Inflammatory mediators that are released in response to environmental irritants, immune responses, and infectious agents increase airway responsiveness by producing bronchospasm, increasing mucus secretion, and producing injury to the mucosal lining of the airways.

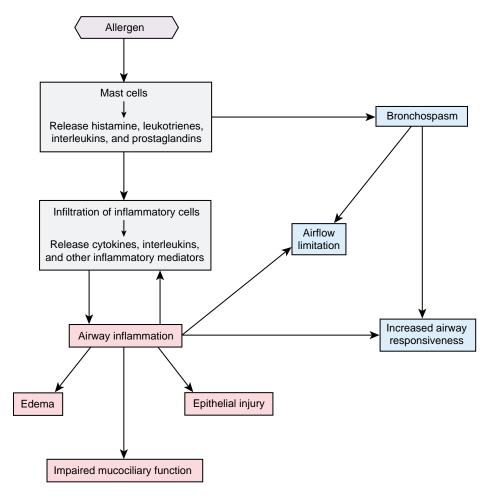
an asthma episode during the past 12 months.⁵ Of the 26 million Americans with diagnoses of asthma, 8.6 million are younger than 18 years. In the general population, asthma prevalence rates increased 102% between 1980 and 1994.⁵ There also has been a reported increase in incidence and mortality associated with asthma during the past several decades.

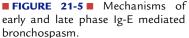
The National Heart, Lung, and Blood Institute's Second Expert Panel on the Management of Asthma defined bronchial asthma as "a chronic inflammatory disorder of the airways in which many cells and cellular elements play a role, in particular, mast cells, eosinophils, T lymphocytes, and epithelial cells."⁶ This inflammatory process produces recurrent episodes of airway obstruction, characterized by wheezing, breathlessness, chest tightness, and a cough that often is worse at night and in the early morning. These episodes, which usually are reversible either spontaneously or with treatment, also cause an associated increase in bronchial responsiveness to a variety of stimuli.⁶

Pathogenesis

In susceptible persons, an asthma attack can be triggered by a variety of stimuli that do not normally cause symptoms. Based on their mechanism of response, these triggers can be divided into two categories—bronchospastic or inflammatory. Bronchospastic triggers depend on the existing level of airway responsiveness. They do not normally increase airway responsiveness but produce symptoms in persons who already are predisposed to bronchospasm. Bronchospastic triggers include cold air, exercise, emotional upset, and exposure to bronchial irritants such as cigarette smoke. Inflammatory triggers exert their effects through the inflammatory response. They cause inflammation and prime the sensitive airways so they are hyperresponsive to nonallergic stimuli. The mechanisms whereby these two types of triggers produce an asthmatic attack can be further described as the early or acute response versus the late phase response.⁷ The acute or early response results in immediate bronchoconstriction on exposure to an inhaled antigen or irritant (Fig. 21-5). The symptoms of the acute response, which usually develop within 10 to 20 minutes, are caused by the release of chemical mediators from IgE-sensitized mast cells. In the case of airborne antigens, the reaction occurs when antigen binds to sensitized mast cells on the mucosal surface of the airways. Mediator release results in the infiltration of inflammatory cells and opening of the mucosal intercellular junctions and enhancement of antigen movement to the more prevalent submucosal mast cells (Fig. 21-5). In addition, there is bronchoconstriction caused by direct stimulation of parasympathetic receptors, mucosal edema caused by increased vascular permeability, and increased mucus secretions.

The *late phase response* develops 4 to 8 hours after exposure to an asthmatic trigger.^{7,8} The late phase response involves inflammation and increased airway responsiveness that prolong





the asthma attack and set into motion a vicious cycle of exacerbations. Typically, the response reaches a maximum within a few hours and may last for days or even weeks. An initial trigger in the late phase response causes the release of inflammatory mediators from mast cells, macrophages, and epithelial cells. These substances induce the migration and activation of other inflammatory cells (*e.g.*, basophils, eosinophils, neutrophils), which then produce epithelial injury and edema, changes in mucociliary function and reduced clearance of respiratory tract secretions, and increased airway responsiveness (see Fig. 21-6). Responsiveness to cholinergic mediators often is heightened, suggesting changes in parasympathetic control of airway function. Chronic inflammation can lead to airway remodeling, with more permanent changes in airway resistance.⁶

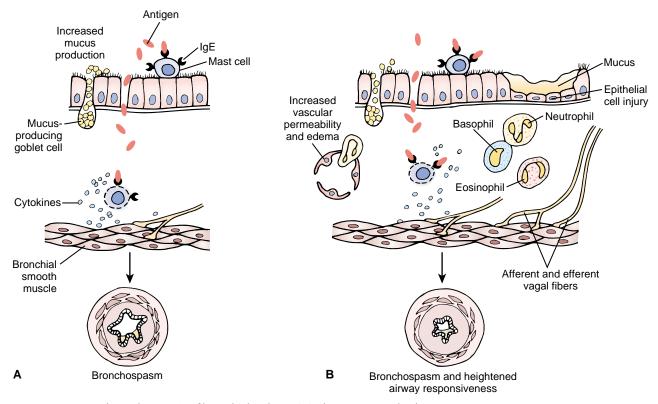
Causes. A number of factors can contribute to an asthmatic attack, including allergens, respiratory tract infections, hyperventilation, cold air, exercise, drugs and chemicals, hormonal changes and emotional upsets, airborne pollutants, and gastroesophageal reflux.

Inhalation of allergens is a common cause of asthma. Usually, this type of asthma has its onset in childhood or adolescence and is seen in persons with a family history of atopic allergy (see Chapter 10). Persons with allergic asthma often have other allergic disorders, such as hay fever, hives, and eczema. Attacks are related to exposure to specific allergens. Among airborne allergens implicated in perennial (year-around) asthma are house dust mite allergens, cockroach allergens, animal danders, and the fungus *Alternaria*.

Respiratory tract infections, especially those caused by viruses, may produce their effects by causing epithelial damage and stimulating the production of IgE antibodies directed toward the viral antigens. In addition to precipitating an asthmatic attack, viral respiratory infections increase airway responsiveness to other asthma triggers that may persist for weeks beyond the original infection.

Exercise-induced asthma occurs in 40% to 90% of persons with bronchial asthma.⁹ The cause of exercise-induced asthma is unclear. It has been suggested that during exercise, bronchospasm may be caused by the loss of heat and water from the tracheobronchial tree because of the need for conditioning (*i.e.*, warming and humidification) of large volumes of air. The response is commonly exaggerated when the person exercises in a cold environment.

Inhaled irritants, such as tobacco smoke and strong odors, are thought to induce bronchospasm by way of irritant receptors and a vagal reflex. Exposure to parental smoking has been reported to increase asthma severity in children.¹⁰ High doses of irritant gases such as sulfur dioxide, nitrogen dioxide, and ozone may induce inflammatory exacerbations of airway responsive-



■ FIGURE 21-6 ■ The pathogenesis of bronchial asthma. (A) The acute or early phase response. On exposure to an antigen, the immediate reaction is triggered by an IgE-mediated release of chemical mediators from sensitized mast cells. The release of chemical mediators results in increased mucus secretion, opening of mucosal intercellular junctions with increased antigen exposure of submucosal mast cells, and bronchoconstriction. Inhaled antigen produces release of chemical mediators from sensitized mast cells. (B) The late phase response involves release of inflammatory mediators from mast cells, macrophages, basophils, neutrophils, and eosinophils; epithelial cell injury and edema, decreased mucociliary function, and accumulation of mucus; and increased airway responsiveness.

ness (*e.g.*, smog-related asthma). Occupational asthma is stimulated by fumes and gases (*e.g.*, epoxy resins, plastics, toluene), organic and chemical dusts (*i.e.*, wood, cotton, platinum), and other chemicals (*e.g.*, formaldehyde) in the workplace.¹¹

There is a small group of persons with asthma in whom aspirin and nonsteroidal anti-inflammatory drugs (NSAIDs) are associated with asthmatic attacks, the presence of nasal polyps, and recurrent episodes of rhinitis.¹² An addition to the list of chemicals that can provoke an asthmatic attack are the sulfites used in food processing and as a preservative added to beer, wine, and fresh vegetables.

Both emotional factors and changes in hormone levels are thought to contribute to an increase in asthma symptoms. Emotional factors produce bronchospasm by way of vagal pathways. They can act as a bronchospastic trigger, or they can increase airway responsiveness to other triggers through noninflammatory mechanisms. The role of sex hormones in asthma is unclear, although there is much circumstantial evidence to suggest they may be important. As many as 40% of women with asthma report a premenstrual increase in asthma symptoms.¹³ Female sex hormones have a regulatory role on β_2 -adrenergic function, and it has been suggested that abnormal regulation may be a possible mechanism for premenstrual asthma.¹³

Symptoms of gastroesophageal reflux are common in both adults and children with asthma, suggesting that reflux of gastric secretions may act as a bronchospastic trigger. Reflux during sleep can contribute to nocturnal asthma.⁶

Clinical Features

Persons with asthma exhibit a wide range of signs and symptoms, from episodic wheezing and feelings of chest tightness to an acute, immobilizing attack. The attacks differ from person to person, and between attacks, many persons are symptom free. Attacks may occur spontaneously or in response to various triggers, respiratory infections, emotional stress, or weather changes. Asthma is often worse at night. Nocturnal asthma attacks usually occur at approximately 4 AM because of the occurrence of the late response to allergens inhaled during the evening and because of circadian variations in bronchial reactivity.¹⁴

During an asthmatic attack, the airways narrow because of bronchospasm, edema of the bronchial mucosa, and mucus plugging. Expiration becomes prolonged because of progressive airway obstruction. The amount of air that can be forcibly expired in 1 second (forced expiratory volume $[FEV_{1,0}]$) and the peak expiratory flow rate (PEF), measured in liters per second, are decreased (see Chapter 19). A fall in the PEF to levels below 50% of the predicted value during an acute asthmatic attack indicates a severe exacerbation and the need for emergency room treatment.6 With a prolonged attack, air becomes trapped behind the occluded and narrowed airways, causing hyperinflation of the lungs and an increase in the residual volume (RV). As a result, more energy is needed to overcome the tension already present in the lungs, and the accessory muscles (i.e., sternocleidomastoid muscles) are used to maintain ventilation and gas exchange. This causes dyspnea and fatigue. Because air is trapped in the alveoli and inspiration is occurring at higher residual lung volumes, the cough becomes less effective. As the condition progresses, the effectiveness of alveolar ventilation declines, and mismatching of ventilation and perfusion occurs, causing hypoxemia and hypercapnia. Pulmonary vascular resistance may increase as a result of the hypoxemia and hyperinflation, leading to a rise in pulmonary artery pressure and increased work demands on the right heart.

The physical signs of bronchial asthma vary with the severity of the attack. A mild attack may produce a feeling of chest tightness, a slight increase in respiratory rate with prolonged expiration, and mild wheezing. A cough may accompany the wheezing. More severe attacks are associated with use of the accessory muscles, distant breath sounds caused by air trapping, and loud wheezing. As the condition progresses, fatigue develops, the skin becomes moist, and anxiety and apprehension are obvious. Dyspnea may be severe, and often the person is able to speak only one or two words before taking a breath. At the point at which airflow is markedly decreased, breath sounds become inaudible with diminished wheezing, and the cough becomes ineffective despite being repetitive and hacking. This point often marks the onset of respiratory failure.

With increased air trapping, a greater negative intrapleural pressure is needed to inflate the lungs. This increased negative pressure, which is transmitted to the heart and blood vessels, causes the systolic blood pressure to fall during inspiration, a condition called *pulsus paradoxus*. It can be detected by using a blood pressure cuff and a mercury manometer (see Chapter 17).

Diagnosis and Management. The diagnosis of asthma is based on a careful history and physical examination, laboratory findings, and pulmonary function studies. Spirometry provides a means for measuring the PEF, FEV_{1.0}, forced vital capacity (FVC), and other indices of lung function (see Chapter 19). The level of airway responsiveness can be measured by inhalation challenge tests using methacholine (a cholinergic agonist), histamine, or exposure to a nonpharmacologic agent such as cold air. The Expert Panel of the National Education and Prevention Program of the National Heart, Lung, and Blood Institute has developed an asthma severity classification system intended for use in directing asthma treatment and identifying persons at high risk for the development of life-threatening asthma attacks⁶ (Table 21-1).

Small, inexpensive, portable meters that measure PEF are available. Although not intended for use in diagnosis of asthma, they can be used in clinics and physicians' offices and in the home to provide frequent measures of flow rates. Day-night (circadian) variations in asthma symptoms and PEF variability can be used to indicate the severity of bronchial hyperreactivity. A *person's best performance* (personal best) is established from readings taken throughout several weeks and is used as a reference to indicate changes in respiratory function.⁶

The treatment of bronchial asthma focuses on control of factors contributing to asthma severity and pharmacologic treatment.⁶ Measures to control factors contributing to asthma severity are aimed at prevention of exposure to irritants and factors that increase asthma symptoms and precipitate asthma exacerbations.

Pharmacologic treatment is used to prevent or treat reversible airway obstruction and airway hyperreactivity caused by the inflammatory process. The medications used in the treatment of asthma include those with bronchodilator and anti-inflammatory actions. The bronchodilators include the β_2 -adrenergic agonists and ipratroprium. The β_2 -adrenergic agonists, which are usually administered by inhalation methods,

TABLE 21-1	Classification of Asthma Severity		
	Symptoms	Nighttime Symptoms	Lung Function
Mild intermittent	Symptoms ≤2 times a week Asymptomatic and normal PEF between exacerbations Exacerbations brief (from a few hours to a few days); intensity may vary	≤2 times a month	FEV _{1.0} or PEF ≥80% predicted PEF variability <20%
Mild persistent	Symptoms >2 times a week but <1 time a day Exacerbations may affect activity	>2 times a month	FEV _{1.0} or PEF ≥80% predicted PEF variability 20%–30%
Moderate persister	It Daily symptoms Daily use of inhaled short-acting β_2 -agonist Exacerbations affect activity Exacerbations ≥ 2 times a week; may last days	>1 time a week	FEV _{1.0} or PEF >60%—<80% predicted PEF variability >30%
Severe persistent	Continual symptoms Limited physical activity Frequent exacerbations	Frequent	FEV _{1.0} or PEF ≤60% predicted PEF variability >30%

FEV_{1.0}, forced expiratory volume in 1 second; PEF, peak expiratory flow rate. (Adapted from National Education and Prevention Program. [1997]. Expert Panel report 2: Guidelines for the diagnosis and management of asthma. National Institutes of Health publication no. 97-4051. Bethesda, MD: National Institutes of Health.)

relax bronchial smooth muscle. Ipratropium is an inhaled anticholinergic drug that blocks the postganglionic efferent vagal pathways that cause bronchoconstriction. The antiinflammatory drugs include the corticosteroids, mast cell stabilizers, and leukotriene modifiers. The corticosteroids, which are often administered by inhalation methods, are considered the most effective anti-inflammatory agents for use in the long-term treatment of asthma. The anti-inflammatory agents sodium cromolyn and nedocromil are used to prevent an asthmatic attack. These agents, which are used prophylactically, act by stabilizing mast cells, thereby preventing release of the inflammatory mediators that cause an asthmatic attack. A newer group of drugs called the leukotriene modifiers are available for use in the treatment of asthma. The leukotrienes are potent biochemical mediators released from mast cells that cause bronchoconstriction, increased mucus secretion, and attraction and activation of inflammatory cells in the airways of people with asthma.

Status Asthmaticus and Fatal Asthma

Status asthmaticus is severe, prolonged asthma that is refractory to conventional methods of therapy. Most asthma deaths have occurred outside the hospital. Persons at highest risk are those with previous exacerbations resulting in respiratory failure, respiratory acidosis, and the need for intubation. Risk factors for fatal asthma are described in Chart 21-1.6 Although the cause of death during an acute asthmatic attack is largely unknown, both cardiac dysrhythmias and asphyxia caused by severe airway obstruction have been implicated. It has been suggested that an underestimation of the severity of the attack may be a contributing factor. Deterioration often occurs rapidly during an acute attack, and underestimation of its severity may lead to a life-threatening delay in seeking medical attention. Frequent and repetitive use of β_2 -agonist inhalers (more than twice in a month) far in excess of the recommended doses may temporarily blunt symptoms and mask the severity of the condition. Lack of access to medical care is another risk factor associated with asthma-related death. Distance, as in rural areas, or lack of financial resources, as in the uninsured or underinsured, may limit access to emergency care.

Bronchial Asthma in Children

Asthma is a leading cause of chronic illness in children and is responsible for a significant number of lost school days. It is the most frequently occurring admitting diagnosis in children's hospitals. As many as 10% to 15% of boys and 7% to 10% of girls have asthma at some time during childhood.¹⁵ Asthma

Risk Factors for Death CHART 21-1 From Asthma

- Past history of sudden severe exacerbations
- Prior intubation for asthma
- Two or more hospitalizations for asthma in the past year
- Three or more emergency care visits for asthma in the past year
- Hospitalization or an emergency care visit for asthma within the past month
- Use of more than two canisters per month of inhaled short-acting β₂-agonist
- Current use of systemic corticosteroids or recent withdrawal from systemic corticosteroids
- Difficulty perceiving airflow obstruction or its severity
- Comorbidity, as from cardiovascular diseases or chronic obstructive pulmonary disease
- Serious psychiatric disease or psychosocial problems
- Low socioeconomic status and urban residence
- Illicit drug use
- Sensitivity to Alternaria

(From National Education and Prevention Program. [1997]. Expert Panel report 2: Guidelines for the diagnosis and management of asthma. National Institutes of Health publication no. 97-4051. Bethesda, MD: National Institutes of Health.)

may have its onset at any age; 30% of children are symptomatic by 1 year of age, and 80% to 90% are symptomatic by 4 to 5 years of age.¹⁵

As with adults, asthma in children commonly is associated with an IgE-related reaction. It has been suggested that IgE directed against respiratory viruses in particular may be important in the pathogenesis of the wheezing illnesses in infants (*i.e.*, bronchiolitis), which often precede the onset of asthma. The respiratory syncytial virus and parainfluenza viruses are the most commonly involved.¹⁵ Other contributing factors include exposure to environmental allergens such as pet danders, dust mite antigens, and cockroach allergens. Exposure to environmental tobacco smoke also may contribute to asthma in children. Of particular concern is the effect of in utero exposure to maternal smoking on lung function in infants and children.¹⁶

The signs and symptoms of asthma in infants and small children vary with the stage and severity of an attack. Because airway patency decreases at night, many children have acute signs of asthma at this time. Often, previously well infants and children experience what may seem to be a cold with rhinorrhea, rapidly followed by irritability, a tight and nonproductive cough, wheezing, tachypnea, dyspnea with prolonged expiration, and use of accessory muscles of respiration. Cyanosis, hyperinflation of the chest, and tachycardia indicate increasing severity of the attack. Wheezing may be absent in children with extreme respiratory distress. The symptoms may progress rapidly and require a trip to the emergency room or hospitalization.

The Expert Panel of the National Heart, Lung, and Blood Institute's National Asthma Education Program has developed guidelines for the management of asthma in infants and children younger than 5 years and for adults and children older than 5 years.⁶ The Panel recommends that adolescents (and younger children when appropriate) be directly involved in developing their asthma management plans. Active participation in physical activities, exercise, and sports should be encouraged. A written asthma management plan should be prepared for the student's school, including plans to ensure reliable, prompt access to medications.⁶

Chronic Obstructive Pulmonary Disease

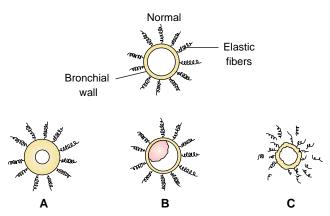
Chronic obstructive pulmonary disease (COPD) denotes a group of respiratory disorders characterized by chronic and recurrent obstruction of airflow in the pulmonary airways. Airflow obstruction usually is progressive, may be accompanied by airway hyperreactivity, and may be partially reversible.¹⁷ The most common cause of COPD is smoking.^{17,18} Thus, the disease is largely preventable. Unfortunately, clinical findings are almost always absent during the early stages of COPD, and by the time symptoms appear, the disease usually is far advanced. For smokers with early signs of airway disease, there is hope that early recognition, combined with appropriate treatment and smoking cessation, may prevent or delay the usually relentless progression of the disease.

The term *chronic obstructive pulmonary disease* encompasses two types of obstructive airway disease: *emphysema*, with enlargement of air spaces and destruction of lung tissue, and *chronic obstructive bronchitis*, with obstruction of small airways. The mechanisms involved in the pathogenesis of COPD usually are multiple and include inflammation and fibrosis of the bronchial wall, hypertrophy of the submucosal glands and hypersecretion of mucus, and loss of elastic lung fibers and alveolar tissue (Fig. 21-7).¹⁸ Inflammation and fibrosis of the bronchial wall, along with excess mucus secretion, obstruct airflow and cause mismatching of ventilation and perfusion. Destruction of alveolar tissue decreases the surface area for gas exchange, and the loss of elastic fibers leads to airway collapse. Normally, recoil of the elastic fibers that were stretched during inspiration provides the force needed to move air out of the lung during expiration. Because the elastic fibers are attached to the airways, they also provide radial traction to hold the airways open during expiration. In persons with COPD, the loss of elastic fibers impairs the expiratory flow rate, increases air trapping, and predisposes to airway collapse.

Emphysema

Emphysema is characterized by a loss of lung elasticity and abnormal enlargement of the air spaces distal to the terminal bronchioles, with destruction of the alveolar walls and capillary beds. Enlargement of the air spaces leads to hyperinflation of the lungs and produces an increase in total lung capacity (TLC). Two of the recognized causes of emphysema are smoking, which incites lung injury, and an inherited deficiency of α_1 -antitrypsin, an antiprotease enzyme that protects the lung from injury. Genetic factors, other than an inherited α_1 -antitrypsin deficiency, also may play a role in smokers who experience COPD at an early age.¹⁹

Emphysema is thought to result from the breakdown of elastin and other alveolar wall components by enzymes, called *proteases*, that digest proteins. These proteases, particularly elastase, which is an enzyme that digests elastin, are released from polymorphonuclear leukocytes (*i.e.*, neutrophils), alveolar macrophages, and other inflammatory cells.¹⁸ Normally, the lung is protected by antiprotease enzymes, including α_1 -antitrypsin. Cigarette smoke and other irritants stimulate the movement of inflammatory cells into the lungs, resulting in increased release of elastase and other proteases. In smokers in whom COPD develops, antiprotease production and release may be inadequate to neutralize the excess protease production



■ FIGURE 21-7 ■ Mechanisms of airflow obstruction in chronic obstructive lung disease. (**Top**) The normal bronchial airway with elastic fibers that provide traction and hold the airway open. (**Bottom**) Obstruction of the airway caused by (**A**) hypertrophy of the bronchial wall, (**B**) inflammation and hypersecretion of mucus, and (**C**) loss of elastic fibers that hold the airway open.

such that the process of elastic tissue destruction goes unchecked (Fig. 21-8).

A hereditary deficiency in α_1 -antitrypsin accounts for approximately 1% of all cases of COPD and is more common in young persons with emphysema.¹⁸ An α_1 -antitrypsin deficiency is inherited as an autosomal recessive disorder. Homozygotes who carry two defective genes have only about 15% to 20% of the normal plasma concentration of α_1 -antitrypsin. It is most common in persons of Scandinavian descent and is rare in Jews, blacks, and the Japanese.²⁰ Smoking and repeated respiratory tract infections, which also decrease α_1 -antitrypsin levels, contribute to the risk of emphysema in persons with an α_1 -antitrypsin deficiency. Laboratory methods are available for measuring α_1 -antitrypsin levels. Human α_1 -antitrypsin is available for replacement therapy in persons with a hereditary deficiency of the enzyme.

There are two commonly recognized types of emphysema: centriacinar and panacinar (Fig. 21-9). The centriacinar type affects the bronchioles in the central part of the respiratory lobule, with initial preservation of the alveolar ducts and sacs²⁰ (Fig. 21-10). It is the most common type of emphysema and is seen predominantly in male smokers. The panacinar type produces initial involvement of the peripheral alveoli and later extends to involve the more central bronchioles. This type of emphysema is more common in persons with α_1 -antitrypsin deficiency. It also is found in smokers in association with centrilobular emphysema. In such cases, panacinar changes are seen in the lower parts of the lung and the centriacinar changes in the upper parts of the lung.

Chronic Bronchitis

In chronic bronchitis, airway obstruction is caused by inflammation of the major and small airways. There is edema and hyperplasia of submucosal glands and excess mucus excretion

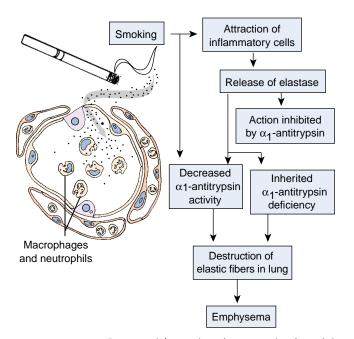
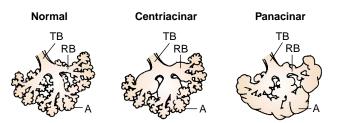


FIGURE 21-8 Protease (elastase)-antiprotease (antitrypsin) mechanisms of emphysema. The effects of smoking and an inherited α_1 -antitrypsin deficiency on the destruction of elastic fibers in the lung and development of emphysema.



■ FIGURE 21-9 ■ Centriacinar and panacinar emphysema. In centriacinar emphysema, the destruction is confined to the terminal (TB) and respiratory bronchioles (RB). In panacinar emphysema, the peripheral alveoli (A) are also involved. (West J.B. [1997]. *Pulmonary pathophysiology* [5th ed., p. 53]. Philadelphia: Lippincott-Raven)

into the bronchial tree. A history of a chronic productive cough of more than 3 months' duration for more than 2 consecutive years is necessary for the diagnosis of chronic bronchitis.^{17,18} Typically, the cough has been present for many years, with a gradual increase in acute exacerbations that produce frankly purulent sputum. Chronic bronchitis without airflow obstruction often is referred to as *simple bronchitis*, and chronic bronchitis with airflow obstruction as *chronic obstructive bronchitis*. The outlook for persons with simple bronchitis is good, compared with the premature morbidity and mortality associated with chronic obstructive bronchitis.

Chronic bronchitis is seen most commonly in middleaged men and is associated with chronic irritation from smok-



■ FIGURE 21-10 ■ Centrilobular emphysema. A whole mount of the left lung of a smoker with mild emphysema shows enlarged air spaces scattered throughout both lobes, which represent destruction of terminal bronchioles in the central part of the pulmonary lobule. These abnormal spaces are surrounded by intact pulmonary parenchyma. (Rubin E., Farber J.L. [1999]. Pathology [3rd ed., p. 628]. Philadelphia: Lippincott Williams & Wilkins)

ing and recurrent infections. In the United States, smoking is the most important cause of chronic bronchitis. Viral and bacterial infections are common in persons with chronic bronchitis and are thought to be a result, rather than a cause, of the problem.

Clinical Features

The mnemonics "pink puffer" and "blue bloater" have been used to differentiate the clinical manifestations of emphysema and chronic obstructive bronchitis. The important features of these two forms of COPD are described in Table 21-2. In practice, differentiation between the two types is not as vivid as presented here. This is because persons with COPD often have some degree of both emphysema and chronic bronchitis.

A major difference between the pink puffers and the blue bloaters is the respiratory responsiveness to the hypoxic stimuli. With pulmonary emphysema, there is a proportionate loss of ventilation and perfusion area in the lung. These persons are pink puffers, or fighters, who are able to overventilate and thus maintain relatively normal blood gas levels until late in the disease. Chronic obstructive bronchitis is characterized by excessive bronchial secretions and airway obstruction that causes mismatching of ventilation and perfusion. Thus, persons with chronic bronchitis are unable to compensate by increasing their ventilation; instead, hypoxemia and cyanosis develop. These are the blue bloaters, or nonfighters.

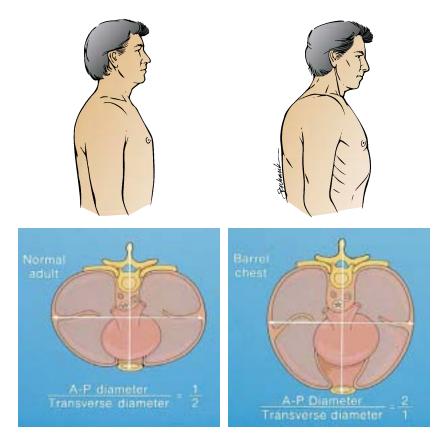
Persons with emphysema have marked dyspnea and struggle to maintain normal blood gas levels with increased breathing effort, including prominent use of the accessory muscles. The seated position, which stabilizes chest structures and allows for maximum chest expansion and use of accessory muscles, is preferred. With loss of lung elasticity and hyperinflation of the lungs, the airways often collapse during expiration because pressure in surrounding lung tissues exceeds airway pressure. Air becomes trapped in lungs, producing an increase in the anteroposterior dimensions of the chest, the so-called *barrel chest* that is typical of persons with emphysema (Fig. 21-11). Expiration often is accomplished through pursed lips. Pursedlip breathing, which increases the resistance to the outflow of air, helps to prevent airway collapse by increasing airway pressure. The work of breathing is greatly increased in persons with emphysema, and eating often is difficult. As a result, there often is considerable weight loss.

Chronic obstructive bronchitis is characterized by shortness of breath with a progressive decrease in exercise tolerance. As the disease progresses, breathing becomes increasingly more labored, even at rest. The expiratory phase of respiration is prolonged, and expiratory wheezes and crackles can be heard on auscultation. In contrast to persons with emphysema, those with chronic obstructive bronchitis are unable to maintain normal blood gases by increasing their breathing effort. Hypoxemia, hypercapnia, and cyanosis develop, reflecting an imbalance between ventilation and perfusion. Hypoxemia, in which arterial PO₂ levels fall below 55 mm Hg, causes reflex vasoconstriction of the pulmonary vessels and further impairment of gas exchange in the lung. Hypoxemia also stimulates red blood cell production, causing polycythemia. As a result, persons with chronic obstructive bronchitis develop pulmonary hypertension and, eventually, right-sided heart failure with peripheral edema (*i.e.*, cor pulmonale).

Chronic Obstructive Lung Disease				
Characteristic	Type A Pulmonary Emphysema ("Pink Puffers")	Type B Chronic Bronchitis ("Blue Bloaters")		
Smoking history	Usual	Usual		
Age of onset	40 to 50 years of age	30 to 40 years of age; disability in middle age		
Clinical features				
Barrel chest (hyperinflation of	Often dramatic	May be present		
the lungs)				
Weight loss	May be severe in advanced disease	Infrequent		
Shortness of breath	May be absent early in disease	Predominant early symptom, insidious in onset, exertional		
Decreased breath sounds	Characteristic	Variable		
Wheezing	Usually absent	Variable		
Rhonchi	Usually absent or minimal	Often prominent		
Sputum	May be absent or may develop late in the course	Frequent early manifestation, frequent infec- tions, abundant purulent sputum		
Cyanosis	Often absent, even late in the disease when there is low PO_2	Often dramatic		
Blood gases	Relatively normal until late in the disease process	Hypercapnia may be present Hypoxemia may be present		
Cor pulmonale	Only in advanced cases	Frequent Peripheral edema		
Polycythemia	Only in advanced cases	Frequent		
Prognosis	Slowly debilitating disease	Numerous life-threatening episodes due to acute exacerbations		

TABLE 21-2 Characteristics of Chronic Bronchitis and Emphysematous Types of Chronic Obstructive Lung Disease

388 Unit Five: Alterations in the Respiratory System



■ FIGURE 21-11 ■ Characteristics of normal chest wall and chest wall in emphysema. The normal chest wall and its cross section are illustrated on the left (A). The barrel-shaped chest of emphysema and its cross section are illustrated on the right (B). (Smeltzer S.C., Bare B.G. [2000]. *Medical-surgical nursing*. [9th ed., p. 454]. Philadelphia: Lippincott Williams & Wilkins)

Persons with combined forms of COPD (*i.e.*, some degree of both emphysema and chronic bronchitis) characteristically seek medical attention in the fifth or sixth decade of life, complaining of cough, sputum production, and shortness of breath. The symptoms typically have existed to some extent for 10 years or longer. The productive cough usually occurs in the morning. Dyspnea becomes more severe as the disease progresses. Frequent exacerbations of infection and respiratory insufficiency are common, causing absence from work and eventual disability.

The late stages of COPD are characterized by pulmonary hypertension, cor pulmonale, recurrent respiratory infections, and chronic respiratory failure. Death usually occurs during an exacerbation of illness associated with infection and respiratory failure.

Diagnosis and Treatment. The diagnosis of COPD is based on a careful history and physical examination, pulmonary function studies, chest radiographs, and laboratory tests. Airway obstruction prolongs the expiratory phase of respiration and affords the potential for impaired gas exchange because of the mismatching of ventilation and perfusion. The FVC is the amount of air that can be forcibly exhaled after maximal inspiration (see Chapter 19, Fig. 19-16). In an adult with normal respiratory function, this should be achieved in 4 to 6 seconds. In patients with COPD, the time required for FVC is increased, the FEV_{1.0} is decreased, and the ratio of FEV_{1.0} to FVC is decreased. In severe disease, the FVC is markedly reduced. Lung volume measurements reveal a marked increase in residual volume (RV), an increase in TLC, and elevation of the RV to TLC ratio. These and other measurements of expiratory flow are determined by spirometry and are used in the diagnosis of COPD.

The treatment of COPD depends on the stage of the disease and often requires an interdisciplinary approach. Smoking cessation is the only measure that slows the progression of the disease.²⁰ Persons in more advanced stages of the disease often require measures to maintain and improve physical and psychosocial functioning, pharmacologic interventions, and oxygen therapy. Respiratory tract infections can prove life threatening to persons with severe COPD. A person with COPD should avoid exposure to others with known respiratory tract infections. Immunization for influenza and pneumococcal infections decreases the likelihood of their occurrence.

Pharmacologic treatment includes the use of bronchodilators, including β_2 -adrenergic agonist drugs; the anticholinergic drug, ipratropium; and theophylline preparations. A long-term pulmonary rehabilitation program can significantly reduce episodes of hospitalization and add measurably to a person's ability to manage and cope with his or her impairment in a positive way.

Oxygen therapy is prescribed for selected persons with significant hypoxemia (arterial $PO_2 < 55 \text{ mm Hg}$). The use of continuous low-flow oxygen decreases dyspnea, helps to prevent pulmonary hypertension, and improves neuropsychological function and activity tolerance. Portable oxygen administration units, which allow mobility and the performance of activities of daily living, usually are used. The overall goal of oxygen therapy is to maintain the hemoglobin oxygen saturation at 89% to 90%, representing an arterial PO_2 of approximately 60 mm Hg (see Chapter 19, Fig. 19-18). Because the ventilatory drive associated with hypoxic stimulation of the peripheral chemoreceptors does not occur until the arterial PO₂ has been reduced to about 60 mm Hg or less, the oxygen flow rate usually is titrated to provide an arterial PO₂ of 55 to 65 mm Hg. Increasing the arterial PO₂ above that level tends to depress ventilation, leading to hypoventilation and carbon dioxide retention.

Bronchiectasis

Bronchiectasis is a chronic obstructive lung disease characterized by an abnormal dilatation of the large bronchi associated with infection and destruction of the bronchial walls (Fig. 21-12). To be diagnosed as bronchiectasis, the dilatation must be permanent, as compared with the reversible bronchial dilatation that sometimes accompanies viral and bronchial pneumonias

The pathogenesis of bronchiectasis can be either obstructive or nonobstructive.²⁰ Obstructive bronchiectasis is confined to a segment of the lung distal to a mechanical obstruction. It is caused by conditions such as tumors, foreign bodies, and mucus plugs in asthma. Nonobstructive bronchiectasis can be either localized or generalized. The use of immunizations and antibiotics has largely eliminated localized bronchiectasis



FIGURE 21-12 Bronchiectasis. The resected upper lobe shows widely dilated bronchi, with thickening of the bronchial walls and collapse and fibrosis of the pulmonary parenchyma. (Rubin E., Farber J.L. [1999]. Pathology [3rd ed., p. 601]. Philadelphia: Lippincott Williams & Wilkins)

caused by childhood bronchopulmonary infections such as measles, pertussis, and other bacterial infections.²¹

Generalized bronchiectasis is attributable largely to inherited impairments of host mechanisms or acquired disorders that permit introduction of infectious organisms into the airways. They include inherited conditions such as cystic fibrosis, in which airway obstruction is caused by impairment of normal mucociliary function; congenital and acquired immunodeficiency states, which predispose to respiratory tract infections; lung infection (e.g., tuberculosis, fungal infections, lung abscess); and exposure to toxic gases that cause airway obstruction.

Generalized bronchiectasis usually is bilateral and most commonly affects the lower lobes. Localized bronchiectasis can affect any area of the lung, the area being determined by the site of obstruction or infection. As the disease progresses, airway obstruction leads to smooth muscle relaxation with dilatation and eventual destruction of the bronchial walls. Infection produces inflammation, impairs mucociliary function, and causes weakening and further dilatation of the walls of the bronchioles. Pooling of secretions produces a vicious cycle of chronic inflammation and development of new infections.

Bronchiectasis is associated with an assortment of abnormalities that profoundly affect respiratory function, including atelectasis, obstruction of the smaller airways, and diffuse bronchitis. Affected persons have fever, recurrent bronchopulmonary infection, coughing, production of copious amounts of foul-smelling, purulent sputum, and hemoptysis. Weight loss and anemia are common. The physiologic abnormalities that occur in bronchiectasis are similar to those seen in chronic bronchitis and emphysema. As in the latter two conditions, chronic bronchial obstruction leads to marked dyspnea and cyanosis. Clubbing of the fingers is common in moderate to advanced bronchiectasis and is not seen in other types of obstructive lung diseases.21

Diagnosis is based on history and imaging studies. The condition often is evident on chest radiographs. High-resolution CT scanning of the chest allows for definitive diagnosis. Treatment consists of early recognition and treatment of infection along with regular postural drainage and chest physical therapy. Persons with this disorder benefit from many of the rehabilitation and treatment measures used for chronic bronchitis and emphysema. Localized bronchiectasis may be treated surgically.

Cystic Fibrosis

Cystic fibrosis is an autosomal recessive disorder involving fluid secretion in the exocrine glands and epithelial lining of the respiratory, gastrointestinal, and reproductive tracts. Most of the clinical manifestations of the disease are related to abnormal secretions that result in obstruction of organ passages such as the respiratory airways and pancreatic ducts. It is the most common fatal hereditary disorder of whites in the United States and is the most common cause of chronic lung disease in children.22

The cystic fibrosis gene, present on the long arm of chromosome 7, encodes the production of a single protein, the cystic fibrosis transmembrane conductance regulator (CFTR),

390 Unit Five: Alterations in the Respiratory System

which functions in chloride transport across cell membranes.^{23,24} Because of defective chloride transport, there is a threefold increase in sodium reabsorption. Water moves out of the extracellular fluid with the sodium, causing exocrine (*e.g.*, mucus) secretions to become exceedingly viscid. The cystic fibrosis gene is rare in African blacks and Asians. Homozygotes (*i.e.*, persons with two defective genes) have all or substantially all of the clinical symptoms of the disease, compared with heterozygotes, who are carriers of the disease but have no recognizable symptoms.

Clinically, cystic fibrosis is manifested by (1) chronic respiratory disease, (2) pancreatic exocrine deficiency, and (3) elevation of sodium chloride in the sweat. Nasal polyps, sinus infections, pancreatitis, and cholelithiasis also are common. Excessive loss of sodium in the sweat predisposes young children to salt depletion episodes. Most males with cystic fibrosis have congenital bilateral absence of the vas deferens with azoospermia.²³

Respiratory manifestations are caused by an accumulation of viscid mucus in the bronchi, impaired mucociliary clearance, and lung infections. Chronic bronchiolitis and bronchitis are the initial lung manifestations, but after months and years, structural changes in the bronchial wall lead to bronchiectasis. Widespread bronchiectasis is common by 10 years of age; large bronchiectatic cysts and abscesses develop in later stages of the disease. Infection and ensuing inflammation are causes of lung destruction in cystic fibrosis. *Staphylococcus aureus* and *Pseudomonas* infections are common. With advanced disease, 80% of persons harbor the *Pseudomonas* organism. New findings suggest that absence of CFTR predisposes to *Pseudomonas* infections, and once established, *Pseudomonas* is not easily cleared from the lungs, producing a cycle of chronic inflammation, tissue damage, and obstruction.

Pancreatic function is abnormal in approximately 80% to 90% of affected persons.²⁴ Steatorrhea, diarrhea, and abdominal pain and discomfort are common. In the newborn, meconium ileus may cause intestinal obstruction. The degree of pancreatic involvement is highly variable. In some children, the defect is relatively mild, and in others the involvement is severe and impairs intestinal absorption.

Early diagnosis and treatment are important in delaying the onset and severity of chronic illness. Diagnosis is based on the presence of respiratory and gastrointestinal manifestations typical of cystic fibrosis, a history of cystic fibrosis in a sibling, or a positive newborn screening test. Newborn screening consists of a test for determination of immunoreactive trypsinogen. Newborns with cystic fibrosis have elevated blood levels of immunoreactive trypsinogen, presumably because of secretory obstruction in the pancreas. Confirmatory tests include the sweat test to detect increased electrolytes in the sweat and genetic tests to detect the presence of the mutant CFTR genes.

The treatment of cystic fibrosis usually consists of replacement of pancreatic enzymes, physical measures to improve the clearance of tracheobronchial secretions (*i.e.*, postural drainage and chest percussion), bronchodilator therapy, and prompt treatment of respiratory tract infections. Lung transplantation is being used as a treatment for persons with end-stage lung disease.

Progress of the disease is variable. Improved medical management has led to longer survival—approximately half of children live beyond 20 years, and approximately one third of the nearly 30,000 persons with cystic fibrosis are adults.

In summary, obstructive ventilatory disorders are characterized by airway obstruction and limitation in expiratory airflow. Bronchial asthma is a chronic inflammatory disorder of the airways, characterized by airway hypersensitivity and episodic attacks of airway narrowing. An asthmatic attack can be triggered by a variety of stimuli. Based on their mechanism of response, these triggers can be divided into two types: bronchospastic and inflammatory. Bronchospastic triggers depend on the level of airway responsiveness. There are two types of responses in persons with asthma: the acute or early response and the late phase response. The acute response results in immediate bronchoconstriction on exposure to an inhaled antigen and usually subsides within 90 minutes. The late phase response usually develops 3 to 5 hours after exposure to an asthmatic trigger; it involves inflammation and increased airway responsiveness that prolong the attack and cause a vicious cycle of exacerbations.

COPD describes a group of conditions characterized by obstruction to airflow in the lungs. Among the conditions associated with COPD are emphysema, chronic bronchitis, and bronchiectasis. Emphysema is characterized by a loss of lung elasticity; abnormal, permanent enlargement of the air spaces distal to the terminal bronchioles; and hyperinflation of the lungs. Chronic bronchitis is caused by inflammation of major and small airways and is characterized by edema and hyperplasia of submucosal glands and excess mucus secretion into the bronchial tree. A history of a chronic productive cough that has persisted for at least 3 months and for at least 2 consecutive years in the absence of other disease is necessary for the diagnosis of chronic bronchitis. Emphysema and chronic bronchitis are manifested by the eventual mismatching of ventilation and perfusion. As the condition advances, signs of respiratory distress and impaired gas exchange become evident, with development of hypercapnia and hypoxemia. Bronchiectasis is a form of COPD that is characterized by an abnormal dilatation of the large bronchi associated with infection and destruction of the bronchial walls.

Cystic fibrosis is an autosomal recessive genetic disorder manifested by chronic lung disease, pancreatic exocrine deficiency, and elevation of sodium chloride in the sweat. Respiratory manifestations are caused by an accumulation of viscid mucus in the bronchi, impaired mucociliary clearance, lung infections, bronchiectasis, and dilatation. Mucus plugs can result in the total obstruction of an airway, causing atelectasis.

INTERSTITIAL LUNG DISEASES

The diffuse interstitial lung diseases are a diverse group of lung disorders that produce similar inflammatory and fibrotic changes in the interstitium or interalveolar septa of the lung. The disorders may be acute or insidious in onset; they may be rapidly progressive, slowly progressive, or static in their course. They include hypersensitivity pneumonitis (see Chapter 10), lung diseases caused by exposure to toxic drugs (*e.g.*, amio-

darone) and radiation, and occupational lung diseases including the pneumoconioses that are caused by the inhalation of inorganic dusts such as silica, coal dust, and asbestos. Some of the most common interstitial lung diseases are caused by exposure to the inhaled dust and particles. In many cases, no specific cause can be found.^{25,26} Examples of interstitial lung diseases and their causes are listed in Chart 21-2.

Current theory suggests that most interstitial lung diseases, regardless of the causes, have a common pathogenesis. It is thought that these disorders are initiated by some type of injury to the alveolar epithelium, followed by an inflammatory process that involves the alveoli and interstitium of the lung. An accumulation of inflammatory and immune cells causes continued damage of lung tissue and the replacement of normal, functioning lung tissue with fibrous scar tissue.

Because the interstitial lung diseases result in a stiff and noncompliant lung, they are commonly classified as fibrotic or restrictive lung disorders. In contrast to obstructive lung diseases, the lungs are stiff and difficult to expand, despite normal functioning airways. Persons with interstitial lung diseases experience dyspnea, tachypnea, and eventual cyanosis, without evidence of wheezing or signs of airway obstruction. Usually there is an insidious onset of breathlessness that initially occurs during exercise and may progress to the point that the person is totally incapacitated. A nonproductive cough may develop, particularly with continued exposure to the in-

CHART 21-2 Causes of Interstitial Lung Diseases* Occupational and Environmental Inhalants Inorganic dusts Asbestosis Silicosis

Coal miner's pneumoconiosis Organic dusts Hypersensitivity pneumonitis Gases and fumes Ammonia, phosgene, sulfur dioxide

Drugs and Therapeutic Agents

Cancer chemotherapeutic agents Busulfan Bleomycin Methotrexate Ionizing radiation

Immunologic Lung Disease

Sarcoidosis Collagen vascular diseases Systemic lupus erythematosus Rheumatoid arthritis Scleroderma Dermatomyositis-polymyositis

Miscellaneous

Postacute respiratory distress syndrome Idiopathic pulmonary fibrosis

*This list is not intended to be inclusive.

KEY CONCEPTS

INTERSTITIAL OR RESTRICTIVE LUNG DISEASES

- Interstitial lung diseases result from inflammatory conditions that affect the interalveolar structures of the lung and produce lung fibrosis and a stiff lung.
- A stiff and noncompliant lung is difficult to inflate, increasing the work of breathing and causing decreased exercise tolerance due to hypoxemia.
- Because of the increased effort needed for lung expansion, persons with interstitial lung disease tend to take small but more frequent breaths.

haled irritant. Typically, a person with a restrictive lung disease breathes with a pattern of rapid, shallow respirations. This tachypneic pattern of breathing, in which the respiratory rate is increased and the tidal volume is decreased, reduces the work of breathing because it takes less work to move air through the airways at an increased rate than it does to stretch a stiff lung to accommodate a larger tidal volume.

Although resting arterial blood gases usually are normal early in the course of the disease, arterial PO₂ levels may fall during exercise, and in cases of advanced disease, hypoxemia often is present, even at rest. In the late stages of the disease, hypercapnia and respiratory acidosis develop. Clubbing of the fingers and toes may develop because of chronic hypoxemia.

The diagnosis of interstitial lung disease requires a careful personal and family history, with particular emphasis on exposure to environmental, occupational, and other injurious agents. Chest radiographs may be used as an initial diagnostic method, and serial chest films often are used to follow the progress of the disease. A biopsy specimen for histologic study and culture may be obtained by surgical incision or bronchoscopy using a fiberoptic bronchoscope. Gallium lung scans often are used to detect and quantify the chronic alveolitis that occurs in interstitial lung disease. Gallium does not localize in normal lung tissue, but uptake of the radionuclide is increased in interstitial lung disease and other diffuse lung diseases.

The treatment goals for persons with interstitial lung disease focus on identifying and removing the injurious agent, suppressing the inflammatory response, preventing progression of the disease, and providing supportive therapy for persons with advanced disease. In general, the treatment measures vary with the type of lung disease. Corticosteroid drugs frequently are used to suppress the inflammatory response. Many of the supportive treatment measures used in the late stages of the disease, such as oxygen therapy and measures to prevent infection, are similar to those discussed for persons with COPD.

In summary, the interstitial lung diseases are characterized by fibrosis and decreased compliance of the lung. They include the occupational lung diseases, lung diseases caused by toxic drugs and radiation, and lung diseases of unknown origin, such as sarcoidosis. These disorders are thought to result from an inflammatory process that begins in the alveoli and extends to involve the interstitial tissues of the lung. Unlike COPD, which affects the airways, interstitial lung diseases affect the supporting collagen and elastic tissues that lie between the airways and blood vessels. These lung diseases decrease lung volumes, reduce the diffusing capacity of the lung, and cause various degrees of hypoxia. Because lung compliance is reduced, persons with this form of lung disease have a rapid, shallow breathing pattern.

PULMONARY VASCULAR DISORDERS

As blood moves through the lung, blood oxygen levels are raised, and carbon dioxide is removed. These processes depend on the matching of ventilation (*i.e.*, gas exchange) and perfusion (*i.e.*, blood flow). This section discusses three major disorders of the pulmonary circulation: pulmonary embolism, pulmonary hypertension, and acute respiratory distress syndrome. Pulmonary edema, another major problem of the pulmonary circulation, is discussed in Chapter 18.

Pulmonary Embolism

Pulmonary embolism develops when a blood-borne substance lodges in a branch of the pulmonary artery and obstructs the flow. The embolism may consist of a thrombus (Fig. 21-13), air that has accidentally been injected during intravenous infusion, fat that has been mobilized from the bone marrow after a fracture or from a traumatized fat depot (see Chapter 42), or



■ FIGURE 21-13 ■ Pulmonary embolism. The main pulmonary artery and its bifurcation have opened to reveal a large saddle embolus. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 289]. Philadelphia: Lippincott Williams & Wilkins)

amniotic fluid that has entered the maternal circulation after rupture of the membranes at the time of delivery.

Almost all pulmonary emboli arise from deep vein thrombosis (DVT) in the lower extremities (see Chapter 15). The presence of thrombosis in the deep veins of the legs or pelvis often is unsuspected until embolism occurs. The effects of emboli on the pulmonary circulation are related to mechanical obstruction of the pulmonary circulation and neurohumoral reflexes causing vasoconstriction. Obstruction of pulmonary blood flow causes reflex bronchoconstriction in the affected area of the lung, wasted ventilation and impaired gas exchange, and loss of alveolar surfactant. Pulmonary hypertension and right heart failure may develop when there is massive vasoconstriction because of a large embolus. Although small areas of infarction may occur, frank pulmonary infarction is uncommon.

Persons at risk for DVT also are at risk for thromboemboli. Among the physiologic factors that contribute to venous thrombosis are venous stasis, venous endothelial injury, and hypercoagulability states. Venous stasis and venous endothelial injury can result from prolonged bed rest, trauma, surgery, childbirth, fractures of the hip and femur, myocardial infarction and congestive heart failure, and spinal cord injury. Persons undergoing orthopedic surgery and gynecologic cancer surgery are at particular risk, as are bedridden patients in an intensive care unit. Cancer cells can produce thrombin and synthesize procoagulation factors, increasing the risk of thromboembolism. Use of oral contraceptive, pregnancy, and hormone replacement therapy are thought to increase the resistance to endogenous anticoagulants. The risk of pulmonary embolism among users of oral contraceptives is approximately three times the risk of nonusers.²⁷ Women who smoke are at particular risk.

The manifestations of pulmonary embolism depend on the size and location of the obstruction. Chest pain, dyspnea, and increased respiratory rate are the most frequent signs and symptoms of pulmonary embolism. Pulmonary infarction often causes pleuritic pain that changes with respiration; it is more severe on inspiration and less severe on expiration. Moderate hypoxemia without carbon dioxide retention occurs as a result of impaired gas exchange. Small emboli that become lodged in the peripheral branches of the pulmonary artery may exert little effect and go unrecognized. However, repeated small emboli often result in a gradual the size of the pulmonary capillary bed, resulting in pulmonary hypertension. Moderate-size emboli often present with breathlessness accompanied by pleuritic pain, apprehension, slight fever, rapid and shallow breathing, and cough productive of blood-streaked sputum. Persons with massive emboli usually present with sudden collapse, crushing substernal chest pain, shock, and sometimes loss of consciousness. The pulse is rapid and weak, the blood pressure is low, the neck veins are distended, and the skin is cyanotic and diaphoretic. Massive pulmonary emboli often are fatal.

The diagnosis of pulmonary embolism is based on clinical signs and symptoms, blood gas determinations, venous thrombosis studies, lung scans, CT scans, and in selected cases, pulmonary angiography. Laboratory studies and radiologic films are useful in ruling out other conditions that might give rise to similar symptoms. The electrocardiogram (ECG) may be used to detect signs of right heart strain resulting in an increase in pulmonary vascular resistance. Because almost all

pulmonary emboli originate from DVT, venous studies such as lower limb compression ultrasonography, impedance plethysmography, and contrast venography often are used as initial diagnostic procedures.

The treatment goals for pulmonary emboli focus on the prevention DVT and the development of thromboemboli, protecting the lungs from exposure to thromboemboli when they occur, and in the case of large and life-threatening pulmonary emboli, sustaining life and restoring pulmonary blood flow. Thrombolytic therapy may be indicated in persons with multiple or large emboli. Restoration of blood flow in persons with life-threatening pulmonary emboli can often be accomplished through the surgical removal of the embolus or emboli.

Prevention focuses on identification of persons at risk, avoidance of venous stasis and hypercoagulability states, and early detection of venous thrombosis. For persons at risk, graded compression elastic stockings and intermittent pneumatic compression (IPC) boots can be used to prevent venous stasis. Pharmacologic prophylaxis involves the use of anticoagulant drugs.

Pulmonary Hypertension

The term *pulmonary hypertension* describes the elevation of pressure in the pulmonary arterial system. The pulmonary circulation is a low-pressure system designed to accommodate varying amounts of blood delivered from the right heart and to facilitate gas exchange. The normal mean pulmonary artery pressure is approximately 15 mm Hg (e.g., 28 systolic/8 diastolic). The main pulmonary artery and major branches are relatively thin-walled, compliant vessels. The distal pulmonary arterioles also are thin walled and have the capacity to dilate, collapse, or constrict, depending on the presence of vasoactive substances released from the endothelial cells of the vessel, neurohumoral influences, flow velocity, oxygen tension, and alveolar ventilation.

Pulmonary hypertension can be caused by an elevation in left atrial pressure, increased pulmonary blood flow, or increased pulmonary vascular resistance. Because of the increased pressure in the pulmonary circulation, pulmonary hypertension increases the workload of the right heart. Although pulmonary hypertension can develop as a primary disorder, most cases develop secondary to some other condition.

Secondary Pulmonary Hypertension

Secondary pulmonary hypertension refers to an increase in pulmonary pressures associated with other disease conditions, usually cardiac or pulmonary. Secondary causes, or mechanisms, of pulmonary hypertension can be divided into four major categories: (1) elevation of pulmonary venous pressure, (2) increased pulmonary blood flow, (3) pulmonary vascular obstruction, and (4) hypoxemia.²⁸ Often more than one factor, such as COPD, heart failure, and sleep apnea, contributes to the elevation in pulmonary pressures.

Elevation of pulmonary venous pressure is common in conditions such as mitral valve stenosis and left ventricular heart failure, in which an elevated left atrial pressure is transmitted to the pulmonary circulation. Continued increases in left atrial pressure can lead to medial hypertrophy and intimal thickening of the small pulmonary arteries, causing sustained hypertension.

Increased pulmonary blood flow results from increased flow through left-to-right shunts in congenital heart diseases such as atrial or ventricular septal defects and patent ductus arteriosus. If the high-flow state is allowed to continue, morphologic changes occur in the pulmonary vessels, leading to sustained pulmonary hypertension. The pulmonary vascular changes that occur with congenital heart disorders are discussed in Chapter 17.

Obstruction of pulmonary blood vessels is most commonly the result of pulmonary emboli. Once initiated, the pulmonary hypertension that develops is self-perpetuating because of hypertrophy and proliferation of vascular smooth muscle.

Hypoxemia is another common cause of pulmonary hypertension. Unlike the vessels in the systemic circulation, most of which dilate in response to hypoxemia and hypercapnia, the pulmonary vessels constrict. The stimulus for constriction is thought to originate in the air spaces near the smaller branches of the pulmonary arteries. In situations in which certain regions of the lung are hypoventilated, the response is adaptive in that it diverts blood flow away from the poorly ventilated areas to more adequately ventilated portions of the lung. However, this effect becomes less beneficial as more and more areas of the lung become poorly ventilated. Pulmonary hypertension is a common problem in persons with advanced COPD. It also may develop at high altitudes in persons with normal lungs. Persons who experience marked hypoxemia during sleep (*i.e.*, those with sleep apnea) may also experience marked elevations in pulmonary arterial pressure.

The signs and symptoms of secondary pulmonary hypertension reflect not only the underlying cause, but the effect that the elevated pressures have on right heart function and oxygen transport. Dyspnea and fatigue are common. Peripheral edema, ascites, and signs of right heart failure (cor pulmonale, to be discussed) develop as the condition progresses.

Diagnosis is based on radiographic findings, echocardiography, and Doppler ultrasonography. Precise measurement of pulmonary pressures can be obtained only through right heart cardiac catheterization. Treatment measures are directed toward the underlying disorder. Vasodilator therapy may be indicated for some persons.

Primary Pulmonary Hypertension

Primary pulmonary hypertension is a relatively rare and rapidly progressive form of pulmonary hypertension that leads to right ventricular failure and death within a few years. Estimates of incidence range from 1 to 2 cases per million people in the general population.²⁹ The disease can occur at any age, and familial occurrences have been reported. Persons with the disorder usually have a steadily progressive downhill course, with death occurring in 3 to 4 years. Overall, the 5-year survival rate of untreated primary pulmonary hypertension is approximately 20%.30

Primary pulmonary hypertension is thought to be associated with a number of factors, including an autosomal dominant genetic predisposition along with an exogenous trigger.³⁰ Triggers include low oxygen levels that occur at high altitudes, exposure to certain drugs, human immunodeficiency virus infection, and autoimmune disorders.

The disorder is characterized by endothelial damage, coagulation abnormalities, and marked intimal fibrosis leading to obliteration or obstruction of the pulmonary arteries and arterioles. Most of the manifestations of the disorder are attributable to increased work demands on the right heart and a decrease in cardiac output. Symptoms are the same as those for secondary hypertension. The most obvious are dyspnea and fatigue that is out of proportion to other signs of a person's well being.

The diagnosis of primary pulmonary hypertension is based on an absence of disorders that cause secondary hypertension and mean pulmonary artery pressures greater than 25 mm Hg at rest or 30 mm Hg with exercise.

Treatment consists of measures to improve right heart function to reduce fatigue and peripheral edema. Supplemental oxygen may be used to increase exercise tolerance. The most widely used drugs for long-term therapy are the calcium channel blockers. Anticoagulant (warfarin) therapy may be used to decrease the risk of thrombosis caused by sluggish pulmonary blood flow. Epoprostenol, a short-acting (half-life, 3 to 5 minutes) analog of the naturally occurring vasodilator prostacyclin (prostaglandin I₂), is increasingly used in the long-term care of persons with the disorder. Because of its short half-life, the drug must be administered by continuous infusion, usually through an indwelling venous catheter with an automatic ambulatory pump. Lung transplantation may be an alternative for persons who do not experience response to other forms of treatment.

Cor Pulmonale

The term cor pulmonale refers to right heart failure resulting from primary lung disease and longstanding primary or secondary pulmonary hypertension. It involves hypertrophy and the eventual failure of the right ventricle. The manifestations of cor pulmonale include the signs and symptoms of the primary lung disease and the signs of right-sided heart failure (see Chapter 18). Signs of right-sided heart failure include venous congestion, peripheral edema, shortness of breath and a productive cough, which becomes worse during periods of worsening failure. Plethora (i.e., redness) and cyanosis and warm, moist skin may result from the compensatory polycythemia and desaturation of arterial blood that accompany chronic lung disease. Drowsiness and altered consciousness may occur as the result of carbon dioxide retention. Management of cor pulmonale focuses on the treatment of the lung disease and the heart failure. Low-flow oxygen therapy may be used to reduce the pulmonary hypertension and polycythemia associated with severe hypoxemia caused by chronic lung disease.

Acute Respiratory Distress Syndrome

Acute respiratory distress syndrome (ARDS), first described in 1967, is a devastating syndrome of acute lung injury. Initially called the *adult respiratory distress syndrome*, it is now called the *acute respiratory distress syndrome* because it also affects children. ARDS affects approximately 150,000 to 200,000 persons each year; at least 50% to 60% of these persons die, despite the most sophisticated intensive care.³¹

ARDS is the final common pathway through which many serious localized and systemic disorders produce diffuse injury to the alveolar-capillary membrane. It may result from a number of conditions, including aspiration of gastric contents, major trauma (with or without fat emboli), sepsis secondary to pulmonary or nonpulmonary infections, acute pancreatitis, hematologic disorders, metabolic events, and reactions to drugs and toxins^{31,32} (Chart 21-3).

Although a number of conditions may lead to ARDS, they all produce similar pathologic lung changes that include diffuse endothelial and epithelial cell injury with increased permeability of the alveolar-capillary membrane (Fig. 21-14). The increased permeability permits fluid, protein, cellular debris and blood cells to move out of the vascular compartment into the interstitium and alveoli of the lung. Alveolar cell damage leads to accumulation of edema fluid, surfactant inactivation, and formation of a hyaline membrane that is impervious to gas exchange. The lungs become very stiff and difficult to inflate. There is increased intrapulmonary shunting of blood, impaired gas exchange, and profound hypoxia. Gas exchange is further compromised by alveolar collapse resulting from abnormalities in surfactant production. When injury to the alveolar epithelium is severe, disorganized epithelial repair may lead to fibrosis.

The pathogenesis of ARDS is unclear. Neutrophils accumulate early in the course of the disorder and are thought to play a role in the pathogenesis of ARDS. Activated neutrophils release a variety of products, including oxidants, proteases, platelet activating factor, and leukotrienes that damage the alveolar capillary endothelium and alveolar cells.

Clinically, the syndrome consists of progressive respiratory distress, an increase in respiratory rate, and signs of respiratory failure. Radiologic findings usually show extensive bilateral consolidation of the lung tissue. Severe hypoxia persists despite increased inspired oxygen levels.

The treatment goals in ARDS are to supply oxygen to vital organs and provide supportive care until the condition causing the pathologic process has been reversed and the lungs have had a chance to heal. Assisted ventilation using high concentrations of oxygen may be required to overcome the hypoxia.

CHART 21-3 Conditions in Which ARDS Can Develop*

Aspiration Near drowning Aspiration gastric contents

Drugs, Toxins, Therapeutic Agents

Heroin Inhaled gases (*e.g.*, smoke, ammonia) Oxygen Radiation

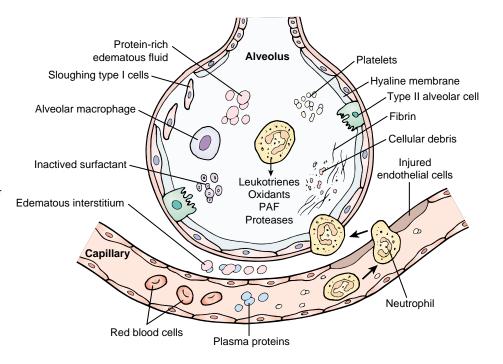
Infections

Gram-negative septicemia Other bacterial infections Viral infections

Trauma and Shock

Burns Fat embolism Chest trauma

*This list is not intended to be inclusive.



■ FIGURE 21-14 ■ The mechanism of lung changes in ARDS. Injury and increased permeability of the alveolar capillary membrane allow fluid, protein, cellular debris, platelets, and blood cells to move out of the vascular compartment and enter the interstitium and alveoli. Activated neutrophils release a variety of products that damage the alveolar cells and lead to edema, surfactant inactivation, and formation of a hyaline membrane.

Positive end-expiratory pressure breathing, which increases the pressure in the airways during expiration, may be used to assist in reinflating the collapsed areas of the lung and to improve the matching of ventilation and perfusion.

In summary, pulmonary vascular disorders include pulmonary embolism and pulmonary hypertension. Pulmonary embolism develops when a blood-borne substance lodges in a branch of the pulmonary artery and obstructs blood flow. The embolus can consist of a thrombus, air, fat, or amniotic fluid. The most common form is a thromboembolus arising from the deep venous channels of the lower extremities. Pulmonary hypertension is the elevation of pulmonary arterial pressure. It can be caused by an elevated left atrial pressure, increased pulmonary blood flow, or increased pulmonary vascular resistance secondary to lung disease. The term *cor pulmonale* describes right heart failure caused by primary pulmonary disease and longstanding pulmonary hypertension.

ARDS is a devastating syndrome of acute lung injury resulting from a number of serious localized and systemic disorders that damage the alveolar-capillary membrane of the lung. It results in interstitial edema of lung tissue, an increase in surface tension caused by inactivation of surfactant, collapse of the alveolar structures, a stiff and noncompliant lung that is difficult to inflate, and impaired diffusion of the respiratory gases with severe hypoxia that is resistant to oxygen therapy.

RESPIRATORY FAILURE

Respiratory failure is a condition in which the lungs fail to oxygenate the blood adequately and prevent carbon dioxide retention. It is not a specific disease but the result of a number of conditions that impair ventilation, compromise the matching of ventilation and perfusion, or disrupt blood flow in the lung. These conditions include impaired ventilation caused by impaired function of the respiratory center, weakness and paralysis of the respiratory muscles, chest wall deformities, airway obstruction, and disease of the airways and lungs. It may occur in previously healthy persons as the result of acute disease or trauma involving the respiratory system, or it may develop in the course of a chronic neuromuscular or respiratory disease. The causes of respiratory failure are summarized in Chart 21-4.

Alterations in Blood Gases

The term *hypoxia* refers to a reduction in oxygen supply to the tissues; *hypoxemia*, to a low level of oxygen in the blood; and *hypercapnia* (sometimes referred to as *hypercarbia*), to excess carbon dioxide in the blood. The abbreviation PO₂ often is used to indicate the partial pressure of oxygen in arterial blood, and the abbreviation PCO₂, the partial pressure of carbon dioxide. The common manifestations of respiratory failure are hypoxemia and hypercapnia. There is no absolute definition of the levels of arterial PO₂ and PCO₂ that indicate respiratory failure. As a general rule, *respiratory failure* refers to an arterial PO₂ level of 50 mm Hg or less and an arterial PCO₂ level greater than 50 mm Hg. However, these values are not reliable when dealing with persons who have chronic lung disease because many of these persons are alert and functioning with blood gas levels outside this range.

Mechanisms of Altered Gas Exchange

Various types of respiratory failure are associated with different degrees of hypoxemia or hypercapnia.³³ For example, pure hypoventilation, as occurs in conditions such as drug overdose, favors retention of CO₂. Severe ventilation-perfusion mismatching that is inadequate to maintain the PO₂ results in

CHART 21-4	Causes	of Respiratory Failure*
Impaired Ven	tilation	
Upper airway of Infection (e. Foreign boo Laryngospa Tumors Weakness or p Brain injury Drug overde Guillain-Bar Muscular dy Spinal cord Chest wall inju	g., epiglot ly sm aralysis of ose ré syndron rstrophy injury	titis) respiratory muscles
,	,	entilation and Perfusion
Chronic obstru Restrictive lung Severe pneum Atelectasis	ictive pulm g disease	
Impaired Diff	usion	
Pulmonary ede Acute respirate		syndrome
*This list is not in	tended to be	inclusive.

hypoxemia that is more severe in relation to hypercapnia. This pattern of respiratory failure is seen most commonly in persons with advanced COPD. In interstitial lung disease, there is severe hypoxemia but no hypercapnia because of increased ventilation. In ARDS, the arterial PCO_2 is typically

KEY CONCEPTS

DISORDERS OF BLOOD GASES IN RESPIRATORY FAILURE

- Respiratory failure represents failure of the lungs to adequately oxygenate the blood and prevent carbon dioxide retention.
- Hypoxemia results from decreased concentration of oxygen in the inspired air, airway diseases that impair ventilation, respiratory disorders that impair ventilation and/or perfusion, and cardiovascular disorders that impair movement of blood through the respiratory portions of the lung.
- Carbon dioxide retention is characteristic of conditions that produce hypoventilation.
- Conditions such as acute respiratory distress syndrome that impede the diffusion of gases in the lung impair the oxygenation of blood but do not interfere with the elimination of carbon dioxide.

low, but hypoxemia is severe. In this case, administration of oxygen can produce an increase in PO_2 without producing an increase in PCO_2 .

Hypoventilation. Hypoventilation occurs when the volume of "fresh" air moving into and out of the lung is significantly reduced. Hypoventilation is commonly caused by conditions outside the lung such as depression of the respiratory center (*e.g.*, drug overdose), diseases of the nerves supplying the respiratory muscles (*e.g.*, Guillain-Barré syndrome), disorders of the respiratory muscles (*e.g.*, muscular dystrophy), or thoracic cage disorders (*e.g.*, severe scoliosis or crushed chest).

Hypoventilation has two important effects on arterial blood gases. First, hypoventilation almost always causes an increase in PCO₂. The rise in PCO₂ is directly related to the level of ventilation. For example, the PCO₂ doubles when alveolar ventilation is reduced by one half. Thus, the PCO₂ level is a good diagnostic measure for hypoventilation. Second, hypoxemia caused by hypoventilation can be easily corrected by increasing the oxygen content of the inspired air.³³

Ventilation-Perfusion Mismatching. The mismatching of ventilation and perfusion occurs when areas of the lung are ventilated but not perfused or when areas are perfused but not ventilated. Usually the hypoxemia that is seen in situations of ventilation-perfusion mismatching is more severe in relation to hypercapnia than that seen in hypoventilation. Severe mismatching of ventilation and perfusion often is seen in persons with advanced COPD. These disorders contribute to the retention of carbon dioxide by reducing the effective alveolar ventilation, even when total ventilation is maintained. This occurs because a region of the lung is not perfused and gas exchange cannot take place or because an area of the lung is not being ventilated. Maintaining a high ventilation rate effectively prevents hypercapnia but also increases the work of breathing.

The hypoxemia associated with ventilation-perfusion disorders often is exaggerated by conditions such as hypoventilation and decreased cardiac output. For example, sedation can cause hypoventilation in persons with severe COPD, resulting in further impairment of ventilation. Likewise, a decrease in cardiac output because of myocardial infarction can exaggerate the ventilation-perfusion impairment in a person with mild pulmonary edema.

The beneficial effect of oxygen administration on PO_2 levels in ventilation-perfusion disorders depends on the degree of mismatching that is present. Because oxygen administration increases the diffusion gradient in ventilated portions of the lung, it usually is effective in raising arterial PO_2 levels but may decrease the respiratory drive and produce an increase in PCO_2 .

Impaired Diffusion. Diffusion impairment describes a condition in which gas exchange between the alveoli and the red blood cells is impeded because of an increase in the distance for diffusion or a decrease in the permeability of the alveolar capillary membrane to movement of gases. Impaired diffusion is caused by conditions such as interstitial lung disease, ARDS, pulmonary edema, and pneumonia. These conditions alter the thickness or permeability of the alveolar capillary membrane, rather than the time that the blood spends in pulmonary capillaries.

Hypoxemia resulting from impaired diffusion can be partially or completely corrected by the administration of 100% oxygen. This occurs because the increase in alveolar oxygen establishes a large alveolar-capillary diffusion gradient that overcomes the resistance to diffusion.

Shunt. Shunt occurs when blood reaches the arterial system without passing through the ventilated portion of the lung. Most shunts, such as those that occur with congenital heart disease, are extrapulmonary. However, a completely unventilated portion of the lung, as occurs with atelectasis, can result in the shunting of blood in the pulmonary circulation. Administration of oxygen usually increases arterial oxygen levels with intrapulmonary shunting but not with extrapulmonary shunting, such as occurs in children with intracardiac shunts. Because unoxygenated venous blood is being mixed with oxygenated blood in persons with intrapulmonary shunts, the rise in PO₂ levels depends on the degree of shunt.

Hypoxemia

Hypoxemia refers to a reduction in blood oxygen levels. Hypoxemia can result from an inadequate amount of oxygen in the air, disease of the respiratory system, or alterations in circulatory function. The mechanisms whereby respiratory disorders lead to a significant reduction in PO₂ are hypoventilation, impaired diffusion of gases, shunt, and mismatching of ventilation and perfusion.33 Another cause of hypoxemia, reduction of the partial pressure of oxygen in the inspired air, occurs only under special circumstances, such as at high altitudes. Often more than one mechanism contributes to hypoxemia in a person with respiratory or cardiac disease.

Clinical Manifestations. Hypoxemia produces its effects through tissue hypoxia and the compensatory mechanisms that the body uses to adapt to the lowered oxygen level. Body tissues vary considerably in their vulnerability to hypoxia; those with the greatest need are the nervous system and heart. Cessation of blood flow to the cerebral cortex results in loss of consciousness within 10 to 20 seconds. If the PO₂ of the tissues falls below a critical level, aerobic metabolism ceases and anaerobic metabolism takes over, with the formation and release of lactic acid.

The signs and symptoms of hypoxemia can be grouped into two categories: those resulting from impaired function of vital centers and those resulting from activation of compensatory mechanisms. Mild hypoxemia produces few manifestations. There may be slight impairment of mental performance and visual acuity and sometimes hyperventilation. This is because hemoglobin saturation still is approximately 90% when the PO₂ is only 60 mm Hg (see Chapter 19, Fig. 19-18). More pronounced hypoxemia may produce personality changes, restlessness, agitated or combative behavior, muscle incoordination, euphoria, impaired judgment, delirium, and, eventually, stupor and coma. Recruitment of sympathetic compensatory mechanisms produces tachycardia, cool skin (i.e., peripheral vasoconstriction), diaphoresis, and a mild increase in blood pressure. Profound acute hypoxemia can cause convulsions, retinal hemorrhages, and permanent brain damage. Hypotension and bradycardia often are preterminal events in persons with hypoxemia, indicating the failure of compensatory mechanisms.

In conditions of chronic hypoxemia, the manifestations may be insidious in onset and attributed to other causes, particularly in chronic lung disease. There may be a gradual increase in red blood cells (polycythemia) and development of pulmonary hypertension. Polycythemia increases the red blood cell concentration and the oxygen-carrying capacity of the blood. Pulmonary vasoconstriction occurs as a local response to alveolar hypoxia; it increases pulmonary arterial pressure and improves the matching of ventilation and blood flow.

Cyanosis. Cyanosis refers to the bluish discoloration of the skin and mucous membranes that results from an excessive concentration of reduced or deoxygenated hemoglobin in the small blood vessels. It usually is most marked in the lips, nail beds, ears, and cheeks. The degree of cyanosis is modified by the amount of cutaneous pigment, skin thickness, and the state of the cutaneous capillaries. Cyanosis is more difficult to distinguish in persons with dark skin and in areas of the body with increased skin thickness.

Although cyanosis may be evident in persons with respiratory failure, it often is a late sign. A concentration of approximately 5 g/dL of deoxygenated hemoglobin is required in the circulating blood for cyanosis to occur.³⁴ The absolute quantity of reduced hemoglobin, rather than the relative quantity, is important in producing cyanosis. Persons with anemia and low hemoglobin levels are less likely to exhibit cyanosis (because they have less hemoglobin to deoxygenate), even though they may be relatively hypoxic because of their decreased ability to transport oxygen, than persons who have high hemoglobin concentrations. Someone with a high hemoglobin level because of polycythemia may be cyanotic without being hypoxic.

Cyanosis can be divided into two types: central or peripheral. Central cyanosis is evident in the tongue and lips. It is caused by an increased amount of deoxygenated hemoglobin in the arterial blood. Peripheral cyanosis occurs in the extremities and on the tip of the nose or ears. It is caused by slowing of blood flow to an area of the body, with increased extraction of oxygen from the blood. It results from vasoconstriction and diminished peripheral blood flow, as occurs with cold exposure, shock, congestive heart failure, and peripheral vascular disease.

Diagnosis. Diagnosis of hypoxemia is based on clinical observation and diagnostic tests to determine blood oxygen levels. The analysis of arterial blood gases provides a direct measure of the oxygen content of the blood and is a good indicator of the lungs' ability to oxygenate the blood. Noninvasive measurement of hemoglobin saturation can be obtained using the pulse oximeter. The pulse oximeter uses light-emitting diodes and light-receiving sensors to quantify the light absorbed by oxygenated/deoxygenated hemoglobin in arterial blood.⁸⁵ Sensors that can be placed on the ear, finger, toe, or forehead are available. Pulse oximetry cannot distinguish between oxygen-carrying hemoglobin and carbon monoxide-carrying hemoglobin, a factor that should be considered when treating persons with carbon monoxide poisoning. It is also inaccurate in persons with pronounced vasoconstriction.

Hypercapnia

Hypercapnia refers to an increase in the carbon dioxide content of the blood. The diagnosis of hypercapnia is based on physiologic manifestations, arterial pH, and blood gas levels. The carbon dioxide level in the arterial blood, or PCO₂, is proportional to carbon dioxide production and inversely related to alveolar ventilation. The diffusing capacity of carbon dioxide is 20 times that of oxygen; therefore, hypercapnia is observed only in situations of hypoventilation sufficient to cause hypoxia.

Causes. Hypercapnia can occur in a number of disorders that cause hypoventilation or mismatching of ventilation and perfusion. Hypoventilation is a cause of hypercapnia in respiratory failure caused by depression of the respiratory center in drug overdose, neuromuscular diseases such as Guillain-Barré syndrome, or chest wall deformities such as those seen with severe scoliosis. Hypercapnia caused by ventilation-perfusion inequalities is seen most commonly in persons with COPD.

The respiratory center, which controls the activity of the muscles of respiration, is a crucial determinant of ventilation and elimination of carbon dioxide. It is composed of widely dispersed groups of neurons located in the medulla oblongata and pons (see Chapter 19). The activity of the respiratory center is regulated by chemoreceptors that monitor changes in the chemical composition of the blood. The most important chemoreceptors in terms of the minute-by-minute control of ventilation are the central chemoreceptors that respond to changes in the hydrogen ion (H⁺) concentration of the cerebrospinal fluid. Although the blood-brain barrier is impermeable to H^+ ions, CO_2 crosses it with ease. The CO_{21} in turn, reacts with water to form carbonic acid, which dissociates to form H^+ and bicarbonate (HCO₃⁻) ions. When the CO₂ content of the blood rises, CO₂ crosses the blood-brain barrier, liberating H⁺ ions that stimulate the central chemoreceptors. The excitation of the respiratory center caused by CO₂ is greatest during the first 1 to 2 days that blood levels are elevated, but it gradually declines during the next 1 to 2 days.34

In persons with respiratory problems that cause chronic hypoxia and hypercapnia, the peripheral chemoreceptors become the driving force for ventilation. These chemoreceptors, which are located in the bifurcation of the common carotid arteries and in the aortic arch, respond to changes in PO₂. Administration of high-flow oxygen to these persons can abolish the input from these peripheral receptors, causing a decrease in alveolar ventilation and a further increase in PCO₂ levels.

Respiratory muscle fatigue can contribute to carbon dioxide retention in persons with various primary respiratory diseases and in those with neuromuscular disorders. In these persons, respiratory muscle fatigue develops when energy requirements exceed the energy supply. A number of factors increase energy requirements or decrease the energy supply. The energy demands of the respiratory muscles are increased by high levels of ventilation or by factors that increase the work of breathing, such as high levels of airway resistance. The energy supply depends on blood flow and the oxygen content of the blood. Low cardiac output, anemia, and decreased oxygen saturation contribute to a decreased energy supply and increase the likelihood of respiratory muscle fatigue. With malnutrition, the energy stores of the muscles are diminished, and there may be structural changes in the muscle as well. Electrolyte imbalances, especially hypokalemia and hypophosphatemia, contribute to respiratory muscle weakness.³⁶

Clinical Manifestations. Hypercapnia affects many body functions, including renal, neural, and cardiovascular function, as well as the acid-base balance. Elevated levels of PCO_2 produce a decrease in pH and respiratory acidosis (see Chapter 6). The body normally compensates for an increase in PCO_2 by increasing renal bicarbonate retention. As long as the pH is in an acceptable range, the main complications of hypercapnia are those resulting from the accompanying hypoxia. Because the body compensates for chronic increases in blood levels of carbon dioxide, persons with chronic hypercapnia may not have symptoms until the PCO_2 is markedly elevated.

Carbon dioxide has a direct vasodilatory effect on many blood vessels and a sedative effect on the nervous system. When the cerebral vessels are dilated, headache develops. The conjunctivae are hyperemic and the skin flushed. Hypercapnia has nervous system effects similar to those of an anesthetic, thus the term *carbon dioxide narcosis*. There is progressive somnolence, disorientation, and if the condition is untreated, coma. Mild to moderate increases in blood pressure are common. Air hunger and rapid breathing occur when alveolar PCO₂ levels rise to approximately 60 to 75 mm Hg; as PCO₂ levels reach 80 to 100 mm Hg, the person becomes lethargic and sometimes becomes semicomatose. Anesthesia and death can result when PCO₂ levels reach 100 to 150 mm Hg.

Treatment of Respiratory Failure

The treatment of respiratory failure focuses on correcting the problems causing impaired gas exchange when possible and on relieving the hypoxemia and hypercapnia. A number of treatment modalities are available, including treatment of drug overdose, use of bronchodilating drugs, and antibiotics to treat severe respiratory tract infections. The insertion of an artificial airway and use of mechanical ventilation may be necessary.

Therapy for hypercapnia is directed at decreasing the work of breathing and improving the ventilation-perfusion balance. Intermittent rest therapy, such as nocturnal negative-pressure ventilation, applied to patients with hypercapnia and chronic obstructive disease or chest wall disease may be effective in increasing the strength and endurance of the respiratory muscle and improving the PCO₂. Respiratory muscle retraining aimed at improving the respiratory muscles, their endurance, or both has been used to improve exercise tolerance and diminish the likelihood of respiratory fatigue.

Hypoxemia is usually treated with oxygen therapy. Oxygen may be delivered by nasal cannula or mask. It also may be administered directly into an endotracheal or tracheostomy tube in persons who are being ventilated. The oxygen must be humidified as it is being administered. The concentration of oxygen that is being administered (usually determined by the flow rate) is based on the PO₂. The rate must be carefully monitored in persons with chronic lung disease because increases in PO₂ above 60 mm Hg are likely to depress the ventilatory drive. There also is the danger of oxygen toxicity with high concentrations of oxygen. Continuous breathing of oxygen at high concentrations can lead to diffuse parenchymal lung injury.

In summary, the lungs enable inhaled air to come in proximity to the blood flowing through the pulmonary capillaries so that the exchange of gases between the internal environment of the body and the external environment can take place. Respiratory failure is a condition in which the lungs fail adequately to oxygenate the blood or prevent undue retention of carbon dioxide. The causes of respiratory failure are many. It may arise acutely in persons with previously healthy lungs, or it may be superimposed on chronic lung disease. Respiratory failure is defined as a PO₂ of 50 mm Hg or less and a PCO₂ of 50 mm Hg or more.

Hypoxia refers to an acute or chronic reduction in tissue oxygenation. It can occur as the result of hypoventilation, diffusion impairment, shunt, and ventilation-perfusion impairment. Acute hypoxemia incites sympathetic nervous system responses such as tachycardia and produces symptoms that are similar to those of alcohol intoxication. In conditions of chronic hypoxia, the manifestations may be insidious in onset and attributed to other causes, particularly in chronic lung disease. The development of cyanosis requires a concentration of 5 g/dL of deoxygenated hemoglobin.

Hypercapnia refers to an increase in carbon dioxide levels. In the clinical setting, four factors contribute to hypercapnia: alterations in carbon dioxide production, disturbance in the gas exchange function of the lungs, abnormalities in respiratory function of the chest wall and respiratory muscles, and changes in neural control of respiration. The manifestations of hypercapnia consist of those associated with the vasodilation of blood vessels, including those in the brain, and depression of the central nervous system (*e.g.*, carbon dioxide narcosis).

REVIEW QUESTIONS

Differentiate among the causes and manifestations of spontaneous pneumothorax, secondary pneumothorax, and tension pneumothorax.

Describe the causes and manifestations of atelectasis.

Describe the physiology of bronchial smooth muscle as it relates to the early phase and late phase responses in the pathogenesis of bronchial asthma.

Explain the distinction between chronic bronchitis and emphysema in terms of pathology and clinical manifestations.

Describe the genetic abnormality responsible for cystic fibrosis and state the disorder's effect on lung function.

Compare the physiology changes that occur with COPD and interstitial lung diseases.

State the most common cause of pulmonary embolism and the clinical manifestations of the disorder.

Describe the pathophysiology of pulmonary arterial hypertension and three causes of secondary pulmonary hypertension. Describe the pathologic lung changes that occur in acute respiratory distress syndrome and relate them to the clinical manifestations of the disorder.

Define the terms hypoxia, hypoxemia, and hypercapnia and characterize the mechanisms whereby respiratory disorders cause hypoxemia and hypercapnia.

Compare the manifestations of hypoxia and hypercapnia.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- Chestnut M.S., Prendergast T.J. (2002). Lung. In Tierney L.M., McPhee S.J., Papadakis M.A. (Eds.), *Current medical diagnosis and treatment* (41st ed., pp. 350–355). New York: Lange Medical Books/ McGraw-Hill.
- Romero S. (2000). Nontraumatic chylothorax. Current Opinion in Pulmonary Medicine 6, 287–291.
- Sahn S.A., Heffner J.E. (2000). Spontaneous pneumothorax. New England Journal of Medicine 342, 868–874.
- Light R.W. (1995). Diseases of the pleura, mediastinum, chest wall, and diaphragm. In George R.B., Light R.W., Matthay M.A., et al. (Eds.), *Chest medicine* (3rd ed., pp. 501–520). Baltimore: Williams & Wilkins.
- 5. American Lung Association. (2000). Asthma statistics. [On-line.] Available: http://lungusa.org/data.
- 6. National Asthma Education and Prevention Program. (1997,2002). Expert Panel report 2: Guidelines for the diagnosis and management of asthma, and Guidelines for the diagnosis and management of asthma— Update on selected topics 2002. Bethesda, MD: National Institutes of Health, National Heart, Lung, and Blood Institute. Available: http://www.nhlbi.nih.gov/guidelines/asthma.
- Cotran R.S., Kumar V., Collins T. (1999). Robbins pathologic basis of disease (6th ed., p. 713). Philadelphia: W.B. Saunders.
- Busse W.W., Lemanske R.F. (2001). Asthma. New England Journal of Medicine 344, 350–362.
- McFadden E.R., Gilbert I.A. (1994). Exercise-induced asthma. New England Journal of Medicine 330, 1362–1366.
- Young S., LeSouef P.N., Geelhoed G.C., et al. (1991). The influence of a family history of asthma and parental smoking on airway responsiveness in early infancy. *New England Journal of Medicine* 324, 1168–1173.
- 11. Chan-Yeung M., Malo J. (1995). Occupational asthma. New England Journal of Medicine 333, 107–112.
- 12. Babu K.S., Salvi S.S. (2000). Aspirin and asthma. *Chest* 118, 1470–1476.
- Tan K.S., McFarlane L.C., Lipworth B.J. (1997). Loss of normal cyclical B₂ adrenoreceptor regulation and increased premenstrual responsiveness to adenosine monophosphate in stable female asthmatic patients. *Thorax* 52, 608–611.
- Dubuske D.M. (1994). Asthma: Diagnosis and management of nocturnal symptoms. *Comprehensive Therapy* 20, 628–639.
- Sly M. (2000). Allergic disorders. In Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.), *Nelson textbook of pediatrics* (16th ed., pp. 664– 685). Philadelphia: W.B. Saunders.
- Gilliland F.D., Berhane K., McConnell R., et al. (2000). Maternal smoking during pregnancy, environmental tobacco smoke exposure and childhood lung function. *Thorax* 55, 271–276.

400 Unit Five: Alterations in the Respiratory System

- 17. American Thoracic Society. (1995). Standards for the diagnosis and care of patients with chronic obstructive pulmonary disease. American Journal of Respiratory Critical Care Medicine 152, S77-S120.
- 18. Barnes P.J. (2000). Medical progress: Chronic obstructive pulmonary disease. New England Journal of Medicine 343, 269-280.
- 19. Silverman E.K., Chapman H.A., Drazen J.M., et al. (1998). Genetic epidemiology of severe, early-onset chronic obstructive pulmonary disease. American Journal of Critical Care Medicine 157, 1770-1778.
- 20. Travis W.D., Farber J.L., Rubin E. (1999). The respiratory system. In Rubin E., Farber J.L. (Eds.), Pathology (3rd ed., pp. 600-601, 623-630). Philadelphia: Lippincott Williams & Wilkins.
- 21. Mysliwiec V., Pina J.S. (1999). Bronchiectasis: The "other" obstructive lung disease. Postgraduate Medicine 106, 123-131.
- 22. Cystic Fibrosis Foundation. (2000). Facts about cystic fibrosis. [Online.] Available: http://www.cff.org/facts.htm.
- 23. Rubin E., Farber J.L. (1999). Developmental and genetic diseases. In Rubin E., Farber J.L. (Eds.), Pathology (3rd ed., pp. 246-249). Philadelphia: Lippincott Williams & Wilkins.
- Stern R.C. (1997). The diagnosis of cystic fibrosis. New England Journal 24. of Medicine 336, 487-491.
- 25. Reynolds H.V., Matthay R.A. (1995). Diffuse interstitial and alveolar inflammatory diseases. In George R.B., Light R.W., Matthay M.A., et al. (Eds.), Chest medicine (3rd ed., pp. 303-323). Baltimore: Williams & Wilkins.

- 26. Kobzik L. (1999). The lung. In Cotran R.S., Kumar V., Collins T. (Eds.), Robbins pathologic basis of disease (6th ed., pp. 697-755). Philadelphia: W.B. Saunders.
- 27. Goldhaber S.Z. (1998). Pulmonary embolism. New England Journal of Medicine 339, 93-104.
- 28. Richardi M.J., Rubenfire M. (1999). How to manage secondary pulmonary hypertension. Postgraduate Medicine 105 (2), 183-190.
- 29. Rubin L.J. (1997). Primary pulmonary hypertension. New England Journal of Medicine 336, 111-117.
- 30. Richardi M.J., Rubenfire M. (1999). How to manage primary pulmonary hypertension. Postgraduate Medicine 105 (2), 45-56.
- 31. Kollef M.H., Schuster D.P. (1995). The acute respiratory distress syndrome. New England Journal of Medicine 332, 27-36.
- 32. Ware L.B., Matthay M.A. (2000). The acute respiratory distress syndrome. New England Journal of Medicine 342, 1334-1348.
- 33. West J.B. (1997). Pulmonary pathophysiology: The essentials (5th ed., pp. 18-40, 132-142, 151). Philadelphia: Lippincott-Raven.
- 34. Guyton A.C., Hall J.E. (2000). Textbook of medical physiology (10th ed., pp. 477-478, 491, 804). Philadelphia: W.B. Saunders.
- 35. St. John R.E., Thomson P.D. (1999). Noninvasive respiratory monitoring. Critical Care Nursing Clinics of North America 11, 423-434.
- 36. Weinberger S.E., Schwartzstein R.M., Weiss J.W. (1989). Hypercapnia. New England Journal of Medicine 321, 1223–1230.

Alterations in the Urinary System

Control of Kidney Function

Kidney Structure and Function Gross Structure and Location The Nephron The Glomerulus Tubular Components of the Nephron **Nephron Blood Supply Urine Formation Glomerular Filtration** Tubular Reabsorption and Secretion **Regulation of Renal Blood Flow** Neural and Humoral Control Mechanisms Autoregulation **Elimination Functions of the Kidney Renal Clearance** Regulation of Sodium and Potassium Elimination **Regulation of pH** pH-Dependent Elimination of Organic Ions **Uric Acid Elimination Urea Elimination Drug Elimination Endocrine Functions of the Kidney** The Renin-Angiotensin-Aldosterone Mechanism **Erythropoietin** Vitamin D

CHAPTER

Tests of Renal Function Urine Tests Glomerular Filtration Rate Blood Tests Serum Creatinine Blood Urea Nitrogen

 It is no exaggeration to say that the composition of the blood is determined not so much by what the mouth takes in as by what the kidneys keep.
 Homer Smith, From Fish to Philosopher

he kidneys are remarkable organs. Each is smaller than a person's fist, but in a single day the two organs process approximately 1700 L of blood and combine its waste products into approximately 1.5 L of urine. As part of their function, the kidneys filter physiologically essential substances, such as sodium and potassium ions, from the blood and selectively reabsorb those substances that are needed to maintain the normal composition of internal body fluids. Substances that are not needed for this purpose or are in excess pass into the urine. In regulating the volume and composition of body fluids, the kidneys perform excretory and endocrine functions. The renin-angiotensin mechanism participates in the regulation of blood pressure and the maintenance of circulating blood volume, and erythropoietin stimulates red blood cell production.

KIDNEY STRUCTURE AND FUNCTION

Gross Structure and Location

The kidneys are paired, bean-shaped organs that lie outside the peritoneal cavity in the back of the upper abdomen, one on each side of the vertebral column at the level of the 12th thoracic to 3rd lumbar vertebrae (Fig. 22-1). The right kidney normally is situated lower than the left, presumably because of the position of the liver. In the adult, each kidney is approximately 10 to 12 cm long, 5 to 6 cm wide, and 2.5 cm deep and weighs approximately 113 to 170 g. The medial border of the kidney is indented by a deep fissure called the hilus. It is here that blood vessels and nerves enter and leave the kidney. The ureters, which connect the kidneys with the bladder, also enter the kidney at the hilus.

The kidney is a multilobular structure, composed of up to 18 lobes. Each lobule is composed of nephrons, which are the functional units of the kidney. Each nephron has a glomerulus that filters the blood and a system of tubular structures that selectively reabsorb material from the filtrate back into the blood and secrete materials from the blood into the filtrate as urine is being formed.

On longitudinal section, a kidney can be divided into an outer cortex and an inner medulla (Fig. 22-2). The cortex, which is reddish-brown, contains the glomeruli and convoluted tubules of the nephron and blood vessels. The medulla consists of light-colored, cone-shaped masses-the renal pyramids—that are divided by the columns of the cortex (i.e., columns of Bertin) that extend into the medulla. Each pyramid,

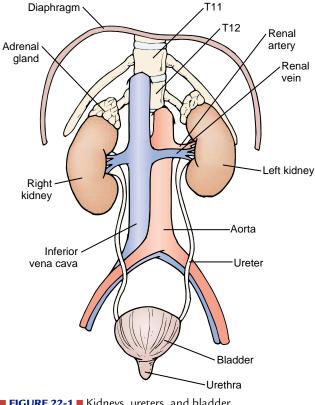


FIGURE 22-1 Kidneys, ureters, and bladder.

topped by a region of cortex, forms a lobe of the kidney. The apices of the pyramids form the papillae, which are perforated by the openings of the collecting ducts. The renal pelvis is a wide, funnel-shaped structure at the upper end of the ureter. It is made up of the calices or cuplike structures that drain the upper and lower halves of the kidney.

The kidney is ensheathed in a fibrous external capsule and surrounded by a mass of fatty connective tissue, especially at its ends and borders. The adipose tissue protects the kidney from mechanical blows and assists, together with the attached blood vessels and fascia, in holding the kidney in place. Although the kidneys are relatively well protected, they may be bruised by blows to the loin or by compression between the lower ribs and the ilium. Because the kidneys are outside the peritoneal cavity, injury and rupture do not produce the same threat of peritoneal involvement as rupture of organs such as the liver or spleen.

Each kidney is supplied by a single renal artery that arises on either side of the aorta. As the renal artery approaches the kidney, it divides into five segmental arteries that enter the hilus of the kidney. In the kidney, each segmental artery subdivides and branches several times. The smallest branches, the intralobular arteries, give rise to the afferent arterioles that supply the glomeruli (Fig. 22-3).

The Nephron

Each kidney is composed of more than 1 million tiny, closely packed functional units called nephrons. Each nephron consists of a glomerulus, where blood is filtered, and a tubular component. Here, water, electrolytes, and other substances needed to maintain the constancy of the internal environment are reabsorbed into the bloodstream while other unneeded materials are secreted into the tubular filtrate for elimination (see Fig. 22-4).

The Glomerulus

The glomerulus consists of a compact tuft of capillaries encased in a thin, double-walled capsule, called Bowman's capsule. Blood flows into the glomerular capillaries from the afferent arteriole and flows out of the glomerular capillaries into the efferent arteriole, which leads into the peritubular capillaries. Fluid and particles from the blood are filtered through the capillary membrane into a fluid-filled space in Bowman's capsule, called Bowman's space. The portion of the blood that is filtered into the capsule space is called the filtrate. The mass of capillaries and its surrounding epithelial capsule are collectively referred to as the renal corpuscle (Fig. 22-5A). The glomerular capillary membrane is composed of three layers: the capillary endothelial layer, the basement membrane, and the single-celled capsular epithelial layer (see Fig. 22-5B). The endothelial layer lines the glomerulus and interfaces with blood as it moves through the capillary. This layer contains many small perforations, called fenestrations.

The epithelial layer that covers the glomerulus is continuous with the epithelium that lines Bowman's capsule. The cells of the epithelial layer have unusual octopus-like structures that possess a large number of extensions, or foot processes (i.e., podocytes), which are embedded in the basement membrane. These foot processes form slit pores through which

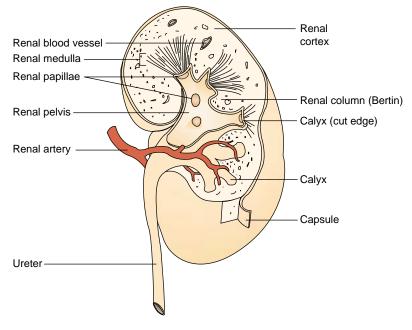
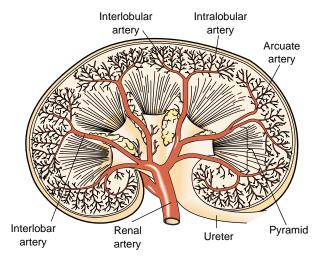


FIGURE 22-2 Internal structure of the kidney.

the glomerular filtrate passes. The *basement membrane* consists of a homogeneous acellular meshwork of collagen fibers, glycoproteins, and mucopolysaccharides (see Fig. 22-5C). Because the endothelial and the epithelial layers of the glomerular capillary have porous structures, the basement membrane determines the permeability of the glomerular capillary membrane. The spaces between the fibers that make up the basement membrane represent the pores of a filter and determine the size-dependent permeability barrier of the glomerulus. The size of the pores in the basement membrane normally prevents red blood cells and plasma proteins from passing through the glomerular membrane into the filtrate. There is evidence that the epithelium plays a major role in producing the basement membrane components, and it is



■ **FIGURE 22-3** ■ Simplified illustration of the arterial supply of the kidney. (Cormack D.H. [1987]. *Ham's histology* [9th ed.]. Philadelphia: J.B. Lippincott)

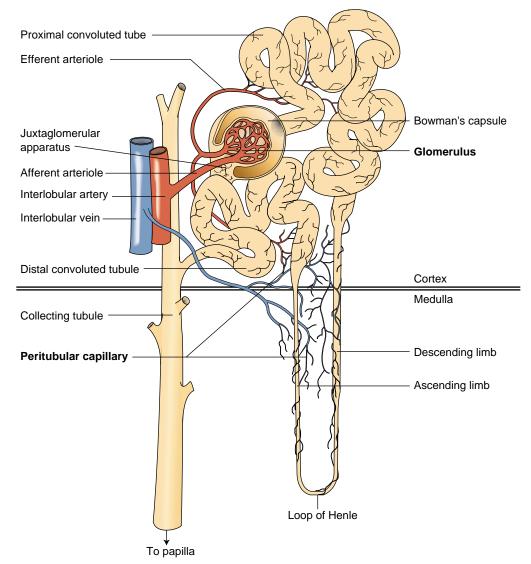
probable that the epithelial cells are active in forming new basement membrane material throughout life. Alterations in the structure and function of the glomerular basement membrane are responsible for the leakage of proteins and blood cells into the filtrate that occurs in many forms of glomerular disease.

Another important component of the glomerulus is the mesangium. In some areas, the capillary endothelium and the basement membrane do not completely surround each capillary. Instead, the mesangial cells, which lie between the capillary tufts, provide support for the glomerulus in these areas (see Fig. 22-5B). The mesangial cells produce an intercellular substance similar to that of the basement membrane. This substance covers the endothelial cells where they are not covered by basement membrane. The mesangial cells possess (or can develop) phagocytic properties and remove macromolecular materials that enter the intercapillary spaces. Mesangial cells also exhibit contractile properties in response to neurohumoral substances and are thought to contribute to the regulation of blood flow through the glomerulus. In normal glomeruli, the mesangial area is narrow and contains only a small number of cells. Mesangial hyperplasia and increased mesangial matrix occur in a number of glomerular diseases.

Tubular Components of the Nephron

The nephron tubule is divided into four segments: a highly coiled segment called the *proximal convoluted tubule*, which drains Bowman's capsule; a thin, looped structure called the *loop of Henle*; a distal coiled portion called the *distal convoluted tubule*; and the final segment called the *collecting tubule*, which joins with several tubules to collect the filtrate (Fig. 22-4). The filtrate passes through each of these segments before reaching the pelvis of the kidney.

Nephrons can be roughly grouped into two categories. Approximately 85% of the nephrons originate in the superficial part of the cortex and are called *cortical nephrons*. They have



■ FIGURE 22-4 ■ Nephron, showing the glomerular and tubular structures along with the blood supply.

KEY CONCEPTS

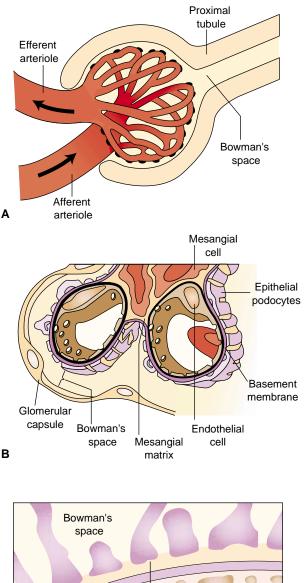
THE NEPHRON

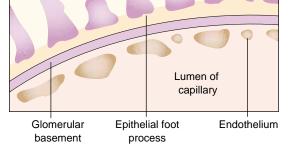
- The nephron, which contains a glomerulus and tubular structures, is the functional unit of the kidney.
- Each nephron is closely associated with two capillary beds: the glomerulus, where water-soluble nutrients, wastes, and other small particles are filtered from the blood, and the peritubular capillaries that surround the tubular structures.
- Tubular structures process the glomerular (urine) filtrate, selectively reabsorbing substances from the tubular fluid into the peritubular capillaries and secreting substances from the peritubular capillaries into the urine filtrate.

short, thick loops of Henle that penetrate only a short distance into the medulla. The remaining 15% are called *juxtamedullary nephrons*. They originate deeper in the cortex and have longer and thinner loops of Henle that penetrate the entire length of the medulla. The juxtamedullary nephrons are largely concerned with urine concentration.

The proximal tubule is a highly coiled structure that dips toward the renal pelvis to become the descending limb of the loop of Henle. The ascending loop of Henle returns to the region of the renal corpuscle, where it becomes the distal tubule. The distal convoluted tubule, which begins at the juxtaglomerular complex, is divided into two segments: the *diluting segment* and the *late distal tubule*. The late distal tubule fuses with the collecting tubule. Like the distal tubule, the collecting duct is divided into two segments: the *cortical collecting tubule* and the *inner medullary collecting tubule*.

Throughout its course, the tubule is composed of a single layer of epithelial cells resting on a basement membrane. The structure of the epithelial cells varies with tubular function.





C membrane

■ FIGURE 22-5 ■ Renal corpuscle. (A) Structures of the glomerulus. (B) Position of the mesangial cells in relation to the capillary loops and Bowman's capsule. (C) Cross-section of the glomerular membrane, showing the position of the endothelium, basement membrane, and epithelial foot processes.

The cells of the proximal tubule have a fine villous structure that increases the surface area for reabsorption; they also are rich in mitochondria, which support active transport processes. The epithelial layer of the thin segment of the loop of Henle has few mitochondria, indicating minimal metabolic activity and passive reabsorptive function.

Nephron Blood Supply

The nephron is supplied by two capillary systems, the glomerulus and the peritubular capillary network (Fig. 22-4). The glomerulus is a unique, high-pressure capillary filtration system located between two arterioles-the afferent and the efferent arterioles-that selectively dilate or constrict to regulate glomerular capillary pressure and consequently their filtration. The peritubular capillary network is a low-pressure reabsorptive system that originates from the efferent arteriole. These capillaries surround all portions of the tubules, an arrangement that permits rapid movement of solutes and water between the fluid in the tubular lumen and the blood in the capillaries. The medullary nephrons are supplied with two types of capillaries: the peritubular capillaries, which are similar to those in the cortex, and the vasa recta, which are long, straight capillaries. The vasa recta accompany the long loops of Henle in the medullary portion of the kidney to assist in exchange of substances flowing in and out of that portion of the kidney and play an important role in concentrating the urine. The peritubular capillaries rejoin to form the venous channels by which blood leaves the kidneys and empties into the inferior vena cava.

Although nearly all the blood flow to the kidneys passes through the cortex, less than 10% is directed to the medulla and only approximately 1% goes to the papillae. Under conditions of decreased perfusion or increased sympathetic nervous system stimulation, blood flow is redistributed away from the cortex toward the medulla. This redistribution of blood flow decreases glomerular filtration while maintaining the urine concentrating ability of the peritubular capillaries, a factor that is important during conditions such as shock.

Urine Formation

Urine formation involves the filtration of blood by the glomerulus to form an *ultrafiltrate of urine* and the tubular reabsorption of electrolytes and nutrients needed to maintain the constancy of the internal environment while eliminating waste materials.

Glomerular Filtration

Urine formation begins with the filtration of essentially protein-free plasma through the glomerular capillaries into Bowman's space. The movement of fluid through the glomerular capillaries is determined by the same factors (*i.e.*, capillary filtration pressure, colloidal osmotic pressure, and capillary permeability) that affect fluid movement through other capillaries in the body (see Chapter 6). The glomerular filtrate has a chemical composition similar to plasma, but it contains almost no proteins because large molecules do not readily cross the glomerular wall. Approximately 125 mL of filtrate is formed each minute. This is called the *glomerular filtration rate* (*GFR*). This rate can vary from a few milliliters per minute to as high as 200 mL/minute.

The location of the glomerulus between two arterioles allows for maintenance of a high-pressure filtration system. The capillary filtration pressure (approximately 60 mm Hg) in the glomerulus is approximately two to three times higher than that of other capillary beds in the body. The filtration pressure and the GFR are regulated by the constriction and relaxation of the afferent and efferent arterioles. Constriction of the efferent arteriole increases resistance to outflow from the glomeruli and increases the glomerular pressure and the GFR. Constriction of the afferent arteriole causes a reduction in the renal blood flow, glomerular filtration pressure, and GFR. The afferent and the efferent arterioles are innervated by the sympathetic nervous system and also are sensitive to vasoactive hormones, such as angiotensin II. During periods of strong sympathetic stimulation, such as occurs during shock, constriction of the afferent arteriole causes a marked decrease in renal blood flow and thus glomerular filtration pressure. Consequently, urine output can fall almost to zero.

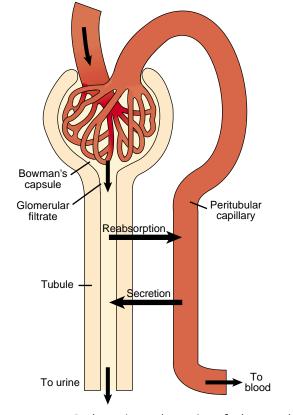
Tubular Reabsorption and Secretion

From Bowman's capsule, the glomerular filtrate moves into the tubular segments of the nephron. In its movement through the lumen of the tubular segments, the glomerular filtrate is changed considerably by the tubular transport of water and solutes. Tubular transport can result in reabsorption of substances from the tubular fluid into the blood or secretion of substances from the blood into the tubular fluid (Fig. 22-6). Segments of the renal tubule are adapted to reabsorb or secrete specific substances, using particular modes of transport.

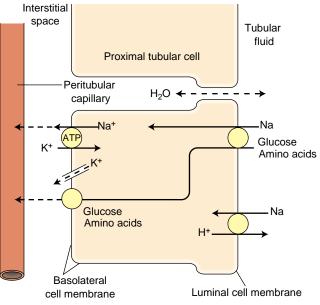
The basic mechanisms of transport across the tubular epithelial cell membrane are similar to those of other cell membranes in the body and include active and passive transport mechanisms (see Chapter 1). Water and urea are passively absorbed along concentration gradients. Sodium, potassium, chloride, calcium, and phosphate ions, as well as urate, glucose, and amino acids are reabsorbed using primary or secondary active transport mechanisms to move across the tubular membrane. Some substances, such as hydrogen, potassium, and urate ions, are secreted into the tubular fluids. Under normal conditions, only approximately 1 mL of the 125 mL of glomerular filtrate that is formed each minute is excreted in the urine. The other 124 mL is reabsorbed in the tubules. This means that the average output of urine is approximately 60 mL/hour.

Renal tubular cells have two membrane surfaces through which substances must pass as they are reabsorbed from the tubular fluid. The side of the cell that is in contact with the tubular lumen and tubular filtrate is called the *luminal membrane*. The outside membrane that lies adjacent to the interstitial fluid and the peritubular capillaries is called the *basolateral membrane*. In most cases, substances move from the tubular filtrate into the tubular cell along a concentration gradient, but they require facilitated transport or carrier systems to move across the basolateral membrane into the interstitial fluid, where they are absorbed into the peritubular capillaries.

The bulk of energy used by the kidney is for active transport mechanisms that facilitate sodium reabsorption and cotransport of other electrolytes and substances such as glucose and amino acids (Fig. 22-7). Cotransport uses a carrier system in which two substances move in the same direction. The active transport of one substance such as sodium is coupled to the movement of a second substance such as glucose or an amino acid. A few substances, such as hydrogen, are secreted into the tubule using countertransport, in which the movement of one substance, such as sodium, enables the movement of a second substance in the opposite direction.



■ **FIGURE 22-6** ■ Reabsorption and secretion of substances between the renal tubules and peritubular capillaries.

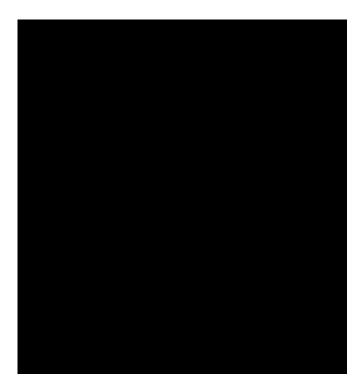


■ FIGURE 22-7 ■ Mechanism for secondary active transport or cotransport of glucose and amino acids in the proximal tubule. The energy-dependent sodium-potassium pump on the basal lateral surface of the cell maintains a low intracellular gradient that facilitates the downhill movement of sodium and glucose or amino acids (cotransport) from the tubular lumen into the tubular cell and then into the peritubular capillary.

Proximal Tubule. Although tubular transport occurs throughout the renal tubule, most of it occurs in the proximal tubule. Approximately 65% of all reabsorptive and secretory processes that occur in the tubular system take place in the proximal tubule. There is almost complete reabsorption of nutritionally important substances, such as glucose, amino acids, lactate, and water-soluble vitamins. Electrolytes, such as sodium, potassium, chloride, and bicarbonate, are 65% to 80% reabsorbed. As these solutes move into the tubular cells, their concentration in the tubular lumen decreases, providing a concentration gradient for the osmotic reabsorption of water and urea. The proximal tubule is highly permeable to water, and the osmotic movement of water occurs so rapidly that the concentration difference of solutes on either side of the membrane seldom is more than a few milliosmoles.

Many substances, such as glucose, are freely filtered in the glomerulus and reabsorbed by energy-dependent cotransport carrier mechanisms. The maximum amount of substance that these transport systems can reabsorb per unit time is called the transport maximum. The transport maximum is related to the number of carrier proteins that are available for transport and usually is sufficient to ensure that all of a filtered substance such as glucose can be reabsorbed, rather than being eliminated in the urine. The plasma level at which the substance appears in the urine is called the *renal threshold* (Fig. 22-8). Under some circumstances, the amount of substance filtered in the glomerulus exceeds the transport maximum. For example, when the blood glucose level is elevated in uncontrolled diabetes mellitus, the amount that is filtered in the glomerulus often exceeds the transport maximum (approximately 320 mg/minute), and glucose spills into the urine.

The Loop of Henle. The loop of Henle is divided into three segments: the thin descending segment, the thin ascending segment, and thick ascending segment. The loop of Henle, taken

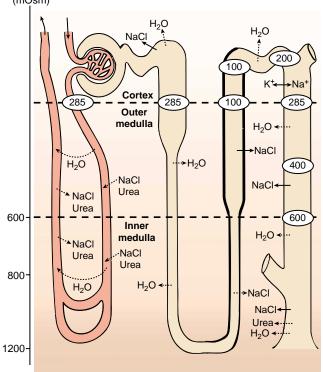


as whole, always reabsorbs more sodium and chloride than water. This is in contrast to the proximal tubule, which reabsorbs sodium and water in equal proportions.

The thin descending limb is highly permeable to water and moderately permeable to urea, sodium, and other ions. As the urine filtrate moves down the descending limb, water moves out of the filtrate into the surrounding interstitium (Fig. 22-9). Thus, the osmolality of the filtrate reaches its highest point at the elbow of the loop of Henle. In contrast to the descending limb, the ascending limb of the loop of Henle is impermeable to water. In this segment, solutes are reabsorbed, but water cannot follow; as a result, the tubular filtrate becomes more and more dilute, often reaching an osmolality of 100 mOsm/kg of H₂O as it enters the distal convoluted tubule, compared with the 285 mOsm/kg of H₂O in plasma. This allows for excretion of free water from the body. For this reason, it is often called the *diluting segment*.

The thick segment of the loop of Henle begins in the ascending limb where the epithelial cells become thickened. The beginning of the thick ascending limb marks the border between the outer and inner medulla; thus, the thick ascending limb is found only in the cortex and outer medulla. As with the thin ascending limb, this segment is relatively impermeable to





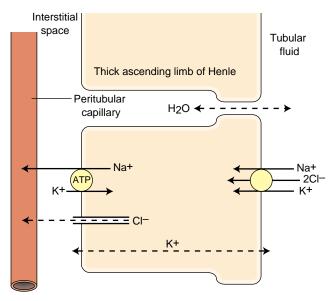
■ FIGURE 22-9 ■ Summary of movements of ions, urea, and water in the kidney during production of a maximally concentrated urine (1200 mOsm/kg H₂O). *Solid arrows* indicate active transport; *dashed arrows* indicate passive transport. The heavy outlining along the ascending limb of Henle's loop indicates decreased water permeability in that tubule segment. Note the osmotic gradient in the medulla from the outer to the inner medulla. (Modified from Rhoades R.A., Tanner G.A. [1996]. *Medical physiology* [p. 441]. Boston: Little, Brown)

408 Unit Six: Alterations in the Urinary System

water. The epithelium of the thick segment has a Na⁺-K⁺-2Cl⁻ cotransport system that moves these ions out of the urine filtrate into the surrounding interstitium and peritubular capillaries (Fig. 22-10). Approximately 20% to 25% of the filtered load of sodium, potassium, and chloride is reabsorbed in the thick loop of Henle. This transport system is selectively blocked by diuretic agents known as *loop diuretics*. The Na⁺-K⁺-2Cl⁻ cotransport system also provides the electrochemical gradient needed for the passive reabsorption of the divalent magnesium and calcium ions. Thus, the inhibition of sodium transport in the thick loop of Henle by the loop diuretics causes an increase in the urinary excretion of these divalent ions in addition to sodium chloride.

In approximately one fifth of the juxtamedullary nephrons, the loops of Henle and special hairpin-shaped capillaries called the *vasa recta* descend into the medullary portion of the kidney. A countercurrent mechanism controls water and solute movement so that water is kept out of the area surrounding the tubule and sodium and urea are retained (see Fig. 22-9). The term countercurrent refers to a flow of fluids in opposite directions in adjacent structures. In this case, there is an exchange of solutes between the adjacent descending and ascending loops of Henle and between the ascending and descending sections of the vasa recta. Because of these exchange processes, a high concentration of osmotically active particles (approximately 1200 mOsm/kg of H_2O) collects in the interstitium of the kidney medulla. The presence of these osmotically active particles in the interstitium surrounding the medullary collecting tubules facilitates the antidiuretic hormone (ADH)-mediated reabsorption of water.

Distal Convoluted Tubule. Like the thick ascending loop of Henle, the distal convoluted tubule is relatively impermeable to water, and reabsorption of sodium chloride from this segment further dilutes the tubular fluid. Sodium reabsorption occurs through a sodium and chloride cotransport mechanism. Approximately 10% of filtered sodium chloride is reabsorbed



■ **FIGURE 22-10** Sodium, chloride, and potassium reabsorption in the thick segment of the loop of Henle.

in this section of the tubule. Unlike the thick ascending loop of Henle, neither calcium nor magnesium is passively absorbed in this segment of the tubule. Instead, calcium ions are actively reabsorbed in a process that is largely regulated by parathyroid hormone and possibly by vitamin D. The thiazide diuretics exert their action by inhibiting sodium chloride reabsorption in this segment of the renal tubules.

Distal Tubule and Cortical Collecting Tubule. The late distal tubule and the cortical collecting tubule constitute the site where aldosterone exerts its action on sodium and potassium reabsorption. Although responsible for only 2% to 5% of so-dium chloride reabsorption, this site is largely responsible for determining the final sodium concentration of the urine. The late distal tubule with the cortical collecting tubule also is the major site for regulation of potassium excretion by the kidney. When the body is confronted with a potassium excess, as occurs with a diet high in potassium content, the amount of potassium secreted at this site may exceed the amount filtered in the glomerulus.

The mechanism for sodium reabsorption and potassium secretion by this section of the kidney is distinct from other tubular segments. This tubular segment is composed of two types of cells, the *intercalated cells*, where potassium is reabsorbed and hydrogen is secreted, and the *principal cells*, where aldosterone exerts its action. The secretion of hydrogen ions into the tubular fluid by the intercalated cells is accompanied by the reabsorption of bicarbonate ions. The intercalated cells can also reabsorb potassium ions. The principal cells reabsorb sodium and facilitate the movement of potassium into the urine filtrate. Under the influence of aldosterone, sodium moves from the urine filtrate into principal cells; from there it moves into the surrounding interstitial fluid and peritubular capillaries. Potassium moves from the peritubular capillaries into the principal cells and then into the urine filtrate.

Medullary Collecting Duct. The epithelium of the inner medullary collecting duct is well designed to resist extreme changes in the osmotic or pH characteristics of tubular fluid, and it is here that the urine becomes highly concentrated, highly diluted, highly alkaline, or highly acidic. During periods of water excess or dehydration, the kidneys play a major role in maintaining water balance.

ADH exerts its effect in the medullary collecting ducts. ADH maintains extracellular volume by returning water to the vascular compartment and leads to the production of a concentrated urine by removing water from the tubular filtrate. Osmoreceptors in the hypothalamus sense the increase in osmolality of extracellular fluids and stimulate the release of ADH from the posterior pituitary gland (see Chapter 6). The permeability of the collecting ducts to water is determined mainly by the concentration of ADH. In exerting its effect, ADH, also known as vasopressin, binds to vasopressin receptors on the blood side of the tubular cells. Binding of ADH to the vasopressin receptors leads to the opening of water channels on the luminal side of the tubular cells, producing a marked increase in water permeability. After the permeability of the collecting tubules has been established, water moves out of the tubular lumen and into the hyperosmotic interstitium of the medullary area, where it enters the peritubular capillaries for return to the vascular system. In the absence of ADH, the water channels are closed, the tubular cells lose their water permeability, and a dilute urine is formed.

Regulation of Renal Blood Flow

In the adult, the kidneys are perfused with 1000 to 1300 mL of blood per minute, or 20% to 25% of the cardiac output. This large blood flow is mainly needed to ensure a sufficient GFR for the removal of waste products from the blood, rather than for the metabolic needs of the kidney. Feedback mechanisms intrinsic to the kidney normally keep blood flow and GFR constant despite changes in arterial blood pressure.

Neural and Humoral Control Mechanisms

The kidney is richly innervated by the sympathetic nervous system. Increased sympathetic activity causes constriction of the afferent and efferent arterioles and thus a decrease in renal blood flow. Intense sympathetic stimulation such as occurs in shock and trauma can produce marked decreases in renal blood flow and GFR, even to the extent of causing blood flow to cease altogether.

Several humoral substances, including angiotensin II, ADH, and endothelins, produce vasoconstriction of renal vessels. The endothelins are a group of peptides released from damaged endothelial cells in the kidney and other tissues. Although not thought to be important regulators of renal blood flow during everyday activities, endothelin I may play a role in reduction of blood flow in conditions such as postischemic acute renal failure (see Chapter 24).

Other substances such as dopamine, nitric oxide, and prostaglandins (*i.e.*, E_2 and I_2) produce vasodilation. Nitric oxide, a vasodilator produced by the vascular endothelium, appears to be important in preventing excessive vasoconstriction of renal blood vessels and allowing normal excretion of sodium and water. Prostaglandins are a group of mediators of cell function that are produced locally and exert their effects locally. Although prostaglandins do not appear to be of major importance in regulating renal blood flow and GFR under normal conditions, they may protect the kidneys against the vasoconstricting effects of sympathetic stimulation and angiotensin II. Salicylates and the nonsteroidal antiinflammatory drugs that inhibit prostaglandin synthesis may cause reduction in renal blood flow and GFR under certain conditions.

Autoregulation

The constancy of renal blood flow is maintained by a process called autoregulation (see Chapter 14). Normally, autoregulation of blood flow is designed to maintain blood flow at a level consistent with the metabolic needs of the tissues. In the kidney, autoregulation of blood flow also must allow for precise regulation of water and solute secretion. For autoregulation to occur, the resistance to blood flow through the kidneys must be varied in direct proportion to the arterial pressure. The exact mechanisms responsible for the intrarenal regulation of blood flow are unclear. One of the proposed mechanisms is a direct effect on vascular smooth muscle that causes the blood vessels to relax when there is an increase in blood pressure, and to constrict when there is a decrease in pressure. A second proposed mechanism is the feedback regulation exerted by the juxtaglomerular complex.

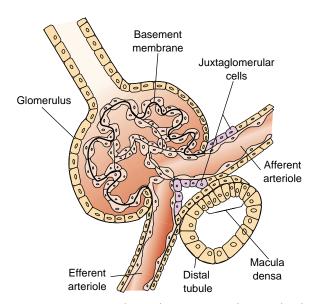
The Juxtaglomerular Complex. The juxtaglomerular complex is thought to represent a feedback control system that links changes in the GFR with renal blood flow. The juxtaglomerular complex is located at the site where the distal tubule extends back to the glomerulus and then passes between the afferent and efferent arteriole (Fig. 22-11). The distal tubular site that is nearest the glomerulus is characterized by densely nucleated cells called the macula densa.

In the adjacent afferent arteriole, the smooth muscle cells of the media are modified as special secretory cells called juxtaglomerular cells. These cells contain granules of inactive renin, an enzyme that functions in the conversion of angiotensinogen to angiotensin. Renin functions by means of angiotensin II to produce vasoconstriction of the efferent arteriole as a means of preventing serious decreases in the glomerular filtration rate (see Chapter 16). Angiotensin II also increases sodium reabsorption indirectly by stimulating aldosterone secretion from the adrenal gland and directly by increasing sodium reabsorption by the proximal tubule cells.

Because of its location between the afferent and efferent arteriole, the juxtaglomerular complex is thought to play an essential feedback role in linking the level of arterial blood pressure and renal blood flow to the GFR and the composition of the distal tubular fluid. The juxtaglomerular complex monitors the systemic blood pressure by sensing the stretch of the afferent arteriole, and it monitors the concentration of sodium chloride in the tubular filtrate as it passes through the macula densa. This information is then used in determining how much renin should be released to keep the arterial blood pressure in its normal range and maintain a relatively constant GFR.

Elimination Functions of the Kidney

The functions of the kidney focus on elimination of water, waste products, excess electrolytes, and unwanted substances from the blood. Blood tests can provide valuable information



■ FIGURE 22-11 ■ Juxtaglomerular apparatus, showing the close contact of the distal tubule with the afferent arteriole, the macula densa, and the juxtaglomerular cells.

about the kidneys' ability to remove metabolic wastes from the blood and maintain normal electrolyte and pH composition of the blood. As renal function declines, serum levels of substances such urea, creatinine, phosphate, and potassium increase. The effect of renal failure on the concentration of serum electrolytes and metabolic end products is discussed in Chapter 24.

Renal Clearance

Renal clearance is the volume of plasma that is completely cleared each minute of any substance that finds its way into the urine. It is determined by the ability of the substance to be filtered in the glomeruli and the capacity of the renal tubules to reabsorb or secrete the substance. Every substance has its own clearance rate, the units of which are expressed as the volume of plasma that is cleared per unit time. It can be determined by measuring the amount of a substance that is excreted in the urine (*i.e.*, urine concentration × urine flow rate in milliliters per minute) and dividing by its plasma concentration. Inulin, a large polysaccharide, is freely filtered in the glomeruli and neither reabsorbed nor secreted by the tubular cells. After intravenous injection, the amount that appears in the urine is equal to the amount that is filtered in the glomeruli (i.e., the clearance rate is equal to the GFR). Because of these properties, inulin can be used as a laboratory measure of the GFR.

Creatinine is a product of creatine metabolism in muscles; its formation and release are relatively constant and proportional to the amount of muscle mass present. Creatinine is freely filtered in the glomeruli, is not reabsorbed from the tubules into the blood, and is only minimally secreted into the tubules from the blood; therefore, its serum values depend closely on the GFR. Serum creatinine levels are often used as a measure of renal function. A normal serum creatinine level usually indicates normal renal function.

Some substances, such as urea, are freely filtered in the glomeruli, but the volume that is cleared from the plasma is less than the GFR, indicating that at least some of the substance is being reabsorbed. At normal plasma levels, glucose has a clearance of zero because it is reabsorbed in the tubules and none appears in the urine.

Regulation of Sodium and Potassium Elimination

Elimination of sodium and potassium is regulated by the GFR and by humoral agents that control reabsorption. Aldosterone functions in the regulation of sodium and potassium elimination. Atrial natriuretic peptide (ANP) contributes to the regulation of sodium elimination.

Aldosterone. Sodium reabsorption in the distal tubule and collecting duct is highly variable and depends on the presence of aldosterone, a hormone secreted by the adrenal gland. In the presence of aldosterone, almost all the sodium in the distal tubular fluid is reabsorbed, and the urine essentially becomes sodium free. In the absence of aldosterone, virtually no sodium is reabsorbed from the distal tubule. The remarkable ability of the distal tubular and collecting duct cells to alter sodium reabsorption in relation to changes in aldosterone allows the kidneys to excrete urine with sodium levels that range from a few tenths of a gram to 40 g per day. Like sodium, potassium is reabsorbed from and secreted into the tubular fluid. The se-

cretion of potassium into the tubular fluid occurs in the distal tubule and, like that of sodium, is regulated by aldosterone. Only approximately 70 mEq of potassium is delivered to the distal tubule each day, but the average person consumes this much and more potassium in the diet. Excess potassium that is not filtered in the glomerulus and delivered to the collecting tubule therefore must be secreted (*i.e.*, transported from the blood) into the tubular fluid for elimination from the body. In the absence of aldosterone (as in Addison's disease; see Chapter 31), potassium secretion becomes minimal. In these circumstances, potassium reabsorption exceeds secretion, and blood levels of potassium increase.

Atrial Natriuretic Peptide. Atrial natriuretic peptide, discovered in 1981, is a hormone believed to have an important role in salt and water excretion by the kidney. It is synthesized in muscle cells of the atria of the heart and released when the atria are stretched. The actions of ANP include vasodilation of the afferent and efferent arterioles, which results in an increase in renal blood flow and glomerular filtration rate. ANP also inhibits sodium reabsorption from the collecting tubules through its inhibition of aldosterone secretion and through direct action on the tubular cells. It also inhibits ADH release from the posterior pituitary gland, thereby increasing excretion of water by the kidneys. ANP also has vasodilator properties. Whether these effects are sufficient to produce long-term changes in blood pressure is uncertain.

Regulation of pH

The kidneys regulate body pH by conserving base bicarbonate (HCO_3^{-}) and eliminating hydrogen ions (H^+) . Neither the blood buffer systems nor the respiratory control mechanisms for carbon dioxide elimination can eliminate H⁺ from the body. This is accomplished by the kidneys. The average North American diet results in the liberation of 40 to 80 mmol of H+ each day. Virtually all the H⁺ excreted in the urine is secreted into the tubular fluid by means of tubular secretory mechanisms. The lowest tubular fluid pH that can be achieved is 4.4 to 4.5. The ability of the kidneys to excrete H⁺ depends on buffers in the urine that combine with the H⁺. The three major urine buffers are HCO_3^- , phosphate (HPO₄⁻), and ammonia (NH_3) . The HCO₃⁻ ions, which are present in the urine filtrate, combine with H⁺ ions that have been secreted into the tubular fluid; this results in the formation of carbon dioxide and water. The carbon dioxide is then absorbed into the tubular cells, and bicarbonate is regenerated. The HPO_4^- ion is a metabolic end product that is filtered into the tubular fluid; it combines with a secreted H⁺ ion and is not reabsorbed. Ammonia is synthesized in tubular cells by deamination of the amino acid glutamine; it diffuses into the tubular fluid and combines with the H⁺ ion. An important aspect of this buffer system is that the deamination process increases when the body's H⁺ ion concentration remains elevated for 1 to 2 days. These mechanisms for pH regulation are described more fully in Chapter 6.

pH-Dependent Elimination of Organic Ions

The proximal tubule actively secretes large amounts of different organic anions. Foreign anions (*e.g.*, salicylates, penicillin) and endogenously produced anions (*e.g.*, bile acids, uric acid) are actively secreted into the tubular fluid. Most of the anions that are secreted use the same transport system, allowing the kidneys to rid the body of many different drugs and environmental agents. Because the same transport system is shared by different anions, there is competition for transport such that elevated levels of one substance tend to inhibit the secretion of other anions. The proximal tubules also possess an active transport system for organic cations that is analogous to that for organic ions.

Uric Acid Elimination

Uric acid is a product of purine metabolism (see Chapter 43). Excessively high blood levels (*i.e.*, hyperuricemia) can cause gout, and excessive levels in the urine can cause kidney stones. Uric acid is freely filtered in the glomerulus and is reabsorbed and secreted into the proximal tubules. Uric acid is one of the anions that uses the previously described anion transport system in the proximal tubule. Tubular reabsorption normally exceeds secretion, and the net effect is removal of uric acid from the filtrate. Although the rate of reabsorption exceeds secretion, the secretory process is homeostatically controlled to maintain a constant plasma level. Many persons with elevated uric acid levels.

Uric acid uses the same transport systems as other anions, such as aspirin, sulfinpyrazone, and probenecid. Small doses of aspirin compete with uric acid for secretion into the tubular fluid and reduce uric acid secretion, and large doses compete with uric acid for reabsorption and increase uric acid excretion in the urine. Because of its effect on uric acid secretion, aspirin is not recommended for treatment of gouty arthritis. Thiazide and loop diuretics (*i.e.*, furosemide and ethacrynic acid) also can cause hyperuricemia and gouty arthritis, presumably through a decrease in extracellular fluid volume and enhanced uric acid reabsorption.

Urea Elimination

Urea is an end product of protein metabolism. The normal adult produces 25 to 30 g/day; the quantity rises when a highprotein diet is consumed, when there is excessive tissue breakdown, or in the presence of gastrointestinal bleeding. With gastrointestinal bleeding, the blood proteins are broken down to form ammonia in the intestine; the ammonia is then absorbed into the portal circulation and converted to urea by the liver before being released into the bloodstream. The kidneys, in their role as regulators of blood urea nitrogen (BUN) levels, filter urea in the glomeruli and then reabsorb it in the tubules. This enables maintenance of a normal BUN, which is in the range of 8 to 20 mg/dL. During periods of dehydration, the blood volume and GFR drop, and BUN levels increase. The renal tubules are permeable to urea, which means that the longer the tubular fluid remains in the kidneys, the greater is the reabsorption of urea into the blood. Only small amounts of urea are reabsorbed into the blood when the GFR is high, but relatively large amounts of urea are returned to the blood when the GFR is reduced.

Drug Elimination

Many drugs are eliminated in the urine. These drugs are selectively filtered in the glomerulus and reabsorbed or secreted into the tubular fluid. Only drugs that are not bound to plasma proteins are filtered in the glomerulus and therefore able to be eliminated by the kidneys.

Many drugs are weak acids or weak bases and are present in the renal tubular fluid partly as water-soluble ions and partly as nonionized lipid-soluble molecules. The nonionized lipid-soluble form of a drug diffuses more readily through the lipid membrane of the tubule and then back into the bloodstream. The water-soluble ionized form remains in the urine filtrate. The ratio of ionized to nonionized drug depends on the pH of the urine. For example, aspirin is highly ionized in alkaline urine and in this form is rapidly excreted in the urine. Aspirin is largely nonionized in acid urine and is reabsorbed, rather than excreted. Alkaline or acid diuresis may be used to increase elimination of drugs in the urine, particularly in situations of drug overdose.

Endocrine Functions of the Kidney

In addition to their function in regulating body fluids and electrolytes, the kidneys function as an endocrine organ in that they produce chemical mediators that travel through the blood to distant sites where they exert their actions. The kidneys participate in control of blood pressure by way of the renin-angiotensin mechanism, in calcium metabolism by activating vitamin D, and in regulating red blood cell production through the synthesis of erythropoietin.

The Renin-Angiotensin-Aldosterone Mechanism

The renin-angiotensin-aldosterone mechanism plays an important part in the short-term and long-term regulation of blood pressure (see Chapter 16). Renin is synthesized and stored in the juxtaglomerular cells of the kidney. This enzyme is released in response to a decrease in renal blood flow or a change in the composition of the distal tubular fluid, or as the result of sympathetic nervous system stimulation. Renin itself has no direct effect on blood pressure. Rather, it acts enzymatically to convert a circulating plasma protein called angiotensinogen to angiotensin I. Angiotensin I, which has few vasoconstrictor properties, leaves the kidneys and enters the circulation; as it is circulated through the lungs, *angiotensinconverting enzyme* catalyzes the conversion of angiotensin I to angiotensin II. Angiotensin II is a potent vasoconstrictor, and it acts directly on the kidneys to decrease salt and water

KEY CONCEPTS

FUNCTIONS OF THE KIDNEY

- Long-term regulation of blood pressure is facilitated through the kidney's activation of the reninangiotensin system and regulation of sodium and water balance.
- The activation of vitamin D, which is important for intestinal absorption of calcium, occurs in the kidney.
- The kidney synthesizes erythropoietin, which stimulates bone marrow production of red blood cells.

excretion. Both mechanisms have relatively short periods of action. Angiotensin II also stimulates aldosterone secretion by the adrenal gland. Aldosterone acts on the distal tubule to increase sodium reabsorption and exerts a longer-term effect on the maintenance of blood pressure. Renin also functions by means of angiotensin II to produce constriction of the efferent arteriole as a means of preventing a serious decrease in glomerular filtration pressure.

Erythropoietin

Erythropoietin is a polypeptide hormone that regulates the differentiation of red blood cells in the bone marrow (see Chapter 13). Between 89% and 95% of erythropoietin is formed in the kidneys. The synthesis of erythropoietin is stimulated by tissue hypoxia, which may be brought about by anemia, residence at high altitudes, or impaired oxygenation of tissues caused by cardiac or pulmonary disease. Persons with endstage kidney disease often are anemic because of an inability of the kidneys to produce erythropoietin. This anemia usually is managed by the administration of epoetin-alfa, a synthetic form of erythropoietin produced through DNA technology, to stimulate erythropoiesis.

Vitamin D

Activation of vitamin D occurs in the kidneys. Vitamin D increases calcium absorption from the gastrointestinal tract and helps to regulate calcium deposition in bone. It also has a weak stimulatory effect on renal calcium absorption. Although vitamin D is not synthesized and released from an endocrine gland, it often is considered as a hormone because of its pathway of molecular activation and mechanism of action.

It exists in several forms: natural vitamin D (cholecalciferol), which results from ultraviolet irradiation of the skin, and synthetic vitamin D (ergocalciferol), which is derived from irradiation of ergosterol. The active form of vitamin D is 1,25-dihydroxycholecalciferol. Cholecalciferol and ergocalciferol must undergo chemical transformation to become active: first to 25-hydroxycholecalciferol in the liver and then to 1,25-dihydroxycholecalciferol in the kidneys. Persons with end-stage renal disease are unable to transform vitamin D to its active form and must rely on pharmacologic preparations of the active vitamin (calcitriol) for maintaining mineralization of their bones.

In summary, the kidneys perform excretory and endocrine functions. In the process of excreting wastes, the kidneys filter the blood and then selectively reabsorb those materials that are needed to maintain a stable internal environment. The kidneys rid the body of metabolic wastes, regulate fluid volume, regulate the concentration of electrolytes, assist in maintaining acid-base balance, aid in the regulation of blood pressure through the renin-angiotensinaldosterone mechanism and control of extracellular fluid volume, regulate red blood cell production through erythropoietin, and aid in calcium metabolism by activating vitamin D.

The kidneys selectively eliminate water, waste products, excess electrolytes, and other substances that are not needed to maintain the constancy of the internal environment. Renal clearance is the volume of plasma that is completely cleared each minute of any substance that finds its way into the urine. It is determined by the ability of the substance to be filtered in the glomeruli and the capacity of the renal tubules to reabsorb or secrete the substance. The GFR is the amount of filtrate that is formed each minute as blood moves through the glomeruli. It is regulated by the arterial blood pressure and renal blood flow in the normally functioning kidney. The juxtaglomerular complex is thought to represent a feedback control system that links changes in the GFR with renal blood flow.

In addition to their function in regulating body fluids and electrolytes, the kidneys function as an endocrine organ in that they produce chemical mediators that travel through the blood to distant sites where they exert their actions. The kidneys participate in control of blood pressure by way of the renin-angiotensin mechanism, in calcium metabolism by activating vitamin D, and in regulating red blood cell production through the synthesis of erythropoietin.

TESTS OF RENAL FUNCTION

The functions of the kidney are to filter the blood and selectively reabsorb those substances that are needed to maintain the constancy of body fluids and excrete metabolic wastes. Laboratory tests of the urine and blood can provide valuable information about kidney pathology and the adequacy of renal function.

Urine Tests

Urine is a clear, amber-colored fluid that is approximately 95% water and 5% dissolved solids. The kidneys normally produce approximately 1.5 L of urine each day. Normal urine contains metabolic wastes and few or no plasma proteins, blood cells, or glucose molecules.

Urine tests can be performed on a single urine specimen or on a 24-hour urine specimen. First-voided morning specimens are useful for qualitative protein and specific gravity testing. A freshly voided specimen is most reliable. Urine specimens that have been left standing may contain lysed red blood cells, disintegrating *casts*, and rapidly multiplying bacteria.

Casts are molds of the distal nephron lumen. A gel-like substance called *Tamm-Horsfall mucoprotein*, which is formed in the tubular epithelium, is the major protein constituent of urinary casts. Casts composed of this gel but devoid of cells are called *hyaline casts*. These casts develop when the protein concentration of the urine is high (as in nephrotic syndrome), urine osmolality is high, and urine pH is low. The inclusion of granules or cells in the matrix of the protein gel leads to the formation of various other types of casts.

Because of the glomerular capillary filtration barrier, less than 150 mg of protein is excreted in the urine during 24 hours in a healthy person. Qualitative and quantitative tests to determine urinary protein content are important tools to assess the extent of glomerular disease. pH-sensitive reagent strips are used to test for the presence of proteins, whereas immunoassay methods are used to test for microalbuminuria (30 to 300 mg albumin/24 hours).

The specific gravity (or osmolality) of urine varies with its concentration of solutes. Urine specific gravity provides a valuable index of the hydration status and functional ability of the kidneys. Healthy kidneys can produce a concentrated urine with a specific gravity of 1.030 to 1.040. During periods of marked hydration, the specific gravity can approach 1.000. With the loss of nephrons and diminished renal function, there is a loss of renal concentrating ability, and the urine specific gravity may fall to levels of 1.006 to 1.010 (usual range is 1.010 to 1.025 with normal fluid intake). These low levels are particularly significant if they occur during periods that follow a decrease in water intake (e.g., during the first urine specimen on arising in the morning). The ability to concentrate urine also depends on the availability of and renal response to ADH. The urine specific gravity is decreased when ADH levels are decreased, such as in diabetes insipidus, and it is increased when ADH levels are inappropriately elevated, such as in the syndrome of inappropriate ADH.

Glomerular Filtration Rate

The GFR provides a gauge of renal function. It can be measured clinically by collecting timed samples of blood and urine. *Creatinine*, a product of creatine metabolism by the muscle, is filtered by the kidneys but not reabsorbed in the renal tubule. Creatinine levels in the blood and urine can be used to measure GFR. The clearance rate for creatinine is the amount that is completely cleared by the kidneys in 1 minute. The formula is expressed as C = UV/P, in which C is the clearance rate (mL/minute), U is the urine concentration (mg/dL), V is the urine volume excreted (mL/minute or 24 hours), and P is plasma concentration (mg/dL).

Normal creatinine clearance is 115 to 125 mL/minute. This value is corrected for body surface area, which reflects the muscle mass where creatinine metabolism takes place. The test may be done on a 24-hour basis, with blood being drawn when the urine collection is completed. In another method, two 1-hour urine specimens are collected, and a blood sample is drawn in between.

Blood Tests

Blood tests can provide valuable information about the kidneys' ability to remove metabolic wastes from the blood and maintain normal electrolyte and pH composition of the blood. Normal blood values are listed in Table 22-1. Serum levels of potassium, phosphate, BUN, and creatinine increase in renal failure. Serum pH, calcium, and bicarbonate levels decrease in renal failure. The effect of renal failure on the concentration of serum electrolytes and metabolic end products is discussed in Chapter 24.

Serum Creatinine

Serum creatinine levels reflect the glomerular filtration rate. Because these measurements are easily obtained and relatively inexpensive, they often are used as a screening measure of renal function. Creatinine is a product of creatine metabolism in muscles; its formation and release are relatively constant and proportional to the amount of muscle mass present. Creatinine is freely filtered in the glomeruli, is not reabsorbed from the tubules into the blood, and is only minimally secreted into the

TABLE 22-1	Normal Blood			
Chemistry Levels				
Substance	Normal Value*			
Blood urea nitroge Creatinine	en 8.0–20.0 mg/dL (2.9–7.1 mmol/L) 0.6–1.2 mg/dL (50–100 μmol/L)			
Sodium	135–145 mEq/L (135–148 mmol/L)			
Potassium	3.5–5 mEq/L (3.5–5 mmol/L)			
(CO ₂ content)				
Phosphate	2.5–4.5 mg/dL (0.77–1.45 mmol/L)			
Uric acid pH	1.4–7.4 mg/dL (0.154–0.42 mmol/L) 7.35–7.45			
Sodium Chloride Potassium Carbon dioxide (CO ₂ content) Calcium Phosphate Uric acid	135–145 mEq/L (135–148 mmol/L) 98–106 mEq/L (98–106 mmol/L) 3.5–5 mEq/L (3.5–5 mmol/L) 24–29 mEq/L (24–29 mmol/L) 8.5–10.5 mg/dL (2.1–2.6 mmol/L) 2.5–4.5 mg/dL (0.77–1.45 mmol/L) 1.4–7.4 mg/dL (0.154–0.42 mmol/L)			

*Values may vary among laboratories, depending on the method of analysis used.

tubules from the blood; therefore, its blood values depend closely on the GFR.

The normal creatinine value is approximately 0.6 mg/dL of blood for a woman with a small frame, approximately 1.0 mg/dL of blood for a normal adult man, and approximately 1.2 mg/dL of blood for a muscular man. There is an age-related decline in creatinine clearance in many elderly persons because muscle mass and the GFR decline with age (see Chapter 24). A normal serum creatinine level usually indicates normal renal function. In addition to its use in calculating the GFR, the serum creatinine level is used in estimating the functional capacity of the kidneys (Fig. 22-12). If the serum creatinine value doubles, the GFR—and renal function—probably has fallen to one half of its normal state. A rise in the serum creatinine level to three times its normal value suggests that there is a 75% loss of renal function, and with creatinine values of 10 mg/dL or more, it can be assumed that approximately 90% of renal function has been lost.

Blood Urea Nitrogen

Urea is formed in the liver as a by-product of protein metabolism and is eliminated entirely by the kidneys. Therefore,

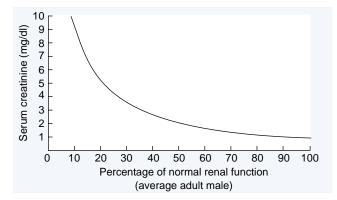


FIGURE 22-12 Relation between the percentage of renal function and serum creatinine levels.

414 Unit Six: Alterations in the Urinary System

BUN is related to the GFR but, unlike creatinine, also is influenced by protein intake, gastrointestinal bleeding, and hydration status. Increased protein intake and gastrointestinal bleeding increase urea by means of protein metabolism. In gastrointestinal bleeding, the blood is broken down by the intestinal flora, and the nitrogenous waste is absorbed into the portal vein and transported to the liver, where it is converted to urea. During dehydration, elevated BUN levels result from increased concentration. Approximately two thirds of renal function must be lost before a significant rise in the BUN level occurs.

The BUN is less specific for renal insufficiency than creatinine, but the *BUN–creatinine ratio* may provide useful diagnostic information. The ratio normally is approximately 10:1. Ratios greater than 15:1 represent prerenal conditions, such as congestive heart failure and upper gastrointestinal tract bleeding, that produce an increase in BUN but not in creatinine. A ratio of less than 10:1 occurs in persons with liver disease and in those who receive a low-protein diet or chronic dialysis because BUN is more readily dialyzable than creatinine.

In summary, urinalysis and blood tests that measure levels of by-products of metabolism and electrolytes provide information about renal function. Serum creatinine reflects the glomerular filtration rate and can be used as an estimate of renal function. Measurements of BUN, which is formed in liver as a by-product of protein metabolism and eliminated almost entirely by the kidney, are also a measure of renal function.

REVIEW QUESTIONS

Describe the location and gross structure of the kidney and explain why kidney injury does not produce peritonitis.

Explain the structure and function of the capillary structures of the nephron (glomerulus and peritubular capillaries) and the tubular components of the nephron in terms of filtering and reabsorbing nutrients, eliminating waste products, and maintaining the acid-base and electrolyte composition of the extracellular fluid.

Describe the kidney blood supply and mechanisms for regulating blood flow.

Use the concept of the countercurrent mechanism and the actions of the antidiuretic hormone to explain how the kidney produces a concentrated or dilute urine.

Characterize the function of the juxtaglomerular complex.

Explain the endocrine functions of the kidney.

Use the concepts of glomerular filtration, tubular reabsorption, and tubular secretion to explain why serum creatinine is a better indicator of the glomerular filtration rate than blood urea nitrogen.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

BIBLIOGRAPHY

- Cormack D.H. (1993). Essential histology (pp. 322–333). Philadelphia: J.B. Lippincott.
- Guyton A.C., Hall J.E. (2000). Textbook of medical physiology (10th ed., pp. 279–311). Philadelphia: W.B. Saunders.
- Koeppen B.M., Stanton B.A. (1997). Renal physiology (2nd ed.). St. Louis: Mosby.
- Price C.P., Finney H. (2000). Developments in the assessment of glomerular filtration rate. *Clinica Chimica Acta* 297, 55–66.

Rahn K.H., Heidenreich S., Bruckner D. (1999). How to assess glomerular function and damage in humans. *Journal of Hypertension* 17, 309–317.

- Rhoades R.A., Tanner G.A. (1996). *Medical physiology* (pp. 417–445). Boston: Little, Brown.
- Smith H. (1953). From fish to philosopher (p. 4). Boston: Little, Brown.
- Vander A.J. (1995). Renal physiology (5th ed.). New York: McGraw-Hill.

Alterations in Renal Function

Congenital and Hereditary Disorders of the Kidney Agenesis and Hypoplasia **Disorders of Kidney Position and Form Polycystic Kidney Disease Obstructive Disorders Renal Calculi Types of Stones Manifestations Diagnosis and Treatment Hydronephrosis** Urinary Tract Infections **Etiologic Factors Host Defenses Pathogen Virulence Obstruction and Reflux Catheter-Induced Infection Manifestations Diagnosis and Treatment Infections in Special Populations** Urinary Tract Infections in Women Urinary Tract Infections in Children Urinary Tract Infections in the Elderly **Disorders of Glomerular Function** Mechanisms of Glomerular Injury Nephritic Syndromes Acute Proliferative Glomerulonephritis **Rapidly Progressive Glomerulonephritis** Nephrotic Syndrome Membranous Glomerulonephritis Minimal Change Disease (Lipoid Nephrosis) Focal Segmental Glomerulosclerosis Chronic Glomerulonephritis **Glomerular Lesions Associated With Systemic Disease Diabetic Glomerulosclerosis** Hypertensive Glomerular Disease

CHAPTER

2

Tubulointerstitial Disorders Pyelonephritis Drug-Related Nephropathies Neoplasms Wilms' Tumor Renal Cell Carcinoma

ore than 20 million North Americans have diseases of the kidneys and urinary tract. Each year, more than million people receive diagnoses of acute urinary tract disorders, and approximately 50,000 die of these diseases.¹

The kidneys are subject to many of the same types of disorders that affect other body structures, including developmental defects, infections, altered immune responses, and neoplasms. The kidneys filter blood from all parts of the body, and although many forms of kidney disease originate in the kidneys, others develop secondary to disorders such as hypertension, diabetes mellitus, and systemic lupus erythematosus (SLE).

CONGENITAL AND HEREDITARY DISORDERS OF THE KIDNEY

Some abnormality of the kidneys and ureters occurs in approximately 3% to 4% of newborn infants.² Anomalies in shape and position are the most common. Less common are disorders involving a decrease in renal mass (*e.g.*, agenesis, hypogenesis) or a change in renal structure (*e.g.*, renal cysts). Many fetal anomalies can be detected before birth by ultrasonography. In the normal fetus, the kidneys can be visualized as early as 12 weeks.

Agenesis and Hypoplasia

The kidneys begin to develop early in the fifth week of gestation and start to function approximately 3 weeks later. Formation of urine is thought to begin in the 9th to 12th weeks of gestation; by the 32nd week, fetal production of urine reaches approximately 28 mL/hour.³ Urine is the main constituent of amniotic fluid. The relative amount of amniotic fluid can provide information about the status of fetal renal function.

416 Unit Six: Alterations in the Urinary System

The term *dysgenesis* refers to a failure of an organ to develop normally. *Agenesis* is the complete failure of an organ to develop. Total agenesis of both kidneys is incompatible with extrauterine life. Infants are stillborn or die shortly after birth of pulmonary hypoplasia. Newborns with renal agenesis often have characteristic facial features, termed *Potter's syndrome*.⁴ The eyes are widely separated and have epicanthic folds, the ears are low set, the nose is broad and flat, the chin is receding, and limb defects often are present.^{4,5} Other causes of neonatal renal failure with the Potter phenotype include cystic renal dysplasia, obstructive uropathy, and autosomal recessive polycystic disease. Unilateral agenesis is an uncommon anomaly that is compatible with life if no other abnormality is present. The opposite kidney usually is enlarged as a result of compensatory hypertrophy.

In *renal hypoplasia*, the kidneys do not develop to normal size. Like agenesis, hypoplasia more commonly affects only one kidney. When both kidneys are affected, there is progressive development of renal failure. It has been suggested that true hypoplasia is extremely rare; most cases probably represent acquired scarring caused by vascular, infectious, or other kidney diseases, rather than an underlying developmental failure.^{5,6}

In pregnancies that involve infants with nonfunctional kidneys or outflow obstruction of the kidneys, the amount of amniotic fluid is small—a condition called *oligohydramnios*. The cause of fetal death in these infants is thought to be cord compression caused by the oligohydramnios.²

Disorders of Kidney Position and Form

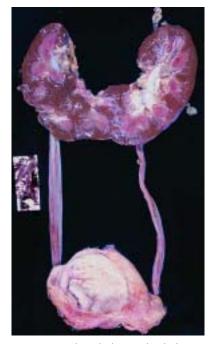
The development of the kidneys during embryonic life can result in kidneys that lie outside their normal position, usually just above the pelvic brim or within the pelvis. Because of the abnormal position, kinking of the ureters and obstruction of urine flow may occur.

One of the most common alterations in kidney form is an abnormality called a *horseshoe kidney*. This abnormality occurs in approximately 1 in every 500 to 1000 persons.^{4,5} In this disorder, the upper or lower poles of the two kidneys are fused, producing a horseshoe-shape structure that is continuous along the midline of the body anterior to the great vessels. Most horseshoe kidneys are fused at the lower pole⁶ (Fig. 23-1). The condition usually does not cause problems unless there is an associated defect in the renal pelvis or other urinary structures that obstruct urine flow.

Polycystic Kidney Disease

Renal cysts are fluid-filled sacs or segments of a dilated nephron. The cysts may be single or multiple and can vary in size from microscopic to several centimeters in diameter. The most common form of renal cystic disease is polycystic kidney disease, which is the result of a hereditary trait. It is one of the most common hereditary diseases in the United States, affecting more than 600,000 Americans.¹ There are two types of inherited polycystic disease: autosomal recessive and autosomal dominant.

Autosomal recessive polycystic kidney disease, which is present at birth, is rare compared with the adult variety.^{5,6} The disorder is inherited as a recessive trait, meaning that both parents are carriers of the gene and that there is a one in four chance of



■ FIGURE 23-1 ■ Horseshoe kidney. The kidneys are fused at the lower poles. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 865]. Philadelphia: Lippincott Williams & Wilkins)

the parents having another child with the disorder. Because the condition is present at birth, it formerly was called *infantile* or *childhood polycystic disease*. The condition is bilateral, and significant renal dysfunction usually is present, accompanied by variable degrees of liver fibrosis and portal hypertension. The disorder can be diagnosed by ultrasonography. There is no known treatment for the disease. Approximately 75% of infants die during the perinatal period, often because the large kidneys compromise expansion of the lungs.⁶ Some children may present with less severe kidney problems and more severe liver disease.

Autosomal dominant polycystic kidney disease, also called adult polycystic kidney disease, affects children and adults in the prime of life and accounts for 10% of persons who require treatment for end-stage renal disease. This disorder is transmitted as an autosomal dominant trait. There is considerable variability in gene expression, and many affected persons do not have clinical symptoms, or if they do, the symptoms occur later in life.

Three mutant genes have been implicated in the disorder.⁵⁻⁷ A polycystic kidney disease gene called PKD1, located on chromosome 16, is responsible for approximately 85% of cases. It encodes a large membrane protein called *polycystin 1* that has domains similar to proteins involved in cell-to-cell and cell-toextracellular matrix interactions.^{5,7} A second gene, called *PKD2*, which is located on chromosome 4, is responsible for a milder form of the disease. It encodes for a product called *polycystin 2*, which is an integral membrane protein that is similar to certain calcium and sodium channel proteins as well as to a portion of polycystin 1. A third gene, *PKD3*, is responsible for a minority of cases and has yet to be mapped.

How the genetic defects in the polycystin proteins cause cyst formation is largely speculative. It is thought that the membrane proteins may play a role in extracellular matrix interactions that are important in tubular epithelial cell growth and differentiation. In addition, cyst fluids have been shown to harbor mediators that enhance fluid secretion and induce inflammation, resulting in further enlargement of the cysts and the interstitial fibrosis that is characteristic of progressive polycystic kidney disease.

The disease is characterized by tubular dilatation with cyst formation interspersed between normally functioning nephrons. Fluid collects in the cyst while it is still part of the tubular lumen, or it is secreted into the cyst after it has separated from the tubule. As the fluid accumulates, the cysts gradually increase in size, with some becoming as large as 5 cm in diameter. The kidneys of persons with polycystic kidney disease eventually become enlarged because of the presence of multiple cysts (Fig. 23-2). Cysts also may be found in the liver and, less commonly, the pancreas and spleen. Mitral valve prolapse and other valvular heart diseases occur in 20% to 25% of persons, but are largely asymptomatic. Most persons with polycystic disease also have colonic diverticula. One of the most devastating extrarenal manifestations is a weakness in the walls of the cerebral arteries that can lead to aneurysm formation. Approximately 20% of persons with polycystic kidney disease have an associated aneurysm, and subarachnoid hemorrhage is a frequent cause of death.6

The manifestations of polycystic kidney disease include pain from the enlarging cysts that may reach debilitating levels, episodes of gross hematuria from bleeding into a cyst, infected cysts from ascending urinary tract infections, and hypertension resulting from compression of intrarenal blood vessels with activation of the renin-angiotensin mechanism.8 Persons with polycystic kidney disease also are at risk for the development of renal cell carcinoma. The progress of the disease is slow, and end-stage renal failure is uncommon before 40 years of age.

The diagnosis of autosomal polycystic kidney disease can be made by radiologic studies, ultrasonography, and computed tomography (CT). Ultrasonography is particularly useful as a screening test for the disease.

The treatment of polycystic kidney disease is largely supportive. Control of hypertension and prevention of ascending urinary tract infections are important. The cysts may be surgically decompressed in persons with severe, disabling pain.⁸ The procedure permits removal of fluid from the cyst. Although affording pain relief, the procedure does not appear to alter the course of the disease.8 Dialysis and kidney transplantation are reserved for those who have end-stage renal disease.

In summary, approximately 10% of infants are born with potentially significant malformations of the urinary system. These abnormalities can range from bilateral renal agenesis, which is incompatible with life, to hypogenesis of one kidney, which usually causes no problems unless the function of the remaining kidney is impaired. The developmental process can result in kidneys that lie outside their normal position. Because of the abnormal position, kinking of the ureters and obstruction of urine flow can occur.

Renal cystic disease is a condition in which there is dilatation of tubular structures with cyst formation. Polycystic kidney disease is an inherited form of renal cystic disease; it can also be inherited as an autosomal recessive or an autosomal dominant trait. Autosomal recessive polycystic kidney disease is rare and usually presents as severe renal dysfunction during infancy. Autosomal dominant polycystic disease usually does not become symptomatic until later in life, often after 40 years of age.



FIGURE 23-2 Adult polycystic disease. The kidney is enlarged. and the parenchyma is almost entirely replaced by cysts of varying size. (Rubin E., Farber J.L. [1999]. Pathology [3rd ed., p. 867]. Philadelphia: Lippincott Williams & Wilkins)

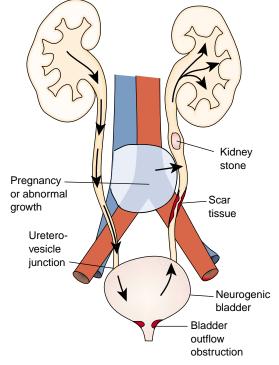
OBSTRUCTIVE DISORDERS

Urinary obstruction can occur in persons of any age and can involve any level of the urinary tract from the urethra to the renal pelvis (Fig. 23-3). The conditions that cause urinary tract obstruction include developmental defects, pregnancy, benign prostatic hyperplasia, tumors, and kidney stones. The causes of urinary tract obstructions are summarized in Table 23-1.

The two most damaging effects of urinary obstruction are stasis of urine, which predisposes to infection and stone formation, and development of backpressure, which interferes with renal blood flow, destroys kidney tissue, and predisposes to hydronephrosis. The destructive effects of urinary obstruction on kidney structures are determined by the degree (*i.e.*, partial vs. complete, unilateral vs. bilateral) and the duration of the obstruction.9

The manifestations of urinary obstruction depend on the site of obstruction, the cause, and the rapidity with which the condition developed. Most commonly, the person has pain, signs and symptoms of urinary tract infection, and manifestations of renal dysfunction, such as an impaired ability to concentrate urine. Changes in urine output may be misleading because output may be normal or even high in cases of partial obstruction.

418 Unit Six: Alterations in the Urinary System



■ FIGURE 23-3 ■ Locations and causes of urinary tract obstruction.

Pain, which often is the factor that causes a person to seek medical attention, is the result of distention of the bladder, collecting system, or renal capsule. Its severity is related most closely to the rate, rather than the degree, of distention. Pain most often occurs with acute obstruction, in which the distention of urinary structures is rapid. This is in contrast to chronic obstruction, in which distention is gradual and may not cause pain. Instead, gradual obstruction may produce only vague abdominal or back discomfort. When pain occurs, it is related to the site of obstruction. Obstruction of the renal pelvis or upper ureter causes pain and tenderness over the flank area. With

TABLE 23-1Causes of UrinaryTract Obstruction				
Level of Obstruction	Cause			
Renal pelvis	Renal calculi			
	Papillary necrosis			
Ureter	Renal calculi			
	Pregnancy			
	Tumors that compress the ureter			
	Ureteral stricture			
	Congenital disorders of the			
	ureterovesical junction and			
	ureteropelvic junction strictures			
Bladder and urethra	Bladder cancer			
	Neurogenic bladder			
	Bladder stones			
	Prostatic hyperplasia or cancer			
	Urethral strictures			
	Congenital urethral defects			

lower levels of obstruction, the pain may radiate to the testes in the male or the labia in the female. With partial obstruction, particularly of the ureteropelvic junction, pain may occur during periods of high fluid intake, when a high rate of urine flow causes an acute distention of the renal pelvis. Because of its visceral innervation, ureteral obstruction may produce reflex impairment of gastrointestinal tract peristalsis and motility with abdominal distention and, in severe cases, paralytic ileus.

Hypertension is an occasional complication of urinary tract obstruction. It is more common in cases of unilateral obstruction in which renin secretion is enhanced, probably secondary to impaired renal blood flow. In these circumstances, removal of the obstruction often leads to a reduction in blood pressure. When hypertension accompanies bilateral obstruction, renin levels usually are normal, and the elevated blood pressure probably is volume related. The relief of bilateral obstruction leads to a loss of volume and a decrease in blood pressure. In some cases, relieving the obstruction does not correct the hypertension.

Renal Calculi

The most common cause of upper urinary tract obstruction is urinary calculi. Although stones can form in any part of the urinary tract, most develop in the kidneys. Approximately 1 million North Americans are hospitalized each year with kidney stones, and an equal number are treated for stones without hospitalization.¹ Men are more frequently affected than women, with a ratio of $4:1.^{10}$

Kidney stones are crystalline structures made up of materials that the kidneys normally excrete in the urine. The etiology of urinary stone formation is complex, and not all aspects are well understood. It is thought to encompass a number of factors, including increases in blood and urinary levels of stone components and interactions among the components; anatomic changes in urinary tract structures; metabolic and endocrine influences; dietary and intestinal absorption factors; and urinary tract infections. To add to the mystery of stone formation is the fact that although both kidneys are exposed to the same urinary constituents, kidney stones tend to form in only one kidney. Three major theories are used to explain stone formation: the saturation theory, the matrix theory, and the inhibitor deficiency theory.^{11,12} One or more of these theories may apply to stone formation in the same person.

KEY CONCEPTS

KIDNEY STONES

- Kidney stones are crystalline structures that form from components of the urine.
- Stones require a nidus to form and a urinary environment that supports continued crystallization of stone components.
- Stone formation is influenced by the concentration of stone components in the urine, the ability of the stone components to complex and form stones, and the presence of substances that inhibit stone formation.

Kidney stones require a nidus, or nucleus, to form and a urinary environment that supports continued precipitation of stone components to grow. The *saturation theory* states that the risk of stone formation is increased when the urine is supersaturated with stone components (*e.g.*, calcium salts, uric acid, magnesium ammonium phosphate, cystine). Supersaturation depends on urinary pH, solute concentration, ionic strength, and complexation. The greater the concentration of two ions, the more likely they are to precipitate. Complexation influences the availability of specific ions. For example, sodium complexes with oxalate, decreasing its free ionic form and increasing the amount that is available to participate in stone formation.

The *matrix theory* proposes that organic materials, such as mucopolysaccharides derived from the epithelial cells that line the tubules, act as a nidus for stone formation. This theory is based on the observation that organic matrix materials can be found in all layers of kidney stones. It is not known whether the matrix material contributes to the initiation of stone formation or if the material is merely entrapped as the stone forms.

Kidney proteins inhibit all phases of crystallization. The *inhibitor theory* suggests that persons who have a deficiency of proteins that inhibit stone formation in their urine are at increased risk for stone formation. Kidney cells produce at least three proteins that are thought to slow the rate of calcium oxalate crystallization: nephrocalcin, Tamm-Horsfall mucoprotein, and uropontin.^{11,13} Nephrocalcin inhibits nucleation, aggregation, and growth of calcium oxalate stones. *Tamm-Horsfall mucoprotein* is thought to exert a minor effect on crystal aggregation. Uropontin inhibits the growth of calcium oxalate crystals.

Urinary obstruction and stagnation of urine predisposes to infection. When present, urinary calculi serve as foreign bodies and contribute to the infection. Once established, the infection is difficult to treat. It often is caused by urea-splitting organisms (*e.g.*, *Proteus*, staphylococci) that increase ammonia production and cause the urine to become alkaline.⁹ Calcium salts precipitate more readily in stagnant alkaline urine; thus, urinary tract obstructions also predispose to stone formation.

Types of Stones

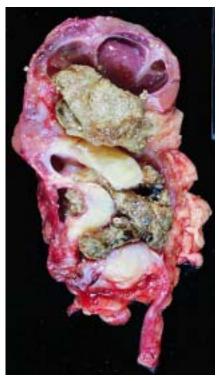
There are four basic types of kidney stones: calcium stones (*i.e.*, oxalate or phosphate), magnesium ammonium phosphate stones, uric acid stones, and cystine stones. The causes and treatment measures for each of these types of renal stones are described in Table 23-2.

Calcium Stones. Most kidney stones (70% to 80%) are calcium stones—calcium oxalate, calcium phosphate, or a combination of the two materials. Calcium stones usually are associated with increased concentrations of calcium in the blood and urine. Excessive bone resorption caused by immobility, bone disease, hyperparathyroidism, and renal tubular acidosis all are contributing conditions. High oxalate concentrations in the blood and urine predispose to formation of calcium oxalate stones. A recent addition to the spectrum of kidney stones are those seen in persons with human immunodeficiency virus (HIV) infection who are being treated with indinavir, a protease inhibitor. The calcium-containing calculi develop in as many as 6% of persons treated with the drug.⁹

Magnesium Ammonium Phosphate Stones. Magnesium ammonium phosphate stones, also called *struvite stones*, form only in alkaline urine and in the presence of bacteria that possess an enzyme called *urease*, which splits the urea in the urine into ammonia and carbon dioxide. The ammonia that is formed takes up a hydrogen ion to become an ammonium ion, increasing the pH of the urine so that it becomes more alkaline. Because phosphate levels are increased in alkaline urine and because magnesium always is present in the urine, struvite stones form. These stones enlarge as the bacterial count grows, and they can increase in size until they fill an entire renal pelvis (Fig. 23-4). Because of their shape, they often are called *staghorn stones*. Staghorn stones almost always are associated with

TABLE 23-2 Composition, Contributing Factors, and Treatment of Kidney Stones			
Type of Stone		Contributing Factors	Treatment
Calcium (oxalate and phosphate)		Hypercalcemia and hypercalciuria Immobilization	Treatment of underlying conditions Increased fluid intake Thiazide diuretics
		Hyperparathyroidism Vitamin D intoxication Diffuse bone disease	
		Milk-alkali syndrome Renal tubular acidosis	Distance sectsistics of foods high is evaluate
		Hyperoxaluria Intestinal bypass surgery	Dietary restriction of foods high in oxalate
Magnesium amm phosphate (stru		Urea-splitting urinary tract infections	Treatment of urinary tract infection Acidification of the urine Increased fluid intake
Uric acid (urate)		Formed in acid urine with pH of approximately 5.5 Gout High-purine diet	Increased fluid intake Allopurinol for hyperuricuria Alkalinization of urine
Cystine		Cystinuria (inherited disorder of amino acid metabolism)	Increased fluid intake Alkalinization of urine

420 Unit Six: Alterations in the Urinary System



■ FIGURE 23-4 ■ Staghorn calculi. The kidney shows hydronephrosis and stones that are casts of the dilated calyces. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 909]. Philadelphia: Lippincott Williams & Wilkins)

urinary tract infections and persistently alkaline urine. Because these stones act as a foreign body, treatment of the infection often is difficult. Struvite stones usually are too large to be passed and require lithotripsy or surgical removal.

Uric Acid Stones. Uric acid stones develop in conditions of gout and high concentrations of uric acid in the urine. Hyperuricosuria also may contribute to calcium stone formation by acting as a nucleus for calcium oxalate stone formation. Unlike radiopaque calcium stones, uric acid stones are not visible on x-ray films. Uric acid stones form most readily in urine with a pH of 5.1 to 5.9.¹¹ Thus, these stones can be treated by raising the urinary pH to 6 to 6.5 with potassium alkali salts.

Cystine Stones. Cystine stones are rare. They are seen in cystinuria, which results from a genetic defect in renal transport of cystine. These stones resemble struvite stones except that infection is unlikely to be present.

Manifestations

One of the major manifestations of kidney stones is pain. Depending on location, there are two types of pain associated with kidney stones: renal colic and noncolicky renal pain.⁹ *Renal colic* is the term used to describe the colicky pain that accompanies stretching of the collecting system or ureter. The symptoms of renal colic are caused by stones 1 to 5 mm in diameter that can move into the ureter and obstruct flow. Classic ureteral colic is manifested by acute, intermittent, and excruciating pain in the flank and upper outer quadrant of the abdomen on the affected side. The pain may radiate to the lower

abdominal quadrant, bladder area, perineum, or scrotum in the male. The skin may be cool and clammy, and nausea and vomiting are common. Noncolicky pain is caused by stones that produce distention of the renal calices or renal pelvis. The pain usually is a dull, deep ache in flank or back that can vary in intensity from mild to severe. The pain may be exaggerated by drinking large amounts of fluid.⁹

Diagnosis and Treatment

Persons with kidney stones often present with acute renal colic, and the diagnosis is based on symptomatology and diagnostic tests, which include urinalysis, abdominal radiographs, and excretory urography. Urinalysis provides information related to hematuria, infection, the presence of stone-forming crystals, and urine pH. At least 90% of stones are radiopaque and readily visible on a plain radiograph of the abdomen. Excretory urography uses an intravenously injected contrast medium that is filtered in the glomeruli to visualize the collecting system and the ureters of the kidneys. Retrograde urography, ultrasonography, and CT scanning also may be used.

Treatment of acute renal colic usually is supportive. Pain relief may be needed during acute phases of obstruction, and antibiotic therapy may be necessary to treat urinary infections. Most stones that are less than 5 mm in diameter pass spontaneously. All urine should be strained during an attack in the hope of retrieving the stone for chemical analysis and determination of stone type.

A major goal of treatment in persons who have passed kidney stones or have had them removed is to prevent their recurrence (see Table 23-2). Adequate fluid intake reduces the concentration of stone-forming crystals in the urine and needs to be encouraged. Depending on the type of stone that is formed, dietary changes, medications, or measures to change the pH of the urine may be used to alter the concentration of stone-forming elements in the urine.

In some cases, stone removal may be necessary. Several methods, including ureteroscopic removal and extracorporeal lithotripsy, are available for removing kidney stones. Ureteroscopic removal involves the passage of an instrument through the urethra into the bladder and then into the ureter. *Extracorporeal shock-wave lithotripsy* uses acoustic shock waves to break up the stones so they can be passed in the urine. All these procedures eliminate the need for an open surgical procedure. Open stone surgery may be required to remove large calculi or those that are resistant to other forms of removal.

Hydronephrosis

Hydronephrosis refers to urine-filled dilatation of the renal pelvis and calices, with accompanying atrophy of renal tissue, caused by obstruction of urine flow. The obstruction may be sudden or insidious and may occur at any level of the urinary tract. The most common causes are congenital disorders of the urethra or ureters, kidney stones, tumors that compress the urethra or ureters, prostatitis, inflammatory conditions of the urethra, and neurogenic bladder.

In situations of marked or complete obstruction, backpressure develops because of a combination of continued glomerular filtration and obstruction of urine flow. Depending on the degree of obstruction, pressure builds up, beginning at the site of obstruction and moving backward from the ureter or renal pelvis into the calices and collecting tubules. Typically, the most severe effects occur at the level of the papillae because these structures are subjected to the greatest pressure. Damage to the nephrons and other functional components of the kidney is caused by compression from increased intrapelvic pressure and ischemia from disturbances in blood flow.

The degree of hydronephrosis depends on the duration, degree, and site of obstruction. Bilateral hydronephrosis occurs only when the obstruction is below the level of the ureters. If the obstruction occurs at the level of the ureters or above, hydronephrosis is unilateral. Unfortunately, unilateral hydronephrosis may remain silent for long periods of time. Often the enlarged kidney is discovered on routine exam. If the obstruction is removed within a few weeks, return of function is possible. Prolonged or severe partial obstruction causes irreversible kidney damage. The kidney eventually is destroyed and appears as a thin-walled shell that is filled with fluid (Fig. 23-5).

In summary, obstruction of urine flow can occur at any level of the urinary tract. Among the causes of urinary tract obstruction are developmental defects, pregnancy, infection and inflammation, kidney stones, neurologic defects, and prostatic hypertrophy. Obstructive disorders produce stasis of urine, increasing the risk of infection and calculi formation and resulting in back pressure that is damaging to kidney structures.

Kidney stones are a major cause of upper urinary tract obstruction. There are four types of kidney stones: calcium (i.e., oxalate and phosphate) stones, which are associated with increased serum calcium levels; magnesium ammonium phosphate (i.e., struvite) stones, which are associated with urinary tract infections; uric acid stones, which are related to elevated uric acid levels; and cystine stones, which are seen in cystinuria. A major goal of treatment for persons who have passed kidney stones or have had them removed is to identify stone composition and prevent their recurrence. Treatment measures depend on stone type and include adequate fluid intake to prevent urine saturation, dietary modification to decrease intake of stone-forming constituents, treatment of urinary tract infections, measures to change urine pH, and the use of diuretics that decrease the calcium concentration of urine.

Hydronephrosis refers to dilation of the renal pelvis and calices, with atrophy of renal tissue, that is caused by obstruction to the outflow of urine. The obstruction may be sudden or insidious in onset and may occur at any level of the urinary tract.

URINARY TRACT INFECTIONS

Urinary tract infections (UTIs) are the second most common type of bacterial infections seen by health care providers (respiratory tract infections are first). UTIs can include several distinct entities, including asymptomatic bacteriuria, symptomatic infections, lower UTIs such as cystitis, and upper UTIs such as pyelonephritis. Because of their ability to cause renal damage, upper UTIs are considered more serious than lower UTIs.



■ FIGURE 23-5 ■ Hydronephrosis. Bilateral urinary tract obstruction has led to conspicuous dilatation of the ureters, pelves, and calyces. The kidney on the right shows severe cortical atrophy. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 910]. Philadelphia: Lippincott Williams & Wilkins)

Etiologic Factors

Most UTIs are caused by *Escherichia coli*. Other common pathogens include *Staphylococcus saprophyticus*, *Proteus mirabilis*, *Klebsiella pneumoniae*, and *Enterococcus* species.^{14–18} Bacteria can enter the kidneys either through the bloodstream or as an ascending infection from the lower urinary tract. Most infections are of the ascending type. Although the distal portion of the urethra often contains pathogens, the urine formed in the kidneys and found in the bladder normally is sterile or free of bacteria. This is because of the *washout phenomenon*, in which urine from the bladder normally washes bacteria out of the urethra. When a UTI occurs, the bacteria that have colonized the urethra, vagina, or perineal area often are responsible.

There is an increased risk of UTI in persons with urinary obstruction and reflux; in people with neurogenic disorders that impair bladder emptying; in women who are sexually active, especially if they use a diaphragm or spermicide for contraception; in postmenopausal women; in men with diseases of the prostate; and in elderly persons. Instrumentation and urinary catheterization are the most common predisposing factors for nosocomial UTIs.

KEY CONCEPTS

URINARY TRACT INFECTIONS

- Urinary tract infections involve both the lower and upper urinary tract structures.
- In lower urinary tract infections, the infecting pathogens tend to propagate in the urine and cause irritative voiding symptoms, often with minimal systemic signs of infection.
- Upper urinary tract infections tend to invade the tissues of the kidney pelvis, inciting an acute inflammatory response with marked systemic manifestations of infection.

Because certain people tend to be predisposed to the development of UTIs, considerable interest has been focused on host-agent interactions and factors such as urinary tract obstruction and reflux that increase the risk of UTI.

Host Defenses

In the development of a UTI, host defenses are matched against the virulence of the pathogen. The host defenses of the bladder have several components, including the washout phenomenon, in which bacteria are removed from the bladder and urethra during voiding; the protective mucin layer that lines the bladder and protects against bacterial invasion; and local immune responses. In the ureters, peristaltic movements facilitate the movement of urine from the renal pelvis through the ureters and into the bladder. Immune mechanisms, particularly secretory immunoglobulin A (IgA), appear to provide an important antibacterial defense. Phagocytic blood cells further assist in the removal of bacteria from the urinary tract.

There has been a growing appreciation of the protective function of the bladder's mucin layer.¹⁴ It is thought that the epithelial cells that line the bladder synthesize protective substances that subsequently become incorporated into the mucin layer that adheres to the bladder wall. One theory proposes that the mucin layer acts by binding water, which then constitutes a protective barrier between the bacteria and the bladder epithelium. Elderly and postmenopausal women produce less mucin than younger women, suggesting that estrogen may play a role in mucin production in women.

Pathogen Virulence

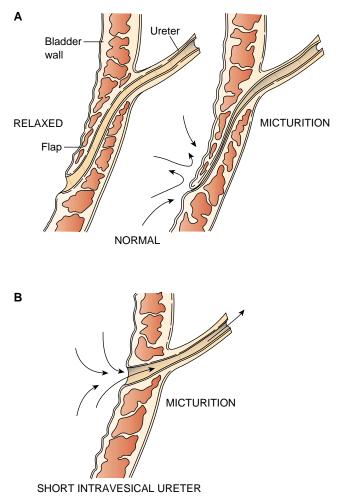
Investigations are focusing on the adherence properties of the bacteria that infect the urinary tract. These bacteria have fine protein filaments that help them adhere to receptors on the lining of urinary tract structures.^{14,15} These filaments are called *fimbriae* or *pili*. Among the factors that contribute to bacterial virulence, the type of fimbriae that the bacteria possess may be the most important. Bacteria with certain types of fimbriae are associated primarily with cystitis, and those with other types are associated with a high incidence of pyelonephritis. The bacteria that bind to carbohydrates that are specific to the surfaces of epithelial cells in this part of the urinary tract.

Obstruction and Reflux

Obstruction and reflux are important contributing factors in the development of UTIs. Any microorganisms that enter the bladder normally are washed out during voiding. When outflow is obstructed, urine remains in the bladder and acts as a medium for microbial growth; the microorganisms in the contaminated urine can then ascend along the ureters to infect the kidneys. The presence of residual urine correlates closely with bacteriuria and with its recurrence after treatment. Another aspect of bladder outflow obstruction and bladder distention is increased intravesicular pressure, which compresses blood vessels in the bladder wall, leading to a decrease in the mucosal defenses of the bladder.

In UTIs associated with stasis of urine flow, the obstruction may be anatomic or functional. Anatomic obstructions include urinary tract stones, prostatic hyperplasia, pregnancy, and malformations of the ureterovesical junction. Functional obstructions include neurogenic bladder, infrequent voiding, detrusor (bladder) muscle instability, and constipation. Reflux occurs when urine from the urethra moves into the bladder (*i.e.*, urethrovesical reflux) or from the bladder into the ureters (*i.e.*, vesicoureteral reflux). In women, *urethrovesical* reflux can occur during activities such as coughing or squatting, in which an increase in intra-abdominal pressure causes the urine to be squeezed into the urethra and then to flow back into the bladder as the pressure decreases. This also can happen when voiding is abruptly interrupted. Because the urethral orifice frequently is contaminated with bacteria, the reflux mechanism may cause bacteria to be drawn back into the bladder.

A second type of reflux mechanism, *vesicoureteral reflux*, occurs at the level of the bladder and ureter. Normally, the distal portion of the ureter courses between the muscle layer and the mucosal surface of the bladder wall, forming a flap. The flap is compressed against the bladder wall during micturition, preventing urine from being forced into the ureter (Fig. 23-6). In



■ FIGURE 23-6 ■ Anatomic features of the bladder and kidney in pyelonephritis caused by vesicoureteral reflux. Bladder. (A) In the normal bladder, the distal portion of the intravesical ureter courses between the mucosa and the muscularis of the bladder. A mucosal flap is thus formed. On micturition, the elevated intravesicular pressure compresses the flap against the bladder wall, thereby occluding the lumen. (B) Persons with a congenitally short intravesical ureter have no mucosal flap, because the entry of the ureter into the bladder approaches a right angle. Thus, micturition forces urine into the ureter. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 903]. Philadelphia: Lippincott Williams & Wilkins.) (Courtesy of Dmitri Karetnikov, artist) persons with vesicoureteral reflux, the ureter enters the bladder at an approximate right angle such that urine is forced into the ureter during micturition.⁶ It is seen most commonly in children with UTIs and is believed to result from congenital defects in length, diameter, muscle structure, or innervation of the submucosal segment of the ureter. Vesicoureteral reflux also is seen in adults with obstruction to bladder outflow, primarily because of increased bladder volume and pressure.

Catheter-Induced Infection

Urinary catheters are tubes made of latex or plastic. They are inserted through the urethra into the bladder for the purpose of draining urine. They are a source of urethral irritation and provide a means for entry of microorganisms into the urinary tract.

Catheter-associated bacteriuria remains the most frequent cause of gram-negative septicemia in hospitalized patients. Studies have shown that bacteria adhere to the surface of the catheter and initiate the growth of a biofilm that then covers the surface of the catheter.¹⁴ The biofilm tends to protect the bacteria from the action of antibiotics and makes treatment difficult. A closed drainage system (*i.e.*, closed to air and other sources of contamination) and careful attention to perineal hygiene (*i.e.*, cleaning the area around the urethral meatus) help to prevent infections in persons who require an indwelling catheter. Careful hand washing and early detection and treatment of UTIs also are essential.

Manifestations

The manifestations of UTI depend on whether the infection involves the lower or upper urinary tract. An acute episode of cystitis (bladder infection) or lower UTI is characterized by frequency of urination (sometimes as often as every 20 minutes), lower abdominal or back discomfort, and burning and pain on urination (*i.e.*, dysuria). Occasionally the urine is cloudy and foul smelling. In adults, fever and other signs of infection usually are absent. If there are no complications, the symptoms disappear within 48 hours of treatment. This type of UTI is common in younger women. The symptoms of cystitis also may represent urethritis caused by *Chlamydia trachomatis, Neisseria gonorrhoeae*, or herpes simplex virus, or vaginitis attributable to *Trichomonas vaginalis* or *Candida* species.

Upper UTIs affect the parenchymal tissues of kidney pelvis (pyelonephritis). They tend to produce more systemic signs of infection than lower UTIs because of closer proximity to the vascular compartment and blood cells (*e.g.*, neutrophils) that incite the inflammatory response. Acute pyelonephritis tends to present with an abrupt onset of shaking chills, moderate to high fever, and constant ache in the loin area of the back that is unilateral or bilateral.^{13,14} Lower urinary tract symptoms, including dysuria, frequency, and urgency, also are common. There may be significant malaise, and the person usually looks and feels ill. Nausea and vomiting may occur along with abdominal pain. Palpation or percussion over the costovertebral angle on the affected side usually causes pain. Pyelonephritis occurs more frequently in children and adults with urinary tract obstructions or other predisposing conditions.

Diagnosis and Treatment

The diagnosis of UTI usually is based on symptoms and on examination of the urine for the presence of microorganisms. When necessary, x-ray films, ultrasonography, and CT and renal scans are used to identify contributing factors, such as obstruction.

Microscopic urine tests are used establish the presence of bacteria and blood cells in the urine.⁹ A commonly accepted criterion for diagnosis of a UTI is the presence of 10⁵ or more bacteria per milliliter of urine.⁹ Colonization usually is defined as the multiplication of microorganisms in or on a host without apparent evidence of invasiveness or tissue injury. Pyuria (the presence of less than five to eight leukocytes per highpower field) indicates a host response to infection, rather than asymptomatic bacterial colonization. A Gram's stain may be done to determine the type of organism that is present (gram positive or gram negative).

Chemical screening (urine dipstick) for markers of infection may provide useful information but is less sensitive than microscopic analysis. A urine culture confirms the presence of pathogenic bacteria in urine specimens, allows for their identification, and permits the determination of their sensitivity to specific antibiotics.

The treatment of acute UTI is based on the type of infection that is present (lower or upper UTI), the pathogen causing the infection, and the presence of contributing host-agent factors. Other considerations include whether the infection is acute, recurrent, or chronic. Most acute lower UTIs are treated successfully with antimicrobial therapy and increased fluid intake. Because there is risk of permanent kidney damage, upper urinary tract infections (*e.g.*, pyelonephritis) are treated more aggressively. Hospitalization may be recommended during the early stages of infection until a response to treatment is observed.¹⁶

Recurrent lower UTIs are those that recur after treatment. They are attributable either to bacterial persistence or reinfection. Bacterial persistence usually is curable by removal of the infectious source (*e.g.*, urinary catheter or infected bladder stones).

Chronic UTIs are more difficult to treat. Because they often are associated with obstructive uropathy or reflux flow of urine, diagnostic tests usually are performed to detect such abnormalities. When possible, the condition causing the reflux flow or obstruction is corrected. Men in particular should be investigated for obstructive disorders or a prostatic focus of infection.

Infections in Special Populations

Urinary Tract Infections in Women

In women, the urethra is short and close to the vagina and rectum, offering little protection against entry of microorganisms into the bladder. There is a peak incidence of these infections in the 15- to 24-year-old age group, suggesting that hormonal and anatomic changes associated with puberty and sexual activity contribute to UTIs.

The role of sexual activity in the development of urethritis and cystitis is a matter of controversy. The anterior urethra usually is colonized with bacteria; urethral massage or sexual intercourse can force these bacteria back into the bladder. Using a diaphragm and spermicide enhances the susceptibility to infection.¹⁷ A nonpharmacologic approach to the treatment of frequent UTIs associated with sexual intercourse is to increase fluid intake before intercourse and to void soon after intercourse. This procedure uses the washout phenomenon to remove bacteria from the bladder. Pregnant women are at increased risk for UTIs. Normal changes in the functioning of the urinary tract that occur during pregnancy predispose to UTIs.¹⁸ These changes involve the collecting system of the kidneys and include dilatation of the renal calices, pelves, and ureters that begins during the first trimester and becomes most pronounced during the third trimester. This dilatation of the upper urinary system is accompanied by a reduction in the peristaltic activity of the ureters that is thought to result from the muscle-relaxing effects of progesterone-like hormones and mechanical obstruction from the enlarging uterus. In addition to the changes in the kidneys and ureters, the bladder becomes displaced from its pelvic position to a more abdominal position, producing further changes in ureteral position.

UTIs are commonly asymptomatic in pregnant women. The complications of asymptomatic UTIs during pregnancy include persistent bacteriuria, acute and chronic pyelonephritis, toxemia of pregnancy, and premature delivery. Evidence suggests that few women become bacteriuric during pregnancy. Rather, it appears that symptomatic UTIs during pregnancy reflect preexisting asymptomatic bacteriuria, and that changes occurring during pregnancy simply permit the prior urinary colonization to lead to symptomatic infection and invasion of the kidneys.¹⁹ Because bacteriuria may occur as an asymptomatic condition in pregnant women, the American College of Obstetricians and Gynecologists recommends that a urine culture be obtained at the first prenatal visit.^{18,19} A repeat culture should be obtained during the third trimester. Women with bacteriuria should be followed up closely, and infections should be properly treated to prevent complications. The choice of antimicrobial agent should address the common infecting organisms and should be safe for the woman and fetus.

Urinary Tract Infections in Children

Urinary tract infections occur in as many as 3% to 5% of female and 1% of male children.⁴ In girls, the average age at first diagnosis is 3 years, which coincides with onset of toilet training. In boys, most UTIs occur during the first year of life; they are more common in uncircumcised than in circumcised boys. Children who are at increased risk for bacteriuria or symptomatic UTIs are premature infants discharged from neonatal intensive care units; children with systemic or immunologic disease or urinary tract abnormalities such as neurogenic bladder or vesicoureteral reflux; those with a family history of UTI or urinary tract anomalies with reflux; and girls younger than 5 years of age with a history of UTI.²⁰

UTIs in children frequently involve the upper urinary tract (pyelonephritis). In children in whom renal development is not complete, pyelonephritis can lead to renal scarring and permanent kidney damage. It has been reported that more than 75% of children younger than 5 years of age with febrile UTIs have pyelonephritis, and that renal scarring occurs in 27% to 64% of children with pyelonephritis.²⁰ Most UTIs that lead to scarring and diminished kidney growth occur in children younger than 4 years, especially infants younger than 1 year of age. The incidence of scarring is greatest in children with gross vesicoureteral reflux or obstruction, in children with recurrent UTIs, and those with a delay in treatment.

Unlike adults, children frequently do not present with the typical signs of a UTI.^{20,21} Many neonates with UTIs have bac-

teremia and may show signs and symptoms of septicemia, including fever, hypothermia, apneic spells, poor skin perfusion, abdominal distention, diarrhea, vomiting, lethargy, and irritability. Older infants may present with feeding problems, failure to thrive, diarrhea, vomiting, fever, and foul-smelling urine. Toddlers often present with abdominal pain, vomiting, diarrhea, abnormal voiding patterns, foul-smelling urine, fever, and poor growth. In older children with lower UTIs, the classic features—enuresis, frequency, dysuria, and suprapubic discomfort—are more common. Fever is a common sign of UTI in children, and the possibility of UTI should be considered in children with unexplained fever.

Diagnosis is based on a careful history of voiding patterns and symptomatology; physical examination to determine fever, hypertension, abdominal or suprapubic tenderness, and other manifestations of UTI; and urinalysis to determine bacteriuria, pyuria, proteinuria, and hematuria. A positive urine culture that is obtained correctly is essential for the diagnosis. Additional diagnostic methods may be needed to determine the cause of the disorder. Vesicoureteral reflux is the most commonly associated abnormality in UTIs, and reflux nephropathy is an important cause of end-stage renal disease in children and adolescents. Children with a relatively uncomplicated first UTI may turn out to have significant reflux. Therefore, even a single documented UTI in a child requires careful diagnosis. Urinary symptoms in the absence of bacteriuria suggest vaginitis, urethritis, sexual molestation, the use of irritating bubble baths, pinworms, or viral cystitis. In adolescent girls, a history of dysuria and vaginal discharge makes vaginitis or vulvitis a consideration.

The approach to treatment is based on the clinical severity of the infection, the site of infection (*i.e.*, lower vs. upper urinary tract), the risk of sepsis, and the presence of structural abnormalities. The immediate treatment of infants and young children is essential. Most infants with symptomatic UTIs and many children with clinical evidence of acute upper UTIs require hospitalization and intravenous antibiotic therapy. Follow-up is essential for children with febrile UTIs to ensure resolution of the infection. Follow-up urine cultures often are done at the end of treatment. Imaging studies often are recommended for all children after first UTIs to detect renal scarring, vesicoureteral reflux, or other abnormalities.^{20,21}

Winary Tract Infections in the Elderly

Urinary tract infections are relatively common in elderly persons. It has been reported that 5% to 20% of the elderly living at home have bacteriuria. These numbers increase to 15% to 25% for the elderly living in nursing homes or extended-care facilities.²²

Most of these infections follow invasion of the urinary tract by the ascending route. Several factors predispose elderly persons to UTIs: immobility resulting in poor bladder emptying, bladder outflow obstruction caused by prostatic hyperplasia or kidney stones, bladder ischemia caused by urine retention, senile vaginitis, constipation, and diminished bactericidal activity of urine and prostatic secretions. Added to these risks are other health problems that necessitate instrumentation of the urinary tract. UTIs develop in 1% of ambulatory patients after a single catheterization and within 3 to 4 days in essentially all patients with indwelling catheters.²³

Elderly persons with bacteriuria have varying symptoms, ranging from the absence of symptoms to the presence of typ-

ical UTI symptoms. Even when symptoms of lower UTIs are present, they may be difficult to interpret because elderly persons without UTIs commonly experience urgency, frequency, and incontinence. Alternatively, elderly persons may have vague symptoms such as anorexia, fatigue, weakness, or change in mental status. Even with more serious upper UTIs (*e.g.*, pyelonephritis), the classic signs of infection such as fever, chills, flank pain, and tenderness may be altered or absent in elderly persons.²² Sometimes, no symptoms occur until the infection is far advanced.

In summary, UTI is the second most common type of bacterial infection seen by health care professionals. Infections can range from simple bacteriuria to severe kidney infections that cause irreversible kidney damage. Predisposition to infection is determined by host defenses and pathogen virulence. Host defenses include the washout phenomenon associated with voiding, the protective mucin lining of the bladder, and local immune defenses. Pathogen virulence is enhanced by the presence of fimbriae that facilitate adherence to structures in the urinary tract.

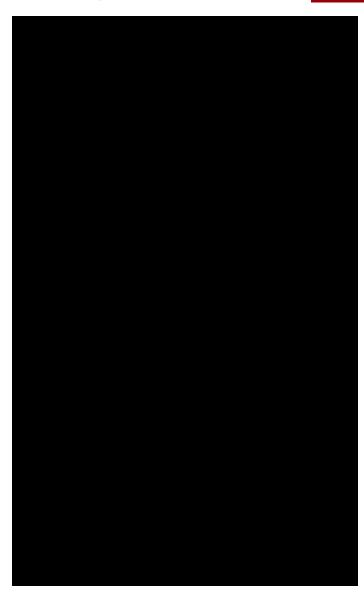
Most UTIs ascend from the urethra and bladder. A number of factors interact in determining the predisposition to development of UTIs, including urinary tract obstruction, urine stasis and reflux, pregnancy-induced changes in urinary tract function, age-related changes in the urinary tract, changes in the protective mechanisms of the bladder and ureters, impaired immune function, and virulence of the pathogen. Urinary tract catheters and urinary instrumentation contribute to the incidence of UTIs. Early diagnosis and treatment of UTIs are essential to preventing permanent kidney damage.

DISORDERS OF GLOMERULAR FUNCTION

The glomeruli are tufts of capillaries that lie between the afferent and efferent arterioles. The capillaries of the glomeruli are arranged in lobules and supported by a stalk consisting of mesangial cells and a basement membrane-like extracellular matrix (Fig. 23-7). The glomerular membrane is composed of three layers: an endothelial layer lining the capillary, a basement membrane, and a layer of epithelial cells forming the outer surface of the capillary and lining Bowman's capsule (see Chapter 22, Fig. 22-5). The epithelial cells are attached to the basement membrane by discrete cytoplasmic extensions, the foot processes (i.e., podocytes). In the glomeruli, blood is filtered, and the urine filtrate formed. The capillary membrane is selectively permeable: it allows water, electrolytes, and dissolved particles, such as glucose and amino acids, to leave the capillary and enter Bowman's space and prevents larger particles, such as plasma proteins and blood cells, from leaving the blood.

Mechanisms of Glomerular Injury

Glomerulonephritis, an inflammatory process that involves glomerular structures, is the leading cause of chronic renal failure in the United States, accounting for one half of persons with end-stage renal disease.²⁴ There are many causes of



KEY CONCEPTS

GLOMERULAR DISEASE

- Glomerular disorders affect the glomerular capillary membrane structures that filter materials from the blood.
- Nephritic syndromes are caused by diseases that produce proliferative inflammatory responses that decrease the permeability of the glomerular capillary membrane.
- The nephrotic syndrome is caused by disorders that increase the permeability of the glomerular capillary membrane, causing massive loss of protein in the urine.

426 Unit Six: Alterations in the Urinary System

glomerular disease. The disease may occur as a primary condition in which the glomerular abnormality is the only disease present, or it may occur as a secondary condition in which the glomerular abnormality results from another disease, such as diabetes mellitus or SLE. An understanding of the various forms of glomerular disease has emerged only recently. Much of this knowledge can be attributed to advances in immunobiology and electron microscopy, development of animal models, and increased use of renal biopsy during the early stages of glomerular disease.

Although little is known about the causative agents or triggering events that produce glomerular disease, most cases of primary and many cases of secondary glomerular disease probably have an immune origin.^{5,6,24} Two types of immune mechanisms have been implicated in the development of glomerular disease: injury resulting from antibodies reacting with fixed glomerular antigens, and injury resulting from circulating antigen-antibody complexes that become trapped in the glomerular membrane (Fig. 23-8). Antigens responsible for development of the immune response may be of endogenous origin, such as DNA in SLE, or they may be of exogenous origin, such as streptococcal membrane antigens in poststreptococcal glomerulonephritis. Frequently, the source of the antigen is unknown.

The cellular changes that occur with glomerular disease include proliferative, sclerotic, and membranous changes. The term *proliferative* refers to an increase in the cellular components of the glomerulus, regardless of origin; *sclerotic* to an increase in the noncellular components of the glomerulus, primarily collagen; and *membranous* to an increase in the thickness of the glomerular capillary wall, often caused by immune complex deposition. Glomerular changes can be *diffuse*, involving all glomeruli and all parts of the glomeruli; *focal*, in which only some glomeruli are affected and others are essentially normal; *segmental*, involving only a certain segment of each glomeruli; or *mesangial*, affecting only the mesangial cell. Figure 23-8 shows changes associated with various types of glomerular disease.

Among the different types of glomerular diseases are the nephritic syndromes, which produce a proliferative inflammatory response; the nephrotic syndromes, which involve increased permeability of the glomerulus; and chronic glomerulonephritis, which represents the chronic phase of a number of glomerular disorders.

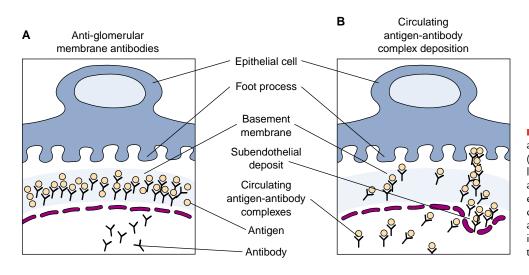
Nephritic Syndromes

Glomerulonephritis is characterized by hematuria with red cell casts, a diminished glomerular filtration rate (GFR), azotemia (presence of nitrogenous wastes in the blood), oliguria, and hypertension. It is caused by diseases that provoke a proliferative inflammatory response of the endothelial, mesangial, or epithelial cells of the glomeruli. The inflammatory process damages the capillary wall, permitting red blood cells to escape into the urine and producing hemodynamic changes that decrease the GFR. The nephritic syndromes include acute proliferative glomerulonephritis and rapidly progressive glomerulonephritis.

Acute Proliferative Glomerulonephritis

The most commonly recognized form of acute glomerulonephritis is diffuse proliferative glomerulonephritis, which follows infections caused by strains of group A β-hemolytic streptococci. Diffuse proliferative glomerulonephritis also may occur after infections by other organisms, including staphylococci and a number of viral agents, such as those responsible for mumps, measles, and chickenpox. With this type of nephritis, the inflammatory response is caused by an immune reaction that occurs when circulating immune complexes become entrapped in the glomerular membrane. Proliferation of the endothelial cells lining the glomerular capillary (i.e., endocapillary form of the disease) and the mesangial cells lying between the endothelium and the epithelium follows (see Fig. 23-8). The capillary membrane swells and becomes permeable to plasma proteins and blood cells. Although the disease is seen primarily in children, adults of any age also can be affected.

The classic case of poststreptococcal glomerulonephritis follows a streptococcal infection by approximately 7 to 12 days the time needed for the development of antibodies.²⁴ Oliguria, which develops as the GFR decreases, is one of the first symptoms. Proteinuria and hematuria follow because of increased glomerular capillary wall permeability. The blood is degraded by materials in the urine, and a cola-colored urine may be the first sign of the disorder. Sodium and water retention gives rise to edema, particularly of the face and hands, and hypertension. Important laboratory findings include an elevated streptococcal exoenzyme (antistreptolysin O) titer, a decline in C3 com-



■ FIGURE 23-8 ■ Immune mechanisms of glomerular disease. (A) Antiglomerular antibodies leave the circulation and interact with antigens that are present in the basement membrane of the glomerulus. (B) Antigenantibody complexes circulating in the blood become trapped as they are filtered in the glomerulus. plement (see Chapter 10), and cryoglobulins (*i.e.*, large immune complexes) in the serum.

Treatment for acute poststreptococcal glomerulonephritis is largely symptomatic. The acute symptoms usually begin to subside in approximately 10 days to 2 weeks, although in some children the proteinuria may persist for several months. The immediate prognosis is favorable, and approximately 95% of children recover spontaneously.⁵ The outlook for adults is less favorable; approximately 60% recover completely. In the remainder of cases, the lesions eventually resolve, but there may be permanent kidney damage.

Rapidly Progressive Glomerulonephritis

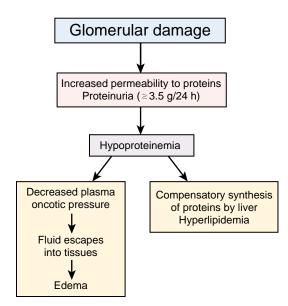
Rapidly progressive glomerulonephritis is a clinical syndrome characterized by signs of severe glomerular injury that does not have a specific cause. As its name indicates, this type of glomerulonephritis is rapidly progressive, often within a matter of months. The disorder involves focal and segmental proliferation of glomerular cells and recruitment of monocytes (macrophages) with formation of crescent-shaped structures that obliterate Bowman's space.⁵ Rapidly proliferative glomerulonephritis may be caused by a number of immunologic disorders, some systemic and others restricted to the kidney. Among the diseases associated with this form of glomerulonephritis are immune complex disorders such as SLE, the small vessel vasculitides (*e.g.*, microscopic polyangiitis), and an immune disorder condition called *Goodpasture's syndrome*.

Goodpasture's syndrome, which is caused by antibodies to the glomerular basement membrane (GBM), accounts for approximately 5% of cases of rapidly progressive glomerulonephritis. It is a rare disease and is associated with a triad of pulmonary hemorrhage, iron-deficiency anemia, and glomerulonephritis. All of these manifestations result from anti-GBM antibody deposition in the lungs and glomeruli. The cause of the disorder is unknown, although influenza infection and exposure to hydrocarbon solvent (found in paints and dyes) have been implicated in some persons, as have various drugs and cancers. There is a high prevalence of certain human leukocyte antigen subtypes (e.g., HLA-DRB1), suggesting a genetic predisposition.⁵ Treatment includes plasmapheresis to remove circulating anti-GBM antibodies and immunosuppressive therapy (i.e., corticosteroids and cyclophosphamide) to inhibit antibody production.

Nephrotic Syndrome

Nephrotic syndrome is not a specific glomerular disease but a constellation of clinical findings that result from increased glomerular permeability to the plasma proteins (Fig. 23-9). The glomerular derangements that occur with nephrosis can develop as a primary disorder or secondary to changes caused by systemic diseases such as diabetes mellitus, amyloidosis, and SLE. Among the primary glomerular lesions leading to nephrotic syndrome are minimal change disease (lipoid nephrosis), focal segmental glomerulosclerosis, and membranous glomerulonephritis. The relative frequency of these causes varies with age. In children younger than 15 years of age, nephrotic syndrome almost always is caused by primary glomerular disease, whereas in adults it often is a secondary disorder.⁵

The nephrotic syndrome is characterized by massive proteinuria (>3.5 g/day) and lipiduria (*e.g.*, free fat, oval bodies,



■ FIGURE 23-9 ■ Pathophysiology of the nephrotic syndrome.

fatty casts), along with an associated hypoalbuminemia (<3 g/dL), generalized edema, and hyperlipidemia (cholesterol >300 mg/dL).^{6,25,26} The initiating event in the development of nephrosis is a derangement in the glomerular membrane that causes increased permeability to plasma proteins. The glomerular membrane acts as a size and charge barrier through which the glomerular filtrate must pass. Any increased permeability allows protein to escape from the plasma into the glomerular filtrate.

Generalized edema, which is a hallmark of nephrosis, results from salt and water retention and a loss of serum albumin below that needed to maintain the colloid osmotic pressure of the vascular compartment.⁵ The sodium and water retention appears to be caused by several factors, including a compensatory increase in aldosterone, stimulation of the sympathetic nervous system, and a reduction in secretion of natriuretic factors. Initially, the edema presents in dependent parts of the body, such as the lower extremities, but becomes more generalized as the disease progresses. Dyspnea caused by pulmonary edema, pleural effusions, and diaphragmatic compromise attributable to ascites can develop in persons with nephrotic syndrome.¹⁵

Although the largest proportion of plasma protein loss is in albumin, globulins also are lost. As a result, persons with nephrosis are particularly vulnerable to infections, particularly those caused by staphylococci and pneumococci.⁵ This decreased resistance to infection probably is related to loss of both immunoglobulins and low-molecular-weight complement components in the urine. Many binding proteins also are lost in the urine. Consequently, the plasma levels of many ions (iron, copper, zinc), hormones (thyroid and sex hormones), and drugs may be low because of decreased binding proteins. Many drugs require protein binding for transport. Hypoalbuminemia reduces the number of available protein binding sites, thereby producing a potential increase in the amount of free (active) drug that is available.²⁷

Thrombotic complications also have evolved as a risk in persons with nephrotic syndrome. These disorders are thought to be related to a disruption in the function of the coagulation system brought about by a loss of coagulation and anticoagulation factors.^{15,27} Renal vein thrombosis, once thought to be a cause of the disorder, is more likely a consequence of the hypercoagulable state.⁵ Other thrombotic complications include deep vein thrombosis and pulmonary emboli.

The hyperlipidemia that occurs in persons with nephrosis is characterized by elevated levels of triglycerides and low-density lipoproteins (LDL). Levels of high-density lipoproteins (HDL) usually are normal. These abnormalities are thought to be related, in part, to increased synthesis of lipoproteins in the liver secondary to a compensatory increase in albumin production. Because of the elevated LDL levels, persons with nephrotic syndrome are at increased risk for development of atherosclerosis.

Membranous Glomerulonephritis

Membranous glomerulonephritis is the most common cause of primary nephrosis in adults, most commonly people in their sixth or seventh decade. The disorders are caused by diffuse thickening of the GBM attributable to deposition of immune complexes. The disorder may be idiopathic or associated with a number of disorders, including autoimmune diseases such as SLE, infections such as chronic hepatitis B, metabolic disorders such as diabetes mellitus and thyroiditis, and use of certain drugs such as gold, penicillamine, and captopril.⁵ Because of the presence of immunoglobulins and complement in the subendothelial deposits, it is thought that the disease represents a chronic antigen-antibody complex-mediated disorder.

The disorder is treated with corticosteroids. Cytotoxic drugs may be added to the treatment regimen. The progress of the disease is variable; approximately one half of persons sustain a slow but progressive loss of renal function.

Minimal Change Disease (Lipoid Nephrosis)

Minimal change disease is characterized by diffuse loss (through fusion) of the foot processes from the epithelial layer of the glomerular membrane. The peak incidence is between 2 and 6 years of age. The cause of minimal change nephrosis is unknown; however, children in whom the disease develops often have a history of recent upper respiratory infections or of receiving routine immunizations.⁵ Although minimal change disease does not progress to renal failure, it can cause significant complications, including predisposition to infection with gram-positive organisms, a tendency toward thromboembolic events, hyperlipidemia, and protein malnutrition. There usually is a dramatic response to corticosteroid therapy.⁵

Focal Segmental Glomerulosclerosis

Focal segmental glomerulosclerosis is characterized by sclerosis (*i.e.*, increased collagen deposition) of some but not all glomeruli, and in the affected glomeruli, only a portion of the glomerular tuft is involved. Although focal segmental sclerosis often is an idiopathic syndrome, it may be associated with reduced oxygen in the blood (*e.g.*, sickle cell disease and cyanotic congenital heart disease), HIV infection, or intravenous drug abuse, or it may be a secondary event reflecting glomerular scarring caused by other forms of glomerulonephritis or reflux nephropathy.^{5,6} The presence of hypertension and decreased renal function distinguishes focal sclerosis from minimal change disease. The disorder usually is treated with corticosteroids. Most persons with focal segmental glomerulosclerosis progress to end-stage renal disease within 5 to 10 years.

Chronic Glomerulonephritis

Chronic glomerulonephritis represents the chronic phase of a number of specific types of glomerulonephritis.⁵ Some forms of glomerulonephritis (*e.g.*, poststreptococcal glomerulonephritis) undergo complete resolution, whereas others progress at variable rates to chronic glomerulonephritis. Some persons who present with chronic glomerulonephritis have no history of glomerular disease. These cases may represent the end result of relatively asymptomatic forms of glomerulonephritis. Histologically, the condition is characterized by small kidneys with sclerosed glomeruli. In most cases, chronic glomerulonephritis develops insidiously and slowly progresses to end-stage renal disease over a period of years (see Chapter 24).

Glomerular Lesions Associated With Systemic Disease

Many immunologic, metabolic, or hereditary systemic diseases are associated with glomerular injury. In some diseases, such as systemic lupus erythematosus and diabetes mellitus, the glomerular involvement may be a major clinical manifestation. The glomerular lesions associated with diabetes mellitus and hypertension are discussed in this chapter.

Diabetic Glomerulosclerosis

Diabetic nephropathy, or kidney disease, is a major complication of diabetes mellitus. It affects approximately 30% of persons with type 1 diabetes and accounts for 20% of deaths in diabetic patients younger than 40 years of age.⁵

The glomerulus is the most commonly affected structure in diabetic nephropathy, evidenced by three glomerular syndromes: non-nephrotic proteinuria, nephrotic syndrome, and renal failure. Widespread thickening of the glomerular capillary basement membrane occurs in almost all persons with diabetes and can occur without evidence of proteinuria.⁵ This is followed by a diffuse increase in mesangial matrix, with mild proliferation of mesangial cells. In nodular glomerulosclerosis, also known as *Kimmelstiel-Wilson syndrome*, there is nodular deposition of hyaline in the mesangial portion of the glomerulus. As the sclerotic process progresses in both the diffuse and nodular forms of glomerulosclerosis, there is complete obliteration of the glomerulus, with impairment of renal function.

Although the mechanisms of glomerular change in diabetes are uncertain, they are thought to represent enhanced or defective synthesis of the glomerular basement membrane and mesangial matrix with an inappropriate incorporation of glucose into the noncellular components of these glomerular structures. Alternatively, hemodynamic changes that occur secondary to elevated blood glucose levels may contribute to the initiation and progression of diabetic glomerulosclerosis. It has been hypothesized that elevations in blood glucose produce an increase in GFR and glomerular intracapillary pressure that leads to an enlargement of glomerular capillary pores by a mechanism that is at least partly mediated by angiotensin II. This enlargement impairs the size-selective function of the membrane so that the protein content of the glomerular filtrate increases, which in turn requires increased endocytosis of protein by the tubular endothelial cells, a process that ultimately leads to nephron destruction and progressive deterioration of renal function.28,29

The clinical manifestations of diabetic glomerulosclerosis are closely linked to those of diabetes. The increased GFR that occurs in persons with early alterations in renal function is associated with microalbuminuria, defined as urinary albumin excretion greater than 30 mg/24 hours and no more than 300 mg/ 24 hours.²⁹ Microalbuminuria is an important predictor of future diabetic nephropathies.^{5,29} In many cases, these early changes in glomerular function can be reversed by careful control of blood glucose levels (see Chapter 32).29 Inhibition of angiotensin by angiotensin-converting enzyme inhibitors (e.g., captopril) has been shown to have a beneficial effect, possibly by reversing increased glomerular pressure.6,29 Hypertension and cigarette smoking have been implicated in the progression of diabetic nephropathy. Thus, control of high blood pressure and smoking cessation are recommended as primary and secondary prevention strategies in persons with diabetes.

Hypertensive Glomerular Disease

Hypertension can be viewed as both a cause and an effect of kidney disease. Most persons with advanced kidney disease have hypertension, and many persons with long-standing hypertension eventually sustain changes in kidney function. Renal failure and azotemia occur in 1% to 5% of persons with long-standing hypertension (see Chapter 16). Hypertension is associated with a number of changes in glomerular structures, including sclerotic changes. As the glomerular vascular structures thicken and perfusion diminishes, the blood supply to the nephron decreases, causing the kidneys to lose some of their ability to concentrate the urine. This may be evidenced by nocturia. Blood urea nitrogen levels also may become elevated, particularly during periods of water deprivation. Proteinuria may occur as a result of changes in glomerular structure.

In summary, diseases of the glomerulus disrupt glomerular filtration and alter the permeability of glomerular capillary membrane to plasma proteins and blood cells. *Glomerulonephritis* is a term used to describe a group of diseases that result in inflammation and injury of the glomerulus. These diseases disrupt the capillary membrane and cause proteinuria, hematuria, pyuria, oliguria, edema, hypertension, and azotemia. Almost all types of glomerulonephritis are caused by immune mechanisms.

Glomerular diseases have been grouped into two categories: the nephritic and the nephrotic syndromes. The nephritic syndrome evokes an inflammatory response in the glomeruli and is characterized by hematuria with red cell casts in the urine, a diminished GFR, azotemia, oliguria, and hypertension. The nephrotic syndrome affects the integrity of the glomerular capillary membrane and is characterized by massive proteinuria, hypoalbuminemia, generalized edema, lipiduria, and hyperlipidemia. Both conditions can lead to progressive loss of glomerular function and eventual development of endstage renal disease. Among the secondary causes of glomerular kidney disease are diabetes and hypertension. Kidney disease is a major complication of diabetes mellitus and is thought to be related to hemodynamic changes associated with defective synthesis of glomerular structures associated with increased blood glucose levels. Hypertension is closely linked with kidney disease, and kidney disease can be a cause or effect of elevated blood pressure.

TUBULOINTERSTITIAL DISORDERS

Several disorders affect renal tubular structures, including the proximal and distal tubules. Most of these disorders also affect the interstitial tissue that surrounds the tubules. These disorders, which sometimes are referred to as *tubulointerstitial disorders*, include acute tubular necrosis (see Chapter 24), pyelonephritis, and drug-induced nephropathies.

Tubulointerstitial renal diseases may be divided into acute and chronic disorders. The acute disorders are characterized by their sudden onset and by signs and symptoms of interstitial edema; they include acute pyelonephritis and acute hypersensitivity reaction to drugs. The chronic disorders produce interstitial fibrosis, atrophy, and mononuclear infiltrates; most persons are asymptomatic until late in the course of the disease. In the early stages, tubulointerstitial diseases commonly are manifested by fluid and electrolyte imbalances that reflect subtle changes in tubular function. These manifestations can include the inability to concentrate urine, as evidenced by polyuria and nocturia; interference with acidification of urine, resulting in metabolic acidosis; and diminished tubular reabsorption of sodium and other substances.⁵

Pyelonephritis

Acute pyelonephritis refers to an inflammation of the kidneys and renal pelves. There are two forms of pyelonephritis: acute and chronic. *Acute pyelonephritis* represents a patchy interstitial suppurative inflammatory process, with abscess formation and tubular necrosis. Infection may occur through bloodstream or ascend from the bladder. Factors that contribute to the development of acute pyelonephritis are catheterization and urinary tract instrumentation, vesicoureteral reflux, pregnancy, and neurogenic bladder. A second less frequent and more serious type of acute pyelonephritis, called *necrotizing papillitis*, is characterized by necrosis of the renal papillae. It is particularly common in persons with diabetes and may also complicate acute pyelonephritis when there is significant urinary tract obstruction.

The onset of acute pyelonephritis typically is abrupt, with chills, fever, headache, back pain, tenderness over the costovertebral angle, and general malaise. It usually is accompanied by symptoms of bladder irritation, such as dysuria, frequency, and urgency. Pyuria occurs but is not diagnostic because it also occurs in lower UTIs. The development of necrotizing papillitis is associated with much poorer prognosis. These persons have evidence of overwhelming sepsis and, often, renal failure.

Acute pyelonephritis is treated with appropriate antimicrobial drugs. Unless obstruction or other complications occur, the symptoms usually disappear within several days. Hospitalization during initial treatment may be necessary. Depending on the cause, recurrent infections are possible.

Chronic pyelonephritis represents a progressive process. There is scarring and deformation of the renal calices and pelvis⁶ (Fig. 23-10). The disorder appears to involve a bacterial infection superimposed on obstructive abnormalities or vesicoureteral reflux. Chronic obstructive pyelonephritis is associated with recurrent bouts of inflammation and scarring, which eventually lead to chronic pyelonephritis. Reflux, which is the most 0 Unit Six: Alterations in the Urinary System



■ FIGURE 23-10 Chronic pyelonephritis. Marked dilatation of calcyces caused by inflammatory destruction of papillae, with atrophy and scarring of the overlying cortex. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 905]. Philadelphia: Lippincott Williams & Wilkins)

common cause of chronic pyelonephritis, results from superimposition of infection on congenital vesicoureteral reflux or intrarenal reflux. Reflux may be unilateral with involvement of a single kidney or bilateral leading to scarring and atrophy of both kidneys with the eventual development of chronic renal insufficiency.

Chronic pyelonephritis may cause many of the same symptoms as acute pyelonephritis, but its onset may be more insidious. Loss of tubular function and the ability to concentrate urine give rise to polyuria and nocturia, and mild proteinuria is common. Severe hypertension often is a contributing factor in the progress of the disease. Chronic pyelonephritis is a significant cause of renal failure. It is thought to be responsible for 11% to 20% of all cases of end-stage renal disease.⁵

Drug-Related Nephropathies

Drug-related nephropathies involve functional or structural changes in the kidneys that occur after exposure to a drug. The kidneys are exposed to a high rate of delivery of any substance in the blood because of their large blood flow and high filtration pressure. The kidneys also are active in the metabolic transformation of drugs and therefore are exposed to a number of toxic metabolites. Some drugs and toxic substances damage the kidneys by causing a decrease in blood flow; others directly damage tubulointerstitial structures; and still others cause damage by producing hypersensitivity reactions.

The tolerance to drugs varies with age and depends on renal function, state of hydration, blood pressure, and the pH of the urine. Because of a decrease in physiologic function, elderly persons are particularly susceptible to kidney damage caused by drugs and toxins. The dangers of nephrotoxicity are increased when two or more drugs capable of producing kidney damage are given at the same time.

Acute drug-related hypersensitivity reactions produce tubulointerstitial nephritis, with damage to the tubules and interstitium. This condition was observed initially in persons who were sensitive to the sulfonamide drugs; currently, it is observed most often with the use of methicillin and other synthetic antibiotics, and with the use of furosemide and the thiazide diuretics in persons sensitive to these drugs. The condition begins approximately 15 days (range, 2-40 days) after exposure to the drug.^{5,6} At the onset, there is fever, eosinophilia, hematuria, mild proteinuria, and in approximately one fourth of cases, a rash. In approximately 50% of cases, signs and symptoms of acute renal failure develop. Withdrawal of the drug commonly is followed by complete recovery, but there may be permanent damage in some persons, usually in older persons. Drug nephritis may not be recognized in its early stage because it is uncommon.

Chronic analgesic nephritis, which is associated with analgesic abuse, causes interstitial nephritis with renal papillary necrosis. When first observed, it was attributed to phenacetin, a then-common ingredient of over-the-counter medications containing aspirin, phenacetin, and caffeine. Although phenacetin is no longer contained in these preparations, it has been suggested that other ingredients, such as aspirin and acetaminophen, also may contribute to the disorder. How much analgesic it takes to produce papillary necrosis is unknown.

Nonsteroidal anti-inflammatory drugs (NSAIDs) also have the potential for damaging renal structures, including medullary interstitial cells. Prostaglandins (particularly PGI₂ and PGE₂) contribute to regulation of tubular blood flow.³⁰ The deleterious effects of NSAIDs on the kidney are thought to result from their ability to inhibit prostaglandin synthesis. Persons who are particularly at risk are the elderly because of age-related changes in renal function, persons who are dehydrated or have a decrease in blood volume, and persons with pre-existing kidney disease or renal insufficiency.

In summary, tubulointerstitial diseases affect the tubules and the surrounding interstitium of the kidneys. These disorders include pyelonephritis and the effects of drugs and toxins. Pyelonephritis, or infection of the kidney and kidney pelvis, can occur as an acute or a chronic condition. Acute pyelonephritis typically is caused by ascending bladder infections or infections that come from the bloodstream; it usually is successfully treated with appropriate antimicrobial drugs. Chronic pyelonephritis is a progressive disease that produces scarring and deformation of the renal calices and pelvis. Druginduced impairment of tubulointerstitial structure and function usually is the result of direct toxic injury, decreased blood flow, or hypersensitivity reactions.

NEOPLASMS

There are two major groups of renal neoplasms: embryonic kidney tumors (*i.e.*, Wilms' tumor), which occur during childhood, and renal cell carcinoma, which is the main kidney cancer in adults.

Wilms' Tumor

Wilms' tumor (*i.e.*, nephroblastoma) is one of the most common primary neoplasms of young children. The median age at time of diagnosis of unilateral Wilms' tumor is approximately 3 years.³¹ It is a mixed tumor, composed of epithelial and mesenchymal embryonic tissue elements.32 An important feature of Wilms' tumor is its association with congenital anomalies, the most frequent being those affecting genitourinary structures.³¹ Deletions involving at least two loci on chromosome 11 have been found in approximately 30% of children with Wilms' tumors. Some familial cases of Wilms' tumor are not associated with identifiable chromosomal deletions or mutations, suggesting a third locus may be involved.

Wilms' tumor usually is a solitary mass that occurs in any part of the kidney. It usually is sharply demarcated and variably encapsulated.6 The tumors grow to a large size, distorting kidney structure. The tumors usually are staged using the Wilms' Tumor Study Group classification.³¹ Stage I tumors are limited to the kidney and can be excised with the capsular surface intact. Stage II tumors extend into the kidney but can be excised. In stage III, extension of the tumor is confined to the abdomen, and in stage IV, hematogenous metastasis most commonly involves the lung. Bilateral kidney involvement occurs in 5% to 10% of cases.

The common presenting signs are a large asymptomatic abdominal mass and hypertension. Approximately 50% of children have abdominal pain, vomiting, or both.³¹ Microscopic and gross hematuria are present in 10% to 25% of children. CT scans are used to confirm the diagnosis.

Treatment involves surgery, chemotherapy, and sometimes radiation therapy. Long-term survival rates have increased to approximately 90% with an aggressive treatment plan.³¹

Renal Cell Carcinoma

Kidney cancer accounts for 2% of all adult cancer incidence and mortality in the United States.³³ The increased use of imaging procedures such as ultrasonography, CT scanning, and magnetic resonance imaging (MRI) has contributed significantly to earlier diagnosis and more accurate staging of kidney cancers.³³

Renal cell carcinoma originates in the renal cortex and accounts for approximately 80% to 85% of kidney tumors, with transitional or squamous cell cancers of the renal pelvis accounting for most of the remaining cancers.³⁴ The cause of renal cell carcinoma remains unclear. It occurs most often in older persons in the sixth to seventh decade. Men are affected twice as frequently as women. Some of these tumors may occur as a result of chronic irritation associated with kidney stones. Epidemiologic evidence suggests a correlation between smoking and kidney cancer. Obesity also is a risk factor, particularly in women.34 Additional risk factors include occupational exposure to petroleum products, heavy metals, and asbestos. The risk of renal cell carcinoma also is increased in persons with acquired cystic kidney disease associated with chronic renal insufficiency.^{34,35} Most cases of renal cell carcinoma occur without a recognizable hereditary pattern. However, there are several rare forms of renal cell cancer that are characterized by an autosomal dominant pattern of inheritance, young age at onset (third and fourth decade), and bilateral or multifocal tumors.34,35

Kidney cancer is largely a silent disorder during its early stages, and symptoms usually denote advanced disease. Presenting features include hematuria, costovertebral pain, presence of a palpable flank mass, polycythemia, and fever. Hematuria, which occurs in 70% to 90% of cases, is the most reliable sign. However, it is intermittent and may be microscopic; as a result, the tumor may reach considerable size before it is detected. In approximately one third of cases, metastases are present at the time of diagnosis.

Kidney cancer is suspected when there are findings of hematuria and a renal mass. Ultrasonography, CT scanning, excretory urography, and renal angiography are used to confirm the diagnosis. MRI with intravenous gadolinium may be used when involvement of the inferior vena cava is suspected.

Surgery (radical nephrectomy with lymph node dissection) is the treatment of choice for all resectable tumors. Nephronsparing surgery may be done when both kidneys are involved or when the contralateral kidney is threatened by an associated disease such as hypertension or diabetes mellitus. Single-agent and combination chemotherapy have been used with limited success. The prognosis depends on the stage of the cancer; the 5-year survival rate associated with stage I tumors is 65% to 85%; for stage IV tumors, it is 0% to 10%.34,35

In summary, there are two major groups of renal neoplasms: embryonic kidney tumors (i.e., Wilms' tumor) that occur during childhood and adult renal cell carcinomas. Wilms' tumor is the most common malignant tumor of children. The most common presenting signs are a large abdominal mass and hypertension. Treatment is surgery, chemotherapy, and sometimes radiation therapy. The long-term survival rate for children with Wilms' tumor is approximately 90% with an aggressive plan of treatment.

Renal cell carcinoma accounts for 2% of all adult cancers. These tumors are characterized by a lack of early warning signs, diverse clinical manifestations, and resistance to chemotherapy and radiation therapy. Because of the lack of early warning signs, the tumors often are far advanced at the time of diagnosis. Diagnostic methods include ultrasonography and CT scans. The treatment of choice is surgical resection.

REVIEW QUESTIONS

Define the terms agenesis, dysgenesis, and hypoplasia as they refer to the development of the kidney.

Describe the genetic basis for renal cystic disease, the pathology of the disorder, and its signs and symptoms.

Describe the effects of urinary tract obstruction on renal structure and function.

Cite three theories that are used to explain the formation of kidney stones.

Explain the mechanisms of pain and infection that occur with kidney stones.

Cite the organisms most responsible for UTIs and state why urinary catheters, obstruction, and reflux predispose to infections.

List three physiologic mechanisms that protect against UTIs.

Compare the manifestations of UTIs in different age groups, including infants, toddlers, adolescents, adults, and older adults.

Use the terms *proliferation, sclerosis, membranous, diffuse, focal, segmental,* and *mesangial* to explain changes in glomerular structure that occur with glomerulonephritis.

Relate the proteinuria, hematuria, pyuria, oliguria, edema, hypertension, and azotemia that occur with glomerulonephritis to changes in glomerular structure.

Cite a definition of tubulointerstitial kidney disease.

Explain the vulnerability of the kidneys to injury caused by drugs and toxins.

Characterize Wilms' tumor in terms of age of onset, possible oncogenic origin, manifestations, and treatment.

Cite the risk factors for renal cell carcinoma, describe the manifestations, and explain why the 5-year survival rate has been so low.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- National Kidney Foundation. (2000). Facts about transplantation and kidney and urologic diseases. [On-line]. Available: http://www.kidney.org.
- Moore K.L., Persaud T.V.N. (1998). The developing human: Clinically oriented embryology (6th ed., pp. 305–315). Philadelphia: W.B. Saunders.
- Stewart C.L., Jose P.A. (1991). Transitional nephrology. Urologic Clinics of North America 18, 143–149.
- Elder J.S. (2000). Urologic disorders in infants and children. In Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.), *Nelson textbook of pediatrics* (16th ed., pp. 619–623, 1621–1623). Philadelphia: W.B. Saunders.
- Cotran R.S., Kumar V., Collins T. (1999). Robbins pathologic basis of disease (6th ed., pp. 936–965, 971–979). Philadelphia: W.B. Saunders.
- Jennette J.C., Spargo B.H. (1999). The kidney. In Rubin E., Farber J.L. (Eds.), *Pathology* (3rd ed., pp. 865–893, 902–907, 913–917). Philadelphia: Lippincott Williams & Wilkins.
- Germino G.G. (1997). Autosomal dominant polycystic kidney disease: A two-hit model. *Hospital Practice* 32 (3), 81–92.
- Gabow P.A. (1993). Autosomal dominant polycystic kidney disease. New England Journal of Medicine 329, 332–342.
- Tanagho E.A. (2000). Urinary obstruction and stasis. In Tanagho E.A., McAninch J.W. (Eds.), *Smith's general urology* (15th ed., pp. 208–219, 291–320). New York: Lange Medical Books/McGraw-Hill.
- Stoller M.L., Presti J.C., Carroll P.R. (2001). Urology. In Tierney L.M., McPhee S.J., Papadakis M.A. (Eds.), *Current medical diagnosis and treatment* (40th ed., pp. 939–943). New York: Lange Medical Books/ McGraw-Hill.
- 11. Coe F.L., Parks J.H., Asplin J.R. (1992). The pathogenesis and treatment of kidney stones. *New England Journal of Medicine* 327, 1141–1152.

- Mandel N. (1996). Mechanisms of stone formation. Seminars in Nephrology 16, 364–374.
- 13. Worchester E.M. (1996). Inhibitors of stone formation. Seminars in Nephrology 16, 474–486.
- McRae S.N., Shortliffe L.M. (2000). Bacterial infections of the genitourinary tract. In Tanagho E.A., McAninch J.W. (Eds.), *Smith's general urology* (15th ed., pp. 237–264). New York: Lange Medical Books/ McGraw-Hill.
- Watnick S., Morrison G. (2001). Kidney. In Tierney L.M., McPhee S.J., Papadakis M.A. (Eds.), *Current medical diagnosis and treatment* (40th ed., pp. 912–923, 932–936). New York: Lange Medical Books/ McGraw-Hill.
- 16. Roberts J.A. (1999). Management of pyelonephritis and upper urinary tract infections. *Urologic Clinics of North America* 26, 753–763.
- Hooton T.M., Scholes D., Hughes J.P., et al. (1996). A prospective study of risk factors for symptomatic urinary tract infections in young women. *New England Journal of Medicine* 335, 468–474.
- Delzell J.E., Lefevre M.L. (2000). Urinary tract infections during pregnancy. American Family Physician 61, 713–721.
- American College of Obstetricians and Gynecologists. (1998). Antimicrobial therapy for obstetric patients (pp. 8–10). ACOG Educational Bulletin no. 245. Washington, DC: Author.
- Shaw K.N., Gorelick M.H. (1999). Urinary tract infections in children. Pediatric Clinics of North America 46, 1111–1122.
- Johnson C.E. (1999). New advances in childhood urinary tract infections. *Pediatrics in Review* 20, 335–342.
- Yoshikawa T.T. (1993). Chronic urinary tract infections in elderly patients. *Hospital Practice* 28(6), 103–118.
- Mouton C.P., Pierce B., Espino D.V. (2001). Common infections in older adults. *American Family Physician* 63, 257–268.
- Hricik D.E., Chung-Park M., Sedor J.R. (1998). Glomerulonephritis. New England Journal of Medicine 339, 888–899.
- Orth S.R., Ritz E. (1998). The nephrotic syndrome. New England Journal of Medicine 339, 1202–1211.
- Jennette J.C., Falk R.J. (1997). Diagnosis and management of glomerular disease. *Medical Clinics of North America* 81(3), 653–675.
- Vincenti F.G., Amend W.J.C. (2000). Diagnosis of medical renal diseases. In Tanagho E.A., McAninch J.W. (Eds.), *Smith's general urology* (15th ed., pp. 594–599). New York: Lange Medical Books/ McGraw-Hill.
- Remuzzi G., Bertani T. (1998). Pathophysiology of progressive nephropathies. *New England Journal of Medicine* 339, 1448–1455.
- Parving H.-H., Østerby R., Ritz E. (2000). Diabetic nephropathy. In Brenner B.M. (Ed.), *Brenner and Rector's the kidney* (6th ed., pp. 1731–1753). Philadelphia: W.B. Saunders.
- Palmer B., Hendrich W.L. (1995). Clinical acute renal failure with nonsteroidal anti-inflammatory drugs. *Seminars in Nephrology* 15, 214–227.
- Anderson P.M. (2000). Neoplasms of the kidney. In Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.), *Nelson textbook of pediatrics* (16th ed., pp. 1554–1556). Philadelphia: W.B. Saunders.
- Schofield D., Cotran R.S. (1999). Diseases of infancy and childhood. In Cotran R.S., Kumar V., Collins T. (Eds.), *Robbins pathologic basis of disease* (6th ed., pp. 487–489). Philadelphia: W.B. Saunders.
- 33. Chow W.H., Devesa S.S., Warren J.L., et al. (1999). Rising incidence of renal cell cancer in the United States. *Journal of the American Medical Association* 281, 1628–1631.
- Motzer R.J., Bander N.H., Nanus D.M. (1997). Medical progress: Renal-cell carcinoma. New England Journal of Medicine 335, 865–875.
- Dreicer R., Williams R.D. (2000). Renal parenchymal neoplasms. In Tanagho E.A., McAninch J.W. (Eds.), *Smith's general urology* (15th ed., pp. 378–394). New York: Lange Medical Books/McGraw-Hill.

Renal Failure

Acute Renal Failure **Types of Acute Renal Failure Prerenal Failure Postrenal Failure Intrinsic Renal Failure Clinical Course Diagnosis and Treatment Chronic Renal Failure Stages of Progression Diminished Renal Reserve Renal Insufficiency Renal Failure End-Stage Renal Disease Clinical Manifestations** Accumulation of Nitrogenous Wastes Disorders of Water, Electrolyte, and Acid-Base Balance Mineral Metabolism and Skeletal Disorders Hematologic Disorders Cardiovascular Disorders **Gastrointestinal Disorders Disorders of Neural Function Altered Immune Function Disorders of Skin Integrity** Sexual Dysfunction **Elimination of Drugs Treatment Medical Management Dialysis and Transplantation Renal Failure in Children and Elderly Persons Chronic Renal Failure in Children Chronic Renal Failure in Elderly Persons**

CHAPTER

76

Renal failure is a condition in which the kidneys fail to remove metabolic end-products from the blood and regulate the fluid, electrolyte, and pH balance of the extracellular fluids. The underlying cause may be renal disease, systemic disease, or urologic defects of nonrenal origin. Renal failure can occur as an acute or a chronic disorder. Acute renal failure is abrupt in onset and often is reversible if recognized early and treated appropriately. In contrast, chronic renal failure is the end result of irreparable damage to the kidneys. It develops slowly, usually over the course of a number of years.

ACUTE RENAL FAILURE

Acute renal failure represents a rapid decline in renal function sufficient to increase blood levels of nitrogenous wastes and impair fluid and electrolyte balance. It is a common threat to seriously ill persons in intensive care units, with a mortality rate ranging from 42% to 88%.¹ Although treatment methods such as dialysis and renal replacement methods are effective in correcting life-threatening fluid and electrolyte disorders, the mortality rate associated with acute renal failure has not changed substantially since the 1960s.^{2,3} This probably is because acute renal failure is seen more often in older persons than before, and because it frequently is superimposed on other life-threatening conditions, such as trauma, shock, and sepsis.

The most common indicator of acute renal failure is *azotemia*, an accumulation of nitrogenous wastes (urea nitrogen, uric acid, and creatinine) in the blood. In acute renal failure the glomerular filtration rate (GFR) is decreased. As a result, excretion of nitrogenous wastes is reduced and fluid and electrolyte balance cannot be maintained. Persons with acute renal failure often are asymptomatic, and the condition is diagnosed by observation of elevations in blood urea nitrogen (BUN) and creatinine.

Types of Acute Renal Failure

Acute renal failure can be caused by several types of conditions, including a decrease in blood flow without ischemic injury; ischemic, toxic, or obstructive tubular injury; and obstruction of urinary tract outflow. The causes of acute renal failure commonly are categorized as prerenal (55% to 60%), postrenal

KEY CONCEPTS

ACUTE RENAL FAILURE

- Acute renal failure is caused by conditions that produce an acute shutdown in renal function.
- It can result from decreased blood flow to the kidney (prerenal failure), disorders that interfere with the elimination of urine from the kidney (postrenal failure), or disorders that disrupt the structures in the kidney (intrinsic or intrarenal failure).
- Acute renal failure, although it causes an accumulation of products normally cleared by the kidney, is a reversible process if the factors causing the condition can be corrected.

(<5%), and intrinsic (35% to 40%).³ Causes of renal failure within these categories are summarized in Chart 24-1.

Prerenal Failure

Prerenal failure, the most common form of acute renal failure, is characterized by a marked decrease in renal blood flow. It is reversible if the cause of the decreased renal blood flow can be identified and corrected before kidney damage occurs. Causes of prerenal failure include profound depletion of vascular volume (*e.g.*, hemorrhage, loss of extracellular fluid volume), im-

CHART 24-1 Causes of Acute Renal Failure				
Prerenal				
Hypovolemia				
Hemorrhage				
Dehydration				
Excessive loss of gastrointestinal tract fluids				
Excessive loss of fluid due to burn injury				
Decreased vascular filling				
Anaphylactic shock				
Septic shock Heart failure and cardiogenic shock				
Decreased renal perfusion due to vasoactive mediators,				
drugs, diagnostic agents				
Intrinsic or Intrarenal				
Acute tubular necrosis				
Prolonged renal ischemia				
Exposure to nephrotoxic drugs, heavy metals,				
and organic solvents				
Intratubular obstruction resulting from hemoglobinuria,				
myoglobinuria, myeloma light chains, or uric acid casts				
Acute renal disease (acute glomerulonephritis,				
pyelonephritis)				
Postrenal				
Bilateral ureteral obstruction				
Bladder outlet obstruction				

paired perfusion caused by heart failure and cardiogenic shock, and decreased vascular filling because of increased vascular capacity (*e.g.*, anaphylaxis or sepsis). Elderly persons are particularly at risk because of their predisposition to hypovolemia and their high prevalence of renal vascular disorders.

Some vasoactive mediators, drugs, and diagnostic agents stimulate intense intrarenal vasoconstriction and induce glomerular hypoperfusion and prerenal failure.3 Examples include hypercalcemia, endotoxins, and radiocontrast agents such as those used for cardiac catheterization.³ Many of these agents also cause acute tubular necrosis (discussed later). In addition, several commonly used classes of drugs impair renal adaptive mechanisms and can convert compensated renal hypoperfusion into prerenal failure. Angiotensin-converting enzyme (ACE) inhibitors reduce the effects of renin on renal blood flow; when combined with diuretics, they may cause prerenal failure in persons with decreased blood flow caused by large-vessel or smallvessel renal vascular disease. Prostaglandins have a vasodilatory effect on renal blood vessels. Nonsteroidal anti-inflammatory drugs (NSAIDs) reduce renal blood flow through inhibition of prostaglandin synthesis. In some persons with diminished renal perfusion, NSAIDs can precipitate prerenal failure.

Normally, the kidneys receive 20% to 25% of the cardiac output.⁴ This large blood supply is required to remove metabolic wastes and regulate body fluids and electrolytes. Fortunately, the normal kidney can tolerate relatively large reductions in blood flow before renal damage occurs. As renal blood flow is reduced, the GFR drops, the amount of sodium and other substances that is filtered by the glomeruli is reduced, and the need for energy-dependent mechanisms to reabsorb these substances is reduced. As the GFR and urine output approach zero, oxygen consumption by the kidney approximates that required to keep renal tubular cells alive.4 When blood flow falls below this level, which is about 20% of normal, ischemic changes occur. Because of their high metabolic rate, the tubular epithelial cells are most vulnerable to ischemic injury. Improperly treated, prolonged renal hypoperfusion can lead to ischemic tubular necrosis with significant morbidity and mortality.

Acute renal failure is manifested by a sharp decrease in urine output and a disproportionate elevation of BUN in relation to serum creatinine levels. The kidney normally responds to a decrease in the GFR with a decrease in urine output. An early sign of prerenal failure is a sharp decrease in urine output. BUN levels also depend on the GFR. A low GFR allows more time for small particles such as urea to be reabsorbed into the blood. Creatinine, which is larger and nondiffusible, remains in the tubular fluid, and the total amount of creatinine that is filtered, although small, is excreted in the urine. Thus, there also is a disproportionate elevation in the ratio of BUN to serum creatinine to greater than 20:1 (normal, approximately 10:1).

Postrenal Failure

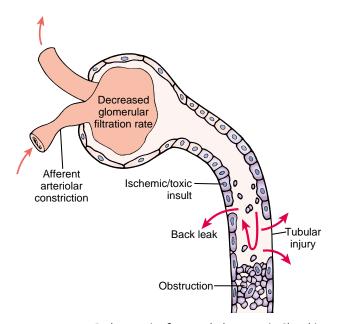
Postrenal failure results from obstruction of urine outflow from the kidneys. The obstruction can occur in the ureter (*i.e.*, calculi and strictures), bladder (*i.e.*, tumors or neurogenic bladder), or urethra (*i.e.*, prostatic hypertrophy). Prostatic hyperplasia is the most common underlying problem. Because both ureters must be occluded to produce renal failure, obstruction of the bladder rarely causes acute renal failure unless one of the kidneys already is damaged or a person has only one kidney. The treatment of acute postrenal failure consists of treating the underlying cause of obstruction so that urine flow can be re-established before permanent nephron damage occurs.

Intrinsic Renal Failure

Intrinsic or intrarenal renal failure results from conditions that cause damage to structures within the kidney-glomerular, tubular, or interstitial. Injury to the tubules is most common and often is ischemic or toxic in origin. The major causes of intrarenal failure are ischemia associated with prerenal failure, toxic insult to the tubular structures of the nephron, and intratubular obstruction. Acute glomerulonephritis and acute pyelonephritis also are intrarenal causes of acute renal failure.

Acute Tubular Necrosis. Acute tubular necrosis (ATN) is characterized by destruction of tubular epithelial cells with acute suppression of renal function (Fig. 24-1).^{3,5} It is the most common cause of intrinsic renal failure. ATN can be caused by a variety of conditions, including acute tubular damage caused by ischemia, the nephrotoxic effects of drugs, tubular obstruction, and toxins from a massive infection. The tubular injury that occurs in ATN frequently is reversible. The process depends on the recovery of the injured cells, removal of the necrotic cells and intratubular casts, and regeneration of renal cells to restore the normal continuity of the tubular epithelium. However, if the ischemia is severe enough to cause cortical necrosis, irreversible renal failure occurs.

Ischemic ATN occurs most frequently in persons who have major surgery, severe hypovolemia, overwhelming sepsis, trauma, and burns.3 Sepsis produces ischemia by provoking a



■ FIGURE 24-1 ■ Pathogenesis of acute tubular necrosis. Sloughing and necrosis of tubular epithelial cells leading to obstruction and increased intraluminal pressure, which reduces glomerular filtration. Afferent arteriolar vasoconstriction, caused in part by tubuloglomular feedback, results in decreased glomerular capillary filtration pressure. Tubular injury and increased intraluminal pressure cause fluid to move from the tubular lumen into the interstitium (backleak). (Modified from Rubin E., Farber J.L. [1999]. Pathology [3rd ed., p. 901]. Philadelphia: Lippincott Williams & Wilkins)

combination of systemic vasodilation and intrarenal hypoperfusion. In addition, sepsis results in the generation of toxins that sensitize renal tubular cells to the damaging effects of ischemia. ATN complicating trauma and burns frequently is multifactorial in origin and caused by the combined effects of hypovolemia and myoglobin or other toxins released from damaged tissue. In contrast to prerenal failure, the GFR does not improve with the restoration of renal blood flow in acute renal failure caused by ischemic ATN.

Nephrotoxic ATN complicates the administration of or exposure to many structurally diverse drugs and other toxic agents. Nephrotoxic agents cause renal injury by inducing varying combinations of renal vasoconstriction, direct tubular damage, or intratubular obstruction. The kidney is particularly vulnerable to nephrotic injury because of its rich blood supply and ability to concentrate toxins to high levels in the medullary portion of the kidney. In addition, the kidney is an important site for metabolic processes that transform relatively harmless agents into toxic metabolites. Pharmacologic agents that are directly toxic to the renal tubule include antimicrobial agents such as the aminoglycosides, chemotherapeutic agents such as cisplatin and ifosfamide, and the radiocontrast agents used during cardiac catheterization and other diagnostic procedures.^{5,6} Radiocontrast media-induced nephrotoxicity is thought to result from direct tubular toxicity and renal ischemia.⁷ The risk of renal damage caused by radiocontrast media is greatest in elderly persons, in persons with diabetes mellitus, and in persons who, for various reasons, are susceptible to kidney disease.3 Heavy metals (e.g., lead, mercury) and organic solvents (e.g., carbon tetrachloride, ethylene glycol) are other nephrotoxic agents.

Myoglobin, hemoglobin, uric acid, and myeloma light chains are the most common cause of ATN attributable to intratubular obstruction. Deposits of immunoglobulins and urine acid crystals usually are seen in the setting of widespread malignancy or massive tumor destruction by therapeutic agents.³ Hemoglobinuria results from blood transfusion reactions and other hemolytic crises. Skeletal and cardiac muscles contain myoglobin, which accounts for their red color. Myoglobin corresponds to hemoglobin in function, serving as an oxygen reservoir in the muscle fibers. Myoglobin normally is not found in the serum or urine. Myoglobinuria most commonly results from muscle trauma but may result from extreme exertion, hyperthermia, sepsis, prolonged seizures, potassium or phosphate depletion, and alcoholism or drug abuse. Both myoglobin and hemoglobin discolor the urine, which may range from the color of tea to red, brown, or black.

Clinical Course

The course of acute renal failure can be divided into three phases: the onset or initiating phase, the maintenance phase, and the recovery or convalescent phase.³ The onset or initiating phase, which lasts hours or days, is the time from the onset of the precipitating event (e.g., ischemic phase of prerenal failure or toxin exposure) until tubular injury occurs.

The maintenance phase is characterized by a marked decrease in the GFR, causing sudden retention of endogenous metabolites such as urea, potassium, sulfate, and creatinine that normally are cleared by the kidneys. The urine output usually is lowest at this point. Fluid retention gives rise to edema, water

436 Unit Six: Alterations in the Urinary System

intoxication, and pulmonary congestion. If the period of oliguria is prolonged, hypertension frequently develops and with it signs of uremia. When untreated, the neurologic manifestations of uremia progress from neuromuscular irritability to seizures, somnolence, coma, and death. Hyperkalemia usually is asymptomatic until serum levels of potassium rise above 6.0 to 6.5 mEq/L, at which point characteristic electrocardiographic changes and symptoms of muscle weakness are seen.

The *recovery phase* is the period during which repair of renal tissue takes place. Its onset usually is heralded by a gradual increase in urine output and a fall in serum creatinine, indicating that the nephrons have recovered to the point where urine excretion is possible. Diuresis often occurs before renal function has fully returned to normal. Consequently, BUN and serum creatinine, potassium, and phosphate levels may remain elevated or continue to rise even though urine output is increased. In some cases, the diuresis may result from impaired nephron function and may cause excessive loss of water and electrolytes. Eventually, renal tubular function is restored with improvement in urine concentrating ability. At about the same time, the BUN and creatinine begin to return to normal. In some cases, mild to moderate kidney damage persists.

Diagnosis and Treatment

Given the high morbidity and mortality rates associated with acute renal failure, attention should be focused on prevention and early diagnosis. This includes assessment measures to identify persons at risk for the development of acute renal failure, including those with pre-existing renal insufficiency and diabetes. Elderly persons are susceptible to all forms of acute renal failure because of the effects of aging on renal reserve.

Careful observation of urine output is essential for persons at risk for the development of acute renal failure. Urine output and urine osmolality or specific gravity should be carefully monitored. One of the earliest manifestations of tubular damage is the inability to concentrate the urine. Further diagnostic information that can be obtained from the urinalysis includes evidence of proteinuria, hemoglobinuria, and casts or crystals in the urine. Blood tests for BUN and creatinine provide information regarding the ability to remove nitrogenous wastes from the blood.

A major concern in the treatment of acute renal failure is identifying and correcting the cause (*e.g.*, improving renal perfusion, discontinuing nephrotoxic drugs). Fluids are carefully regulated in an effort to maintain normal fluid volume and electrolyte concentrations. Adequate caloric intake is needed to prevent the breakdown of body proteins, which increases the need for elimination of nitrogenous wastes. Parenteral hyperalimentation may be used for this purpose. Because secondary infections are a major cause of death in persons with acute renal failure, constant effort is needed to prevent and treat such infections.

Dialysis or continuous renal replacement therapy (CRRT) may be indicated when nitrogenous wastes and the water and electrolyte balance cannot be kept under control by other means.

In summary, acute renal failure is an acute, reversible suppression of kidney function. It is a common threat to seriously ill persons in intensive care units and is associated with a high mortality rate. Acute renal failure is characterized by an accumulation of nitrogenous wastes in the blood (*i.e.*, azo-temia) and alterations in body fluids and electrolytes. Acute renal failure is classified as prerenal, postrenal, or intrarenal in origin. Prerenal failure is caused by decreased blood flow to the kidneys; postrenal failure by obstruction to urine output; and intrinsic renal failure by disorders in the kidney itself. ATN, caused by ischemia or nephrotoxic agents, is a common cause of acute intrinsic renal failure. Acute renal failure typically progresses through three phases: the initiation phase, during which tubular injury is induced; the maintenance phase, during which the GFR falls, nitrogenous wastes accumulate, and urine output decreases; and the recovery or reparative phase, during which the GFR, urine output, and blood levels of nitrogenous wastes return to normal.

Because of the high morbidity and mortality rates associated with acute renal failure, identification of persons at risk is important to clinical decision making. Acute renal failure often is reversible, making early identification and correction of the underlying cause (*e.g.*, improving renal perfusion, discontinuing nephrotoxic drugs) important. Treatment includes the judicious administration of fluids and dialysis or CRRT.

CHRONIC RENAL FAILURE

Unlike acute renal failure, chronic renal failure represents progressive and irreversible destruction of kidney structures. As recently as 1965, many patients with chronic renal failure pro-

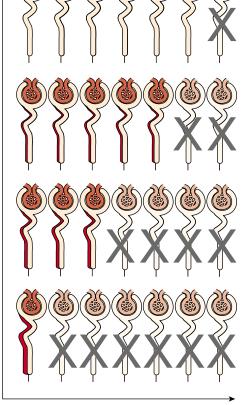
KEY CONCEPTS CHRONIC RENAL FAILURE

- Chronic renal failure represents the end result of conditions that greatly reduce renal function by destroying renal nephrons and producing a marked decrease in the glomerular filtration rate (GFR).
- Signs of renal failure begin to appear as renal function moves from renal insufficiency (GFR 50% to 20% normal), to renal failure (20% to 5% normal), to end-stage renal disease (<5% normal). When the GFR decreases to less than 5% of normal, dialysis or kidney transplantation is necessary for survival.
- The manifestations of chronic renal failure represent the inability of the kidney to perform its normal functions in terms of regulating fluid and electrolyte balance, controlling blood pressure through fluid volume and the renin-angiotensin system, eliminating nitrogenous and other waste products, governing the red blood cell count through erythropoietin synthesis, and directing parathyroid and skeletal function through phosphate elimination and activation of vitamin D.

gressed to the final stages of the disease and then died. The high mortality rate was associated with limitations in the treatment of renal disease and with the tremendous cost of ongoing treatment. In 1972, federal support began for dialysis and transplantation through a Medicare entitlement program.⁸ Technologic advances in renal replacement therapy (*i.e.*, dialysis therapy and transplantation) have improved the outcomes for persons with renal failure. In the United States, there are approximately 400,000 persons with end-stage renal disease who are living today, a product of continued research and advances in treatment methods.⁹

Chronic renal failure can result from a number of conditions that cause permanent loss of nephrons, including diabetes, hypertension, glomerulonephritis, and polycystic kidney disease. Typically, the signs and symptoms of renal failure occur gradually and do not become evident until the disease is far advanced. This is because of the amazing compensatory ability of the kidneys. As kidney structures are destroyed, the remaining nephrons undergo structural and functional hypertrophy, each increasing its function as a means of compensating for those that have been lost (Fig. 24-2). It is only when the few remaining nephrons are destroyed that the manifestations of renal failure become evident.

Glomerular filtration rate



Number of functioning nephrons

■ **FIGURE 24-2** ■ Relation of renal function and nephron mass. Each kidney contains 1 million tiny nephrons. A proportional relation exists between the number of nephrons affected by disease and the resulting glomerular filtration rate. Regardless of cause, chronic renal failure results in progressive deterioration of glomerular filtration, tubular reabsorptive capacity, and endocrine functions of the kidneys. All forms of renal failure are characterized by a reduction in the GFR, reflecting a corresponding reduction in the number of functional nephrons.

Stages of Progression

The rate of nephron destruction differs from case to case, ranging from several months to many years. The progression of chronic renal failure usually occurs in four stages: diminished renal reserve, renal insufficiency, renal failure, and end-stage renal disease.⁵

Diminished Renal Reserve

Diminished renal reserve occurs when the GFR drops to approximately 50% of normal. At this point, the serum BUN and creatinine levels still are normal, and no symptoms of impaired renal function are evident. This is supported by the fact that many persons survive an entire lifetime with only one kidney. Because of the diminished reserve, the risk for development of azotemia increases with an additional renal insult, such as that caused by nephrotoxic drugs.

Renal Insufficiency

Renal insufficiency represents a reduction in the GFR to 20% to 50% of normal. During this stage, azotemia, anemia, and hypertension appear. Signs and symptoms of renal insufficiency do not begin to appear until more than 50% of the function in both kidneys is lost. As nephrons are destroyed, the remaining nephrons compensate for those that are lost by filtering more solute particles from the blood. Because the solute particles are osmotically active, they cause additional water to be lost in the urine. One of the earliest symptoms of renal insufficiency is isosthenuria or polyuria with urine that is almost isotonic with plasma.⁴

Conservative treatment during this stage includes measures to retard the deterioration of renal function and assist the body in managing the effects of impaired function. Because the kidneys have difficulty eliminating the waste products of protein metabolism, a restricted-protein diet usually produces fewer symptoms and slows progression of renal failure. The few remaining nephrons that constitute the functional reserve of the kidneys can be easily disrupted; at that point, renal failure progresses rapidly.

Renal Failure

Renal failure develops when the GFR is less than 20% of normal. At this point, the kidneys cannot regulate volume and solute composition, and edema, metabolic acidosis, and hyperkalemia develop. These alterations affect other body systems to cause neurologic, gastrointestinal, and cardiovascular manifestations.

End-Stage Renal Disease

End-stage renal disease (ESRD) occurs when the GFR is less than 5% of normal. Histologic findings of an end-stage kidney include a reduction in renal capillaries and scarring in the glomeruli. Atrophy and fibrosis are evident in the tubules. The mass of the kidneys usually is reduced. At this final phase of renal failure, treatment with dialysis or transplantation is necessary for survival.

Clinical Manifestations

The clinical manifestations of renal failure include alterations in water, electrolyte, and acid-base balance; mineral and skeletal disorders; anemia and coagulation disorders; hypertension and alterations in cardiovascular function; gastrointestinal disorders; neurologic complications; disorders of skin integrity; and immunologic disorders (Fig. 24-3). *Uremia*, which literally means "urine in the blood," is the term used to describe the clinical manifestations of ESRD. Uremia differs from azotemia, which merely indicates the accumulation of nitrogenous wastes in the blood and can occur without symptoms.

There currently are four target populations that comprise the entire population of persons with chronic renal failure: persons with chronic renal insufficiency, those with ESRD being treated with hemodialysis, those being treated with peritoneal dialysis, and renal transplant recipients. The manifestations of renal failure are determined largely by the extent of renal function that is present (*e.g.*, renal insufficiency, ESRD), coexisting disease conditions, and the type of renal replacement therapy the person is receiving.

Accumulation of Nitrogenous Wastes

The accumulation of nitrogenous wastes is an early sign of renal failure, usually occurring before other symptoms become evident. Urea is one of the first nitrogenous wastes to accumulate in the blood, and the BUN level becomes increasingly elevated as renal failure progresses. The normal concentration of urea in the plasma is usually less than 20 mg/dL. In renal failure, this level may rise to as high as 800 mg/dL. Creatinine, a by-product of muscle metabolism, is freely filtered in the glomerulus and is not reabsorbed in the renal tubules. Creatinine is produced at a relatively constant rate, and any creatinine that is filtered in the glomerulus is lost in the urine, rather than being reabsorbed into the blood. Thus, serum creatinine can be used as an indirect method for assessing the GFR and the extent of renal damage that has occurred in renal failure. Because creatinine is a by-product of muscle metabolism, serum values vary with age and muscle mass. An increase in serum creatinine to three times its normal value suggests that there is a 75% loss of renal function, and with creatinine levels of 10 mg/dL or more, it can be assumed that 90% of renal function has been lost. (See Chapter 22, Fig. 22-12.)

Disorders of Water, Electrolyte, and Acid-Base Balance

The kidneys function in the regulation of extracellular fluid volume. They do this by either eliminating or conserving sodium and water. Chronic renal failure can produce dehydration or fluid overload, depending on the pathology of the renal disease. In addition to volume regulation, the ability of the kidneys to concentrate the urine is diminished. With early losses in renal function, the specific gravity of the urine becomes fixed (1.008 to 1.012) and varies little from voiding to voiding. Polyuria and nocturia are common.

As renal function declines further, the ability to regulate sodium excretion is reduced. There is impaired ability to adjust to a sudden reduction in sodium intake and poor tolerance of an acute sodium overload. Volume depletion with an accompanying decrease in the GFR can occur with a restricted sodium intake or excess sodium loss caused by diarrhea or vomiting. Salt wasting is a common problem in advanced renal failure because of impaired tubular reabsorption of sodium. Increas-

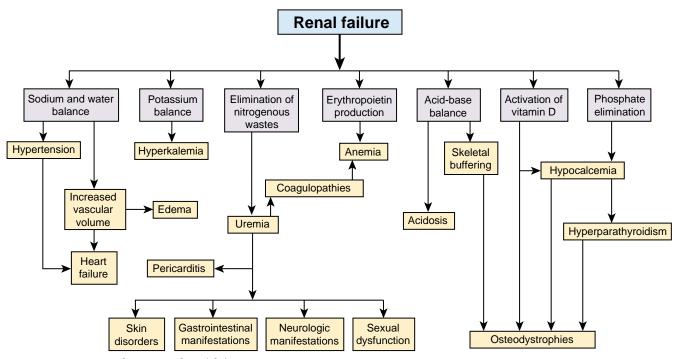


FIGURE 24-3 Manifestations of renal failure.

ing sodium intake in persons with chronic renal failure often improves the GFR and whatever renal function remains. In patients with associated hypertension, the possibility of increasing blood pressure or production of congestive heart failure often excludes supplemental sodium intake.

Approximately 90% of potassium excretion is through the kidneys. In renal failure, potassium excretion by each nephron increases as the kidneys adapt to a decrease in the GFR. As a result, hyperkalemia usually does not develop until renal function is severely compromised. Because of this adaptive mechanism, it usually is not necessary to restrict potassium intake in patients with chronic renal failure until the GFR has dropped below 5 mL/minute. In persons with ESRD, hyperkalemia often results from the failure to follow dietary potassium restrictions and ingestion of medications that contain potassium, or from an endogenous release of potassium, as in trauma or infection.

The kidneys normally regulate blood pH by eliminating hydrogen ions produced in metabolic processes and regenerating bicarbonate. This is achieved through hydrogen ion secretion, sodium and bicarbonate reabsorption, and the production of ammonia, which acts as a buffer for titratable acids (see Chapter 6). With a decline in renal function, these mechanisms become impaired, and metabolic acidosis results. In chronic renal failure, acidosis seems to stabilize as the disease progresses, probably as a result of the tremendous buffering capacity of bone. However, this buffering action is thought to increase bone resorption and contribute to the skeletal defects present in chronic renal failure.

Mineral Metabolism and Skeletal Disorders

Abnormalities of calcium, phosphate, and vitamin D occur early in the course of chronic renal failure.¹⁰ They involve the renal regulation of serum calcium and phosphate levels, activation of vitamin D, and regulation of parathyroid hormone (PTH) levels.

The regulation of serum phosphate levels requires a daily urinary excretion of an amount equal to that absorbed from the diet. With deteriorating renal function, phosphate excretion is impaired, and as a result, serum phosphate levels rise. At the same time, serum calcium levels fall because serum calcium is inversely regulated in relation to serum phosphate levels (see Chapter 6). The drop in serum calcium, in turn, stimulates PTH release, with a resultant increase in calcium resorption from bone. Most persons with ESRD develop secondary hyperparathyroidism, the result of chronic stimulation of the parathyroid glands. Although serum calcium levels are maintained through increased PTH function, this adjustment is accomplished at the expense of the skeletal system and other body organs.

The kidneys regulate vitamin D activity by converting the inactive form of vitamin D [25(OH) vitamin D₃] to its active form $(1,25-OH_2 \text{ vitamin } D_3)$. Decreased levels of active vitamin D lead to a decrease in intestinal absorption of calcium with a resultant increase in parathyroid hormone levels. Vitamin D also regulates osteoblast differentiation, thereby affecting bone matrix formation and mineralization.

Skeletal Disorders. The term renal osteodystrophy is used to describe the skeletal complications of ESRD.^{10,11} Several factors are thought to contribute to the development of renal osteodystrophy, including elevated serum phosphate levels, decreased serum calcium levels, impaired renal activation of vitamin D, and hyperparathyroidism. The skeletal changes that occur with renal failure have been divided into two major types of disorders: high-turnover and low-turnover osteodystrophy.5 Inherent to both of these conditions is abnormal reabsorption and defective remodeling of bone.

High-bone-turnover osteodystrophy, sometimes referred to as osteitis fibrosa,¹¹ is characterized by increased bone resorption and formation, with bone resorption predominating. The disorder is associated with secondary hyperparathyroidism; altered vitamin D metabolism along with resistance to the action of vitamin D; and impaired regulation of locally produced growth factors and inhibitors. There is an increase in both osteoblast and osteoclast numbers and activity. Although the osteoblasts produce excessive amounts of bone matrix, mineralization fails to keep pace, and there is a decrease in bone density and formation of porous and coarse-fibered bone. Cortical bone is affected more severely than cancellous bone. Marrow fibrosis is another component of osteitis fibrosa; it occurs in areas of increased bone cell activity. In advanced stages of the disorder, cysts may develop in the bone, a condition called osteitis fibrosa cystica.¹²

Low-bone-turnover osteodystrophy is characterized by decreased numbers of osteoblasts and low or reduced numbers of osteoclasts, a low rate of bone turnover, and an accumulation of unmineralized bone matrix.12 There are two forms of low-turnover osteodystrophy: osteomalacia and adynamic osteodystrophy. Osteomalacia, sometimes referred to as renal rickets, is characterized by a slow rate of bone formation and defects in bone mineralization. Until the 1980s, osteomalacia in ESRD resulted mainly from aluminum intoxication. Aluminum intoxication causes decreased and defective mineralization of bone by existing osteoblasts and more long-term inhibition of osteoblast differentiation. During the 1970s and 1980s, it was discovered that accumulation of aluminum from water used in dialysis and aluminum salts used as phosphate binders caused osteomalacia and advnamic bone disease.¹³ This discovery led to a change in the composition of dialysis solutions and substitution of calcium carbonate for aluminum salts as phosphate binders. As a result, the prevalence of osteomalacia in persons with ESRD is declining.

The second type of low-turnover osteodystrophy, advnamic osteodystrophy, is characterized by a low number of osteoblasts, the osteoclast number being normal or reduced.¹² In persons with adynamic bone disease, bone remodeling is greatly reduced, and the bone surfaces become hypocellular. Adynamic bone disease is associated with an increased fracture rate. The disease is associated with a "relative hypothyroidism." It has been suggested that hypersecretion of parathyroid hormone may be necessary to maintain normal rates of bone formation in persons with ESRD. Thus, this form of renal osteodystrophy is seen more commonly in persons with ESRD who do not have secondary hyperparathyroidism (i.e., those who have been treated with parathyroidectomy) or have been overtreated with calcium and vitamin D.

The symptoms of renal osteodystrophy, which occur late in the disease, include bone tenderness and muscle weakness. Proximal muscle weakness in the lower extremities is common, making it difficult to get out of a chair or climb stairs.¹¹ Fractures are more common with low-turnover osteomalacia and advnamic renal bone disease.

Early treatment of hyperphosphatemia and hypocalcemia is important to prevent or slow long-term skeletal complications.¹⁴ Activated forms of vitamin D and calcium supplements often are used to facilitate intestinal absorption of calcium, increase serum calcium levels, and reduce parathyroid hormone levels. Milk products and other foods high in phosphorus content are restricted in the diet. Oral phosphatebinding agents such as calcium carbonate may be prescribed to decrease absorption of phosphate from the gastrointestinal tract. Aluminum-containing antacids can contribute to the development of osteodystrophy and their use should be avoided except in acute situations.

Hematologic Disorders

Chronic anemia is the most profound hematologic alteration that accompanies renal failure. Anemia first appears when the GFR falls below 40 mL/minute and is present in most persons with ESRD.¹⁵ Several factors contribute to anemia in persons with chronic renal failure, including a erythropoietin deficiency, uremic toxins, and iron deficiency. The kidneys are the primary site for the production of the hormone *erythropoietin*, which controls red blood cell production. The accumulation of uremic toxins further suppresses red cell production in the bone marrow, and the cells that are produced have a shortened life span. Iron is essential for erythropoiesis. Many persons receiving maintenance hemodialysis also are iron deficient because of blood sampling and accidental loss of blood during dialysis. Other causes of iron deficiency include factors such as anorexia and dietary restrictions that limit intake.

When untreated, anemia causes or contributes to weakness, fatigue, depression, insomnia, and decreased cognitive function. There is also increasing concern regarding the physiologic effects of anemia on cardiovascular function. The anemia of renal failure produces a decrease in blood viscosity and a compensatory increase in heart rate. The decreased blood viscosity also exacerbates peripheral vasodilatation and contributes to decreased vascular resistance. Cardiac output increases in a compensatory fashion to maintain tissue perfusion. Echocardiographic studies after initiation of chronic dialysis have shown ventricular dilatation with compensatory left ventricular hypertrophy.¹⁵ Anemia also limits myocardial oxygen supply, particularly in persons with coronary heart disease, leading to angina pectoris and other ischemic events.¹⁶ Thus, anemia, when coupled with hypertension, may be a major contributing factor to the development of left ventricular dysfunction and congestive heart failure in persons with ESRD.

A remarkable advance in medical management of anemia in persons with ESRD occurred with the availability of recombinant human erythropoietin (rhEPO). Secondary benefits of treating anemia with rhEPO, previously attributed to the correction of uremia, include improvement in appetite, energy level, sexual function, skin color, and hair and nail growth, and reduced cold intolerance. Frequent measurements of hematocrit are necessary. Worsening hypertension and seizures have occurred when the hematocrit was raised too suddenly. Because iron deficiency is common among persons with chronic renal failure, iron supplementation often is needed. Iron can be given orally or intravenously. Intravenous iron may be used for treatment of persons who are not able to maintain adequate iron status with oral iron.

Bleeding disorders, which are manifested by epistaxis, menorrhagia, gastrointestinal bleeding, and bruising of the skin and subcutaneous tissues, are also common among persons with chronic renal failure. Although platelet production often is normal in ESRD, platelet function is impaired. Platelet function improves with dialysis but does not completely normalize, suggesting that uremia contributes to the problem. Anemia may accentuate the problem by changing the position of the platelets with respect to the vessel wall. Normally the red cells occupy the center of the bloodstream, and the platelets are in the skimming layer along the endothelial surface. In anemia, the platelets become dispersed, impairing the plateletendothelial cell adherence needed to initiate hemostasis.¹⁷

Cardiovascular Disorders

Cardiovascular disorders, including hypertension, left ventricular hypertrophy, and pericarditis, are a major cause of morbidity and mortality in patients with ESRD. Hypertension commonly is an early manifestation of chronic renal failure. The mechanisms that produce hypertension in ESRD are multifactorial; they include an increased vascular volume, elevation of peripheral vascular resistance, decreased levels of renal vasodilator prostaglandins, and increased activity of the reninangiotensin system.¹⁸ Early identification and aggressive treatment of hypertension has been shown to slow the rate of renal impairment in many types of renal disease. Treatment involves salt and water restriction and the use of antihypertensive medications to control blood pressure. Many persons with renal insufficiency need to take several antihypertensive medications to control blood pressure (see Chapter 16).

The spectrum of heart disease includes left ventricular hypertrophy and ischemic heart disease. People with ESRD tend to have an increased prevalence of left ventricular dysfunction, both with a depressed left ventricular ejection fraction, as in systolic dysfunction, as well as impaired ventricular filling, as in diastolic failure (see Chapter 18).¹⁹ There are multiple factors that lead to development of left ventricular dysfunction, including extracellular fluid overload, shunting of blood through an arteriovenous fistula for dialysis, and anemia. These abnormalities, coupled with the hypertension that often is present, cause increased myocardial work and oxygen demand, with eventual development of heart failure. Coexisting conditions that have been identified as contributing to the burden of cardiovascular disease include anemia, diabetes mellitus, dyslipidemia, and coagulopathies. Anemia, in particular, has been correlated with the presence of left ventricular hypertrophy.

Pericarditis occurs in approximately 20% of persons receiving chronic dialysis.²⁰ It can result from metabolic toxins associated with the uremic state or from dialysis. The manifestations of uremic pericarditis resemble those of viral pericarditis, with all its complications, including cardiac tamponade (see Chapter 17).

Gastrointestinal Disorders

Anorexia, nausea, and vomiting are common in patients with uremia, along with a metallic taste in the mouth that further depresses the appetite. Early-morning nausea is common. Ulceration and bleeding of the gastrointestinal mucosa may develop, and hiccups are common. A possible cause of nausea and vomiting is the decomposition of urea by intestinal flora, resulting in a high concentration of ammonia. Parathyroid hormone increases gastric acid secretion and contributes to gastrointestinal problems. Nausea and vomiting often improve with restriction of dietary protein and after initiation of dialysis and disappear after kidney transplantation.

Disorders of Neural Function

Many persons with chronic renal failure have alterations in peripheral and central nervous system function. Peripheral neuropathy, or involvement of the peripheral nerves, affects the lower limbs more frequently than the upper limbs. It is symmetric and affects both sensory and motor function. Neuropathy is caused by atrophy and demyelination of nerve fibers, possibly caused by uremic toxins. Restless legs syndrome is a manifestation of peripheral nerve involvement and can be seen in as many as two thirds of patients on dialysis. This syndrome is characterized by creeping, prickling, and itching sensations that typically are more intense at rest. Temporary relief is obtained by moving the legs. A burning sensation of the feet, which may be followed by muscle weakness and atrophy, is a manifestation of uremia.

The central nervous system disturbances in uremia are similar to those caused by other metabolic and toxic disorders. Sometimes referred to as *uremic encephalopathy*, the condition is poorly understood and may result, at least in part, from an excess of toxic organic acids that alter neural function. Electrolyte abnormalities, such as sodium shifts, also may contribute. The manifestations are more closely related to the progress of the uremic disorder than to the level of the metabolic endproducts. Reductions in alertness and awareness are the earliest and most significant indications of uremic encephalopathy. This often is followed by an inability to fix attention, loss of recent memory, and perceptual errors in identifying persons and objects. Delirium and coma occur late in the course; seizures are the preterminal event.

Disorders of motor function commonly accompany the neurologic manifestations of uremic encephalopathy. During the early stages, there often is difficulty in performing fine movements of the extremities; the gait becomes unsteady and clumsy with tremulousness of movement. Asterixis (dorsiflexion movements of the hands and feet) typically occurs as the disease progresses. It can be elicited by having the person hyperextend his or her arms at the elbow and wrist with the fingers spread apart. If asterixis is present, this position causes side-to-side flapping movements of the fingers.

Altered Immune Function

Infection is a common complication and cause of hospitalization and death of patients with chronic renal failure. Immunologic abnormalities decrease the efficiency of the immune response to infection. All aspects of inflammation and immune function may be affected adversely by the high levels of urea and metabolic wastes, including a decrease in granulocyte count, impaired humoral and cell-mediated immunity, and defective phagocyte function. The acute inflammatory response and delayed-type hypersensitivity response are impaired. Although persons with ESRD have normal humoral responses to vaccines, a more aggressive immunization program may be needed. Skin and mucosal barriers to infection also may be defective. In persons who are receiving dialysis, vascular access devices are common portals of entry for pathogens. Many persons with ESRD do not experience fever with infection, making the diagnosis more difficult.

Disorders of Skin Integrity

Skin manifestations are common in persons with ESRD. The skin often is pale because of anemia and may have a sallow, yellow-brown hue. The skin and mucous membranes often are dry, and subcutaneous bruising is common. Skin dryness is caused by a reduction in perspiration caused by the decreased size of sweat glands and the diminished activity of oil glands. Pruritus is common; it results from the high serum phosphate levels and the development of phosphate crystals that occur with hyperparathyroidism. Severe scratching and repeated needlesticks, especially with hemodialysis, break the skin integrity and increase the risk for infection. In the advanced stages of untreated renal failure, urea crystals may precipitate on the skin as a result of the high urea concentration in body fluids. The fingernails may become thin and brittle, with a dark band just behind the leading edge of the nail, followed by a white band. This appearance is known as Terry's nails.

Sexual Dysfunction

Alterations in physiologic sexual responses, reproductive ability, and libido are common. The cause probably is multifactorial and may result from high levels of uremic toxins, neuropathy, altered endocrine function, psychological factors, and medications (*e.g.*, antihypertensive drugs).

Impotence occurs in as many as 56% of male patients receiving dialysis.²¹ Derangements of the pituitary and gonadal hormones, such as decreases in testosterone levels and increases in prolactin and luteinizing hormone levels, are common and cause erectile difficulties and decreased spermatocyte counts. Loss of libido may result from chronic anemia and decreased testosterone levels. Several drugs, such as exogenous testosterone and bromocriptine, have been used in an attempt to return hormone levels to normal.

Impaired sexual function in women is manifested by abnormal levels of progesterone, luteinizing hormone, and prolactin. Hypofertility, menstrual abnormalities, decreased vaginal lubrication, and various orgasmic problems have been described.²² Amenorrhea is common among women who are receiving dialysis therapy.

Elimination of Drugs

The kidneys are responsible for the elimination of many drugs and their metabolites.⁶ Renal failure and its treatment can interfere with the absorption, distribution, and elimination of drugs. The administration of large quantities of phosphatebinding antacids to control hyperphosphatemia and hypocalcemia in patients with advanced renal failure interferes with the absorption of some drugs. Many drugs are bound to plasma proteins, such as albumin, for transport in the body; the unbound portion of the drug is available to act at the various receptor sites and is free to be metabolized. A decrease in plasma proteins, particularly albumin, that occurs in many persons with ESRD results in less protein-bound drug and greater amounts of free drug.

Decreased elimination by the kidneys allows drugs or their metabolites to accumulate in the body and requires that drug dosages be adjusted accordingly. Some drugs contain unwanted nitrogen, sodium, potassium, and magnesium and must be avoided in patients with renal failure. For example, penicillin contains potassium. Nitrofurantoin and ammonium chloride add to the body's nitrogen pool. Many antacids contain magnesium. Because of problems with drug dosing and elimination, persons with renal failure should be cautioned against the use of over-the-counter remedies.

Treatment

During the past several decades, an increasing number of persons have required renal replacement therapy with dialysis or transplantation. The growing volume is largely attributable to the improvement in treatment and more liberal policies regarding who is treated. Between 1980 and 1992, there was a twofold reported increase in treatment for ESRD.²³ In 1998, almost 245,910 persons were receiving dialysis therapy in the United States, and another 13,272 underwent kidney transplantation.²⁴

Medical Management

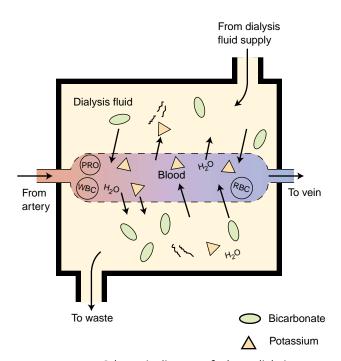
Chronic renal failure can be treated by conservative management of renal insufficiency and by renal replacement therapy with dialysis or transplantation. Conservative treatment consists of measures to prevent or retard deterioration in remaining renal function and to assist the body in compensating for the existing impairment. Interventions that have been shown to significantly retard the progression of chronic renal insufficiency include dietary protein restriction and blood pressure normalization. Various interventions are used to compensate for reduced renal function and correct the resulting anemia, hypocalcemia, and acidosis. These interventions often are used in conjunction with dialysis therapy for patients with ESRD.

Dialysis and Transplantation

Dialysis or renal replacement therapy is indicated when advanced uremia or serious electrolyte imbalances are present. The choice between dialysis and transplantation is dictated by age, related health problems, donor availability, and personal preference. Although transplantation often is the treatment preference, dialysis plays a critical role as a treatment method for ESRD. It is life sustaining for persons who are not candidates for transplantation or who are awaiting transplantation. There are two broad categories of dialysis: hemodialysis and peritoneal dialysis.

Hemodialysis. The basic principles of hemodialysis have remained unchanged throughout the years, although new technology has improved the efficiency and speed of dialysis.²⁵ A hemodialysis system, or artificial kidney, consists of three parts: a blood compartment, a dialysis fluid compartment, and a cellophane membrane that separates the two compartments. The cellophane membrane is semipermeable, permitting all molecules except blood cells and plasma proteins to move freely in both directions-from the blood into the dialyzing solution and from the dialyzing solution into the blood. The direction of flow is determined by the concentration of the substances contained in the two solutions. The waste products and excess electrolytes in the blood normally diffuse into the dialyzing solution. If there is a need to replace or add substances, such as bicarbonate, to the blood, these can be added to the dialyzing solution (Fig. 24-4).

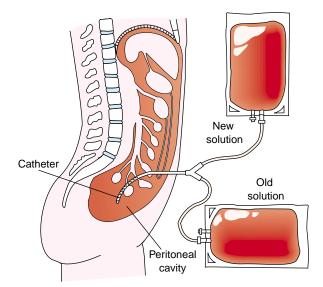
During dialysis, blood moves from an artery through the tubing and blood chamber in the dialysis machine and then



■ FIGURE 24-4 ■ Schematic diagram of a hemodialysis system. The blood compartment and dialysis solution compartment are separated by a cellophane membrane. This membrane is porous enough to allow all the constituents, except the plasma proteins and blood cells, to diffuse between the two compartments.

back into the body through a vein. Access to the vascular system is accomplished through an external arteriovenous shunt (*i.e.*, tubing implanted into an artery and a vein) or, more commonly, through an internal arteriovenous fistula (*i.e.*, anastomosis of a vein to an artery, usually in the forearm). Heparin is used to prevent clotting during the dialysis treatment; it can be administered continuously or intermittently.

Peritoneal Dialysis. Peritoneal dialysis uses the same principles of diffusion, osmosis, and ultrafiltration that apply to hemodialysis. The thin serous membrane of the peritoneal cavity serves as the dialyzing membrane. A silastic catheter is surgically implanted in the peritoneal cavity below the umbilicus to provide access. The catheter is tunneled through subcutaneous tissue and exits on the side of the abdomen (Fig. 24-5). The dialysis process involves instilling a sterile dialyzing solution (usually 2 L) through the catheter during a period of approximately 10 minutes. The solution then is allowed to remain, or dwell, in the peritoneal cavity for a prescribed amount of time, during which the metabolic endproducts and extracellular fluid diffuse into the dialysis solution. At the end of the dwell time, the dialysis fluid is drained out of the peritoneal cavity by gravity into a sterile bag. Glucose in the dialysis solution accounts for water removal. Commercial dialysis solution is available in 1.5%, 2.5%, and 4.25% dextrose concentrations. Solutions with higher dextrose levels increase osmosis, causing more fluid to be removed. The most common method is continuous ambulatory peritoneal dialysis (CAPD), a self-care procedure in which the person manages the dialysis procedure and the type of solution (*i.e.*, dextrose concentration) used at home.



■ FIGURE 24-5 ■ Peritoneal dialysis. A semipermeable membrane, richly supplied with small blood vessels, lines the peritoneal cavity. With dialysate dwelling in the peritoneal cavity, waste products diffuse from the network of blood cells into the dialysate.

Transplantation. Greatly improved success rates have made kidney transplantation the treatment of choice for many patients with chronic renal failure. The availability of donor organs continues to limit the number of transplantations performed each year. Donor organs are obtained from cadavers and living related donors (*e.g.*, parent, sibling). The success of transplantation depends primarily on the degree of histocompatibility, adequate organ preservation, and immunologic management.²⁶

In summary, chronic renal failure results from the destructive effects of many forms of renal disease. Regardless of the cause, the consequences of nephron destruction in ESRD are alterations in the filtration, reabsorption, and endocrine functions of the kidneys. The progression of chronic renal failure usually occurs in four stages: diminished renal reserve, renal insufficiency, renal failure, and ESRD. Renal insufficiency represents a reduction in the GFR to approximately 20% to 50% of normal; renal failure, a reduction to less than 20% to 25% of normal; and ESRD, a decrease in GFR to less than 5% of normal.

End-stage renal disease affects almost every body system. It causes an accumulation of nitrogenous wastes (*i.e.*, azotemia), alters sodium and water excretion, and alters regulation of body levels of potassium, phosphate, calcium, and magnesium. It also causes skeletal disorders, anemia, alterations in cardiovascular function, neurologic disturbances, gastrointestinal dysfunction, and discomforting skin changes.

The treatment of ESRD can be divided into two types: conservative management of renal insufficiency and renal replacement therapy with dialysis or transplantation. Conservative treatment consists of measures to prevent or retard deterioration in remaining renal function and to assist the body in compensating for the existing impairment.

RENAL FAILURE IN CHILDREN AND ELDERLY PERSONS

Although the spectrum of renal disease among children and elderly persons is similar to that of adults, several unique issues affecting these groups warrant further discussion.

Chronic Renal Failure in Children

The true incidence of chronic renal failure in infants and children is unknown. The data indicate that 2500 people in the United States who are younger than 20 years of age begin treatment for chronic renal failure each year; 100 of these children are younger than 2 years of age.²⁷ The most common cause of chronic renal failure in children is glomerulonephritis and congenital malformations, such as renal hypoplasia or dysplasia, obstructive uropathy, and reflux nephropathy.²⁸

Features of renal disease that are marked during childhood include severe growth impairment, developmental delay, delay in sexual maturation, bone abnormalities, and development of psychosocial problems.^{27,29} Critical growth periods occur during the first 2 years of life and during adolescence. Physical growth and cognitive development occur at a slower rate as consequences of renal disease, especially among children with congenital renal disease. Puberty usually occurs at a later age in children with renal failure, partly because of endocrine abnormalities. Renal osteodystrophy is more common and extensive in children than in adults because of the presence of open epiphyses. As a result, metaphyseal fractures, bone pain, impaired bone growth, short stature, and osteitis fibrosa cystica occur with greater frequency. Some hereditary renal diseases, such as medullary cystic disease, have patterns of skeletal involvement that further complicate the problems of renal osteodystrophy. Factors related to impaired growth include deficient nutrition, anemia, renal osteodystrophy, chronic acidosis, and cases of nephrotic syndrome that require high-dose corticosteroid therapy.

The success of treatment depends on the level of bone maturation at the initiation of therapy. Nutrition is believed to be the most important determinant during infancy. During childhood, growth hormone is important, and gonadotropic hormones become important during puberty.³⁰ Parental heights provide a means of assessing growth potential. For many children, catch-up growth is important because a growth deficit frequently is established during the first months of life. Recombinant human growth hormone therapy has been used to improve growth in children with ESRD.³⁰ Success of treatment depends on the level of bone maturation at the initiation of therapy.

All forms of renal replacement therapy can be safely and reliably used for children. Children typically are treated with CAPD or transplantation to optimize growth and development.²⁸ An alternative to CAPD is continuous cyclic peritoneal dialysis. The procedure reverses the schedule of CAPD by providing the exchanges at night, rather than during the day. The exchanges are performed automatically during sleep by a simple cycler machine. Renal transplantation is considered the best alternative for children.³¹ Early transplantation in young children is regarded as the best way to promote physical growth, improve cognitive function, and foster psychosocial development.³¹ Immunosuppressive therapy in children is similar to that used in adults. All of these immunosuppressive agents have side effects, including increased risk of infection. Corticosteroids, which have been the mainstay of chronic immunosuppressive therapy for decades, carry the risk of hypertension, orthopedic complications (especially aseptic necrosis), cataracts, and growth retardation.

Chronic Renal Failure in Elderly Persons

Since the mid-1980s, there have been increasing numbers of elderly persons accepted to ESRD programs. In 2001, 20.1% of persons being treated for ESRD were 65 to 74 years of age, and 14.4% were older than 75 years of age.³² Among elderly persons, the presentation and course of renal failure may be altered because of age-related changes in the kidneys and concurrent medical conditions.

Normal aging is associated with a decline in the GFR and subsequently with reduced homeostatic regulation under stressful conditions.³³ This reduction in GFR makes elderly persons more susceptible to the detrimental effects of nephrotoxic drugs, such as radiographic contrast compounds. The reduction in GFR related to aging is not accompanied by a parallel increase in the serum creatinine level because the serum creatinine level, which results from muscle metabolism, is significantly reduced in elderly persons because of diminished muscle mass and other age-related changes. Evaluation of renal function in elderly persons should include a measurement of creatinine clearance along with the serum creatinine level.

The prevalence of chronic disease affecting the cerebrovascular, cardiovascular, and skeletal systems is higher in this age group. Because of concurrent disease, the presenting symptoms of renal disease in elderly persons may be less typical than those observed in younger adults. For example, congestive heart failure and hypertension may be the dominant clinical features with the onset of acute glomerulonephritis, whereas oliguria and discolored urine more often are the first signs in younger adults. The course of renal failure may be more complicated in older patients with numerous chronic diseases.

Treatment options for chronic renal failure in elderly patients with ESRD include hemodialysis and peritoneal dialysis. Neither hemodialysis nor peritoneal dialysis has proven to be superior in the elderly. Many transplantation centers have increased the age for acceptance on transplant waiting lists, so renal transplantation provides another option. The general reduction in T-cell function that occurs with aging has been suggested as a beneficial effect that increases transplant graft survival.

In summary, there is approximately a 2% per year incidence of renal failure in children, most frequently resulting from congenital malformations and glomerulonephritis. Problems associated with renal failure in children include growth impairment, delay in sexual maturation, and more extensive bone abnormalities than in adults. Although all forms of renal replacement therapy can be safely and reliably used for children, CAPD or transplantation optimize growth and development. Adults 65 years of age and older account for close to one half of the new cases of ESRD each year. Normal aging is associated with a decline in the GFR, which makes elderly persons more susceptible to the detrimental effects of nephrotoxic drugs and other conditions that compromise renal function. Treatment options for chronic renal failure in elderly patients are similar to those for younger persons.

REVIEW QUESTIONS

Compare acute and chronic renal failure in terms of reversibility.

Differentiate the prerenal, intrinsic, and extrarenal forms of acute renal failure in terms of the mechanisms of development and manifestations.

Cite the two most common causes of acute tubular necrosis and describe the course of the disease in terms of the initiation, maintenance, and recovery phases.

State the definitions of renal impairment, renal insufficiency, renal failure, and end-stage renal disease.

List the common problems associated with end-stage renal disease, including alterations in fluid and electrolyte balance and disorders of skeletal, hematologic, cardiovascular, immune, neurologic, skin, and sexual function, and explain their physiologic significance.

Describe the scientific principles underlying dialysis treatment, and compare hemodialysis with peritoneal dialysis.

List the causes of renal failure in children and describe the special problems of children with ESRD.

Explain why renal failure is so common in the elderly, and describe measures to prevent or delay the onset of ESRD in this population.

connection-J

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- 1. Levy E.M., Viscose C.M., Horwitz R.I. (1996). The effect of acute renal failure on mortality: A cohort analysis. *Journal of the American Medical Association* 275, 1489–1494.
- Thadhani R., Pascual M., Bonventre J.V. (1996). Acute renal failure. New England Journal of Medicine 334, 1448–1460.
- Brady H.R., Brenner B.M., Clarkson M.R., et al. (2000). Acute renal failure. In Brenner B.M. (Ed.), *Brenner and Rector's the kidney* (6th ed., pp. 1201–1247). Philadelphia: W.B. Saunders.
- 4. Guyton A., Hall J.E. (2000). *Textbook of medical physiology* (10th ed., pp. 369–371, 373–378). Philadelphia: W.B. Saunders.
- Cotran R.S., Kumar V., Collins T. (1999). Robbins pathologic basis of disease (6th ed., pp. 932–933, 969–971, 1229). Philadelphia: W.B. Saunders.
- Bailie G.R. (1996). Acute renal failure. In Young L.Y., Koda-Kimble M.A. (Eds.), Applied therapeutics: The clinical use of drugs (6th ed., pp. 29-6–29-17). Vancouver, WA: Applied Therapeutics.

- 7. Gerlach A.T., Pickworth K.K. (2000). Contrast medium-induced nephrotoxicity: Pathophysiology and prevention. Pharmacotherapy 20, 540 - 548
- 8. Rettig R.A. (1996). The social contract and the treatment of permanent renal failure. Journal of the American Medical Association 274, 1123-1126
- 9. National Kidney and Urological Information Center. (2001). Kidney and urologic disease statistics for the United States. [On-line]. Available: http://www.niddk.nih.gov/health/kidney/pubs/kstats/kstats.htm.
- 10. Llach F., Bover J. (2000). Renal osteodystrophies. In Brenner B.M. (Ed.), Brenner and Rector's the kidney (6th ed., pp. 2103-2135). Philadelphia: W.B. Saunders.
- 11. Hrusks K.A., Teitelbaum S.L. (1995). Renal osteodystrophy. New England Journal of Medicine 333, 166–174.
- 12. Couttenve M.M., D'Haese P.C., Verschoren W.J., et al. (1999). Low bone turnover in patients with renal failure. Kidney International 56 (7 Suppl. 73), S70-S76.
- 13. Brenner B.M., Lazarus J.M. (1991). Chronic renal failure. In Wilson J.D., Braunwald E., Isselbacher K.J., et al. (Eds.), Harrison's principles of internal medicine (12th ed., pp. 1150-1156). New York: McGraw-Hill.
- 14. Drüeke T.B. (2001). Control of secondary hyperthyroidism by vitamin D derivatives. American Journal of Kidney Diseases 37 (1 Suppl. 2), \$58-\$61.
- 15. Tong E.M., Nissenson A.R. (2001). Erythropoietin and anemia. Seminars in Nephrology 21, 190-203.
- 16. Besarab A., Levin A. (2000). Defining a renal anemia management period. American Journal of Kidney Diseases 36 (6 Suppl. 3), S13-S23.
- 17. Eberst M.E., Berkowitz L.R. (1993). Hemostasis in renal disease: Pathophysiology and management. American Journal of Medicine 96, 168 - 179.
- 18. Preston R.A., Singer I., Epstein M. (1996). Renal parenchymal hypertension. Archives of Internal Medicine 156, 602-611.
- 19. Al-Ahmad A., Sarnak M.J., Salem D.N., et al. (2001). Cause and management of heart failure in patients with chronic renal disease. Seminars in Nephrology 21, 3-12.
- 20. Gunukula S., Spodick D.H. (2001). Pericardial disease in renal failure. Seminars in Nephrology 21, 52-56.

- 21. Foulks C.J., Cushner H.M. (1986). Sexual dysfunction in the male dialysis patient: Pathogenesis, evaluation, and therapy. American Journal of Kidney Diseases 8, 211-212.
- 22. Rickus M.A. (1987). Sexual dysfunction in the female ESRD patient. American Nephrology Nurses' Association Journal 14, 185–186.
- 23. Agodaoa L.Y., Eggers P.W. (1995). Renal replacement therapy in the United States: Data from the United States Renal Data System. American Journal of Kidney Diseases 25, 119–133.
- 24. National Kidney and Urologic Diseases Information Clearinghouse. (2001). Kidney and urologic diseases statistics in the United States. [Online]. Available: http://www.niddk.gov/health/kidney/pubs/kustats/ kustats htm
- 25. Daelemans R.A., D'Haese P.C., BeBroe M.E. (2001). Dialysis. Seminars in Nephrology 21, 204-212.
- 26. Ramanathan V., Goral S., Helderman J.H. (2001). Renal transplantation. Seminars in Nephrology 21, 213-219.
- 27. Hanna J.D., Krieg R.J., Scheinman J.I., et al. (1996). Effects of uremia on growth in children. Seminars in Nephrology 16, 230-241.
- 28. Bergstein J.M. (2000). Renal failure. In Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.), Nelson textbook of pediatrics (16th ed., pp. 1605-1617). Philadelphia: W.B. Saunders.
- 29. Abitbol C., Chan J.C.M., Trachtman H., et al. (1996). Growth in children with moderate renal insufficiency: Measurement, evaluation, and treatment. Journal of Pediatrics 129, S3-S7.
- 30. Haffner D., Schaffer F., Nissel R., et al. (Study Group for Growth Hormone Treatment in Chronic Renal Failure). (2000). Effect of growth hormone treatment on the adult height of children with chronic renal failure. New England Journal of Medicine 343, 923-930.
- 31. Urizar R.E. (2000). Renal transplantation. In Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.), Nelson textbook of pediatrics (16th ed., pp. 1612-1617). Philadelphia: W.B. Saunders.
- 32. National Kidney Foundation. (2000). End stage renal disease. [Online]. Available: http://www.kidney.org/general/news/ esrd/cfm.
- 33. Choudhury D., Raj D.S.D., Palmer B., et al. (2000). Effect of aging on renal function and disease. In Brenner B.M. (Ed.), Brenner and Rector's the kidney (6th ed., pp. 2187-2210). Philadelphia: W.B. Saunders.

Alterations in Urine Elimination

Control of Urine Elimination Bladder Structure Micturition Neural Control of Bladder Function Spinal Cord Centers Pontine Micturition Center Cortical and Subcortical Centers Development of Bladder Control in Children **Evaluation of Bladder Function Alterations in Bladder Function** Urinary Obstruction and Stasis Neurogenic Bladder Disorders Spastic Bladder: Failure to Store Urine Flaccid Bladder: Failure to Empty Urine Urinary Incontinence **Stress Incontinence** Urge Incontinence/Overactive Bladder **Overflow Incontinence** Other Causes of Incontinence **Diagnosis and Treatment** Special Needs of Elderly Persons **Cancer of the Bladder Diagnosis and Treatment**

CHAPTER

Ithough the kidneys control the formation of urine and regulate the composition of body fluids, it is the bladder that stores urine and controls its elimination from the body. Alterations in the storage and expulsion functions of the bladder can result in incontinence, with its accompanying social and hygienic problems, or obstruction of urine flow, which has deleterious effects on ureteral and, ultimately, renal function. The discussion in this chapter focuses on normal control of urine elimination, urinary obstruction and stasis, neurogenic bladder, incontinence, and bladder cancer. Urinary tract infections are discussed in Chapter 23.

CONTROL OF URINE ELIMINATION

The urinary bladder, also known as the urinary vesicle, is a freely movable organ located behind the pelvic bone in the male and in front of the vagina in the female. It consists of two parts: the fundus, or body, and the neck, or posterior urethra. In the man, the urethra continues anteriorly through the penis. Urine passes from the kidneys to the bladder through the ureters. The ureters enter the bladder obliquely through the *trigone*, a triangular area located above the bladder neck on the posterior wall of the bladder (Fig. 25-1). There are no valves at the ureteral openings, but because of their oblique position in the bladder wall, the pressure of the urine keeps the ends of the ureters compressed to prevent the backflow of urine from the bladder into the ureters.

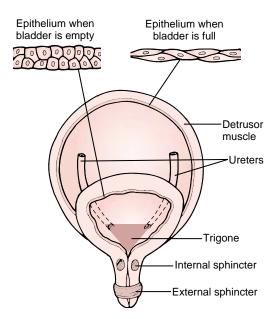
Bladder Structure

The bladder is composed of four layers. The first is an outer serosal layer, which covers the upper surface and is continuous with the peritoneum. The second is a network of smooth muscle fibers called the *detrusor muscle*. The third is a submucosal layer of loose connective tissue, and the fourth is an inner mucosal lining of transitional epithelium.

The tonicity of the urine often is quite different from that of the blood, and the transitional epithelial lining of the bladder acts as an effective barrier to prevent the passage of water between the bladder contents and the blood. The inner elements of the bladder form smooth folds, or rugae. As the bladder expands during filling, these rugae spread out to form a single layer without disrupting the integrity of the epithelial lining.

The detrusor muscle is the muscle of micturition (passage of urine). When it contracts, urine is expelled from the bladder. The abdominal muscles play a secondary role in micturition. Their contraction increases intra-abdominal pressure, which further increases intravesicular pressure.

Muscles in the bladder neck, sometimes referred to as the *internal sphincter*, are a continuation of the detrusor muscle. They run down obliquely behind the proximal urethra, forming the posterior urethra in males and the entire urethra in females. When the bladder is relaxed, these circular muscle fibers



■ FIGURE 25-1 ■ Diagram of the bladder, showing the detrusor muscle, ureters, trigone area, and urethral orifice. Note the flattening of epithelial cells when the bladder is full and the wall is stretched.

are closed and act as a sphincter. When the detrusor muscle contracts, the sphincter is pulled open by the changes that occur in bladder shape. In the female, the urethra (2.5 to 3.5 cm) is shorter than that in the male (16.5 to 18.5 cm) and usually affords less resistance to urine outflow.

Another muscle important to bladder function is the *external sphincter*, a circular muscle composed of striated muscle fibers that surrounds the urethra distal to the base of the bladder. The external sphincter operates as a reserve mechanism to stop micturition when it is occurring and to maintain continence in the face of unusually high bladder pressure. The skeletal muscles of the pelvic floor also contribute to the support of the bladder and the maintenance of continence.

Micturition

To maintain continence, or retention of urine, the bladder must function as a low-pressure storage system; the pressure in the bladder must remain lower than urethral pressure. To ensure that this condition is met, the increase in intravesicular pressure that accompanies bladder filling is almost imperceptible. An increase in bladder volume from 10 to 400 mL may be accompanied by only a 5 cm H₂O increase in pressure.¹ Sustained elevations in intravesicular pressures (>40 to 50 cm H₂O) often are associated with vesicoureteral reflux (*i.e.*, backflow of urine from the bladder into the ureter) and the development of ureteral dilatation (see Chapter 23). Although the pressure in the bladder is maintained at low levels, sphincter pressure remains high (45 to 65 cm H₂O) as a means of preventing loss of urine as the bladder fills.

Micturition involves both sensory and motor functions associated with bladder emptying. When the bladder is distended to 150 to 300 mL, the sensation of fullness is transmitted to the spinal cord and then to the cerebral cortex, allowing for conscious inhibition of the micturition reflex. During the act of micturition, the detrusor muscle of the bladder fundus and bladder neck contract down on the urine; the ureteral orifices are forced shut; the bladder neck is widened and shortened as it is pulled up by the globular muscles in the bladder fundus; the resistance of the internal sphincter in the bladder neck is decreased; and the external sphincter relaxes as urine moves out of the bladder.

Neural Control of Bladder Function

The control of bladder emptying is unique in that it involves both involuntary autonomic nervous system (ANS) reflexes and some voluntary control. The ANS and its neuromediators play a central role in micturition. The somatic nervous system controls the external sphincter and contributes to the voluntary control of bladder emptying.

The parasympathetic nervous system produces contraction of the detrusor muscle and bladder emptying. Parasympathetic innervation of the bladder is mediated by cholinergic (muscarinic) receptors and the neurotransmitter acetylcholine.²

Sympathetic innervation, although not essential to the act of micturition, allows the bladder to store a large volume without the involuntary escape of urine—a mechanism that is consistent with the fight-or-flight function subserved by the sympathetic nervous system. The bladder is supplied with α_1 - and β_2 -adrenergic receptors. The β_2 -adrenergic receptors are found in the detrusor muscle; they produce relaxation of the detrusor muscle, increasing the bladder volume at which the micturition reflex is triggered. The α_1 -adrenergic receptors are found in the trigone area, including the intramural ureteral musculature, bladder neck, and internal sphincter. The activation of α_1 receptors produces contraction of these muscles. Sympathetic activity ceases when the micturition reflex is activated. During male ejaculation, which is mediated by the sympathetic nervous system, the musculature of the trigone area and that of the

KEY CONCEPTS

BLADDER FUNCTION

- The storage and emptying of urine involve both involuntary (autonomic nervous system) and voluntary (somatic nervous system) control.
- The parasympathetic nervous system promotes bladder emptying. It produces contraction of the smooth muscle of the bladder wall and relaxation of the internal sphincter.
- The sympathetic nervous system promotes bladder filling. It produces relaxation of the smooth muscle of the bladder wall and contraction of the internal sphincter.
- The striated muscles in the external sphincter and pelvic floor, which are innervated by the somatic nervous system, provide for the voluntary control of urination and maintenance of continence.

bladder neck and prostatic urethra contracts and prevents the backflow of seminal fluid into the bladder.

There are three main levels of neurologic control for bladder function: (1) the spinal cord reflex centers, (2) the micturition center in the pons, and (3) the cortical and subcortical centers.

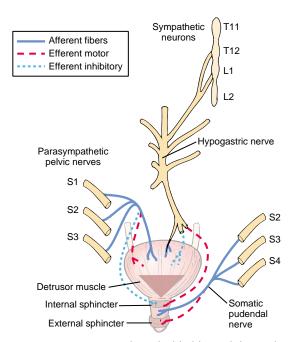
Spinal Cord Centers

The centers for reflex control of bladder emptying or micturition are located in the sacral (S1 through S4) and thoracolumbar (T11 through L2) segments of the spinal cord (Fig. 25-2). The parasympathetic motor neurons that innervate the detrusor muscle of the bladder are located in the sacral segments of the spinal cord; their axons travel to the bladder by way of the *pelvic nerve*. Motor neurons for the external sphincter also are located in the sacral segments of the spinal cord. These motor neurons receive their control from the motor cortex by way of the corticospinal tract and send impulses to the external sphincter through the *pudendal nerve*. The bladder neck and trigone area of the bladder are innervated by the sympathetic fibers from the thoracolumbar segments of the spinal cord.

The afferent or sensory input, which detects stretch of the bladder wall, is carried to the central nervous system (CNS) by way of fibers that travel with the parasympathetic (pelvic), somatic (pudendal), and sympathetic (hypogastric) nerves. The pelvic nerve carries sensory fibers from the stretch receptors in the bladder wall; the pudendal nerve carries sensory fibers from the external sphincter and pelvic muscles; and the hypogastric nerve carries sensory fibers from the trigone area.

Pontine Micturition Center

The immediate coordination of the normal micturition reflex occurs in the micturition center in the pons, facilitated by descending input from the forebrain and ascending input from the reflex centers in the spinal cord^{3,4} (Fig. 25-3). This center coordinates the activity of the detrusor muscle and the external



■ FIGURE 25-2 ■ Nerve supply to the bladder and the urethra.

sphincter. As bladder filling occurs, and the muscle stretch receptors are activated, ascending spinal afferents relay this information to the micturition center, which also receives important descending information from the forebrain concerning behavioral cues for bladder emptying. Descending pathways from the pontine micturition center produce coordinated inhibition of neurons controlling the internal and external sphincters. The onset of urinary flow through the urethra causes reflex contraction of the bladder.

Cortical and Subcortical Centers

Cortical brain centers enable inhibition of the micturition center in the pons and conscious control of urination. Neural influences from the subcortical centers in the basal ganglia, which are conveyed by extrapyramidal pathways, modulate the contractile response. They modify and delay the detrusor contractile response during filling and then modulate the expulsive activity of the bladder to facilitate complete emptying.

Development of Bladder Control in Children

In infants and young children, micturition is an involuntary act that is triggered by a spinal cord reflex; when the bladder fills to a given capacity, the detrusor muscle contracts and the external sphincter relaxes. As the bladder grows and increases in capacity, the tone of the external sphincter muscle increases. At 2 to 3 years of age, the child becomes conscious of the need to urinate and can learn to contract the pelvic muscles to maintain closure of the external sphincter and delay urination. As the nervous system continues to mature, inhibition of involuntary detrusor muscle activity takes place. After the child achieves continence, micturition becomes voluntary.

Evaluation of Bladder Function

Bladder structure and function can be assessed by a number of methods.⁵ Reports or observations of frequency, hesitancy, straining to void, and a weak or interrupted stream are suggestive of outflow obstruction. Palpation and percussion provide information about bladder distention.

Postvoided residual (PVR) urine volume provides information about bladder emptying. It can be estimated by abdominal palpation and percussion. Catheterization and ultrasonography can be used to obtain specific measurements of PVR. A PVR value of less than 50 mL is considered adequate bladder emptying, and more than 200 mL indicates inadequate bladder emptying.⁶

Pelvic examination is used in women to assess perineal skin condition, perivaginal muscle tone, genital atrophy, pelvic prolapse (*e.g.*, cystocele, rectocele, uterine prolapse), pelvic mass, or other conditions that may impair bladder function. Bimanual examination (*i.e.*, pelvic and abdominal palpation) can be used to assess PVR volume. Rectal examination is used to test for perineal sensation, sphincter tone, fecal impaction, and rectal mass. It is used to assess the contour of the prostate in men.

Urine tests provide information about kidney function and urinary tract infections. The presence of bacteriuria or pyuria suggests urinary tract infection and the possibility of urinary tract obstruction. Blood tests (*i.e.*, blood urea nitrogen and creatinine) provide information about renal function.

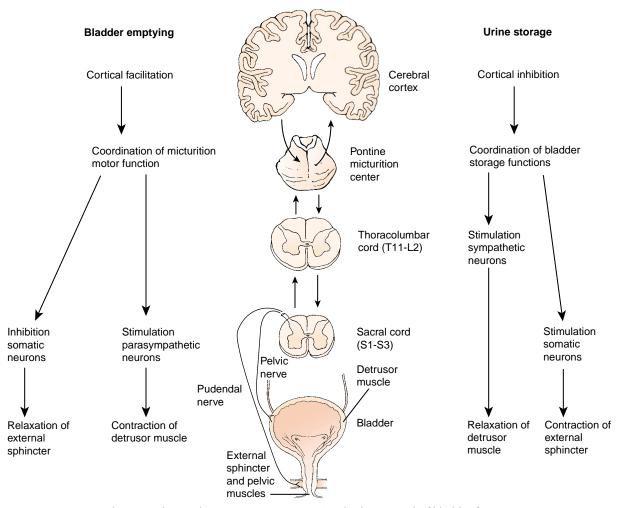


FIGURE 25-3 Pathways and central nervous system centers involved in control of bladder function. Efferent pathways for micturition (**left**) and urine storage (**right**).

Bladder structures can be visualized indirectly by taking x-ray films of the abdomen and by using excretory urography, which involves the use of a radiopaque dye, computed tomographic (CT) scanning, magnetic resonance imaging (MRI), or ultrasonography. Cystoscopy enables direct visualization of the urethra, bladder, and ureteral orifices. The ultrasound bladder scan provides a noninvasive method for estimating bladder volume. Urodynamic studies may be used to study bladder function and voiding problems.

In summary, although the kidneys function in the formation of urine and the regulation of body fluids, it is the bladder that stores and controls the elimination of urine. Micturition is a function of the peripheral ANS, subject to facilitation or inhibition from higher neurologic centers. The parasympathetic nervous system controls the motor function of the bladder detrusor muscle and the tone of the internal sphincter; its cell bodies are located in the sacral spinal cord and communicate with the bladder through the pelvic nerve. Efferent sympathetic control originates at the level of segments T11 through L2 of the spinal cord and produces relaxation of the detrusor muscle and contraction of the internal sphincter. Skeletal muscle found in the external sphincter and the pelvic muscles that support the bladder are supplied by the pudendal nerve, which exits the spinal cord at the level of segments S2 through S4. The micturition center in the brain stem coordinates the action of the detrusor muscle and the external sphincter, whereas cortical centers permit conscious control of micturition.

Bladder function can be evaluated using urodynamic studies that measure bladder, urethral, and abdominal pressures; urine flow characteristics; and skeletal muscle activity of the external sphincter.

ALTERATIONS IN BLADDER FUNCTION

Alterations in bladder function include urinary obstruction with retention or stasis of urine and urinary incontinence with involuntary loss of urine. Although the two conditions have almost opposite effects on urination, they can have similar causes. Both can result from structural changes in the bladder, urethra, or surrounding organs or from impairment of neurologic control of bladder function.

Urinary Obstruction and Stasis

In lower urinary tract obstruction and stasis, urine is produced normally by the kidneys but is retained in the bladder. Obstructions are classified according to cause (congenital or acquired), degree (partial or complete), duration (acute or chronic), and level (upper or lower urinary tract).⁷ Because it has the potential to produce vesicoureteral reflux and cause kidney damage, urinary obstruction and stasis is a serious disorder.

Congenital narrowing of the external meatus (*i.e.*, meatal stenosis) is more common in boys, and obstructive disorders of the posterior urethra are more common in girls. Another common cause of congenital obstruction is the damage to sacral nerves that occurs in spina bifida and meningomyelocele.

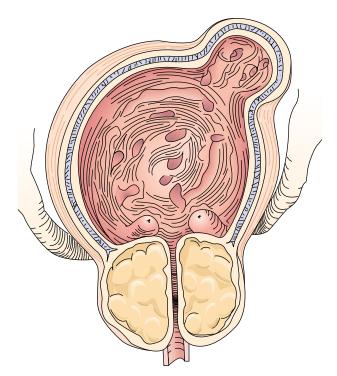
The acquired causes of lower urinary tract obstruction and stasis are numerous. In males, the most important cause of urinary obstruction is external compression of the urethra caused by the enlargement of the prostate gland. Gonorrhea and other sexually transmitted diseases contribute to the incidence of infection-produced urethral strictures. Bladder tumors and secondary invasion of the bladder by tumors arising in structures that surround the bladder and urethra can compress the bladder neck or urethra and cause obstruction. Constipation and fecal impaction can compress the urethra and produce urethral obstruction. This can be a particular problem in elderly persons.

The body compensates for the obstruction of urine outflow with mechanisms designed to prevent urine retention. These mechanisms can be divided into two stages: a compensatory stage and a decompensatory stage.⁷ The degree to which these changes occur and their effect on bladder structure and urinary function depend on the extent of the obstruction, the rapidity with which it occurs, and the presence of other contributing factors, such as neurologic impairment and infection.

During the early stage of obstruction, the bladder begins to hypertrophy and becomes hypersensitive to afferent stimuli arising from bladder filling. The ability to suppress urination is diminished, and bladder contraction can become so strong that it virtually produces bladder spasm. There is urgency, sometimes to the point of incontinence, and frequency during the day and at night.

With continuation and progression of the obstruction, compensatory changes begin to occur. There is further hypertrophy of the bladder muscle, the thickness of the bladder wall may double, and the pressure generated by detrusor contraction can increase from a normal 20 to 40 cm H_2O to 50 to 100 cm H_2O to overcome the resistance from the obstruction. As the force needed to expel urine from the bladder increases, compensatory mechanisms may become ineffective, causing muscle fatigue before complete emptying can be accomplished. After a few minutes, voiding can again be initiated and completed, accounting for the frequency of urination.

The inner bladder surface forms smooth folds. With continued outflow obstruction, this smooth surface is replaced with coarsely woven structures (*i.e.*, hypertrophied smooth muscle fibers) called *trabeculae*. Small pockets of mucosal tissue, called *cellules*, commonly develop between the trabecular ridges. These pockets form diverticula when they extend between the actual fibers of the bladder muscle (Fig. 25-4). Because the diverticula have no muscle, they are unable to contract and expel their urine into the bladder, and secondary infections caused by stasis are common.



■ FIGURE 25-4 ■ Destructive changes of the bladder wall with development of diverticula caused by benign prostatic hypertrophy.

Along with hypertrophy of the bladder wall, there is hypertrophy of the trigone area and the interureteric ridge, which is located between the two ureters. This causes backpressure on the ureters, the development of hydroureters, vesicourethral reflux, and eventually, kidney damage (see Chapter 23). Stasis of urine predisposes to urinary tract infections.

When compensatory mechanisms no longer are effective, signs of decompensation begin to appear. The period of detrusor muscle contraction becomes too short to completely expel the urine, and residual urine remains in the bladder. At this point, the symptoms of obstruction—frequency of urination, hesitancy, a need to strain to initiate urination, a weak and small stream, and termination of the stream before the bladder is completely emptied—become pronounced. The amount of residual urine may increase to 1000 to 3000 mL, and overflow incontinence occurs. There also may be acute retention of urine. The signs of urine retention are summarized in Chart 25-1.

CHART 25-1 Signs of Outflow Obstruction and Urine Retention

Bladder distention Hesitancy Straining when initiating urination Small and weak stream Frequency Feeling of incomplete bladder emptying Overflow incontinence

KEY CONCEPTS

NEUROGENIC BLADDER DISORDERS

- Neurogenic disorders of the bladder commonly are manifested by spastic bladder dysfunction, in which there is failure to store urine, or as flaccid bladder dysfunction, in which bladder emptying is impaired.
- Spastic bladder dysfunction results from neurologic lesions above the level of the sacral cord that allow neurons in the micturition center to function reflexively without control from higher central nervous system centers.
- Flaccid bladder dysfunction results from neurologic disorders affecting the motor neurons in the sacral cord or peripheral nerves that control detrusor muscle contraction and bladder emptying.

The immediate treatment of lower urinary obstruction and stasis is directed toward relief of bladder distention. This usually is accomplished through urinary catheterization (discussed later in this chapter). Constipation or fecal impaction should be corrected. Long-term treatment is directed toward correcting the problem causing the obstruction.

Neurogenic Bladder Disorders

The urinary bladder is unique in that it is probably the only ANS innervated visceral organ that is under CNS control. The neural control of bladder function can be interrupted at any level. It can be interrupted at the level of the peripheral nerves that connect the bladder to the reflex micturition center in the sacral cord, the ascending and descending tracts in the spinal cord, the pontine micturition center, or the cortical centers that are involved in voluntary control of micturition.^{8,9} Neurogenic disorders of bladder function commonly are manifested in one of two ways: failure to store urine (spastic bladder dysfunction) or failure to empty (flaccid bladder dysfunction). The neural pathways involved in the emptying and storage functions of the bladder are illustrated in Figure 25-3. Spastic bladder dysfunction usually results from neurologic lesions located above the level of the sacral micturition reflexes, whereas flaccid bladder dysfunction results from lesions at the level of sacral reflexes or the peripheral nerves that innervate the bladder. In addition to disorders of detrusor muscle function, disruption of micturition occurs when the neurologic control of external sphincter function is disrupted. Some disorders, such as stroke and Parkinson's disease, may affect both the storage and emptying functions of the bladder. Table 25-1 describes the characteristics of neurogenic bladder according to the level of the lesion.

Spastic Bladder: Failure to Store Urine

Failure to store urine results from conditions that cause reflex bladder spasm and a decrease in bladder volume. It commonly is caused by conditions that produce partial or extensive neural damage above the micturition reflex center in the sacral cord. As a result, bladder function is regulated by segmental reflexes, without control from higher brain centers. The degree of bladder spasticity and dysfunction depends on the level and extent of neurologic dysfunction. Usually both the ANS neurons controlling bladder function and the somatic neurons controlling the function of the striated muscles in the external sphincter are affected. In some cases, there is a detrusor-sphincter dyssynergia with uncoordinated contraction and relaxation of the detrusor and external sphincter muscles. The most common causes of spastic bladder dysfunction are spinal cord lesions such as spinal cord injury, herniated intervertebral disk, vascular lesions, tumors, and myelitis. Other neurologic conditions that affect voiding are stroke, multiple sclerosis, and brain tumors.

TABLE 25-1 Types and Characteristics of Neurogenic Bladder				
Level of Lesion	Change in Bladder Function	Common Causes		
Sensory cortex, motor cortex, or corticospinal tract	Loss of ability to perceive bladder filling; low-volume, physiologically normal micturition that occurs suddenly and is difficult to inhibit	Stroke and advanced age		
Basal ganglia or extrapyramidal tract	Detrusor contractions are elicited suddenly without warning and are difficult to control; bladder contraction is shorter than normal and does not produce full bladder emptying	Parkinson's disease		
Pontine micturition center or com- municating tracts in the spinal cord	Storage reflexes are provoked during filling, and external sphincter responses are heightened; uninhibited bladder contractions occur at a lower volume than normal and do not continue until the bladder is emptied; antagonistic activity occurs between the detrusor muscle and the external sphincter	Spinal cord injury		
Sacral cord or nerve roots	Areflexic bladder fills but does not contract; loss of external sphinc- ter tone occurs when the lesion affects the α-adrenergic motor neurons or pudendal nerve	Injury to sacral cord or spinal roots		
Pelvic nerve	Increased filling and impaired sphincter control cause increased intravesicular pressure	Radical pelvic surgery		
Autonomic peripheral sensory pathways	Bladder overfilling occurs owing to a loss of ability to perceive bladder filling	Diabetic neuropathies, multiple sclerosis		

452 Unit Six: Alterations in the Urinary System

Bladder Dysfunction Caused by Spinal Cord Injury. One of the most common types of spinal cord lesions is spinal cord injury (see Chapter 38). The immediate and early effects of spinal cord injury on bladder function are quite different from those that follow recovery from the initial injury. During the period immediately after spinal cord injury, a state of spinal shock develops, during which all the reflexes, including the micturition reflex, are depressed. During this stage, the bladder becomes atonic and cannot contract. Catheterization is necessary to prevent injury to urinary structures associated with overdistention of the bladder. Aseptic intermittent catheterization is the preferred method of catheterization. Depression of reflexes lasts from a few weeks to 6 months (usually 2 to 3 months), after which the spinal reflexes return and become hyperactive.⁹

After the acute stage of spinal cord injury, the micturition response changes from a long-tract reflex to a segmental reflex. Because the sacral reflex arc remains intact, stimuli generated by bladder stretch receptors during filling produce frequent spontaneous contractions of the detrusor muscle. This creates a small, hyperactive bladder subject to high-pressure and short-duration uninhibited bladder contractions. Voiding is interrupted, involuntary, or incomplete. Dilation of the internal sphincter and spasticity of the external sphincter and perineal muscles innervated by upper motoneurons occur, producing resistance to bladder emptying. Hypertrophy of the trigone develops, often leading to vesicoureteral reflux and kidney damage.

Uninhibited Neurogenic Bladder. A mild form of reflex neurogenic bladder, sometimes called *uninhibited bladder*, can develop after a stroke, during the early stages of multiple sclerosis, or as a result of lesions located in the inhibitory centers of the cortex or the pyramidal tract. With this type of disorder, the sacral reflex arc and sensation are retained, the urine stream is normal, and there is no residual urine. However, bladder capacity is diminished because of increased detrusor muscle tone and external sphincter spasticity.

Detrusor-Sphincter Dyssynergia. Depending on the level of the lesion, the coordinated activity of the detrusor muscle and the external sphincter may be affected. Lesions that affect the micturition center in the pons or impair communication between this center and spinal cord centers interrupt the coordinated activity of the detrusor muscle and the external sphincter. This is called *detrusor-sphincter dyssynergia*. Instead of relaxing during micturition, the external sphincter becomes more constricted. This condition can lead to elevated intravesicular pressures, vesicoureteral reflux, and kidney damage.

Treatment. Among the methods used to treat spastic bladder and detrusor-sphincter dyssynergia are the use of anticholinergic medications to decrease bladder hyperactivity and urinary catheterization to produce bladder emptying (discussed later). A sphincterotomy (surgical resection of the external sphincter) or implantable urethral stent may be used to decrease outflow resistance in persons whose condition cannot be managed with medications and catheterization procedures. An alternative to surgical resection of the external sphincter is the injection of botulinum-A toxin to produce paralysis of the striated muscles in the external sphincter. The effects of the injection last from 3 to 9 months, after which the injection must be repeated.⁸

Flaccid Bladder: Failure to Empty Urine

Failure to empty the bladder can be caused by flaccid bladder dysfunction, peripheral neuropathies that interrupt afferent or efferent communication between the bladder and the spinal cord, or conditions that prevent relaxation of the external sphincter.

Flaccid Bladder Dysfunction. Detrusor muscle areflexia, or flaccid neurogenic bladder, occurs when there is injury to the micturition center of the sacral cord, the cauda equina, or the sacral nerve roots that supply the bladder.⁹ Atony of the detrusor muscle and loss of the perception of bladder fullness permits overstretching of the detrusor muscle and contributes to weak and ineffective bladder contractions. External sphincter tone and perineal muscle tone are diminished. Voluntary urination does not occur, but fairly efficient emptying usually can be achieved by increased intra-abdominal pressure or manual suprapubic pressure. Among the causes of flaccid neurogenic bladder are spina bifida and meningomyelocele.

Bladder Dysfunction Caused by Peripheral Neuropathies. In addition to CNS lesions and conditions that disrupt bladder function, disorders of the peripheral (pelvic, pudendal, and hypogastric) neurons that supply the bladder can occur. These neuropathies can selectively interrupt sensory or motor pathways or involve both pathways.

Bladder atony and dysfunction is a frequently occurring complication of diabetes mellitus. The disorder initially affects the sensory axons of the urinary bladder without involvement of the pudendal nerve. This leads to large residual volumes after micturition, sometimes complicated by infection.¹⁰ There frequently is a need for straining, accompanied by hesitation, weakness of the stream, dribbling, and a sensation of incomplete bladder emptying.11 The chief complications are vesicoureteral reflux and ascending urinary tract infection. Because persons with diabetes are already at risk for development of glomerular disease, reflux can have serious effects on kidney function. Treatment consists of client education, including the need for frequent voidings (e.g., every 3 to 4 hours while awake), use of abdominal compression to effect more complete bladder emptying, and intermittent catheterization when necessary.11

Nonrelaxing External Sphincter. Another condition that affects micturition and bladder emptying is the nonrelaxing external sphincter.¹² This condition usually is related to a delay in maturation, developmental regression, psychomotor disorders, or locally irritative lesions. Inadequate relaxation of the external sphincter can be the result of anxiety or depression. Any local irritation can produce spasms of the sphincter by means of afferent sensory input from the pudendal nerve; included are vaginitis, perineal inflammation, and inflammation or irritation of the urethra. In men, chronic prostatitis contributes to the impaired relaxation of the external sphincter.

Treatment. The goals of treatment for conditions that interfere with bladder emptying focus on preventing bladder overdistention, urinary tract infections, and potentially life-threatening renal damage, and reducing the undesirable social and psychological effects of the disorder. The methods used in treatment of neurogenic bladder disorders are individualized based

on the type of neurologic lesion that is involved; information obtained through the health history, including fluid intake; report or observation of voiding patterns; presence of other health problems; urodynamic studies when indicated; and the ability of the person to participate in the treatment. Treatment methods include catheterization, bladder retraining, pharmacologic manipulation of bladder function, and surgery.

Urinary Incontinence

The Urinary Incontinence Guideline Panel defines urinary incontinence as an involuntary loss of urine that is sufficient to be a problem.¹³ This panel was convened by the Agency for Health Care Policy and Research in 1992 and again in 1996 for the purpose of developing specific guidelines to improve the care of persons with urinary incontinence.⁶

Urinary incontinence affects approximately 13 million Americans. Many body functions decline with age, and incontinence, although not a normal accompaniment of the aging process, is seen with increased frequency in elderly persons. For noninstitutionalized persons older than 60 years, the prevalence of urinary incontinence ranges from 15% to 35%, with women affected twice as often as men.⁶ The increase in health problems often seen in elderly persons probably contributes to the greater frequency of incontinence. Despite the prevalence of incontinence, most affected persons do not seek help for it, primarily because of embarrassment or because they are not aware that help is available.

Incontinence can be caused by a number of conditions. It can occur without the person's knowledge; at other times, the person may be aware of the condition but be unable to prevent it. The Urinary Incontinence Guideline Panel has identified four main types of incontinence: stress incontinence, urge incontinence, overflow incontinence, and mixed incontinence, which is a combination of stress and urge incontinence.⁶ Recently, the term *overactive bladder* has been designated as a term to replace *urge incontinence*.¹⁴ Table 25-2 summarizes the characteristics of stress, urge incontinence/overactive bladder, and overflow incontinence.

Incontinence may occur as a transient and correctable phenomenon, or it may not be totally correctable and occur with

TABLE 25-2 Types and Characteristics of Urinary Incontinence				
Туре	Characteristics			
Stress	Involuntary loss of urine associated with activities, such as coughing, that in- crease intra-abdominal pressure			
Overactive bladder/ urge incontinence	Urgency and frequency associated with hyperactivity of the detrusor muscle; may or may not involve involuntary loss of urine			
Overflow	Involuntary loss of urine when intravesi- cular pressure exceeds maximal ure- thral pressure in the absence of detrusor activity			

KEY CONCEPTS

INCONTINENCE

- Incontinence represents the involuntary loss of urine due to increased bladder pressures (overactive bladder with urge incontinence or overflow incontinence) or decreased ability of the vesicourethral sphincter to prevent the escape of urine (stress incontinence).
- Overactive bladder with urge incontinence is caused by neurogenic or myogenic disorders that result in hyperactive bladder contractions.
- Overflow incontinence results from overfilling of the bladder with escape of urine.
- Stress incontinence is caused by the decreased ability of the vesicourethral sphincter to prevent the escape of urine during activities, such as lifting and coughing, that raise bladder pressure above the sphincter closing pressure.

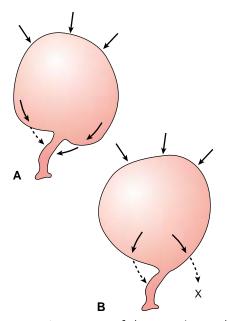
various degrees of frequency. Among the transient causes of urinary incontinence are confusional states; medications that alter bladder function or perception of bladder filling and the need to urinate; diuretics and conditions that increase bladder filling; restricted mobility; and stool impaction.¹⁵

Stress Incontinence

Stress incontinence is the involuntary loss of urine during coughing, laughing, sneezing, or lifting that increases intraabdominal pressure. The most common cause is hypermobility and significant displacement of the urethra during exertion.

In women, the angle between the bladder and the posterior proximal urethra (i.e., urethrovesical junction) is important to continence. This angle normally is 90 to 100 degrees, with at least one third of the bladder base contributing to the angle when not voiding¹⁶ (Fig. 25-5). During the first stage of voiding, this angle is lost as the bladder descends. In women, diminution of muscle tone associated with normal aging, childbirth, or surgical procedures can cause weakness of the pelvic floor muscles and result in stress incontinence by obliterating the critical posterior urethrovesical angle. In these women, loss of the posterior urethrovesical angle, descent and funneling of the bladder neck, and backward and downward rotation of the bladder occur, so that the bladder and urethra are already in an anatomic position for the first stage of voiding. Any activity that causes downward pressure on the bladder is sufficient to allow the urine to escape involuntarily.

Another cause of stress incontinence is intrinsic urethral deficiency, which may result from congenital sphincter weakness, as occurs with meningomyelocele. It also may be acquired as a result of trauma, irradiation, or sacral cord lesion. Stress incontinence in men may result from a congenital defect or from trauma or surgery to the bladder outlet, as occurs with prostatectomy. Neurologic dysfunction, as occurs with impaired sympathetic innervation of the bladder neck, impaired pelvic nerve



■ FIGURE 25-5 ■ Importance of the posterior urethrovesical (PU-V) angle to the continence mechanism. (A) In the presence of the normal PU-V angle, sudden changes in intra-abdominal pressure are transmitted optimally (indicated by the *arrows* with dotted lines) to all sides of the proximal urethra. In this way, intraurethral pressure is maintained higher than the simultaneously elevated intravesicular pressure. This prevents loss of urine with sudden stress. (B) Loss of the PU-V angle results in displacement of the vesicle neck to the most dependent portion of the bladder, preventing the equal transmission of sudden increases in intraabdominal pressure to the lumen of the proximal urethra. Thus, the pressure in the region of the vesicle neck rises considerably more than the intraurethral pressure just beyond it, and stress incontinence occurs. (Green J.T., Jr. [1968]. *Obstetrical and Gynecological Survey* 23, 603. Reprinted with permission)

innervation to the intrinsic sphincter, or impaired pudendal nerve innervation to the external sphincter, may be a contributing factor.

Urge Incontinence/Overactive Bladder

The Urinary Incontinence Guideline Panel has defined urge incontinence as the involuntary loss of urine associated with a strong desire to void (urgency).⁶ To expand the number and types of patients eligible for clinical trials, the U.S. Food and Drug Administration adopted the term *overactive bladder* to describe the clinical syndrome that describes urge incontinence, as well as urgency, frequency, dysuria, and nocturia.¹⁴ Overactive bladder, which often is associated with urge incontinence, can occur without incontinence.¹⁷

Although some cases of overactive bladder result from specific conditions such as acute or chronic urinary tract infections, in many cases the cause is unknown. Regardless of the primary cause of overactive bladder, two types of mechanisms are thought to contribute to its symptomatology: those involving CNS control of bladder sensation and emptying (neurogenic) and those involving the smooth muscle of the bladder itself (myogenic).¹⁷

The CNS functions as an on-off switching circuit for voluntary control of bladder function. Therefore, neurologic damage to central inhibitory pathways or sensitization of peripheral afferent terminals in the bladder may trigger bladder overactivity caused by uncontrolled voiding reflexes. Neurogenic causes of overactive bladder include stroke, Parkinson's disease, and multiple sclerosis. Other neurogenic causes of overactive bladder include increased peripheral afferent activity or increased peripheral sensitivity to efferent impulses.

Myogenic causes of overactive bladder result from changes in the excitability of smooth muscle of the bladder itself. One example is overactive bladder associated with bladder outlet obstruction. It is hypothesized that the sustained increase in intravesicular pressure that occurs with the outlet obstruction causes a partial destruction of the efferent nerve endings that control bladder excitability.¹⁷ The result is urgency and frequency of urination caused by spontaneous bladder contractions resulting from detrusor muscle hyperexcitability. Disorders of detrusor muscle structure and excitability also can occur as the result of the aging process or disease conditions such as diabetes mellitus. Overactive bladder symptoms usually are exaggerated by incomplete bladder emptying.

In persons with overactive bladder, urge incontinence may occur because the interval between knowing the bladder needs to be emptied and being able to stop it from emptying may be less than the time needed to reach the lavatory. Musculoskeletal disorders, such as arthritis and joint instability, also may prevent an otherwise continent person from reaching the toilet in time. Drugs such as hypnotics, tranquilizers, and sedatives can interfere with the conscious inhibition of voiding, leading to urge incontinence. Diuretics, particularly in elderly persons, increase the flow of urine and may contribute to incontinence, particularly in persons with diminished bladder capacity and in those who have difficulty reaching the toilet.

Treatment methods for overactive bladder include the use of behavioral methods and pharmacologic agents. Behavioral methods include fluid management, modification of voiding frequency, and bladder retraining.¹⁸ Bladder retraining and biofeedback techniques seek to re-establish cortical control over bladder function by having the person ignore urgency and respond only to cortical signals during waking hours. Two newer anticholinergic medications, tolterodine and extended-release oxybutynin, may be used to inhibit detrusor muscle hyperactivity and thereby increase overall bladder capacity.^{18,19}

Overflow Incontinence

Overflow incontinence is an involuntary loss of urine that occurs when intravesicular pressure exceeds the maximal urethral pressure because of bladder distention in the absence of detrusor activity. It can occur with retention of urine caused by nervous system lesions or obstruction of the bladder neck. With this type of incontinence, the bladder is distended and small amounts of urine are passed, particularly at night. In males, one of the most common causes of obstructive incontinence is enlargement of the prostate gland. Another cause that commonly is overlooked is fecal impaction (*i.e.*, dry, hard feces in the rectum). When a large bolus of stool forms in the rectum, it can push against the urethra and block the flow of urine.

Other Causes of Incontinence

Another cause of incontinence is decreased bladder compliance or distensibility. This abnormal bladder condition may result from radiation therapy, radical pelvic surgery, or interstitial cystitis. Many persons with this disorder have severe urgency related to bladder hypersensitivity that results in loss of bladder elasticity, such that any small increase in bladder volume or detrusor function causes a sharp increase in bladder pressure and severe urgency.

Treatment with drugs such as diuretics may cause the bladder to fill more rapidly than usual, making it difficult to reach the bathroom in time if there are problems with mobility or if a bathroom is not readily available. Night sedation may cause a person to sleep through the signal that normally would waken a person so he or she could get up and empty the bladder and avoid wetting the bed.

Diagnosis and Treatment

Urinary incontinence is a frequent and major health problem. It increases social isolation, frequently leads to institutionalization of elderly persons, and predisposes to infections and skin breakdown.

Urinary incontinence is not a single disease but a symptom with many possible causes. As a symptom, it requires full investigation to establish its cause. This usually is accomplished through a careful history, physical examination, blood tests, and urinalysis. A voiding record (*i.e.*, diary) may be used to determine the frequency, timing, amount of voiding, and other factors associated with the incontinence.^{13,14} Because many drugs affect bladder function, a full drug history is essential. Estimation of PVR volume is recommended for all persons with incontinence. Provocative stress testing, such as having a person relax and then cough vigorously, may be done when stress incontinence is suspected. Urodynamic studies may be needed to provide information about urinary pressures and urine flow rates.

Treatment or management depends on the type of incontinence, accompanying health problems, and the person's age. Exercises to strengthen the pelvic muscles and surgical correction of pelvic relaxation disorders often are used for women with stress incontinence. The α -adrenergic agonist drugs, such as pseudoephedrine, increase sympathetic relaxation of the detrusor muscle and internal sphincter tone and may be used in treating stress incontinence.^{6,16}

Noncatheter devices to obstruct urine flow or collect urine as it is passed may be used when urine flow cannot be controlled. Indwelling catheters, although a solution to the problem of urinary incontinence, usually are considered only after all other treatment methods have failed. In some types of incontinence, such as that associated with spinal cord injury or meningomyelocele, self-catheterization provides the means for controlling urine elimination.

Surgical intervention may be considered when other treatment methods have proved ineffective. Three types of surgical procedures are used: procedures that increase outlet resistance, surgeries that decrease detrusor muscle instability, and operations that remove outflow obstruction to reduce overflow incontinence and detrusor muscle instability.⁶ A minimally invasive procedure for the treatment of stress incontinence is periurethral injection of a bulking agent (glutaraldehyde cross-linked bovine collagen or carbon-coated beads). Both of these agents typically require multiple treatment sessions to achieve cure.¹⁴

Surgically implanted artificial sphincters are available for use in males and females. These devices consist of an inflatable cuff that surrounds the proximal urethra. The cuff is connected by tubing to an implanted fluid reservoir and an inflation bulb.

Special Needs of Elderly Persons

Urinary incontinence is a common problem in elderly persons. An estimated 15% to 35% of community-dwelling elders and 50% of institutionalized elders have severe urinary incontinence.^{20,21} Many factors contribute to incontinence in elderly persons, a number of which can be altered.

Physiologically, detrusor muscle function tends to decline with aging so there is a trend toward a reduction in the strength of bladder contraction and impairment in emptying that leads to larger PVR volumes.⁴ It has been proposed that many of these changes are attributable to degenerative detrusor muscle changes, rather than neurologic changes, as was once thought. The combination of involuntary detrusor contraction (detrusor hyperactivity) leading to urge incontinence along with impaired contractile function leads to incomplete bladder emptying.

Pelvic relaxation disorders are more frequent in older than in younger women, and prostatic hypertrophy is more common in older than in younger men. Many elderly persons have difficulty getting to the toilet in time. This can be caused by arthritis that makes walking or removing clothing difficult or by failing vision that makes trips to the bathroom precarious, especially in new and unfamiliar surroundings.

Medication prescribed for other health problems may prevent a healthy bladder from functioning normally. Potent, fast-acting diuretics are known for their ability to cause urge incontinence. Psychoactive drugs, such as tranquilizers and sedatives, may diminish normal attention to bladder cues. Impaired thirst or limited access to fluids predisposes to constipation with urethral obstruction and overflow incontinence and to concentrated and infected urine, which increases bladder excitability.

Treatment may involve changes in the physical environment so that the older person can reach the bathroom more easily or remove clothing more quickly. Habit training with regularly scheduled toileting—usually every 2 to 4 hours often is effective. Many elderly persons who void on a regular schedule can gradually increase the interval between toileting while improving their ability to suppress bladder instability. The treatment plan may require dietary changes to prevent constipation or a plan to promote adequate fluid intake to ensure adequate bladder filling and prevent urinary stasis and symptomatic urinary tract infections.

In summary, alterations in bladder function include urinary obstruction with retention of urine, neurogenic bladder, and urinary incontinence with involuntary loss of urine. Urine retention occurs when the outflow of urine from the bladder is obstructed because of urethral obstruction or impaired bladder innervation. Urethral obstruction causes bladder irritability, detrusor muscle hypertrophy, trabeculation and the formation of diverticula, development of hydroureters, and eventually, renal failure.

Neurogenic bladder is caused by interruption in the innervation of the bladder. It can result in spastic bladder dysfunction caused by failure of the bladder to fill or flaccid bladder dysfunction caused by failure of the bladder to empty. Spastic bladder dysfunction usually results from neurologic lesions that are above the level of the sacral micturition reflex center; flaccid bladder dysfunction results from lesions at the level of the sacral micturition reflexes or peripheral innervation of the bladder. A third type of neurogenic disorder involves a nonrelaxing external sphincter.

Urinary incontinence is the involuntary loss of urine in amounts sufficient to be a problem. It may manifest as stress incontinence, in which the loss of urine occurs as a result of coughing, sneezing, laughing, or lifting; overactive bladder, characterized by frequency and urgency associated with hyperactive bladder contractions; or overflow incontinence, which results when intravesicular pressure exceeds the maximal urethral pressure because of bladder distention. Other causes of incontinence include a small, contracted bladder or external environmental conditions that make it difficult to access proper toileting facilities.

The treatment of urinary obstruction, neurogenic bladder, and incontinence requires careful diagnosis to determine the cause and contributing factors. Treatment methods include correction of the underlying cause, such as obstruction caused by prostatic hyperplasia; pharmacologic methods to improve bladder and external sphincter tone; behavior methods that focus on bladder and habit training; exercises to improve pelvic floor function; and the use of catheters and urine collection devices.

CANCER OF THE BLADDER

Bladder cancer is the most common form of urinary tract cancer in the United States, accounting for more than 56,500 new cases and 12,600 deaths each year.²² Whites are twice as likely to have bladder cancer as African Americans.²² It occurs most commonly in people in their late seventh decade.²³ When detected and treated early, the chances for survival are very good. The 5-year survival rate for early noninvasive bladder cancers is approximately 94%.

Approximately 90% of bladder cancers are derived from the transitional (urothelial) cells that line the bladder.^{24,25} These tumors can range from low-grade noninvasive tumors to high-grade tumors that invade the bladder wall and metastasize frequently. The low-grade tumors, which may recur after resection, have an excellent prognosis, with only a small number (2% to 10%) progressing to higher-grade tumors.²⁶ The high-grade tumors tend to have greater invasive and metastatic potential and are potentially fatal in approximately 60% of cases within 10 years of diagnosis.²⁴

Although the cause of bladder cancer is unknown, evidence suggests that its origin is related to local influences, such as carcinogens that are excreted in the urine and stored in the bladder. These include the breakdown products of aniline dyes used in the rubber and cable industries. Smoking also deserves attention. Fifty percent to 80% of all bladder cancers in men are associated with cigarette smoking. Chronic bladder infections and bladder stones also increase the risk of bladder cancer. Bladder cancer occurs more frequently among persons harboring the parasite *Schistosoma haematobium* in their bladders. The parasite is endemic in Egypt and Sudan. It is not known whether the parasite excretes a carcinogen or produces its effects through irritation of the bladder.

The most common sign of bladder cancer is painless hematuria.^{24–26} Gross hematuria is a presenting sign in 75% of persons with the disease, and microscopic hematuria is present in most others. Frequency, urgency, and dysuria occasionally accompany the hematuria. Because hematuria often is intermittent, the diagnosis may be delayed. Periodic urine cytology is recommended for all persons who are at high risk for the development of bladder cancer because of exposure to urinary tract carcinogens. Ureteral invasion leading to bacterial and obstructive renal disease and dissemination of the cancer are potential complications and ultimate causes of death. The prognosis depends on the histologic grade of the cancer and the stage of the disease at the time of diagnosis.

Diagnosis and Treatment

Diagnostic methods include cytologic studies, excretory urography, cystoscopy, and biopsy. Ultrasonography, CT scans, and MRI are used as aids for staging the tumor. Cytologic studies performed on biopsy tissues or cells obtained from bladder washings may be used to detect the presence of malignant cells.

The treatment of bladder cancer depends on the extent of the lesion and the health of the patient. Endoscopic resection usually is done for diagnostic purposes and may be used as a treatment for superficial lesions. Diathermy (*i.e.*, electrocautery) may be used to remove the tumors. Segmental surgical resection may be used for removing a large single lesion. When the tumor is invasive, cystectomy with resection of the pelvic lymph nodes frequently is the treatment of choice. In males, the prostate and seminal vesicles often are removed as well. Cystectomy requires urinary diversion, an alternative reservoir, usually created from the ileum (*e.g.*, an ileal loop), which is designed to collect the urine. External beam radiation is an alternative to radical cystectomy in some persons with deeply infiltrating bladder cancers.

Although a number of chemotherapeutic drugs have been used in the treatment of bladder cancer, no chemotherapeutic regimens for the disease have been established. Perhaps of more importance is the increasing use of intravesicular chemotherapy, in which the cytotoxic drug is instilled directly into the bladder, thereby avoiding the side effects of systemic therapy.²⁶ The intervesicular administration of bacillus Calmette-Guérin (BCG) vaccine, made from a strain of *Mycobacterium bovis*, may be used to reduce the disease recurrence or progression. The vaccine is thought to act as a nonspecific stimulator of cellmediated immunity. Instillation into the bladder decreases the risk of systemic dissemination.²⁵

In summary, cancer of the bladder is the most common cause of urinary tract cancer in the United States. Bladder cancers fall into two major groups: low-grade noninvasive tumors and high-grade invasive tumors that are associated with metastasis and a worse prognosis. Although the cause of cancer of the bladder is unknown, evidence suggests that carcinogens excreted in the urine may play a role. Microscopic and gross, painless hematuria are the most common presenting signs of bladder cancer. The methods used in the treatment of bladder cancer depend on the cytologic grade of the tumor and the lesion's degree of invasiveness. The methods include surgical removal of the tumor, radiation therapy, and chemotherapy. In many cases, chemotherapeutic or immunotherapeutic agents can be instilled directly into the bladder, thereby avoiding the side effects of systemic therapy.

REVIEW QUESTIONS

Trace the innervation of the bladder and control of micturition from the detrusor muscle and external sphincter, the micturition centers in the sacral and thoracolumbar cord, the pontine micturition center, and the cerebral cortex.

Describe the causes of and compensatory changes that occur with urinary tract obstruction.

Differentiate lesions that produce storage dysfunction associated with spastic bladder from those that produce emptying dysfunction associated with flaccid bladder in terms of the level of the lesions and their effects on bladder function.

Define *incontinence* and list the categories of this condition.

Discuss the difference between superficial and invasive bladder cancer in terms of bladder involvement, extension of the disease, and prognosis.

State the most common sign of bladder cancer.

connection-J

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- Berne R.M., Levy M.N. (2000). Principles of physiology (3rd ed., p. 413). St. Louis: C.V. Mosby.
- Dmochowski R.R., Appell R.A. (2000). Advances in pharmacologic management of overactive bladder. Urology 56 (Suppl. 6A), 41–49.
- Kandel E.R., Schwartz J.H., Jessel T.M. (2000). Principles of neural science (4th ed.). New York: McGraw-Hill.
- Fowler C.J. (1999). Neurological disorders of micturition and their treatment. Brain 122, 1213–1231.
- Tanagho E.A. (2000). Urodynamic studies. In Tanagho E.A., McAninch J.W. (Eds.), Smith's general urology (15th ed., pp. 516–537). New York: Lange Medical Books/McGraw-Hill.
- Fantl J.A., Newman D.K., Colling J., et al., for the Public Health Service, Agency for Health Care Policy and Research. (1996). Urinary

incontinence in adults: Acute and chronic management. Clinical Practice Guideline no. 2, 1996 update. AHCPR publication no. 96–0682. Rockville, MD: U.S. Department of Health and Human Services.

- Tanagho E.A. (2000). Urinary obstruction and stasis. In Tanagho E.A., McAninch J.W. (Eds.), *Smith's general urology* (15th ed., pp. 208–220). New York: Lange Medical Books/McGraw-Hill.
- Elliott D.S., Boone T.B. (2000). Recent advances in management of neurogenic bladder. Urology 56 (Suppl. 6A), 76–81.
- Tanagho E.A., Lue T.F. (2000). Neuropathic bladder disorders. In Tanagho E.A., McAninch J.W. (Eds.), *Smith's general urology* (15th ed., pp. 498–515). New York: Lange Medical Books/McGraw-Hill.
- 10. Said G. (1996). Diabetic neuropathy. Journal of Neurology 243, 431-440.
- 11. Bays H.E., Pfiefer M.A. (1988). Peripheral diabetic neuropathy. *Medical Clinics of North America* 72, 1439–1464.
- 12. Thon W., Altwein J.E. (1984). Voiding dysfunction. Urology 23, 323.
- Urinary Incontinence Guideline Panel. (1992). Urinary incontinence in adults: Clinical practice guidelines. AHCPR publication no. 92-0038. Rockville, MD: Agency for Health Care Policy and Research, Public Health Service, U.S. Department of Health and Human Services.
- Culligan P.J., Heit M. (2000). Urinary incontinence in women: Evaluation and management. *American Family Physician* 62, 2433– 2452.
- 15. Gray M., Burns S.M. (1996). Continence management. *Critical Care Clinics of North America* 8, 29–38.
- Green T.H. (1975). Urinary stress incontinence: Differential diagnosis, pathophysiology, and management. American Journal of Obstetrics and Gynecology 122, 368–382.
- Dmochowski R.R., Appell R.A. (2000). Advancements in pharmacologic management of overactive bladder. *Urology* 56 (Suppl. 6A), 41–49.
- Wein A.J. (2001). Putting overactive bladder into clinical perspective. Patient Care for the Nurse Practitioner (Spring Suppl.), 1–5.
- Roberts R.R. (2001). Current management strategies for overactive bladder. Patient Care for the Nurse Practitioner (Spring Suppl.), 22–30.
- Lee S.Y., Phanumus D., Fields S.D. (2000). Urinary incontinence: A primary guide to managing acute and chronic symptoms in older adults. *Geriatrics* 55(11), 65–71.
- Weiss B.D. (1998). Diagnostic evaluation of urinary incontinence in geriatric patients. *American Family Physician* 57, 2675–2684, 2688– 2690.
- American Cancer Society. (2002). Cancer facts & figures 2002. [Online]. Available: http://www.cancer.org.
- 23. Lee R., Droller M.J. (2000). The natural history of bladder cancer. *Urologic Clinics of North America* 27(1), 1–13.
- Cotran R.S., Kumar V., Collins T. (1999). Robbins pathologic basis of disease (6th ed., pp. 1003–1008). Philadelphia: W.B. Saunders.
- Murphy G.P. (2001). Urologic and male genital cancer. In Lenhard R.E., Osteen R.T., Gansler T (Eds.), *Clinical oncology* (pp. 408–415). Atlanta: American Cancer Society.
- Carroll P. (2000). Urothelial carcinoma: Cancers of the bladder, ureter, and renal pelvis. In Tanagho E.A., McAninch J.W. (Eds.), *Smith's general urology* (15th ed., pp. 355–376). New York: Lange Medical Books/McGraw-Hill.

Seven

CHAPTER

Alterations in the Gastrointestinal System

Structure and Function of the Gastrointestinal System

Structure and Organization of the **Gastrointestinal Tract Upper Gastrointestinal Tract Esophagus** Stomach Middle Gastrointestinal Tract Lower Gastrointestinal Tract **Gastrointestinal Wall Structure** Innervation and Motility Innervation **Enteric Nervous System** Autonomic Nervous System Swallowing and Esophageal Motility **Gastric Motility Small Intestinal Motility Colonic Motility Defecation** Hormonal and Secretory Function **Gastrointestinal Hormones Gastrointestinal Secretions Salivary Secretions Gastric Secretions** Intestinal Secretions

Digestion and Absorption Carbohydrates Fats Proteins Anorexia, Nausea, and Vomiting Anorexia Nausea Vomiting

Swith its lumen inside the body and its wall acting as an interface between the internal and external environments. The wall does not normally allow harmful agents to enter the body, nor does it permit body fluids and other materials to escape. The process of digestion and absorption of nutrients requires an intact and healthy gastrointestinal tract epithelial lining that can resist the effects of its own digestive secretions. The process also involves movement of materials through the gastrointestinal tract at a rate that facilitates absorption, and it requires the presence of enzymes for the digestion and absorption of nutrients.

As a matter of semantics, the gastrointestinal tract also is referred to as the *digestive tract*, the *alimentary canal*, and at times, the *gut*. The intestinal portion also may be called the *bowel*. For the purposes of this text, the salivary glands, the liver, and the pancreas, which produce secretions that aid in digestion, are considered *accessory organs*.

STRUCTURE AND ORGANIZATION OF THE GASTROINTESTINAL TRACT

In the digestive tract, food and other materials move slowly along its length as they are systematically broken down into ions and molecules that can be absorbed into the body. In the large intestine, unabsorbed nutrients and wastes are collected for later elimination. Although the gastrointestinal tract is located inside the body, it is a long, hollow tube, the lumen (*i.e.*, hollow center) of which is an extension of the external environment. Nutrients do not become part of the internal environment until they have passed through the intestinal wall and have entered the blood or lymph channels.

For simplicity and understanding, the digestive system can be divided into four parts (Fig. 26-1). The upper part—the mouth, esophagus, and stomach—acts as an intake source and receptacle through which food passes and in which initial digestive processes take place. The middle portion consists of the small intestine—the duodenum, jejunum, and ileum. Most digestive and absorptive processes occur in the small intestine. The lower segment—the cecum, colon, and rectum—serves as a storage channel for the efficient elimination of waste. The fourth part consists of the accessory organs—the salivary glands, liver, and pancreas. These structures produce digestive secretions that help dismantle foods and regulate the use and storage of nutrients. The discussion in this chapter focuses on the first three parts of the gastrointestinal tract. The liver and pancreas are discussed in Chapter 28.

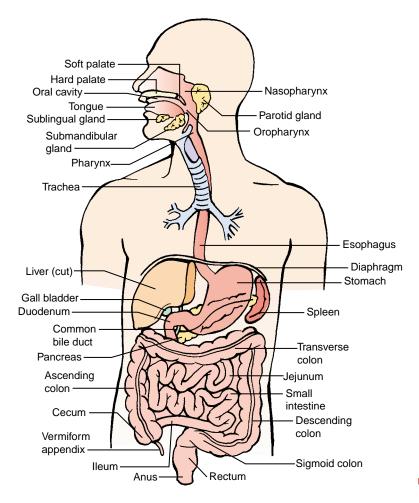
Upper Gastrointestinal Tract

The mouth forms the entryway into the gastrointestinal tract for food; it contains the teeth, used in the mastication of food, and the tongue and other structures needed to direct food toward the pharyngeal structures and the esophagus.

Esophagus

The esophagus is a tube that connects the oropharynx with the stomach. The esophagus begins at the lower end of the pharynx. It is a muscular, collapsible tube, approximately 25 cm (10 in) long, that lies behind the trachea. The muscular walls of the upper third of the esophagus are skeletal-type striated muscle; these muscle fibers are gradually replaced by smooth muscle fibers until, at the lower third of the esophagus, the muscle layer is entirely smooth muscle.

The esophagus functions primarily as a conduit for passage of food from the pharynx to the stomach, and the structures of its walls are designed for this purpose: the smooth muscle layers provide the peristaltic movements needed to move food along its length, and the epithelial layer secretes mucus, which protects its surface and aids in lubricating food. There are sphincters at either end of the esophagus: an upper esophageal



■ **FIGURE 26-1** ■ The digestive system.

STRUCTURE AND FUNCTION OF THE GASTROINTESTINAL TRACT

- The gastrointestinal tract is a long, hollow tube that extends from the mouth to the anus; food and fluids that enter the gastrointestinal tract do not become part of the internal environment until they have been broken down and absorbed into the blood or lymph channels.
- The wall of the gastrointestinal tract is essentially a five-layered tube: an inner mucosal layer; a supporting submucosal layer of connective tissue; a fourth and fifth layer of circular and longitudinal smooth muscle that functions to propel its contents in a proximal-to-distal direction; and an outer, twolayered peritoneum that encloses and prevents friction between the continuously moving segments of the intestine.
- The nutrients contained in ingested foods and fluids must be broken down into molecules that can be absorbed across the wall of the intestine. Gastric acids and pepsin from the stomach begin the digestive process: bile from the liver, digestive enzymes from the pancreas, and brush border enzymes break carbohydrates, fats, and proteins into molecules that can be absorbed from the intestine.

sphincter and a lower esophageal sphincter. The upper esophageal, or pharyngoesophageal, sphincter consists of a circular layer of striated muscle. The lower esophageal, or gastroesophageal, sphincter is an area approximately 3 cm above the junction with the stomach. The circular muscle in this area normally remains tonically contracted, creating a zone of high pressure that serves to prevent reflux of gastric contents into the esophagus. During swallowing, there is "receptive relaxation" of the lower esophageal sphincter, which allows easy propulsion of the esophageal contents into the stomach. The lower esophageal sphincter passes through an opening, or hiatus, in the diaphragm as it joins with the stomach, which is located in the abdomen. The portion of the diaphragm that surrounds the lower esophageal sphincter helps to maintain the zone of high pressure needed to prevent reflux of stomach contents into the esophagus.

Stomach

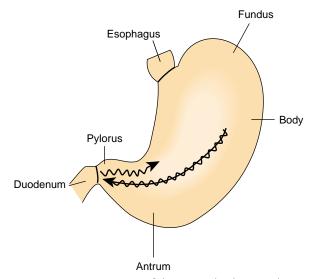
The stomach is a pouchlike structure that lies in the upper part of the abdomen and serves as a food storage reservoir during the early stages of digestion. Although the residual volume of the stomach is only approximately 50 mL, it can increase to almost 1000 mL before the intraluminal pressure begins to rise. The esophagus opens into the stomach through an opening called the *cardiac orifice*, so named because of its proximity to the heart. The part of the stomach that lies above and to the left of the cardiac orifice is called the *fundus*, the central portion is called the *body*, the orifice encircled by a ringlike muscle that opens into the small intestine is called the *pylorus*, and the portion between the body and pylorus is called the *antrum* (Fig. 26-2). The presence of a true pyloric sphincter is a matter of controversy. Regardless of whether an actual sphincter exists, contractions of the smooth muscle in the pyloric area control the rate of gastric emptying.

Middle Gastrointestinal Tract

The small intestine, which forms the middle portion of the digestive tract, consists of three subdivisions: the duodenum, the jejunum, and the ileum. The duodenum, which is approximately 22 cm (10 in) long, connects the stomach to the jejunum and contains the opening for the common bile duct and the main pancreatic duct. Bile and pancreatic juices enter the intestine through these ducts. It is in the jejunum and ileum, which together are approximately 7 m (23 ft) long and must be folded onto themselves to fit into the abdominal cavity, that food is digested and absorbed.

Lower Gastrointestinal Tract

The large intestine, which forms the lower gastrointestinal tract, is approximately 1.5 m (4.5 to 5 ft) long and 6 to 7 cm (2.4 to 2.7 in) in diameter. It is divided into the cecum, colon, rectum, and anal canal. The cecum is a blind pouch that projects down at the junction of the ileum and the colon. The ileocecal valve lies at the upper border of the cecum and prevents the return of feces from the cecum into the small intestine. The appendix arises from the cecum approximately 2.5 cm (1 in) from the ileocecal valve. The colon is further divided into ascending, transverse, descending, and sigmoid portions. The ascending colon extends from the cecum to the undersurface of the liver, where it turns abruptly to form the right colic (hepatic) flexure. The transverse colon crosses the upper half of the abdominal cavity from right to left and then curves sharply downward beneath the lower end of the spleen, forming the left colic



■ **FIGURE 26-2** ■ Structures of the stomach, showing the pacemaker area and the direction of chyme movement resulting from peristaltic contractions.

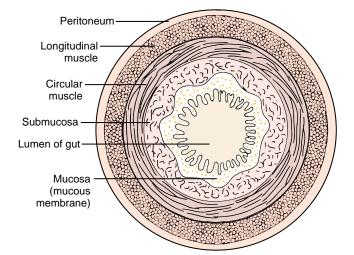
(splenic) flexure. The descending colon extends from the colic flexure to the rectum. The rectum extends from the sigmoid colon to the anus. The anal canal passes between the two medial borders of the levator ani muscles. Powerful sphincter muscles guard against fecal incontinence.

Gastrointestinal Wall Structure

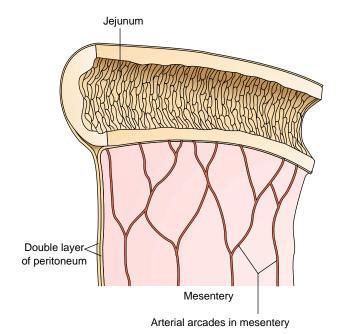
The digestive tract is essentially a five-layered tube (Fig. 26-3). The inner luminal layer, or *mucosal layer*, is so named because its cells produce mucus that lubricates and protects the inner surface of the alimentary canal. The epithelial cells in this layer have a rapid turnover rate and are replaced every 4 to 5 days. Approximately 250 g of these cells are shed each day in the stool. Because of the regenerative capabilities of the mucosal layer, injury to this layer of tissue heals rapidly without leaving scar tissue. The *submucosal layer* consists of connective tissue. This layer contains blood vessels, nerves, and structures responsible for secreting digestive enzymes. The third and fourth layers, the *circular* and *longitudinal muscle layers*, facilitate movement of the contents of the gastrointestinal tract. The outer layer, the *peritoneum*, is loosely attached to the outer wall of the intestine.

The peritoneum is the largest serous membrane in the body, having a surface area approximately equal to that of the skin. The peritoneum consists of two continuous layers—the parietal and the visceral peritoneum. The *parietal peritoneum* comes in contact with and is loosely attached to the abdominal wall, whereas the *visceral peritoneum* invests the viscera such as the stomach and intestines. A thin layer of serous fluid separates the parietal and visceral peritoneum, forming a potential space called the *peritoneal cavity*. The serous fluid forms a moist and slippery surface that prevents friction between the continuously moving abdominal structures. In certain pathologic states, the amount of fluid in the potential space of the peritoneal cavity is increased, causing a condition called *ascites*.

The jejunum and ileum are suspended by a double-layered fold of peritoneum called the *mesentery* (Fig. 26-4). The mesentery contains the blood vessels, nerves, and lymphatic vessels

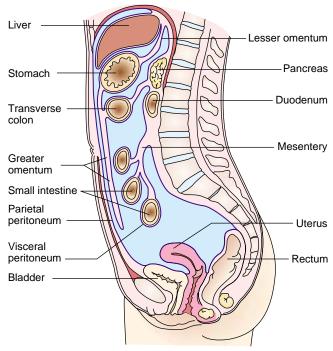


■ FIGURE 26-3 ■ Transverse section of the digestive system. (Thomson J.S. [1977]. *Core textbook of anatomy*. Philadelphia: J.B. Lippincott)



■ **FIGURE 26-4** ■ The attachment of the mesentery to the small bowel. (Thomson J.S. [1977]. *Core textbook of anatomy*. Philadel-phia: J.B. Lippincott)

that supply the intestinal wall. The mesentery is gathered in folds that attach to the dorsal abdominal wall along a short line of insertion, giving a fan-shaped appearance, with the intestines at the edge. A filmy, double fold of peritoneal membrane called the *greater omentum* extends from the stomach to cover the transverse colon and folds of the intestine (Fig. 26-5).



■ **FIGURE 26-5** ■ Reflections of the peritoneum as seen in sagittal section.

The greater omentum protects the intestines from cold. It always contains some fat, which in obese persons can be a considerable amount. The omentum also controls the spread of infection from gastrointestinal contents. In the case of infection, the omentum adheres to the inflamed area so that the infection is less likely to enter the peritoneal cavity. The lesser omentum extends between the transverse fissure of the liver and the lesser curvature of the stomach.

In summary, the gastrointestinal tract is a long, hollow tube, the lumen of which is an extension of the external environment. The digestive tract can be divided into four parts: an upper part, consisting of the mouth, esophagus, and stomach; a middle part, consisting of the small intestine; a lower part, consisting of the cecum, colon, and rectum; and the accessory organs, consisting of the salivary glands, the liver, and the pancreas. Throughout its length, except for the mouth, throat, and upper esophagus, the gastrointestinal tract is composed of five layers: an inner mucosal layer, a submucosal layer, a layer of circular smooth muscle fibers, a layer of longitudinal smooth muscle fibers, and an outer serosal layer that forms the peritoneum and is continuous with the mesentery.

INNERVATION AND MOTILITY

The motility of the gastrointestinal tract propels food products and fluids along its length, from mouth to anus, in a manner that facilitates digestion and absorption. Except in the pharynx and upper third of the esophagus, smooth muscle provides the contractile force for gastrointestinal motility (the actions of smooth muscle are discussed in Chapter 1). The rhythmic movements of the digestive tract are selfperpetuating, much like the activity of the heart, and are influenced by local, humoral (*i.e.*, blood-borne), and neural influences. The ability to initiate impulses is a property of the smooth muscle itself. Impulses are conducted from one muscle fiber to another.

The smooth muscle movements of the gastrointestinal tract are tonic and rhythmic. The *tonic movements* are continuous movements that last for minutes or even hours. Tonic contractions occur at sphincters. The *rhythmic movements* consist of intermittent contractions that are responsible for mixing and moving food along the digestive tract. *Peristaltic movements* are rhythmic propulsive movements that occur when the smooth muscle layer constricts, forming a contractile band that forces the intraluminal contents forward. During peristalsis, the segment that lies distal to, or ahead of, the contracted portion relaxes, and the contents move forward with ease. Normal peristalsis always moves in the direction from the mouth toward the anus.

Innervation

Gastrointestinal function is controlled by the *enteric nervous sys*tem, which lies entirely within the wall of the gastrointestinal tract, and by the parasympathetic and sympathetic divisions of the autonomic nervous system (ANS).

Enteric Nervous System

The intramural neurons (*i.e.*, those contained within the wall of the gastrointestinal tract) consist of two networks: the myenteric and submucosal plexuses. Both plexuses are aggregates of ganglionic cells that extend along the length of the gastrointestinal wall. The myenteric (Auerbach's) plexus is located between the circular muscle and longitudinal muscle layers, and the submucosal (Meissner's) plexus between the mucosal layer and the circular muscle layers (Fig. 26-6). The activity of the neurons in the myenteric and submucosal plexuses is regulated by local influences, input from the ANS, and by interconnecting fibers that transmit information between the two plexuses.

The myenteric plexus consists mainly of a linear chain of interconnecting neurons that extend the full length of the gastrointestinal tract. Because it extends all the way down the intestinal wall and because it lies between the two muscle layers, it is concerned mainly with motility along the length of the gut. The submucosal plexus, which lies between the mucosal and circular muscle layers of the intestinal wall, is mainly concerned with controlling the function of each segment of the intestinal tract. It integrates signals received from the mucosal layer into local control of motility, intestinal secretions, and absorption of nutrients.

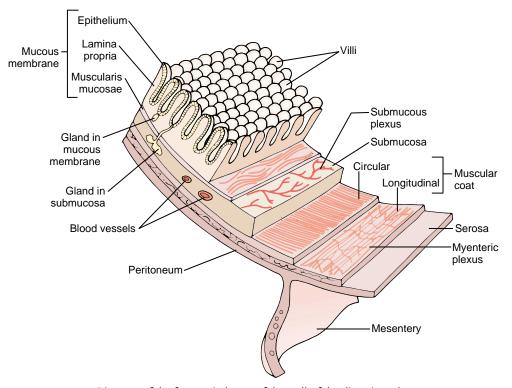
Intramural plexus neurons also communicate with receptors in the mucosal and muscle layers. Mechanoreceptors monitor the stretch and distention of the gastrointestinal tract wall, and chemoreceptors monitor the chemical composition (*i.e.*, osmolality, pH, and digestive products of protein and fat metabolism) of its contents. These receptors can communicate directly with ganglionic cells in the intramural plexuses or with visceral afferent fibers that influence ANS control of gastrointestinal function.

Autonomic Nervous System

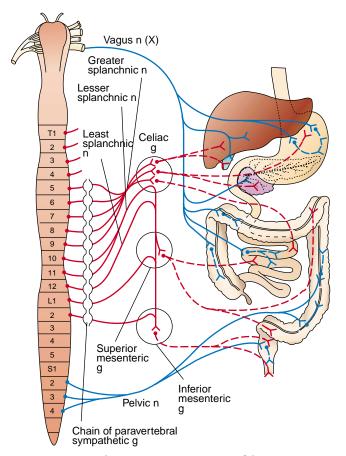
The gastrointestinal tract is innervated by both the sympathetic and parasympathetic nervous systems. Parasympathetic innervation to the stomach, small intestine, cecum, ascending colon, and transverse colon occurs by way of the vagus nerve (Fig. 26-7). The remainder of the colon is innervated by parasympathetic fibers that exit the sacral segments of the spinal cord by way of the pelvic nerve. Preganglionic parasympathetic fibers can synapse with intramural plexus neurons, or they can act directly on intestinal smooth muscle. Most parasympathetic fibers are excitatory. Numerous vagovagal reflexes influence motility and secretions of the digestive tract.

Sympathetic innervation of the gastrointestinal tract occurs through the thoracic chain of sympathetic ganglia and the celiac, superior mesenteric, and inferior mesenteric ganglia. Sympathetic control of gastrointestinal function is largely mediated by altering the activity of neurons in the intramural plexuses. The sympathetic nervous system exerts several effects on gastrointestinal function. It controls mucus secretion by the mucosal glands, reduces motility by inhibiting the activity of intramural plexus neurons, enhances sphincter function, and increases the vascular smooth muscle tone of the blood vessels that supply the gastrointestinal tract. The sympathetic fibers that supply the lower esophageal, pyloric, and internal and external anal sphincters are largely excitatory, but their role in controlling these sphincters is poorly understood.

Unit Seven: Alterations in the Gastrointestinal System



■ FIGURE 26-6 ■ Diagram of the four main layers of the wall of the digestive tube: mucosa, submucosa, muscular, and serosa (below the diaphragm).



■ **FIGURE 26-7** ■ The autonomic innervation of the gastrointestinal tract (g, ganglion; n, nerve).

Swallowing and Esophageal Motility

The movement of foods and fluids in the gastrointestinal tract begins with chewing, in the case of solid foods, and swallowing. Chewing begins the digestive process; it breaks the food into particles of a size that can be swallowed, lubricates it by mixing it with saliva, and mixes starch-containing food with salivary amylase. Although chewing usually is considered a voluntary act, it can be carried out involuntarily by a person who has lost the function of the cerebral cortex.

The swallowing reflex is a rigidly ordered sequence of events that results in the propulsion of food from the mouth to the stomach through the esophagus. Although swallowing is initiated as a voluntary activity, it becomes involuntary as food or fluid reaches the pharynx. Sensory impulses for the reflex begin at tactile receptors in the pharynx and esophagus and are integrated with the motor components of the response in an area of the reticular formation of the medulla and lower pons called the swallowing center. The motor impulses for the oral and pharyngeal phases of swallowing are carried in the trigeminal (V), glossopharyngeal (IX), vagus (X), and hypoglossal (XII) cranial nerves, and impulses for the esophageal phase are carried by the vagus nerve. Diseases that disrupt these brain centers or their cranial nerves disrupt the coordination of swallowing and predispose an individual to food and fluid lodging in the trachea and bronchi, leading to risk of asphyxiation or aspiration pneumonia.

Swallowing consists of three phases: an oral, or voluntary phase; a pharyngeal phase; and an esophageal phase. During the *oral phase*, the bolus is collected at the back of the mouth so the tongue can lift the food upward until it touches the posterior wall of the pharynx. At this point, the *pharyngeal phase* of swallowing is initiated. The soft palate is pulled upward, the palatopharyngeal folds are pulled together so that food does not enter the nasopharynx, the vocal cords are pulled together, and the epiglottis is moved so that it covers the larynx. Respiration is inhibited, and the bolus is moved backward into the esophagus by constrictive movements of the pharynx. Although the striated muscles of the pharynx are involved in the second stage of swallowing, it is an involuntary stage.

The third phase of swallowing is the *esophageal stage*. As food enters the esophagus and stretches its walls, local and central nervous system reflexes that initiate peristalsis are triggered. There are two types of peristalsis: primary and secondary. Primary peristalsis is controlled by the swallowing center in the brain stem and begins when food enters the esophagus. Secondary peristalsis is partially mediated by smooth muscle fibers in the esophagus and occurs when primary peristalsis is inadequate to move food through the esophagus. Peristalsis begins at the site of distention and moves downward. Before the peristaltic wave reaches the stomach, the lower esophageal sphincter relaxes to allow the bolus of food to enter the stomach. The pressure in the lower esophageal sphincter normally is greater than that in the stomach, an important factor in preventing the reflux of gastric contents.

Gastric Motility

The stomach serves as a reservoir for ingested solids and liquids. Motility of the stomach results in the churning and grinding of solid foods and regulates the emptying of the gastric contents, or chyme, into the duodenum. Peristaltic mixing and churning contractions begin in a pacemaker area in the middle of the stomach and move toward the antrum (see Fig. 26-2). They occur at a frequency of three to five contractions per minute, each with a duration of 2 to 20 seconds. As the peristaltic wave approaches the antrum, it speeds up, and the entire terminal 5 to 10 cm of the antrum contracts, occluding the pyloric opening. Contraction of the antrum reverses the movement of the chyme, returning the larger particles to the body of the stomach for further churning and kneading. Because the pylorus is contracted during antral contraction, the gastric contents are emptied into the duodenum between contractions.

Although the pylorus does not contain a true anatomic sphincter, it does function as a physiologic sphincter to prevent the backflow of gastric contents and allow them to flow into the duodenum at a rate commensurate with the ability of the duodenum to accept them. This is important because the regurgitation of bile salts and duodenal contents can damage the mucosal surface of the antrum and lead to gastric ulcers. Likewise, the duodenal mucosa can be damaged by the rapid influx of highly acid gastric contents.

Like other parts of the gastrointestinal tract, the stomach is richly innervated by the enteric nervous system and its connections with the sympathetic and parasympathetic nervous systems. Axons from the intramural plexuses innervate the smooth muscles and glands of the stomach. Parasympathetic innervation is provided by the vagus nerve and sympathetic innervation by the celiac ganglia. The emptying of the stomach is regulated by hormonal and neural mechanisms. The hormones cholecystokinin and gastric inhibitory peptide, which are thought to control gastric emptying, are released in response to the pH and the osmolar and fatty acid composition of the chyme. Local and central circuitry are involved in the neural control of gastric emptying. Afferent receptor fibers synapse with the neurons in the intramural plexus or trigger intrinsic reflexes by means of vagal or sympathetic pathways that participate in extrinsic reflexes.

Disorders of gastric motility can occur when the rate is too slow or too fast (see Chapter 27). A rate that is too slow leads to gastric retention. It can be caused by obstruction or gastric atony. Gastric atony can occur as a complication of visceral neuropathies in diabetes mellitus. Surgical procedures that disrupt vagal activity also can result in gastric atony. Abnormally fast emptying occurs in the dumping syndrome, which is a consequence of certain types of gastric operations. This condition is characterized by the rapid dumping of highly acidic and hyperosmotic gastric secretions into the duodenum and jejunum.

Small Intestinal Motility

The small intestine is the major site for the digestion and absorption of food; its movements are mixing and propulsive. Regular peristaltic movements begin in the duodenum near the entry sites of the common duct and the main hepatic duct. A series of local pacemakers maintains the frequency of intestinal contraction. The peristaltic movements (approximately 12 per minute in the jejunum) become less frequent as they move further from the pylorus, becoming approximately 9 per minute in the ileum.

The peristaltic contractions produce segmentation waves and propulsive movements of the small intestine. With segmentation waves, slow contractions of circular muscle occlude the lumen and drive the contents forward and backward. Most of the contractions that produce segmentation waves are local events involving only 1 to 4 cm at a time. They function mainly to mix the chyme with the digestive enzymes from the pancreas and to ensure adequate exposure of all parts of the chyme to the mucosal surface of the intestine, where absorption takes place. The frequency of segmenting activity increases after a meal, presumably stimulated by receptors in the stomach and intestine.

Propulsive movements occur with synchronized activity in a section 10 to 20 cm long. They are accomplished by contraction of the proximal portion of the intestine with the sequential relaxation of its distal, or anal, portion. After material has been propelled to the ileocecal junction by peristaltic movement, stretching of the distal ileum produces a local reflex that relaxes the sphincter and allows fluid to squirt into the cecum.

Motility disturbances of the small bowel are common, and auscultation of the abdomen can be used to assess bowel activity. Inflammatory changes increase motility. In many instances, it is not certain whether changes in motility occur because of inflammation or occur secondary to toxins and unabsorbed materials. Delayed passage of materials in the small intestine also can be a problem. Transient interruption of intestinal motility often occurs after gastrointestinal surgery. Intubation with suction often is required to remove the accumulating intestinal contents and gases until activity is resumed.

Colonic Motility

The storage function of the colon dictates that movements in this section of the gut are different from those in the small intestine. Movements in the colon are of two types. First are the segmental mixing movements, called *haustrations*, so named because they occur within sacculations called *haustra*. These movements produce a local digging-type action, which ensures that all portions of the fecal mass are exposed to the intestinal surface. Second are the propulsive mass movements, in which a large segment of the colon (≥ 20 cm) contracts as a unit, moving the fecal contents forward as a unit. Mass movements last approximately 30 seconds, followed by a 2- to 3-minute period of relaxation, after which another contraction occurs. A series of mass movements lasts only for 10 to 30 minutes and may occur only several times a day. Defecation normally is initiated by the mass movements.

Defecation

Defecation is controlled by the action of two sphincters, the internal and external anal sphincters. The internal sphincter is a several-centimeters long circular thickening of smooth muscle that lies inside the anus. The external sphincter, which is composed of striated voluntary muscle, surrounds the internal sphincter. Defecation is controlled by defecation reflexes. One of these reflexes is the intrinsic myenteric reflex mediated by the local enteric nervous system. It is initiated by distention of the rectal wall, with initiation of reflex peristaltic waves that spread through the descending colon, sigmoid colon, and rectum. A second defecation reflex, the parasympathetic reflex, is integrated at the level of the sacral cord. When the nerve endings in the rectum are stimulated, signals are transmitted first to the sacral cord and then reflexly back to the descending colon, sigmoid colon, rectum, and anus by way of the pelvic nerves (Fig. 26-7). These impulses greatly increase peristaltic movements as well as relax the internal sphincter.

To prevent involuntary defecation from occurring, the external anal sphincter, which is supplied by nerve fibers in the pudendal nerve, is under the conscious control of the cortex. As afferent impulses arrive at the sacral cord, signaling the presence of a distended rectum, messages are transmitted to the cortex. If defecation is inappropriate, the cortex initiates impulses that constrict the external sphincter and inhibit efferent parasympathetic activity. Normally, the afferent impulses in this reflex loop fatigue easily, and the urge to defecate soon ceases. At a more convenient time, contraction of the abdominal muscles compresses the contents in the large bowel, reinitiating afferent impulses to the cord.

In summary, motility of the gastrointestinal tract propels food products and fluids along its length from mouth to anus. Although the activity of gastrointestinal smooth muscle is selfpropagating and can continue without input from the nervous system, its rate and strength of contractions are requlated by a network of intramural neurons that receive input from the ANS and local receptors that monitor wall stretch and the chemical composition of luminal contents. Parasympathetic innervation occurs by means of the vagus nerve and nerve fibers from sacral segments of the spinal cord; it increases gastrointestinal motility. Sympathetic activity occurs by way of thoracolumbar output from the spinal cord, its paravertebral ganglia, and celiac, superior mesenteric, and inferior mesenteric ganglia. Sympathetic stimulation enhances sphincter function and reduces motility by inhibiting the activity of intramural plexus neurons.

HORMONAL AND SECRETORY FUNCTION

Each day, approximately 7000 mL of fluid is secreted into the gastrointestinal tract (Table 26-1). Approximately 50 to 200 mL of this fluid leaves the body in the stool; the remainder is reabsorbed in the small and large intestines. These secretions are mainly water and have sodium and potassium concentrations similar to those of extracellular fluid. Because water and electrolytes for digestive tract secretions are derived from the extracellular fluid compartment, excessive secretion or impaired absorption can lead to extracellular fluid deficit.

The secretory activity of the gut is influenced by local, humoral, and neural influences. Neural control of gastrointestinal secretory activity is mediated through the ANS. Secretory activity, like motility, is increased with parasympathetic stimulation and inhibited with sympathetic activity. Many of the local influences, including pH, osmolality, and chyme, consistently act as stimuli for neural and humoral mechanisms.

Gastrointestinal Hormones

The gastrointestinal tract is the largest endocrine organ in the body. It produces hormones that pass from the portal circulation into the general circulation and then back to the digestive tract, where they exert their actions. Among the hormones produced by the gastrointestinal tract are gastrin, secretin, and cholecystokinin. These hormones influence motility and the secretion of electrolytes, enzymes, and other hormones. The gastrointestinal tract hormones and their functions are summarized in Table 26-2.

The primary function of *gastrin* is the stimulation of gastric acid secretion. Gastrin also has a trophic, or growth-producing, effect on the mucosa of the small intestine, colon, and oxyntic (acid-secreting) gland area of the stomach. Removal of the tissue that produces gastrin results in atrophy of these structures. This atrophy can be reversed by the administration of exogenous gastrin.

Secretin is secreted by S cells in the mucosa of the duodenum and jejunum in an inactive form called *prosecretin*. When an acid chyme with a pH of less than 4.5 to 5.0 enters the intestine, secretin is activated and absorbed into the blood. Secretin causes the pancreas to secrete large quantities of fluid with a high bicarbonate concentration and low chloride concentration.

TABLE 26-1	Secretions of		
the Gastrointestinal Tract			
Secretions Amount Daily (mL)			
Salivary	1200		
Gastric	2000		
Pancreatic	1200		
Biliary	700		
Intestinal	2000		
Total	7100		

TABLE 26-2	Major Gastrointestinal Hormones and Their Actions		
Hormone	Site of Secretion	Stimulus for Secretion	Action
Cholecystokinin	Duodenum, jejunum	Products of protein digestion and long chain fatty acids	Stimulates contraction of the gallbladder; stimulates secretion of pancreatic enzymes; slows gastric emptying
Gastrin	Antrum of the stomach, duodenum	Vagal stimulation; epinephrine; neutral amino acids; calcium-containing fluids such as milk; and alcohol. Secretion is inhibited by acid contents in the an- trum of the stomach (below pH 2.5)	Stimulates secretion of gastric acid and pep- sinogen; increases gastric blood flow; stimu- lates gastric smooth muscle contraction; stimulates growth of gastric, small intestine, and colon mucosa
Secretin	Duodenum	Acid pH or chyme entering duodenum (below pH 3.0)	Stimulates secretion of bicarbonate-containing solution by pancreas and liver

The primary function of *cholecystokinin* is stimulation of pancreatic enzyme secretion. It potentiates the action of secretin, increasing the pancreatic bicarbonate response to low circulating levels of secretin. In addition to its effects on the pancreas, cholecystokinin also stimulates biliary secretion of fluid and bicarbonate and it regulates gallbladder contraction and gastric emptying.

Two other hormones that contribute to gastrointestinal function are gastric inhibitory peptide and motilin. *Gastric inhibitory peptide*, which is released from the intestinal mucosa in response to increased concentration of glucose and fats, inhibits gastric acid secretion, gastric motility, and gastric emptying. *Motilin*, which stimulates intestinal motility and contributes to the control of the interdigestive actions of the intestinal neurons, is released from the upper small intestine.

Gastrointestinal Secretions

Salivary Secretions

Saliva is secreted by the salivary glands. The salivary glands consist of the parotid, submaxillary, sublingual, and buccal glands. Saliva has three functions. The first is protection and lubrication. Saliva is rich in mucus, which protects the oral mucosa and coats the food as it passes through the mouth, pharynx, and esophagus. The sublingual and buccal glands produce only mucus-type secretions. The second function of saliva is its protective antimicrobial action. The saliva cleans the mouth and contains the enzyme lysozyme, which has an antibacterial action. Third, saliva contains ptyalin and amylase, which initiate the digestion of dietary starches. Secretions from the salivary glands are primarily regulated by the ANS. Parasympathetic stimulation increases flow, and sympathetic stimulation decreases flow. The dry mouth that accompanies anxiety attests to the effects of sympathetic activity on salivary secretions.

Mumps, or parotitis, is an infection of the parotid glands. Although most of us associate mumps with the contagious viral form of the disease, inflammation of the parotid glands can occur in the seriously ill person who does not receive adequate oral hygiene and who is unable to take fluids orally. Potassium iodide increases the secretory activity of the salivary glands, including the parotid glands. In a small percentage of persons, parotid swelling may occur in the course of treatment with this drug.

Gastric Secretions

In addition to mucus-secreting cells that line the entire surface of the stomach, the stomach mucosa has two types of glands: oxyntic (or gastric) glands and pyloric glands. The *oxyntic glands* are located in the proximal 80% (body and fundus) of the stomach. They secrete hydrochloric acid, pepsinogen, intrinsic factor, and mucus. The *pyloric glands* are located in the distal 20%, or antrum, of the stomach. The pyloric glands secrete mainly mucus, some pepsinogen, and the hormone gastrin.

The oxyntic gland area of the stomach is composed of glands and pits (Fig. 26-8). The surface area and gastric pits are lined with mucus-producing epithelial cells. The bases of the gastric pits contain the parietal (or oxyntic) cells, which secrete hydrochloric acid and intrinsic factor, and the chief (peptic) cells, which secrete large quantities of pepsinogen. There are approximately 1 billion parietal cells in the stomach; together they produce and secrete approximately 20 mEq of hydrochloric acid in several hundred milliliters of gastric juice each hour. Gastric intrinsic factor, which is produced by the parietal cells, is necessary for the absorption of vitamin B₁₂. The pepsinogen that is secreted by the chief cells is rapidly converted to pepsin when exposed to the low pH of the gastric juices.

One of the important characteristics of the gastric mucosa is resistance to the highly acid secretions that it produces. When the gastric mucosa is damaged by aspirin, nonsteroidal antiinflammatory drugs (NSAIDs), ethyl alcohol, or bile salts, this resistance is disrupted, and hydrogen ions move into the mucosal cells. As the hydrogen ions accumulate in the mucosal cells, intracellular pH decreases, enzymatic reactions become impaired, and cellular structures are disrupted. The result is local ischemia and tissue necrosis. The mucosal surface is further protected by prostaglandins.

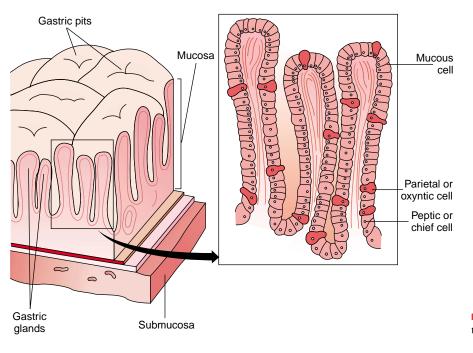
Parasympathetic stimulation (through the vagus nerve) and gastrin increase gastric secretions. Histamine increases gastric acid secretions. Gastric acid secretion and its relation to peptic ulcer are discussed in Chapter 27.

Intestinal Secretions

The small intestine secretes digestive juices and receives secretions from the liver and pancreas (see Chapter 28). An extensive array of mucus-producing glands, called *Brunner's glands*,

Dr Murtadha Alshareifi e-Library

468 Unit Seven: Alterations in the Gastrointestinal System



■ **FIGURE 26-8** ■ Gastric pit from body of the stomach.

are concentrated at the site where the contents from the stomach and secretions from the liver and pancreas enter the duodenum. These glands secrete large amounts of alkaline mucus that protect the duodenum from the acid content in the gastric chyme and from the action of the digestive enzymes. The activity of Brunner's glands is strongly influenced by ANS activity. For example, sympathetic stimulation causes a marked decrease in mucus production, leaving this area more susceptible to irritation.

In addition to mucus, the intestinal mucosa produces two other types of secretions. The first is a serous fluid (pH 6.5 to 7.5) secreted by specialized cells (*i.e.*, crypts of Lieberkühn) in the intestinal mucosal layer. This fluid, which is produced at the rate of 2000 mL/day, acts as a vehicle for absorption. The second type of secretion consists of surface enzymes that aid absorption. These enzymes are the peptidases, or enzymes that separate amino acids, and the disaccharidases, or enzymes that split sugars.

The large intestine usually secretes only mucus. ANS activity strongly influences mucus production in the bowel, as in other parts of the digestive tract. During intense parasympathetic stimulation, mucus secretion may increase to the point that the stool contains large amounts of obvious mucus. Although the bowel normally does not secrete water or electrolytes, these substances are lost in large quantities when the bowel becomes irritated or inflamed.

In summary, the secretions of the gastrointestinal tract include saliva, gastric juices, bile, and pancreatic and intestinal secretions. Each day, more than 7000 mL of fluid is secreted into the digestive tract; all but 50 to 200 mL of this fluid is reabsorbed. Water, derived from the extracellular fluid compartment, is the major component of gastrointestinal tract secretions. Neural, humoral, and local mechanisms contribute to the control of these secretions. The parasympathetic nervous system increases secretion, and sympathetic activity exerts an inhibitory effect. In addition to secreting fluids containing digestive enzymes, the gastrointestinal tract produces and secretes hormones, such as gastrin, secretin, and cholecystokinin, that contribute to the control of gastrointestinal function.

DIGESTION AND ABSORPTION

Digestion and absorption occur mainly in the small intestine. The stomach is a poor absorptive structure, and only a few lipid-soluble substances, including alcohol, are absorbed from the stomach.

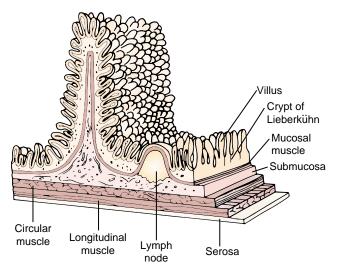
Digestion is the process of dismantling foods into their constituent parts. Digestion requires hydrolysis, enzyme cleavage, and fat emulsification. Hydrolysis is breakdown of a compound that involves a chemical reaction with water. The importance of hydrolysis to digestion is evidenced by the amount of water (7 to 8 L) that is secreted into the gastrointestinal tract daily. The intestinal mucosa is impermeable to most large molecules. Most proteins, fats, and carbohydrates must be broken down into smaller particles before they can be absorbed. Although some digestion of carbohydrates and proteins begins in the stomach, digestion takes place mainly in the small intestine. The breakdown of fats to free fatty acids and monoglycerides takes place entirely in the small intestine. The liver, with its production of bile, and the pancreas, which supplies a number of digestive enzymes, play important roles in digestion.

Absorption is the process of moving nutrients and other materials from the external environment of the gastrointestinal tract into the internal environment. Absorption is accomplished by active transport and diffusion. The absorptive function of the large intestine focuses mainly on water reabsorption. A number of substances require a specific carrier or transport system. For example, vitamin B_{12} is not absorbed in the absence of intrinsic factor, which is secreted by the parietal cells of the stomach. Transport of amino acids and glucose occurs mainly in the presence of sodium. Water is absorbed passively along an osmotic gradient.

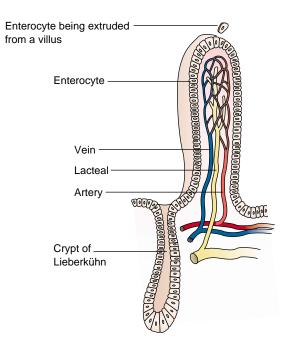
The distinguishing characteristic of the small intestine is its large surface area, which in the adult is estimated to be approximately 250 m². Anatomic features that contribute to this enlarged surface area are the circular folds that extend into the lumen of the intestine and the villi, which are finger-like projections of mucous membrane, numbering as many as 25,000, that line the entire small intestine (Fig. 26-9). Each villus is equipped with an arrangement of blood vessels for the absorption of fluid and dissolved material into the portal blood and a central lacteal for absorption into the lymph (Fig. 26-10). Fats rely largely on the lymphatics for absorption.

Each villus is covered with cells called *enterocytes* that contribute to the absorptive and digestive functions of the small bowel, and goblet cells that provide mucus. The crypts of Lieberkühn are glandular structures that open into the spaces between the villi. The enterocytes have a life span of approximately 4 to 5 days; their replacement cells differentiate from progenitor cells located in the area of the crypts. The maturing enterocytes migrate up the villus and eventually are extruded from the tip.

The enterocytes secrete enzymes that aid in the digestion of carbohydrates and proteins. These enzymes are called *brush border enzymes* because they adhere to the border of the villus structures. In this way they have access to the carbohydrates and protein molecules as they come in contact with the absorptive surface of the intestine. This mechanism of secretion places the enzymes where they are needed and eliminates the need to produce enough enzymes to mix with the entire contents that fill the lumen of the small bowel. The digested mol-



■ FIGURE 26-9 ■ The mucous membrane of the small intestine. Note the numerous villi on a circular fold.



■ FIGURE 26-10 ■ A single villus from the small intestine.

ecules diffuse through the membrane or are actively transported across the mucosal surface to enter the blood or, in the case of fatty acids, the lacteal. These molecules are then transported through the portal vein or lymphatics into the systemic circulation.

Carbohydrates

Carbohydrates must be broken down into monosaccharides, or single sugars, before they can be absorbed from the small intestine. The average daily intake of carbohydrate in the American diet is approximately 350 to 400 g. Starch makes up approximately 50% of this total, sucrose (*i.e.*, table sugar) approximately 30%, lactose (*i.e.*, milk sugar) approximately 6%, and maltose approximately 1.5%.

Digestion of starch begins in the mouth with the action of amylase. Pancreatic secretions also contain an amylase. Amylase breaks down starch into several disaccharides, including maltose, isomaltose, and α -dextrins. The brush border enzymes convert the disaccharides into monosaccharides that can be absorbed (Table 26-3). Sucrose yields glucose and fructose, lactose is converted to glucose and galactose, and maltose is converted to two glucose molecules. When the disaccharides are not broken down to monosaccharides, they cannot be absorbed but remain as osmotically active particles in the contents of the digestive system, causing diarrhea. Persons who are deficient in lactase, the enzyme that breaks down lactose, experience diarrhea when they drink milk or eat dairy products.

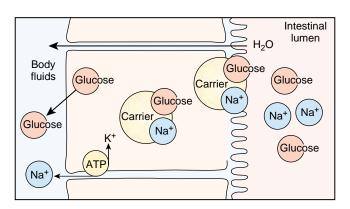
Fructose is transported across the intestinal mucosa by facilitated diffusion, which does not require energy expenditure. In this case, fructose moves along a concentration gradient. Glucose and galactose are transported by way of a sodiumdependent carrier system that uses adenosine triphosphate and the Na⁺/K⁺-ATPase pump as an energy source (Fig. 26-11).

TABLE 26-3	TABLE 26-3 Enzymes Used in Digestion of Carbohydrates			
Dietary Carbol	nydrates	Enzyme	Monosaccharides Produced	
Lactose		Lactase	Glucose and galactose	
Sucrose		Sucrase	Fructose and glucose	
Starch		Amylase	Maltose, maltotriase, and α -dextrins	
Maltose and m	naltotriose	Maltase	Glucose and glucose	
α-Dextrins		α -Dextrimase	Glucose and glucose	

Water absorption from the intestine is linked to absorption of osmotically active particles, such as glucose and sodium. It follows that an important consideration in facilitating the transport of water across the intestine (and decreasing diarrhea) after temporary disruption in bowel function is to include sodium and glucose in the fluids that are taken.

Fats

The average adult eats approximately 60 to 100 g of fat daily, principally as triglycerides containing long-chain fatty acids. These triglycerides are broken down by pancreatic lipase. Bile salts act as a carrier system for the fatty acids and fat-soluble vitamins A, D, E, and K by forming micelles, which transport these substances to the surface of intestinal villi, where they are absorbed by the lacteal. The major site of fat absorption is the upper jejunum. Medium-chain triglycerides, with 6 to 10 carbon atoms in their structures, are absorbed better than longer-chain fatty acids because they are more completely broken down by pancreatic lipase and they form micelles more easily. Because they are easily absorbed, medium-chain triglycerides often are used in the treatment of persons with malabsorption syndrome. The absorption of vitamins A, D, E, and K, which are fat-soluble vitamins, requires bile salts.



■ FIGURE 26-11 ■ The hypothetical sodium-dependent transport system for glucose. Both sodium and glucose must attach to the transport carrier before either can be transported into the cell. The concentration of glucose builds up in the intestinal cell until a diffusion gradient develops, causing glucose to move into the body fluids. Sodium is transported out of the cell by the energydependent Na⁺/K⁺-ATPase pump. This creates the gradient needed to operate the transport system. Water is passively absorbed along a concentration gradient generated by the absorption of solutes. Fat that is not absorbed in the intestine is excreted in the stool. *Steatorrhea* is the term used to describe fatty stools. It usually indicates that there is 20 g or more of fat in a 24-hour stool sample. Normally, a chemical test is done on a 72-hour stool collection, during which time the diet is restricted to 80 to 100 g of fat per day.

Proteins

Proteins from the diet must be broken into amino acids to be absorbed. Protein digestion begins in the stomach with the action of pepsin. Pepsinogen, the enzyme precursor of pepsin, is secreted by the chief cells in response to a meal and acid pH. Acid in the stomach is required for the conversion of pepsinogen to pepsin. Pepsin is inactivated when it enters the intestine by the alkaline pH.

Proteins are broken down further by pancreatic enzymes, such as trypsin, chymotrypsin, carboxypeptidase, and elastase. As with pepsin, the pancreatic enzymes are secreted as precursor molecules. Trypsinogen, which lacks enzymatic activity, is activated by an enzyme located on the brush border cells of the duodenal enterocytes. Activated trypsin activates additional trypsinogen molecules and other pancreatic precursor proteolytic enzymes. The amino acids are liberated intramurally or on the surface of the villi by brush border enzymes that degrade proteins into peptides that are one, two, or three amino acids long. Similar to glucose, many amino acids are transported across the mucosal membrane in a sodium-linked process that uses the Na⁺/K⁺- ATPase pump as an energy source.

In summary, the digestion and absorption of foodstuffs take place in the small intestine. Digestion is the process of dismantling foods into their constituent parts. Digestion requires hydrolysis, enzyme cleavage, and fat emulsification. Proteins, fats, carbohydrates, and other components of the diet are broken down into molecules that can be transported from the intestinal lumen into the body fluids. Absorption is the process of moving nutrients and other materials from the external environment of the gastrointestinal tract into the internal environment. Brush border enzymes break carbohydrates into monosaccharides that can be transported across the intestine into the bloodstream. The digestion of proteins begins in the stomach with the action of pepsin and is further facilitated in the intestine by the pancreatic enzymes, such as trypsin, chymotrypsin, carboxypeptidase, and elastase. The absorption of glucose and amino acids is facilitated by a sodium-dependent transport system. Fat in the diet is broken down by pancreatic lipase into triglycerides containing

medium- and long-chain fatty acids. Bile salts form micelles that transport these substances to the surface of intestinal villi, where they are absorbed.

ANOREXIA, NAUSEA, AND VOMITING

Anorexia, nausea, and vomiting are physiologic responses that are common to many gastrointestinal disorders. These responses are protective to the extent that they signal the presence of disease and, in the case of vomiting, remove noxious agents from the gastrointestinal tract. They also can contribute to impaired intake or loss of fluids and nutrients.

Anorexia

Anorexia represents a loss of appetite. Several factors influence appetite. One is hunger, which is stimulated by contractions of the empty stomach. Appetite or the desire for food intake is regulated by the hypothalamus and other associated centers in the brain (see Chapter 29). Smell plays an important role, as evidenced by the fact that appetite can be stimulated or suppressed by the smell of food. Loss of appetite is associated with emotional factors, such as fear, depression, frustration, and anxiety. Many drugs and disease states cause anorexia. For example, in uremia the accumulation of nitrogenous wastes in the blood contributes to the development of anorexia. Anorexia often is a forerunner of nausea, and most conditions that cause nausea and vomiting also produce anorexia.

Nausea

Nausea is an ill-defined and unpleasant subjective sensation. It is the conscious sensation resulting from stimulation of the medullary vomiting center that often precedes or accompanies vomiting. Nausea usually is preceded by anorexia, and stimuli such as foods and drugs that cause anorexia in small doses usually produce nausea when given in larger doses. A common cause of nausea is distention of the duodenum or upper small intestinal tract. Nausea frequently is accompanied by autonomic nervous system manifestations such as watery salivation and vasoconstriction with pallor, sweating, and tachycardia. Nausea may function as an early warning signal of a pathology.

Vomiting

Vomiting, or emesis, is the sudden and forceful oral expulsion of the contents of the stomach. It usually is preceded by nausea. The contents that are vomited are called vomitus. Vomiting, as a basic physiologic protective mechanism, limits the possibility of damage from ingested noxious agents by emptying the contents of the stomach and portions of the small intestine. Nausea and vomiting may represent a total-body response to drug therapy, including overdosage, cumulative effects, toxicity, and side effects.

Vomiting involves two functionally distinct medullary centers: the vomiting center and the chemoreceptor trigger zone. The act of vomiting is integrated by the vomiting center, which is located in the dorsal portion of the reticular formation of the medulla near the sensory nuclei of the vagus. The chemoreceptor trigger zone is located in a small area on the floor of the fourth ventricle, where it is exposed to both blood and cerebrospinal fluid. It is thought to mediate the emetic effects of blood-borne drugs and toxins.

The act of vomiting consists of taking a deep breath, closing the airways, and producing a strong, forceful contraction of the diaphragm and abdominal muscles along with relaxation of the gastroesophageal sphincter. Respiration ceases during the act of vomiting. Vomiting may be accompanied by dizziness, light-headedness, decrease in blood pressure, and bradycardia.

The vomiting center receives input from the gastrointestinal tract and other organs; from the cerebral cortex; from the vestibular apparatus, which is responsible for motion sickness; and from the chemoreceptor trigger zone, which is activated by many drugs and endogenous and exogenous toxins. Hypoxia exerts a direct effect on the vomiting center, producing nausea and vomiting. This direct effect probably accounts for the vomiting that occurs during periods of decreased cardiac output, shock, environmental hypoxia, and brain ischemia caused by increased intracranial pressure. Inflammation of any of the intra-abdominal organs, including the liver, gallbladder, or urinary tract, can cause vomiting because of the stimulation of the visceral afferent pathways that communicate with the vomiting center. Distention or irritation of the gastrointestinal tract also causes vomiting through the stimulation of visceral afferent neurons.

Several neurotransmitters and receptor subtypes are implicated as neuromediators in nausea and vomiting. Dopamine, serotonin, and opioid receptors are found in the gastrointestinal tract and in the vomiting and chemoreceptor trigger zone. Dopamine antagonists, such as prochlorperazine, depress vomiting caused by stimulation of the chemoreceptor trigger zone. Serotonin is believed to be involved in the nausea and emesis associated with cancer chemotherapy and radiation therapy. Serotonin antagonists (e.g., granisetron and ondansetron) are effective in treating the nausea and vomiting associated with these stimuli. Motion sickness appears to be a central nervous system (CNS) response to vestibular stimuli. Norepinephrine and acetylcholine receptors are located in the vestibular center. The acetylcholine receptors are thought to mediate the impulses responsible for exciting the vomiting center; norepinephrine receptors may have a stabilizing influence that resists motion sickness. Many of the motion sickness drugs (e.g., dimenhydrinate) have a strong CNS anticholinergic effect and act on the receptors in the vomiting center and areas related to the vestibular system.

In summary, the signs and symptoms of many gastrointestinal tract disorders are manifested by anorexia, nausea, and vomiting. Anorexia, or loss of appetite, may occur alone or may accompany nausea and vomiting. Nausea, which is an ill-defined, unpleasant sensation, signals the stimulation of the medullary vomiting center. It often precedes vomiting and frequently is accompanied by autonomic responses, such as salivation and vasoconstriction with pallor, sweating, and tachycardia. The act of vomiting, which is integrated by the vomiting center, involves the forceful oral expulsion of the gastric contents. It is a basic physiologic mechanism that rids the gastrointestinal tract of noxious agents.

REVIEW QUESTIONS

Describe the physiologic function of the digestive system in terms of transport of food, digestion, absorption of nutrients, and storage of waste products.

Approximately 7000 mL of fluid is secreted into the gastrointestinal tract, with only about 200 mL being eliminated in the stool. Explain the source, composition, function, and removal of this fluid from the gastrointestinal tract.

Describe the site of gastric acid and pepsin production and secretion in the stomach.

Describe the function of the gastric mucosal barrier.

Relate the characteristics of the small intestine to its absorptive function and explain the function of intestinal brush border enzymes in the digestion and absorption of carbohydrates, fats, and proteins.

Explain the relationship between the enteric nervous system and the autonomic nervous system in the regulation of gastrointestinal function.

Describe the innervation of the reflexive and voluntary components of defecation.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

BIBLIOGRAPHY

Berne R.M., Levy M.N. (1997). Principles of physiology (3rd ed., pp. 354– 400). St. Louis: C.V. Mosby.

- Ganong W.F. (1999). *Review of medical physiology* (19th ed., 459–491). Stanford: Appleton & Lange.
- Gershon M.D. (1999). The enteric nervous system: A second brain. Hospital Practice 34 (7), 31–52.
- Guyton A.C., Hall J.E. (2000). Textbook of medical physiology (10th ed., pp. 718–770). Philadelphia: W.B. Saunders.

Johnson L.R. (2001). Gastrointestinal physiology (6th ed.). St. Louis: C.V. Mosby.

Alterations in Gastrointestinal Function

Disorders of the Esophagus Dysphagia Esophageal Diverticulum Gastroesophageal Reflux Disease Gastroesophageal Reflux in Children **Cancer of the Esophagus Disorders of the Stomach Gastric Mucosal Barrier Gastritis** Acute Gastritis **Chronic Gastritis** Ulcer Disease Peptic Ulcer Disease Zollinger-Ellison Syndrome **Stress Ulcers Cancer** of the Stomach **Disorders of the Small and Large Intestines** Irritable Bowel Syndrome Inflammatory Bowel Disease Crohn's Disease **Ulcerative Colitis** Infectious Colitis Clostridium Difficile Colitis Escherichia Coli O157:H7 Infection **Diverticular Disease Appendicitis Alterations in Intestinal Motility** Diarrhea Constipation Intestinal Obstruction **Peritonitis**

CHAPTER

Alterations in Intestinal Absorption Malabsorption Syndrome Neoplasms Adenomatous Polyps Colorectal Cancer

Gause of death in the United States, nor do they receive the same publicity as heart disease and cancer. However, according to government reports, digestive diseases rank third in the total economic burden of illness, resulting in considerable human suffering, personal expenditures for treatment, lost working hours, and a drain on the nation's economy. It has been estimated that 60 to 70 million Americans have digestive disease.¹ Even more important is the fact that proper nutrition or a change in health practices could prevent or minimize many of these disorders.

DISORDERS OF THE ESOPHAGUS

The esophagus is a tube that connects the oropharynx with the stomach. It lies posterior to the trachea and larynx and extends through the mediastinum, intersecting the diaphragm at the level of the 11th thoracic vertebra. The esophagus functions primarily as a conduit for passage of food from the pharynx to the stomach, and the structure of its wall is designed for this purpose: the smooth muscle layers provide the peristaltic movements needed to move food along its length, and the epithelial layer secretes mucus, which protects its surface and aids in lubricating food.

Dysphagia

The act of swallowing depends on the coordinated action of the tongue and pharynx. These structures are innervated by cranial nerves V, IX, X, and XII. *Dysphagia* refers to difficulty in swallowing. If swallowing is painful, it is referred to as *odynophagia*. Dysphagia can result from altered nerve function or from disorders that produce narrowing of the esophagus. Lesions of the central nervous system (CNS), such as a stroke, often involve

the cranial nerves that control swallowing. Strictures and cancer of the esophagus and strictures resulting from scarring can reduce the size of the esophageal lumen and make swallowing difficult. Scleroderma, an autoimmune disease that causes fibrous replacement of tissues in the muscularis layer of the gastrointestinal tract, is another important cause of dysphagia.² Persons with dysphagia usually report choking, coughing, or an abnormal sensation of food sticking in the back of the throat or upper chest when they swallow.

In a condition called *achalasia*, the lower esophageal sphincter fails to relax; food that has been swallowed has difficulty passing into the stomach, and the esophagus above the lower esophageal sphincter becomes enlarged. One or several meals may lodge in the esophagus and pass slowly into the stomach. There is danger of aspiration of esophageal contents into the lungs when the person lies down.

Esophageal Diverticulum

A diverticulum of the esophagus is an outpouching of the esophageal wall caused by a weakness of the muscularis layer. An esophageal diverticulum tends to retain food. Reports that the food stops before it reaches the stomach are common, as are reports of gurgling, belching, coughing, and foul-smelling breath. The trapped food may cause esophagitis and ulceration. Because the condition usually is progressive, correction of the defect requires surgical intervention.

Gastroesophageal Reflux Disease

The term *reflux* refers to backward or return movement. In the context of gastroesophageal reflux, it refers to the backward movement of gastric contents into the esophagus, a condition that causes heartburn. Often referred to as *gastroesophageal reflux disease* (GERD), it probably is the most common disorder originating in the gastrointestinal tract. Most persons experience heartburn occasionally as a result of reflux. Such symptoms usually occur soon after eating, are short lived, and seldom cause more serious problems. However, for some persons, persistent heartburn can represent reflux disease with esophagitis.

The lower esophageal sphincter regulates the flow of food from the esophagus into the stomach. Both internal and external mechanisms function in maintaining the antireflux function of the lower esophageal sphincter.^{3,4} Relaxation of the lower esophageal sphincter is a brain stem reflex that is mediated by the vagus nerve in response to a number of afferent stimuli. Transient relaxation with reflux is common after meals. Gastric distension and meals high in fat increase the frequency of relaxation. Refluxed material normally is returned to the stomach by secondary peristaltic waves in the esophagus, with swallowed saliva neutralizing and washing away the refluxed acid.

GERD is thought to be associated with a weak or incompetent lower esophageal sphincter that allows reflux to occur, the irritant effects of the refluxate, and decreased clearance of the refluxed acid from the esophagus after it has occurred.^{5,6} Delayed gastric emptying also may contribute to reflux by increasing gastric volume and pressure with greater chance for reflux. Esophageal mucosal injury is related to the destructive nature of the refluxate and the amount of time it is in contact with mucosa. Acidic gastric fluids (pH < 4.0) are particularly damaging. There is controversy regarding the importance of hiatal hernia (*i.e.*, herniation of the stomach through an enlarged hiatus in the diaphragm) in the pathogenesis of reflux disease. Small hiatal hernias are common and considered to be of no significance in asymptomatic people. However, in cases of severe erosive esophagitis where gastroesophageal reflux and large hiatal hernia coexist, the hernia may retard esophageal acid clearance and contribute to the disorder.^{3,5}

Reflux esophagitis involves mucosal injury to the esophagus, hyperemia, and inflammation. The most common symptom of gastroesophageal reflux is heartburn. It frequently is severe, occurring 30 to 60 minutes after eating. It often is made worse by bending at the waist and recumbency and usually is relieved by sitting upright. The severity of heartburn is not indicative of the extent of mucosal injury; only a small percentage of people who report heartburn have mucosal injury. Often, the heartburn occurs during the night. Antacids give prompt, although transient relief. Other symptoms include belching and chest pain. The pain usually is located in the epigastric or retrosternal area and often radiates to the throat, shoulder, or back. Because of its location, the pain may be confused with angina. The reflux of gastric contents also may produce respiratory symptoms such as wheezing, chronic cough, and hoarseness. There is considerable evidence linking gastroesophageal reflux with bronchial asthma.7 The proposed mechanisms of reflux-associated asthma and chronic cough include aspiration, laryngeal injury, and vagal-mediated bronchospasm.

Persistent gastroesophageal reflux produces a cycle of mucosal damage that predisposes to strictures and a condition called *Barrett's esophagus*. Strictures are caused by a combination of scar tissue, spasm, and edema. They produce narrowing of the esophagus and can cause dysphagia. Barrett's esophagus is characterized by a reparative process in which the squamous mucosa that normally lines the esophagus gradually is replaced by columnar epithelium resembling that in the stomach or intestines.^{5,6} It is associated with increased risk for development of esophageal cancer.

The diagnosis of gastroesophageal reflux depends on a history of reflux symptoms and selective use of diagnostic methods, including radiographic studies using a contrast medium such as barium, esophagoscopy, and ambulatory esophageal pH monitoring.⁸

The treatment of gastroesophageal reflux usually focuses on conservative measures. These measures include avoidance of positions and conditions that increase gastric reflux.⁸ Avoidance of large meals and foods that reduce lower esophageal sphincter tone (*e.g.*, caffeine, fats, chocolate), alcohol, and smoking is recommended. Sleeping with the head elevated helps to prevent reflux during the night. Weight loss usually is recommended in overweight people.

Antacids or a combination of antacids and alginic acid also are recommended for mild disease. Alginic acid produces a foam when it comes in contact with gastric acid; if reflux occurs, the foam, rather than acid, rises into the esophagus. Histamine type 2 receptor-blocking drugs, which inhibit gastric acid production, often are recommended when additional treatment is needed. Proton pump inhibitors, which block the final pathway for acid secretion, may be used for persons who continue to have daytime symptoms, recurrent strictures, or large esophageal ulcerations. Surgical treatment may be indicated in some people.

Gastroesophageal Reflux in Children

Gastroesophageal reflux is a common problem in infants. The small reservoir capacity of an infant's esophagus coupled with frequent spontaneous reductions in sphincter pressure contributes to reflux. Regurgitation of at least one episode a day occurs in as many as half of infants 0 to 3 months of age. By 6 months of age it becomes less frequent, and it abates by 2 years of age as the child assumes a more upright posture and eats solid foods.^{9,10} Although many infants have minor degrees of reflux, complications occur in 1 of every 300 to 500 children.^{9,10} The condition occurs more frequently in children with cerebral palsy, Down's syndrome, and other causes of developmental delay.

In most cases, infants with simple reflux are thriving and healthy, and symptoms resolve between 9 and 24 months of age. Pathologic reflux is classified into three categories: (1) regurgitation and malnutrition, (2) esophagitis, and (3) respiratory problems. Symptoms of esophagitis include evidence of pain when swallowing, hematemesis, anemia caused by esophageal bleeding, heartburn, irritability, and sudden or inconsolable crying. Parents often report feeding problems in their infants.9 These infants often are irritable and demonstrate early satiety. Sometimes the problems progress to actual resistance to feeding. Tilting of the head to one side and arching of the back may be noted in children with severe reflux. The head positioning is thought to represent an attempt to protect the airway or reduce the pain-associated reflux. Sometimes regurgitation is associated with dental caries and recurrent otalgia. The ear pain is thought to occur through referral from the vagus nerve in the esophagus to the ear. A variety of respiratory symptoms are caused by damage to the respiratory mucosa when gastric reflux enters the esophagus. Reflux may cause laryngospasm, apnea, and bradycardia. A relationship between reflux and acute life-threatening events or sudden infant death syndrome has been proposed. However, the association remains a matter of controversy, and the linkage may be coincidental.9,10

Rumination is the repetitive gagging, regurgitation, mouthing, and reswallowing of regurgitated material. Although the cause of the disorder is unknown, it often is associated with mental retardation or altered interaction with the environment (*e.g.*, lack of stimulation in newborn intensive care units or because of altered relationships with caregivers). As with pure reflux, rumination may produce severe esophagitis with signs of iron deficiency anemia, failure to thrive, and head tilting.¹⁰

Diagnosis of gastroesophageal reflux in infants and children often is based on parental and clinical observations. The diagnosis may be confirmed by esophageal pH probe studies or barium fluoroscopic esophagography. In severe cases, esophagoscopy may be used to demonstrate reflux and obtain a biopsy.

Various treatment methods are available for infants and children with gastroesophageal reflux. Small, frequent feedings are recommended because of the association between gastric volume and transient relaxation of the esophagus. Thickening an infant's feedings with cereal tends to decrease the volume of reflux, decrease crying and energy expenditure, and increase the calorie density of the formula.^{9,10} In infants, positioning on the left side seems to decrease reflux. In older infants and children, raising the head of the bed and keeping the child upright may help. Medications usually are not added to the treatment regimen until pathologic reflux has been documented by diagnostic testing.

Chapter 27: Alterations in Gastrointestinal Function

Cancer of the Esophagus

Carcinoma of the esophagus accounts for approximately 6% of all gastrointestinal cancers. This disease is more common in persons older than 50 years, with a male-to-female ratio of approximately 2:1.¹¹

There are two types of esophageal cancers: squamous cell and adenocarcinomas. Fewer than 50% of esophageal tumors are squamous cell cancers. These cancers are associated more commonly with dietary and environmental influences.^{11,12} Most squamous cell cancers in the United States and Europe are attributable to alcohol and tobacco use. Adenocarcinomas, which typically arise from Barrett's esophagus, account for more than 50% of esophageal cancers. The incidence of this type of cancer appears to be increasing. Adenocarcinomas typically are located in the distal esophagus and may invade the adjacent upper part of the stomach. Endoscopic surveillance in people with Barrett's esophagus may detect adenocarcinoma at an earlier stage, when it is more amenable to curative surgical resection.¹²

Dysphagia is by far the most frequent complaint of persons with esophageal cancer. It is apparent first with ingestion of bulky food, later with soft food, and finally with liquids. Unfortunately, it is a late manifestation of the disease. Weight loss, anorexia, fatigue, and pain on swallowing also may occur.

Treatment includes surgical resection, which provides a means of cure when done in early disease and palliation when done in late disease. Irradiation is used as a palliative treatment. Chemotherapy sometimes is used before surgery to decrease the size of the tumor, and it may be used along with irradiation and surgery in an effort to increase survival.¹²

The prognosis for persons with cancer of the esophagus, although poor, has improved. However, even with modern forms of therapy, the long-term survival is limited because, in many cases, the disease has already metastasized by the time the diagnosis is made.

In summary, the esophagus is a tube that connects the oropharynx with the stomach; it functions primarily as a conduit for passage of food from the pharynx to the stomach. Dysphagia refers to difficulty in swallowing; it can result from altered nerve function or from disorders that produce narrowing of the esophagus. A diverticulum of the esophagus is an outpouching of the esophageal wall caused by a weakness of the muscularis layer.

Gastrointestinal reflux refers to the backward movement of gastric contents into the esophagus, a condition that causes heartburn. Although most persons experience occasional esophageal reflux and heartburn, persistent reflux can cause esophagitis. Gastroesophageal reflux is a common problem in infants. Reflux commonly corrects itself with age, and symptoms abate in most children by 2 years of age. Although many infants have minor degrees of reflux, some infants and small children have significant reflux that interferes with feeding, causes esophagitis, and results in respiratory symptoms and other complications.

Carcinoma of the esophagus, which accounts for 6% of all cancers, is more common in persons older than 50 years, and the male-to-female ratio is approximately 2:1. There are two types of esophageal cancer: squamous cell and adenocarcinoma. Most squamous cell cancers are attributable to alcohol and tobacco use, whereas adenocarcinomas are more closely linked to esophageal reflux and Barrett's esophagus.

DISORDERS OF THE STOMACH

The stomach is a reservoir for contents entering the digestive tract. While in the stomach, food is churned and mixed with hydrochloric acid and pepsin before being released into the small intestine. Normally, the mucosal surface of the stomach provides a barrier that protects it from the hydrochloric acid and pepsin contained in gastric secretions. Disorders of the stomach include gastritis, peptic ulcer, and gastric carcinoma.

Gastric Mucosal Barrier

The stomach lining usually is impermeable to the acid it secretes, a property that allows the stomach to contain acid and pepsin without having its wall digested. Several factors contribute to the protection of the gastric mucosa, including an impermeable epithelial cell surface covering, mechanisms for the selective transport of hydrogen and bicarbonate ions, and the characteristics of gastric mucus.¹³ These mechanisms are collectively referred to as the *gastric mucosal barrier*.

KEY CONCEPTS

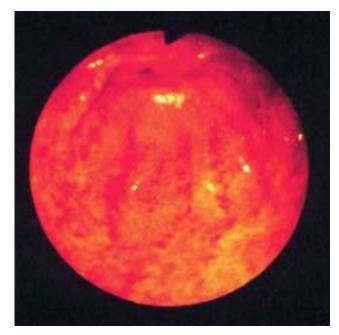
DISRUPTION OF THE GASTRIC MUCOSA AND ULCER DEVELOPMENT

- The stomach is protected by tight cellular junctions, a protective mucus layer, and prostaglandins that serve as chemical messengers to protect the stomach lining by improving blood flow, increasing bicarbonate secretion, and enhancing mucus production.
- Two of the major causes of gastric irritation and ulcer formation are aspirin or nonsteroidal antiinflammatory drugs (NSAIDs) and infection with *Helicobacter pylori*.
- Aspirin and NSAIDs exert their destructive effects by damaging epithelial cells, impairing mucus production, and inhibiting prostaglandin synthesis.
- H. pylori is an infectious agent that thrives in the acid environment of the stomach and disrupts the mucosal barrier that protects the stomach from the harmful effects of its digestive enzymes.

The gastric epithelial cells are connected by tight junctions that prevent acid penetration, and they are covered with an impermeable hydrophobic lipid layer that prevents diffusion of ionized water-soluble molecules. Aspirin, which is nonionized and lipid soluble in acid solutions, rapidly diffuses across this lipid layer, increasing mucosal permeability and damaging epithelial cells.¹⁴ Gastric irritation and occult bleeding caused by gastric irritation occur in a significant number of persons who take aspirin on a regular basis (Fig. 27-1). Alcohol, which also is lipid soluble, disrupts the mucosal barrier; when aspirin and alcohol are taken in combination, as they often are, there is increased risk of gastric irritation. Bile acids also attack the lipid components of the mucosal barrier and afford the potential for gastric irritation when there is reflux of duodenal contents into the stomach.

Normally, the secretion of hydrochloric acid by the parietal cells of the stomach is accompanied by secretion of bicarbonate ions (HCO_3^-). For every hydrogen ion (H^+) that is secreted, a HCO_3^- is produced, and as long as HCO_3^- production is equal to H^+ secretion, mucosal injury does not occur. Changes in gastric blood flow, as in shock, tend to decrease HCO_3^- production. This is particularly true in situations in which decreased blood flow is accompanied by acidosis. Aspirin and the nonsteroidal anti-inflammatory drugs (NSAIDs) also impair HCO_3^- secretion.

Prostaglandins, chemical messengers derived from cell membrane lipids, play an important role in protecting the gastrointestinal mucosa from injury. The prostaglandins probably exert their effect through improved blood flow, increased bicarbonate ion secretion, and enhanced mucus production. The fact that drugs such as aspirin and the NSAIDs inhibit prostaglandin synthesis may contribute to their ability to produce gastric irritation.¹⁴ Smoking and older age have been asso-



■ FIGURE 27-1 ■ Erosive gastritis. This endoscopic view of the stomach in a patient who was ingesting aspirin reveals acute hemorrhagic lesions. (From Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 683]. Philadelphia: Lippincott Williams & Wilkins)

Chapter 27: Alterations in Gastrointestinal Function

ciated with reduced gastric and duodenal prostaglandin concentrations; these observations may explain the predisposition to ulcer disease in smokers and older persons.¹⁵

Gastritis

Gastritis refers to inflammation of the gastric mucosa. There are many causes of gastritis, most of which can be grouped under the headings of acute or chronic gastritis.

Acute Gastritis

Acute gastritis refers to a transient inflammation of the gastric mucosa. It is most commonly associated with local irritants such as bacterial endotoxins, caffeine, alcohol, and aspirin. Depending on the severity of the disorder, the mucosal response may vary from moderate edema and hyperemia to hemorrhagic erosion of the gastric mucosa.

The complaints of persons with acute gastritis vary. Persons with aspirin-related gastritis can be totally unaware of the condition or may report only heartburn or sour stomach. Gastritis associated with excessive alcohol consumption is a different situation; it often causes transient gastric distress, which may lead to vomiting and, in more severe situations, to bleeding and hematemesis. Gastritis caused by the toxins of infectious organisms, such as the staphylococcal enterotoxins, usually has an abrupt and violent onset, with gastric distress and vomiting ensuing approximately 5 hours after the ingestion of a contaminated food source. Acute gastritis usually is a self-limiting disorder; complete regeneration and healing usually occur within several days.

Chronic Gastritis

Chronic gastritis is a separate entity from acute gastritis. It is characterized by the absence of grossly visible erosions and the presence of chronic inflammatory changes leading eventually to atrophy of the glandular epithelium of the stomach. The changes may become dysplastic and possibly transform into carcinoma. Factors such as chronic alcohol abuse, cigarette smoking, and chronic use of NSAIDs may contribute to the development of the disease.

There are four major types of chronic gastritis: (1) autoimmune gastritis, (2) multifocal atrophic gastritis, (3) *Helicobacter pylori* gastritis, and (4) chemical gastropathy.¹⁶ Autoimmune gastritis is the least common form of chronic gastritis. Most persons with the disorder have circulating antibodies to parietal cells and intrinsic factor, so this form of chronic gastritis is considered to be of autoimmune origin. Autoimmune destruction of the parietal cells leads to hypochlorhydria or achlorhydria, a high intragastric pH, and hypergastrinemia. Pernicious anemia is a megaloblastic anemia that is caused by malabsorption of vitamin B₁₂ caused by a deficiency of intrinsic factor (see Chapter 13). This type of chronic gastritis frequently is associated with other autoimmune disorders, such as Hashimoto's thyroiditis and Addison's disease.

Multifocal atrophic gastritis is a disorder of unknown etiology. It is more common than autoimmune gastritis and is seen more frequently in whites than in other races. It is particularly common in Asia, Scandinavia, and parts of Europe and Latin America. As with autoimmune gastritis, it is associated with reduced gastric acid secretion, but achlorhydria and pernicious anemia are less common.

Chronic autoimmune gastritis and multifocal atrophic gastritis cause few symptoms related directly to gastric changes. Persons with autoimmune chronic gastritis may have signs of pernicious anemia. More important is the development of peptic ulcer and increased risk of peptic ulcer and gastric carcinoma. Approximately 2% to 4% of persons with atrophic gastritis eventually experience gastric carcinoma.¹¹

H. pylori gastritis is the most common type of chronic nonerosive gastritis in the United States. H. pylori are small, curved, gram-negative rods that can colonize the mucus-secreting epithelial cells of the stomach^{16,17} (Fig. 27-2). H. pylori have multiple flagella, which allow them to move through the mucous layer of the stomach, and they secrete urease, which enables them to produce sufficient ammonia to buffer the acidity of their immediate environment. Because the organism adheres only to the mucus-secreting cells of the stomach, it does not usually colonize other parts of the gastrointestinal tract. H. pylori produce an enzyme that degrades mucin and has the capacity to interfere with the local protection of the gastric mucosa against acid. It also may produce toxins that directly damage the mucosa and produce ulceration in other ways. Chronic infection with *H. pylori* can lead to gastric atrophy and intestinal metaplasia. H. pylori also can cause peptic ulcer (to be discussed) and has been linked to the development of gastric adenocarcinoma.

Chemical gastropathy is a chronic gastric injury resulting from reflux of alkaline duodenal contents, pancreatic secretions, and bile into the stomach. It is most commonly seen in persons who have had gastroduodenostomy or gastrojejunostomy surgery. A milder form may occur in persons with gastric ulcer,



■ FIGURE 27-2 ■ Infective gastritis. *H. pylori* appears on silver staining as small, curved rods on the surface of the gastric mucosa. (From Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 687]. Philadelphia: Lippincott Williams & Wilkins)

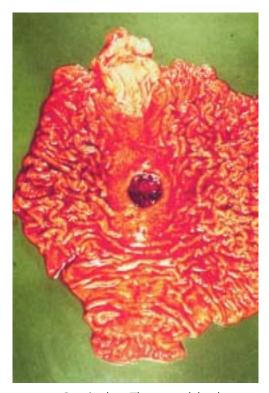
gallbladder disease, or various motility disorders of the distal stomach.

Ulcer Disease

Peptic Ulcer

Peptic ulcer is a term used to describe a group of ulcerative disorders that occur in areas of the upper gastrointestinal tract (*e.g.*, stomach and duodenum) that are exposed to acid-pepsin secretions. Peptic ulcer disease, with its remissions and exacerbations, represents a chronic health problem. Approximately 10% of the population has or will experience peptic ulcer.⁶ Duodenal ulcers occur five times more commonly than do gastric ulcers. Ulcers in the duodenum occur at any age and frequently are seen in early adulthood. Gastric ulcers tend to affect the older age group, with a peak incidence between 55 and 70 years of age. Both types of ulcers affect men three to four times more frequently than women.

A peptic ulcer can affect one or all layers of the stomach or duodenum (Fig. 27-3). The ulcer may penetrate only the mucosal surface, or it may extend into the smooth muscle layers. Occasionally, an ulcer penetrates the outer wall of the stomach or duodenum. Spontaneous remissions and exacerbations are common. Healing of the muscularis layer involves replacement with scar tissue; although the mucosal layers that cover the scarred muscle layer regenerate, the regeneration often is less than perfect, which contributes to repeated episodes of ulceration.



■ FIGURE 27-3 ■ Gastric ulcer. The stomach has been opened to reveal a sharply demarcated, deep peptic ulcer on the lesser curvature. (From Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 693]. Philadelphia: Lippincott Williams & Wilkins)

Since the early 1980s, there has been a radical shift in thinking regarding the cause of peptic ulcer. No longer is peptic ulcer thought to result from a genetic predisposition, stress, or dietary indiscretions. Most cases of peptic ulcer are caused by H. pylori infection.¹⁸ The second most common cause of peptic ulcer is NSAID and aspirin use.18 It has been reported that virtually all persons with duodenal ulcer and 70% of persons with gastric ulcer have H. pylori infection.11 Aspirin and NSAIDs account for 10% to 20% of gastric ulcers and 2% to 5% of duodenal ulcers. Aspirin appears to be the most ulcerogenic of the NSAIDs. Ulcer development in NSAID users is dose dependent, but some risk occurs even with aspirin doses of 325 mg/day.⁶ The pathogenesis of NSAID-induced ulcers is thought to involve mucosal injury and inhibition of prostaglandin synthesis. In contrast to peptic ulcer from other causes, NSAIDinduced gastric injury often is without symptoms, and lifethreatening complications can occur without warning.

The clinical manifestations of uncomplicated peptic ulcer focus on discomfort and pain. The pain, which is described as burning, gnawing, or cramplike, usually is rhythmic and frequently occurs when the stomach is empty—between meals and at 1 or 2 o'clock in the morning. The pain usually is located over a small area near the midline in the epigastrium near the xiphoid and may radiate below the costal margins, into the back, or rarely, to the right shoulder. The pain is usually relieved by food or antacids.

The complications of peptic ulcer include hemorrhage, obstruction, and perforation. Hemorrhage is caused by bleeding from granulation tissue or from erosion of an ulcer into an artery or vein. It occurs in as many as 10% to 20% of persons with peptic ulcer.⁶ Bleeding may be sudden, severe, and without warning, or it may be insidious, producing only occult blood in the stool. As many as 20% of persons with bleeding ulcers have no antecedent symptoms of pain; this is particularly true in persons receiving NSAIDs. Acute hemorrhage is evidenced by the sudden onset of weakness, dizziness, thirst, cold, moist skin, the desire to defecate, and the passage of loose, tarry, or even red stools and emesis of coffee-ground vomitus. Signs of circulatory shock develop, depending on the amount of blood lost.

Obstruction is caused by edema, spasm, or contraction of scar tissue and interference with the free passage of gastric contents through the pylorus or adjacent areas. There is a feeling of epigastric fullness and heaviness after meals. With severe obstruction, there is vomiting of undigested food.

Perforation occurs when an ulcer erodes through all the layers of the stomach or duodenum wall. Perforation develops in approximately 5% of persons with peptic ulcers, usually from ulcers on the anterior wall of the stomach or duodenum.⁶ With perforation, gastrointestinal contents enter the peritoneum and cause peritonitis, or penetrate adjacent structures such as the pancreas. Radiation of the pain into the back, severe night distress, and inadequate pain relief with the eating of foods or taking of antacids in persons with a long history of peptic ulcer may signify perforation.

Diagnosis and Treatment. Diagnostic procedures for peptic ulcer include history taking with an emphasis on aspirin and NSAID use, laboratory tests, radiologic imaging, and endoscopic examination. Laboratory findings of hypochromic anemia and occult blood in the stools indicate bleeding. X-ray

studies with a contrast medium such as barium are used to detect the presence of an ulcer crater and to exclude gastric carcinoma. Endoscopy (*i.e.*, gastroscopy and duodenoscopy) can be used to visualize the ulcer area and obtain biopsy specimens to test for *H. pylori* and exclude malignant disease.

The treatment of peptic ulcer has changed dramatically during the past several years and now aims to eradicate the cause and effect a permanent cure for the disease. Pharmacologic treatment focuses on eradicating *H. pylori* with appropriate antibiotics when present, relieving ulcer symptoms, and healing the ulcer crater. Acid-neutralizing, acid-inhibiting drugs (histamine₂-receptor antagonists and proton pump blockers), and mucosal protective agents are used to relieve symptoms and promote healing of the ulcer crater. Aspirin and NSAID use should be avoided when possible.

The current surgical management of peptic ulcer disease is largely limited to treatment of complications. When surgery is needed, it usually is performed using minimally invasive methods. With bleeding ulcers, hemostasis often can be achieved by endoscopic methods, and endoscopic balloon dilation often is effective in relieving outflow obstruction.

Zollinger-Ellison Syndrome

The Zollinger-Ellison syndrome is a rare condition caused by a gastrin-secreting tumor (gastrinoma). In persons with this disorder, gastric acid secretion reaches such levels that ulceration becomes inevitable.¹⁹ The tumors may be single or multiple; although most tumors are located in the pancreas, a few develop in the submucosa of the stomach or duodenum. More than two thirds of gastrinomas are malignant, and one third have already metastasized at the time of diagnosis.¹¹ The increased gastric secretions cause symptoms related to peptic ulcer. Diarrhea may result from hypersecretion or from the inactivation of intestinal lipase and impaired fat digestion that occurs with a decrease in intestinal pH.

Stress Ulcers

A stress ulcer, sometimes called Curling's ulcer, refers to gastrointestinal ulcerations that develop in relation to major physiologic stress. Persons at high risk for the development of stress ulcers include those with large surface-area burns, trauma, sepsis, acute respiratory distress syndrome, severe liver failure, and major surgical procedures. These lesions occur most often in the fundus of the stomach and proximal duodenum and are thought to result from ischemia, tissue acidosis, and bile salts entering the stomach in critically ill persons with decreased gastrointestinal tract motility.²⁰ Another form of stress ulcer, called Cushing ulcer, consists of gastric, duodenal, and esophageal ulcers arising in persons with intracranial injury, operations, or tumors. They are thought to be caused by hypersecretion of gastric acid resulting from stimulation of vagal nuclei by increased intracranial pressure. These ulcers are associated with a high incidence of perforation.¹¹

Stress ulcers develop in approximately 5% to 10% of persons admitted to hospital intensive care units.¹¹ They usually are manifested by painless upper gastrointestinal tract bleeding. Monitoring and maintaining the gastric pH at 3.5 or higher helps to prevent the development of stress ulcers. H₂-receptor antagonists, proton pump inhibitors, and mucosal protective agents are used in the prevention and treatment of stress ulcers.

Cancer of the Stomach

Stomach cancer is the seventh most frequent cause of cancer mortality in the United States. In 2003, it is estimated that approximately 22,400 Americans will receive a diagnosis of stomach cancer and 12,100 will die of the disease.²¹ The disease is much more common in other countries and regions, principally Japan, central Europe, the Scandinavian countries, South and Central America, republics of the former Soviet Union, China, and Korea. It is the major cause of cancer death worldwide.

Among the factors that increase the risk of gastric cancer are a genetic predisposition, carcinogenic factors in the diet (*e.g.*, *N*-nitroso compounds and benzopyrene found in smoked and preserved foods), autoimmune gastritis, and gastric adenomas or polyps. The incidence of stomach cancer in the United States has decreased fourfold since 1930, presumably because of improved storage of food with decreased consumption of salted, smoked, and preserved foods.^{11,21} Infection with *H. pylori* appears to serve as a cofactor in some types of gastric carcinomas.¹¹

Between 50% and 60% of gastric cancers occur in the pyloric region or adjacent to the antrum. Compared with a benign ulcer, which has smooth margins and is concentrically shaped, gastric cancers tend to be larger, are irregularly shaped, and have irregular margins.

Unfortunately, stomach cancers often are asymptomatic until late in their course. Symptoms, when they do occur, usually are vague and include indigestion, anorexia, weight loss, vague epigastric pain, vomiting, and an abdominal mass.

Diagnosis of gastric cancer is accomplished by means of a variety of techniques, including barium x-ray studies, endoscopic studies with biopsy, and cytologic studies (*e.g.*, Papanicolaou smear) of gastric secretions. Cytologic studies can prove particularly useful as routine screening tests for persons with atrophic gastritis or gastric polyps. Computed tomography (CT) and endoscopic ultrasonography often are used to delineate the spread of a diagnosed stomach cancer.

Surgery in the form of radical subtotal gastrectomy usually is the treatment of choice.²¹ Irradiation and chemotherapy have not proved particularly useful as primary treatment modalities in stomach cancer. These methods usually are used for palliative purposes or to control metastatic spread of the disease.

In summary, disorders of the stomach include gastritis, peptic ulcer, and cancer of the stomach. Gastritis refers to inflammation of the gastric mucosa. Acute gastritis refers to a transient inflammation of the gastric mucosa; it is associated most commonly with local irritants such as bacterial endotoxins, caffeine, alcohol, and aspirin. Chronic gastritis is characterized by the absence of grossly visible erosions and the presence of chronic inflammatory changes leading eventually to atrophy of the glandular epithelium of the stomach. Chronic gastritis increases the risk of stomach cancer.

Peptic ulcer is a term used to describe a group of ulcerative disorders that occur in areas of the upper gastrointestinal tract that are exposed to acid-pepsin secretions, most commonly the duodenum and stomach. There are two main causes of peptic ulcer: *H. pylori* infection and aspirin or NSAID use. The treatment of peptic ulcer focuses on eradication of *H. pylori*, avoidance of gastric irritation from NSAIDs, and conventional pharmacologic treatment directed at symptom relief and ulcer healing.

The Zollinger-Ellison syndrome is a rare condition caused by a gastrin-secreting tumor, in which gastric acid secretion reaches such levels that ulceration becomes inevitable. Stress ulcers occur in relation to major physiologic stresses such as burns and trauma and are thought to result from ischemia, tissue acidosis, and bile salts entering the stomach in critically ill persons with decreased gastrointestinal tract motility. Another form of stress ulcer, resulting from hypersecretion of gastric acid caused by stimulation of vagal nuclei by increased intracranial pressure, may develop in persons with intracranial injury, operations, or tumors.

Although the incidence of cancer of the stomach has declined during the past 50 years, it remains the seventh leading cause of death in the United States. Because there are few early symptoms with this form of cancer, the disease often is far advanced at the time of diagnosis.

DISORDERS OF THE SMALL AND LARGE INTESTINES

There are many similarities in conditions that disrupt the integrity and function of the small and large intestine. The walls of the small and large intestines consist of five layers (see Chapter 26, Fig. 26-3): an outer serosal layer; a muscularis layer, which is divided into a layer of circular and a layer of longitudinal muscle fibers; a submucosal layer; and an inner mucosal layer, which lines the lumen of the intestine. Among the conditions that cause altered intestinal function are irritable bowel disease, inflammatory bowel disease, diverticulitis, appendicitis, alterations in bowel motility (*i.e.*, diarrhea, constipation, and bowel obstruction), malabsorption syndrome, and cancer of the colon and rectum.

Irritable Bowel Syndrome

The term *irritable bowel syndrome* is used to describe a functional gastrointestinal disorder characterized by a variable combination of chronic and recurrent intestinal symptoms not explained by structural or biochemical abnormalities. There is evidence to suggest that 10% to 20% of people in Western countries have the disorder, although most do not seek medical attention.^{22,23}

The condition is characterized by persistent or recurrent symptoms of abdominal pain, altered bowel function, and varying complaints of flatulence, bloatedness, nausea and anorexia, and anxiety or depression. A hallmark of irritable bowel syndrome is abdominal pain that is relieved by defecation and associated with a change in consistency or frequency of stools.

Irritable bowel syndrome is believed to result from dysregulation of intestinal motor and sensory functions modulated by the CNS.²³ Persons with irritable bowel syndrome tend to experience increased motility and abnormal intestinal contractions in response to psychological and physiologic stress. The role that psychological factors play in the disease is uncertain. Although changes in intestinal activity are normal responses to stress, these responses appear to be exaggerated in persons with irritable bowel syndrome. Women tend to be affected more often than men. Menarche often is associated with onset of the disorder. Women frequently notice an exacerbation of symptoms during the premenstrual period, suggesting a hormonal component.

Diagnosis of irritable bowel syndrome is based on continuous or recurrent symptoms of at least 3 months' duration consisting of abdominal pain or discomfort relieved by defecation, a change in the frequency or consistency of stool, and the presence of three or more varying patterns of altered defecation that are present at least 25% of the time.²⁴ These patterns of defecation include altered stool frequency, altered stool form (*i.e.*, hard or loose, watery stool), altered stool passage (*i.e.*, straining, urgency, or feeling of incomplete evacuation), passage of mucus, and bloating or feeling of abdominal discomfort. A history of lactose intolerance should be considered because intolerance to lactose and other sugars may be a precipitating factor in some persons.

The treatment of irritable bowel syndrome focuses on methods of stress management, particularly those related to symptom production. Reassurance is important. Usually, no special diet is indicated, although adequate fiber intake usually is recommended. Avoidance of offending dietary substances such as fatty and gas-producing foods, alcohol, and caffeinecontaining beverages may be beneficial. Various pharmacologic agents, including antispasmodic and anticholinergic drugs, have been used with varying success in treatment of the disorder.

Inflammatory Bowel Disease

The term *inflammatory bowel disease* is used to designate two related inflammatory intestinal disorders: Crohn's disease and ulcerative colitis.^{25,26} The prevalence of these diseases ranges from 300,000 to 500,000. Although the two diseases differ sufficiently to be distinguishable, they have many features in common. Both diseases produce inflammation of the bowel, both lack confirming evidence of a proven causative agent, both have a pattern of familial occurrence, and both can be accompanied by systemic manifestations.¹⁶ The distinguishing characteristics of Crohn's disease and ulcerative colitis are summarized in Table 27-1.

The etiology of Crohn's disease and ulcerative colitis is largely unknown, but several theories have been proposed regarding their causation. The diseases appear to have a familial occurrence, suggesting a hereditary predisposition.¹¹ It is thought that genetic factors predispose to an autoimmune reaction, possibly triggered by a relatively innocuous environmental agent such as a dietary antigen or microbial agent. It also is thought that the diseases may have an infectious origin, such as Chlamydia, atypical bacteria, and mycobacteria.¹¹ Another theory is one of defective immunoregulation, in which the mucosal branch of the immune system is stimulated and then fails to down-regulate. Although psychogenic factors may contribute to the severity and onset of both conditions, it seems unlikely that they are the primary cause.

The clinical manifestations of both Crohn's disease and ulcerative colitis are ultimately the result of activation of inflammatory cells with elaboration of inflammatory mediators that

and Ulcerative Colitis			
Characteristic	Crohn's Disease	Ulcerative Colitis	
Types of inflammation	Granulomatous	Ulcerative and exudative	
Level of involvement	Primarily submucosal	Primarily mucosal	
Extent of involvement	Skip lesions	Continuous	
Areas of involvement	Primarily ileum, secondarily colon	Primarily rectum and left colon	
Diarrhea	Common	Common	
Rectal bleeding	Rare	Common	
Fistulas	Common	Rare	
Strictures	Common	Rare	
Perianal abscesses	Common	Rare	
Development of cancer	Uncommon	Relatively common	

TABLE 27-1	Differentiating	g Characteristics	of Crohn's	Disease
and Ulcerativ		0		

cause nonspecific tissue damage. Both diseases are characterized by remissions and exacerbations of diarrhea, fecal urgency, and weight loss. Acute complications such as intestinal obstruction may develop during periods of fulminant disease.

A number of systemic manifestations have been identified in persons with Crohn's disease and ulcerative colitis. These include axial arthritis affecting the spine and sacroiliac joints and oligoarticular arthritis affecting the large joints of the arms and legs; inflammatory conditions of the eye, usually uveitis; skin lesions, especially erythema nodosum; stomatitis; and autoimmune anemia, hypercoagulability of blood, and sclerosing cholangitis. Occasionally, these systemic manifestations may herald the recurrence of intestinal disease. In children, growth retardation may occur, particularly if the symptoms are prolonged and nutrient intake has been poor.

Crohn's Disease

Crohn's disease is a recurrent, granulomatous type of inflammatory response that can affect any area of the gastrointestinal tract from the mouth to the anus. In nearly 40% of persons with disease, the lesions are restricted to the small intestine; in 30%, only the large bowel is affected; and in the remaining 30%, the large bowel and small bowel are affected.¹¹ It is a slowly progressive, relentless, and often disabling disease. Despite the substantial increase in the prevalence of Crohn's disease in the early 1980s, the distribution of affected sites has not changed substantially. The disease usually strikes adolescents and young adults and is most common among persons of European origin, with considerably higher frequency among Ashkenazi Jews.¹⁶

A characteristic feature of Crohn's disease is the sharply demarcated, granulomatous lesions that are surrounded by normal-appearing mucosal tissue. When the lesions are multiple, they often are referred to as skip lesions because they are interspersed between what appear to be normal segments of the bowel. All the layers of the bowel are involved, with the submucosal layer affected to the greatest extent. The surface of the inflamed bowel usually has a characteristic "cobblestone" appearance resulting from the fissures and crevices that develop and that are surrounded by areas of submucosal edema (Fig. 27-4). There usually is a relative sparing of the smooth muscle layers of the bowel, with marked inflammatory and fibrotic changes of the submucosal layer. The bowel wall, after

a time, often becomes thickened and inflexible; its appearance has been likened to a lead pipe or rubber hose. The adjacent mesentery may become inflamed, and the regional lymph nodes and channels may become enlarged.

The clinical course of Crohn's disease is variable; often, there are periods of exacerbations and remissions, with symptoms being related to the location of the lesions. The principal symptoms include intermittent diarrhea, colicky pain (usually in the lower right quadrant), weight loss, fluid and electrolyte disorders, malaise, and low-grade fever. Because Crohn's disease affects the submucosal layer to a greater extent than the mucosal layer, there is less bloody diarrhea than with ulcerative colitis. Ulceration of the perianal skin is common, largely because of the severity of the diarrhea. The absorptive surface of the intestine may be disrupted; nutritional deficiencies may occur, related to the specific segment of the intestine that is involved. When Crohn's disease occurs in childhood, one of its major manifestations may be retardation of growth and physical development.16

Complications of Crohn's disease include fistula formation, perforation, abdominal abscess formation, and intestinal obstruction (Fig. 27-4). Fistulas are tubelike passages that form connections between different sites in the gastrointestinal tract.



■ FIGURE 27-4 ■ Crohn's disease. The mucosal surface of the colon displays a "cobblestone" appearance owing to the presence of linear ulcerations and edema and inflammation of the intervening tissue. (From Rubin E., Farber J.L. [1999]. Pathology [3rd ed., p. 728]. Philadelphia: Lippincott Williams & Wilkins)

They also may develop between other sites, including the bladder, vagina, urethra, and skin. Fistulas between segments of the gastrointestinal tract may lead to malabsorption, syndromes of bacterial overgrowth, and diarrhea. They also can become infected and cause abscess formation.

Diagnosis and Treatment. The diagnosis of Crohn's disease requires a thorough history and physical examination. Sigmoidoscopy is used for direct visualization of the affected areas and to obtain biopsies. Measures are taken to exclude infectious agents as the cause of the disorder. This usually is accomplished by the use of stool cultures and examination of fresh stool specimens for ova and parasites. In persons suspected of having Crohn's disease, radiologic contrast studies provide a means for determining the extent of involvement of the small bowel and establishing the presence and nature of fistulas. CT scans may be used to detect an inflammatory mass or abscess.

Treatment methods focus on terminating the inflammatory response and promoting healing, maintaining adequate nutrition, and preventing and treating complications. Nutritional deficiencies are common in Crohn's disease because of diarrhea, steatorrhea (fatty stools), and other malabsorption problems. A nutritious diet that is high in calories, vitamins, and proteins is recommended. Elemental diets, which are nutritionally balanced but residue free and bulk free, may be given during the acute phase of the illness. These diets are largely absorbed in the jejunum and allow the inflamed bowel to rest. Total parenteral nutrition (*i.e.*, parenteral hyperalimentation), which is administered intravenously, may be needed when food cannot be absorbed from the intestine.

Several medications have been successful in suppressing the inflammatory reaction, including the corticosteroids, sulfasalazine, and metronidazole. Sulfasalazine is a topically active agent that has a variety of anti-inflammatory effects. Metronidazole is an antibiotic used to treat bacterial overgrowth in the small intestine. Immunosuppressive drugs such as azathioprine and its active derivative, 6-mercaptopurine, also may be used. Infliximab, a monoclonal antibody that targets tumor necrosis factor- α (TNF- α), a mediator of the inflammatory response, may be used for treatment of moderate to severe Crohn's disease that does not respond to standard therapies.²⁷ Surgical resection of damaged bowel, drainage of abscesses, or repair of fistula tracts may be necessary.

Ulcerative Colitis

Ulcerative colitis is a nonspecific inflammatory condition of the colon. The disease begins most often between 20 and 25 years of age, but the condition may affect both younger and older persons.¹¹ Unlike Crohn's disease, which can affect various sites in the gastrointestinal tract, ulcerative colitis is confined to the rectum and colon. The disease usually begins in the rectum and spreads proximally, affecting primarily the mucosal layer, although it can extend into the submucosal layer. The length of proximal extension varies. It may involve the rectum alone, the rectum and sigmoid colon, or the entire colon. The inflammatory process tends to be confluent and continuous, instead of skipping areas, as it does in Crohn's disease.

Characteristic of the disease are the lesions that form in the crypts of Lieberkühn in the base of the mucosal layer (see

Chapter 26, Fig. 26-10). The inflammatory process leads to the formation of pinpoint mucosal hemorrhages, which in time suppurate and develop into crypt abscesses. These inflammatory lesions may become necrotic and ulcerate. Although the ulcerations usually are superficial, they often extend, causing large denuded areas (Fig. 27-5). As a result of the inflammatory process, the mucosal layer often develops tonguelike projections that resemble polyps and thus are called *pseudopolyps*. The bowel wall thickens in response to repeated episodes of colitis.

Diarrhea, which is the characteristic manifestation of ulcerative colitis, varies according to the severity of the disease. There may be as many as 30 to 40 bowel movements a day. Because ulcerative colitis affects the mucosal layer of the bowel, the stools typically contain blood and mucus. Nocturnal diarrhea usually occurs when daytime symptoms are severe. There may be mild abdominal cramping and fecal incontinence. Anorexia, weakness, and fatigability are common.

Ulcerative colitis usually follows a course of remissions and exacerbations. The severity of the disease varies from mild to fulminating. Accordingly, the disease has been divided into three types: mild chronic, chronic intermittent, and acute fulminating. The most common form of the disease is the mild chronic, in which bleeding and diarrhea are mild and systemic signs are minimal or absent. This form of the disease usually can be managed conservatively. The chronic intermittent form continues after the initial attack. Compared with the milder form, more of the colon surface usually is involved with the chronic intermittent form, and there are more systemic signs and complications. In approximately 15% of affected persons, the disease assumes a more fulminant course, involves the entire colon, and manifests with severe, bloody diarrhea, fever, and acute abdominal pain. These persons are at risk for development of toxic megacolon, which is characterized by dilatation of the colon and signs of systemic toxicity. It results from extension of the inflammatory response, with involvement of neural and vascular components of the bowel. Contributing factors include the use of laxatives, narcotics, and



■ FIGURE 27-5 ■ Ulcerative colitis. Prominent erythema and ulceration of the colon begin in the ascending colon and are most severe in the rectosigmoid area. (From Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 731]. Philadelphia: Lippincott Williams & Wilkins)

anticholinergic drugs and the presence of hypokalemia. Cancer of the colon is one of the feared complications of ulcerative colitis.

Diagnosis and Treatment. Diagnosis of ulcerative colitis is based on history and physical examination. The diagnosis usually is confirmed by proctosigmoidoscopy.

Treatment depends on the extent of the disease and severity of symptoms. It includes measures to control the acute manifestations of the disease and prevent recurrence. Some people with mild to moderate symptoms are able to control their symptoms simply by avoiding caffeine, lactose (milk), highly spiced foods, and gas-forming foods. Fiber supplements may be used to decrease diarrhea and rectal symptoms.

The medications used in treatment of ulcerative colitis are similar to those used in the treatment of Crohn's disease. The corticosteroids are used selectively to lessen the acute inflammatory response. Many of these medications can be administered rectally by suppository or enema. Immunosuppressant drugs, such as cyclosporine, may be used to treat persons with severe colitis.

Surgical treatment (*i.e.*, removal of the rectum and entire colon) with the creation of an ileostomy or ileoanal anastomosis may be required for those persons with ulcerative colitis who do not experience a response to conservative methods of treatment.

Infectious Colitis

Two forms of pathogens have emerged as important causes of infectious colitis: *Clostridium difficile* and *Escherichia coli* sero-type O157:H7.

Clostridium Difficile Colitis

C. difficile, which is part of normal flora in 2% to 10% of humans, is a gram-positive spore-forming bacillus that has been implicated as the offending organism in colitis associated with antibiotic therapy.²⁸ The spores are resistant to the acid environment of the stomach and convert to vegetative forms in the colon. Treatment with broad-spectrum antibiotics predisposes to disruption of the normal bacterial flora of the colon, leading to colonization by *C. difficile*. Almost any antibiotic with activity against the gram-negative bacteria of the normal intestinal flora are the most common agents.²⁸ After antibiotic therapy has made the bowel susceptible to infection, colonization by *C. difficile* occurs by the oral-fecal route. *C. difficile* infection usually is acquired in the hospital, where the organism is commonly encountered.

C. difficile, which is not invasive, produces its effects through the release of toxins that cause mucosal damage and inflammation. The infection commonly manifests with diarrhea that is mild to moderate and sometimes is accompanied by lower abdominal cramping. Symptoms usually begin during or shortly after antibiotic therapy has been initiated, although they can be delayed for weeks. In most cases, systemic manifestations are absent, and the symptoms subside after the antibiotic has been discontinued. In the more severe cases, mucosal damage results in the development a pseudomembranous colitis. It is a life-threatening form of the disease characterized by an adherent inflammatory membrane overlying the areas of mucosal injury. Persons with the disease are acutely ill, with lethargy, fever, tachycardia, abdominal pain and distention, and dehydration. The smooth muscle tone of the colon may be lost, resulting in toxic dilatation of the colon. Prompt therapy is needed to prevent perforation of the bowel.

Diagnostic findings include a history of antibiotic use and laboratory tests that confirm the presence of *C. difficile* toxins in the stool. Treatment includes the immediate discontinuation of antibiotic therapy. Specific treatment aimed at eradicating *C. difficile* is used when symptoms are severe or persistent.^{28,29}

Escherichia Coli O157:H7 Infection

E. coli O157:H7 is a strain of *E. coli* found in feces and contaminated milk of healthy dairy and beef cattle, but it also has been found in pork, poultry, and lamb. Infection usually is by food-borne transmission, often by ingesting undercooked hamburger. The organism also can be transferred to nonmeat products such as fruits and vegetables. Personto-person transmission may occur, particularly in nursing homes, day care settings, and hospitals. The very young and the very old are particularly at risk for the infection and its complications.

The infection may cause no symptoms or cause a variety of manifestations, including acute, nonbloody diarrhea, hemorrhagic colitis, hemolytic-uremic syndrome, and thrombotic thrombocytopenic purpura. The infection often presents with abdominal cramping and watery diarrhea and subsequently may progress to bloody diarrhea. The diarrhea commonly lasts 3 to 7 days or longer, with 10 to 12 diarrheal episodes per day. Fever occurs in up to one third of the cases.

An important aspect of the disease is the production of toxins and the ability to produce toxemia. The two complications of the infection, hemolytic-uremic syndrome and thrombotic thrombocytopenic purpura, reflect the effects of toxins. Hemolytic-uremic syndrome is characterized by hemolytic anemia, thrombocytopenia, and renal failure. It occurs predominantly in infants and young children and is the most common cause of acute renal failure in children.³⁰ It has a mortality rate of 5% to 10%, and one third of the survivors are left with permanent disability. Thrombotic thrombocytopenic purpura is manifested by thrombocytopenia, renal failure, fever, and neurologic manifestations. It often is regarded as the severe end of the disease that leads to hemolytic-uremic syndrome plus neurologic problems.

No specific therapy is available for *E. coli* O157:H7 infection. Treatment is largely symptomatic and directed toward treating the effects of complications. Antibiotics have not proved useful and may even be harmful, extending the duration of bloody diarrhea.

Because of the seriousness of the infection and its complications, education of the public about techniques for decreasing primary transmission of the infection from animal sources is important. Undercooked meats and unpasteurized milk are sources of transmission. The FDA recommends a minimal internal temperature of 155°F for cooked hamburger. Food handlers and consumers should be aware of the proper methods for handling uncooked meat to prevent crosscontamination of other foods. Particular attention should be paid to hygiene in day care centers and nursing homes, where the spread of infection to the very young and very old may result in severe complications.³⁰

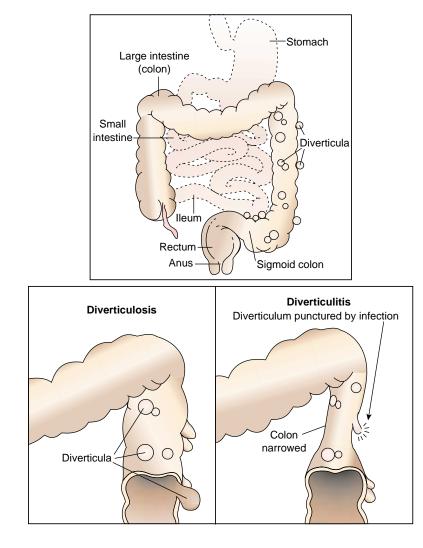
Diverticular Disease

Diverticulosis is a condition in which the mucosal layer of the colon herniates through the muscularis layer.³¹ Often, there are multiple diverticula, and most occur in the sigmoid colon (Fig. 27-6). Diverticular disease is common in Western society, affecting approximately 5% to 10% of the population older than 45 years and almost 80% of those older than 85 years.³¹ Although the disorder is prevalent in the developed countries of the world, it is almost nonexistent in many African nations and underdeveloped countries. This suggests that dietary factors (*e.g.*, lack of fiber content), a decrease in physical activity, and poor bowel habits (*e.g.*, neglecting the urge to defecate), along with the effects of aging, contribute to the development of the disease.

Diverticula, in which the mucosal and submucosal layers herniate through the muscle layers of the colon, are located primarily in the sigmoid colon but can affect any area of the colon. They vary in number from a few to several hundred. The diverticula, which measure up to 1 cm in greatest dimension, are attached to the intestinal lumen by necks of varying size. Hardened feces may be present in the diverticula without causing symptoms.

Most persons with diverticular disease remain asymptomatic. The disease often is found when x-ray studies are done for other purposes. When symptoms do occur, they often are attributed to irritable bowel syndrome or other causes. Illdefined lower abdominal discomfort, a change in bowel habits (*e.g.*, diarrhea, constipation), bloating, and flatulence are common.

Diverticulitis is a complication of diverticulosis in which there is inflammation and gross or microscopic perforation of the diverticulum. One of the most common complaints of diverticulitis is pain in the lower left quadrant, accompanied by nausea and vomiting, tenderness in the lower left quadrant, a slight fever, and an elevated white blood cell count. These symptoms usually last for several days, unless complications occur, and usually are caused by localized inflammation of the diverticula with perforation and development of a small, localized abscess. Complications include perforation with peritonitis, hemorrhage, and bowel obstruction. Fistulas can form, usually involving the bladder (*i.e.*, vesicosigmoid fistula) but sometimes involving the skin, perianal area, or small bowel. Pneumaturia (*i.e.*, air in the urine) is a sign of vesicosigmoid fistula.



■ FIGURE 27-6 ■ (Top) Location of diverticula in the sigmoid colon. (Bottom left) Diverticulosis. (Bottom right) Diverticulitis. (National Digestive Diseases Information Clearinghouse. [1989]. *Clearinghouse fact sheet: Diverticulosis and diverticulitis*. NIH publication 90-1163. Washington, DC: US Department of Health and Human Services)

The diagnosis of diverticular disease is based on history and presenting clinical manifestations. The disease may be confirmed by barium enema x-ray studies (except when acute diverticulitis is suspected), CT scans, and ultrasonographic studies. The treatment goals for diverticular disease focus on prevention of symptoms and complications. This includes increasing the bulk in the diet and bowel retraining so that the person has at least one bowel movement each day. Acute diverticulitis is treated by withholding solid food and administering a broad-spectrum antibiotic. Surgical treatment is reserved for complications.

Appendicitis

Acute appendicitis is extremely common. It is seen most frequently in the 5- to 30-year-old age group, but it can occur at any age. The appendix becomes inflamed, swollen, and gangrenous, and it eventually perforates if not treated. Although the cause of appendicitis is unknown, it is thought to be related to intraluminal obstruction with a fecalith (*i.e.*, hard piece of stool) or to twisting.

Appendicitis usually has an abrupt onset, with pain referred to the epigastric or periumbilical area. This pain is caused by stretching of the appendix during the early inflammatory process. At approximately the same time that the pain appears, there are one or two episodes of nausea. Initially, the pain is vague, but over a period of 2 to 12 hours, it gradually increases and may become colicky. When the inflammatory process has extended to involve the serosal layer of the appendix and the peritoneum, the pain becomes localized to the lower right quadrant. There usually is an elevation in temperature and a white blood cell count greater than 10,000/mm³, with 75% or more polymorphonuclear cells. Palpation of the abdomen usually reveals a deep tenderness in the lower right quadrant, which is confined to a small area approximately the size of the fingertip. It usually is located at approximately the site of the inflamed appendix. Rebound tenderness, which is pain that occurs when pressure is applied to the area and then released, and spasm of the overlying abdominal muscles are common.

Treatment consists of surgical removal of the appendix. Complications include peritonitis, localized periappendiceal abscess formation, and septicemia.

Alterations in Intestinal Motility

The movement of contents through the gastrointestinal tract is controlled by neurons located in the submucosal and myenteric plexuses of the gut (see Chapter 26). The axons from the cell bodies in the myenteric plexus innervate the circular and longitudinal smooth muscle layers of the gut. These neurons receive impulses from local receptors located in the mucosal and muscle layers of the gut and extrinsic input from the parasympathetic and sympathetic nervous systems. As a general rule, the parasympathetic nervous system tends to increase the motility of the bowel, whereas sympathetic stimulation tends to slow its activity.

The colon has sphincters at both ends: the ileocecal sphincter, which separates it from the small intestine, and the anal sphincter, which prevents the movement of feces to the outside of the body. The colon acts as a reservoir for fecal material. Normally, approximately 400 mL of water, 55 mEq of sodium, 30 mEq of chloride, and 15 mEq of bicarbonate are absorbed

KEY CONCEPTS

DISORDERS OF GASTROINTESTINAL MOTILITY

- The enteric nervous system that is incorporated into the wall of the gut controls the basic movement of the gastrointestinal tract, with input from the autonomic nervous system.
- Local irritation and the composition and constituents of gastrointestinal contents influence motility through the submucosal afferent neurons of the enteric nervous system. Gastrointestinal wall distention, chemical irritants, osmotic gradients, and bacterial toxins exert many of their effects on gastrointestinal motility through these afferent pathways.
- Autonomic influences generated by factors such as medications, trauma, and emotional experiences interact with the enteric nervous system to alter gastrointestinal motility.

each day in the colon. At the same time, approximately 5 mEq of potassium is secreted into the lumen of the colon. The amount of water and electrolytes that remains in the stool reflects the absorption or secretion that occurs in the colon. The average adult ingesting a typical American diet evacuates approximately 200 to 300 g of stool each day.

Diarrhea

The usual definition of *diarrhea* is excessively frequent passage of stools. Diarrhea can be acute or chronic. Diarrhea is considered to be chronic when the symptoms persist for 3 weeks in children or adults and 4 weeks in infants. Acute diarrhea affects 500 million children throughout the world and is the leading cause of death of children younger than 4 years of age.³² Although diarrheal disease in the United States is less prevalent than it is in other countries, it places a burden on the health care system. Approximately 220,000 children are hospitalized each year for gastroenteritis.³³

The complaint of diarrhea is a general one and can be related to a number of pathologic and nonpathologic factors. It can be caused by infectious organisms, food intolerance, drugs, or intestinal disease. Acute diarrheas that last less than 4 days are predominantly caused by infectious agents and follow a self-limited course.³⁴ Chronic diarrheas are those that persist for longer than 3 to 4 weeks. They often are caused by conditions such as inflammatory bowel disease, irritable bowel syndrome, malabsorption syndrome, endocrine disorders (hyperthyroidism, diabetic autonomic neuropathy), or radiation colitis.

Diarrhea commonly is divided into two types, large volume and small volume, based on the characteristics of the diarrheal stool. Large-volume diarrhea results from an increase in the water content of the stool, and small-volume diarrhea results from an increase in the propulsive activity of the bowel. Some of the common causes of small- and large-volume diarrhea are summarized in Chart 27-1. Often, diarrhea is a combination of these two types.

CHART 27-1 Causes of Largeand Small-Volume Diarrhea

Large-Volume Diarrhea

Osmotic diarrhea
Saline cathartics
Lactase deficiency
Secretory diarrhea
Acute infectious diarrhea
Failure to absorb bile salts
Fat malabsorption
Chronic laxative abuse
Carcinoid syndrome
Zollinger-Ellison syndrom
u ,
Fecal impaction
u ,
Fecal impaction
Fecal impaction Small-Volume Diarrhea
Fecal impaction Small-Volume Diarrhea Inflammatory bowel disease
Fecal impaction Small-Volume Diarrhea Inflammatory bowel disease Crohn's disease
Fecal impaction Small-Volume Diarrhea Inflammatory bowel disease Crohn's disease Ulcerative colitis
Fecal impaction Small-Volume Diarrhea Inflammatory bowel disease Crohn's disease Ulcerative colitis Infectious disease
Fecal impaction Small-Volume Diarrhea Inflammatory bowel disease Crohn's disease Ulcerative colitis Infectious disease Shigellosis

Large-Volume Diarrhea. Large-volume diarrhea can be classified as secretory or osmotic, according to the cause of the increased water content in the feces. Water is pulled into the colon along an osmotic gradient (*i.e.*, osmotic diarrhea) or is secreted into the bowel by the mucosal cells (*i.e.*, secretory diarrhea). The large-volume form of diarrhea usually is a painless, watery type without blood or pus in the stools.

In osmotic diarrhea, water is pulled into the bowel by the hyperosmotic nature of its contents. It occurs when osmotically active particles are not absorbed. In persons with lactase deficiency, the lactose in milk cannot be broken down and absorbed. Magnesium salts, which are contained in milk of magnesia and many antacids, are poorly absorbed and cause diarrhea when taken in sufficient quantities. Another cause of osmotic diarrhea is decreased transit time, which interferes with absorption. Osmotic diarrhea usually disappears with fasting.

Secretory diarrhea occurs when the secretory processes of the bowel are increased. Most acute infectious diarrheas are of this type. Enteric organisms cause diarrhea by several ways. Some are noninvasive but secrete toxins that stimulate fluid secretion (e.g., pathogenic E. coli or Vibrio cholerae). Others (e.g., Staphylococcus aureus, Bacillus cereus, Clostridium perfringens) invade and destroy intestinal epithelial cells, thereby altering fluid transport so that secretory activity continues while absorption activity is halted.35 Diarrhea with vomiting and fever suggests food poisoning, often caused by staphylococcal enterotoxin. Secretory diarrhea also occurs when excess bile acids remain in the intestinal contents as they enter the colon. This often happens with disease processes of the ileum because bile salts are absorbed there. It also may occur with bacterial overgrowth in the small bowel, which interferes with bile absorption.

Small-Volume Diarrhea. Small-volume diarrhea commonly is associated with acute or chronic inflammation or intrinsic disease of the colon, such as ulcerative colitis or Crohn's disease. Small-volume diarrhea usually is evidenced by frequency and urgency and colicky abdominal pain. It commonly is accompanied by tenesmus (*i.e.*, painful straining at stool), fecal soiling of clothing, and awakening during the night with the urge to defecate.

Diagnosis and Treatment. The diagnosis of diarrhea is based on complaints of frequent stools and a history of accompanying factors such as concurrent illnesses, medication use, and exposure to potential intestinal pathogens. Disorders such as inflammatory bowel disease should be considered. If the onset of diarrhea is related to travel outside the United States, the possibility of traveler's diarrhea must be considered.

Although most acute forms of diarrhea are self-limited and require no treatment, diarrhea can be particularly serious in infants and small children, persons with other illnesses, the elderly, and even previously healthy persons if it continues for any length of time. Thus, the replacement of fluids and electrolytes is considered to be a primary therapeutic goal in the treatment of diarrhea. Oral electrolyte replacement solutions can be given in situations of uncomplicated diarrhea that can be treated at home. Evidence suggests that feeding should be continued during diarrheal illness, particularly in children.^{33,36} It is recommended that children who require rehydration therapy because of diarrhea be fed an age-appropriate diet. Starch and simple proteins are thought to provide cotransport molecules with little osmotic activity, increasing fluid and electrolyte uptake by intestinal cells. When oral rehydration is not feasible or adequate, intravenous fluid replacement may be needed.

Drugs used in the treatment of diarrhea include diphenoxylate and loperamide, which are opium-like drugs. These drugs decrease gastrointestinal motility and stimulate water and electrolyte absorption. Adsorbents, such as kaolin and pectin, adsorb irritants and toxins from the bowel. These ingredients are included in many over-the-counter antidiarrheal preparations. Bismuth subsalicylate can be used to reduce the frequency of unformed stools and increase stool consistency, particularly in cases of traveler's diarrhea. The drug is thought to inhibit intestinal secretion caused by enterotoxic *E. coli* and cholera toxins. Diarrheal medications should not be used in persons with bloody diarrhea, high fever, or signs of toxicity for fear of worsening the disease. Antibiotics are reserved for persons with identified enteric pathogens.

Constipation

Constipation can be defined as the infrequent passage of stools. The difficulty with this definition arises from the many individual variations of function that are normal. What is considered normal for one person (*e.g.*, two or three bowel movements per week) may be considered evidence of constipation by another. The problem increases with age; there is a sharp rise in health care visits for constipation after 65 years of age.

Constipation can occur as a primary problem or as a problem associated with another disease condition. Some common causes of constipation are failure to respond to the urge to defecate, inadequate fiber in the diet, inadequate fluid intake, weakness of the abdominal muscles, inactivity and bed rest, pregnancy, and hemorrhoids. Diseases associated with chronic

Chapter 27: Alterations in Gastrointestinal Function

487

constipation include neurologic diseases such as spinal cord injury, Parkinson's disease, and multiple sclerosis; endocrine disorders such as hypothyroidism; and obstructive lesions in the gastrointestinal tract. Drugs such as narcotics, anticholinergic agents, calcium channel blockers, diuretics, calcium (antacids and supplements), iron supplements, and aluminum antacids tend to cause constipation. Elderly people with long-standing constipation may experience dilation of the rectum, colon, or both. This condition allows large amounts of stool to accumulate with little or no sensation. Constipation, in the context of a change in bowel habits, may be a sign of colorectal cancer.

Diagnosis of constipation usually is based on a history of infrequent stools, straining with defection, the passing of hard and lumpy stools, or the sense of incomplete evacuation with defection.³⁷ Constipation as a sign of another disease condition should be excluded. The treatment of constipation usually is directed toward relieving the cause. Adequate fluid intake and bulk in the diet should be encouraged. Moderate exercise is essential, and persons on bed rest benefit from passive and active exercises. Laxatives and enemas should be used judiciously. They should not be used on a regular basis to treat simple constipation because they interfere with the defecation reflex and enemas may damage the rectal mucosa.

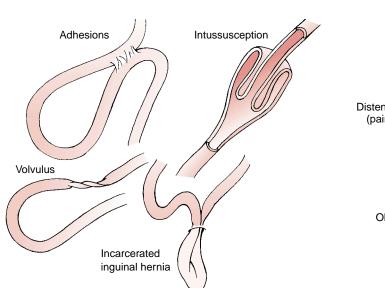
Intestinal Obstruction

Intestinal obstruction designates an impairment of movement of intestinal contents in the usual oral to anal direction. The causes can be categorized as mechanical or paralytic obstruction. Strangulation with necrosis of the bowel may occur and lead to perforation, peritonitis, and sepsis. This is a serious complication and may increase the mortality rate associated with intestinal obstruction to approximately 25% if surgery is delayed.⁶

Mechanical obstruction can result from a number of conditions, intrinsic or extrinsic, that encroach on the patency of the bowel lumen (Fig. 27-7). Major inciting causes include external hernia (*i.e.*, inguinal, femoral, or umbilical) and postoperative adhesions. Less common causes are strictures, tumor, foreign bodies, intussusception, and volvulus. Intussusception involves the telescoping of bowel into the adjacent segment. It is the most common cause of intestinal obstruction in children younger than 2 years.³⁸ The most common form is intussusception of the terminal ileum into the right colon, but other areas of the bowel may be involved. In most cases, the cause of the disorder is unknown. The condition can also occur in adults when an intraluminal mass or tumor acts as a traction force and pulls the segment along as it telescopes into the distal segment. Volvulus refers to a complete twisting of the bowel on an axis formed by its mesentery (see Fig. 27-7). Mechanical bowel obstruction may be a simple obstruction, in which there is no alteration in blood flow, or a strangulated obstruction, in which there is impairment of blood flow and necrosis of bowel tissue.

Paralytic, or adynamic, obstruction results from neurogenic or muscular impairment of peristalsis. Paralytic ileus is seen most commonly after abdominal surgery. It also accompanies inflammatory conditions of the abdomen, intestinal ischemia, pelvic fractures, and back injuries. It occurs early in the course of peritonitis and can result from chemical irritation caused by bile, bacterial toxins, electrolyte imbalances as in hypokalemia, and vascular insufficiency.

The major effects of both types of intestinal obstruction are abdominal distention and loss of fluids and electrolytes (Fig. 27-8). Gases and fluids accumulate in the area; if untreated, the distention resulting from bowel obstruction tends to perpetuate itself by causing atony of the bowel and further distention. Distention is further aggravated by the accumulation of gases. Approximately 70% of these gases are derived from swallowed air. As the process continues, the distention moves proximally (*i.e.*, toward the mouth), involving additional segments of bowel. Either form of obstruction eventually



■ FIGURE 27-7 ■ Causes of mechanical bowel obstruction.

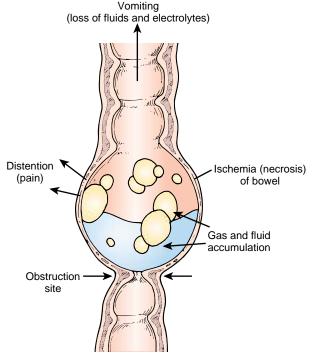


FIGURE 27-8 Pathophysiology of intestinal obstruction.

may lead to strangulation (*i.e.*, interruption of blood flow), gangrenous changes, and ultimately, perforation of the bowel. The increased pressure in the intestine tends to compromise mucosal blood flow, leading to necrosis and movement of blood into the luminal fluids. This promotes rapid growth of bacteria in the obstructed bowel. Anaerobes grow rapidly in this favorable environment and produce a lethal endotoxin.

The manifestations of intestinal obstruction depend on the degree of obstruction and its duration. With acute obstruction, the onset usually is sudden and dramatic. With chronic conditions, the onset often is more gradual. The cardinal symptoms of intestinal obstruction are pain, absolute constipation, abdominal distention, and vomiting. With mechanical obstruction, the pain is severe and colicky, in contrast with the continuous pain and silent abdomen of paralytic ileus. There also is borborygmus (i.e., rumbling sounds made by propulsion of gas in the intestine); audible, high-pitched peristalsis; and peristaltic rushes. Visible peristalsis may appear along the course of the distended intestine. Extreme restlessness and conscious awareness of intestinal movements are experienced along with weakness, perspiration, and anxiety. Should strangulation occur, the symptoms change. The character of the pain shifts from the intermittent colicky pain caused by the hyperperistaltic movements of the intestine to a severe and steady type of pain. Vomiting and fluid and electrolyte disorders occur with both types of obstruction.

Diagnosis of intestinal obstruction usually is based on history and physical findings. Abdominal x-ray studies reveal a gas-filled bowel.

Treatment depends on the cause and type of obstruction. Most cases of adynamic obstruction respond to decompression of the bowel through nasogastric suction and correction of fluid and electrolyte imbalances. Strangulation and complete bowel obstruction require surgical intervention.

Peritonitis

Peritonitis is an inflammatory response of the serous membrane that lines the abdominal cavity and covers the visceral organs. It can be caused by bacterial invasion or chemical irritation. Most commonly, enteric bacteria enter the peritoneum because of a defect in the wall of one of the abdominal organs. The most common causes of peritonitis are perforated peptic ulcer, ruptured appendix, perforated diverticulum, gangrenous bowel, pelvic inflammatory disease, and gangrenous gallbladder. Other causes are abdominal trauma and wounds. Generalized peritonitis, although no longer the overwhelming problem it once was, is still a leading cause of death after abdominal surgery.

The peritoneum has several characteristics that increase its vulnerability to or protect it from the effects of peritonitis. One weakness of the peritoneal cavity is that it is a large, unbroken space that favors the dissemination of contaminants. For the same reason, it has a large surface that permits rapid absorption of bacterial toxins into the blood. The peritoneum is particularly well adapted for producing an inflammatory response as a means of controlling infection. For example, it tends to exude a thick, sticky, and fibrinous substance that adheres to other structures, such as the mesentery and omentum, and that seals off the perforated viscus and aids in localizing the process. Localization is enhanced by sympathetic stimulation that limits intestinal motility. Although the diminished or absent peristalsis that occurs tends to give rise to associated problems, it does inhibit the movement of contaminants throughout the peritoneal cavity.

One of the most important manifestations of peritonitis is the translocation of extracellular fluid into the peritoneal cavity (through weeping or serous fluid from the inflamed peritoneum) and into the bowel as a result of bowel obstruction. Nausea and vomiting cause further losses of fluid. The fluid loss may encourage development of hypovolemia and shock. The onset of peritonitis may be acute, as with a ruptured appendix, or it may have a more gradual onset, as occurs in pelvic inflammatory disease. Pain and tenderness are common symptoms. The pain usually is more intense over the inflamed area. The person with peritonitis usually lies still because any movement aggravates the pain. Breathing often is shallow to prevent movement of the abdominal muscles. The abdomen usually is rigid and sometimes described as boardlike because of reflex muscle guarding. Vomiting is common. Fever, an elevated white blood cell count, tachycardia, and hypotension are common. Hiccups may develop because of irritation of the phrenic nerve. Paralytic ileus occurs shortly after the onset of widespread peritonitis and is accompanied by abdominal distention. Peritonitis that progresses and is untreated leads to toxemia and shock.

Treatment measures for peritonitis are directed toward preventing the extension of the inflammatory response, correcting the fluid and electrolyte imbalances that develop, and minimizing the effects of paralytic ileus and abdominal distention. Surgical intervention may be needed to remove an acutely inflamed appendix or close the opening in a perforated peptic ulcer. Oral fluids are forbidden. Nasogastric suction, which entails the insertion of a tube placed through the nose into the stomach or intestine, is used to decompress the bowel and relieve the abdominal distention. Fluid and electrolyte replacement is essential. These fluids are prescribed on the basis of frequent blood chemistry determinations. Antibiotics are given to combat infection. Narcotics often are needed for pain relief.

Alterations in Intestinal Absorption

Malabsorption is the failure to transport dietary constituents, such as fats, carbohydrates, proteins, vitamins, and minerals, from the lumen of the intestine to the extracellular fluid compartment for transport to the various parts of the body. It can selectively affect a single component, such as vitamin B_{12} or lactose, or its effects can extend to all the substances absorbed in a specific segment of the intestine. When one segment of the intestine is affected, another may compensate. For example, the ileum may compensate for malabsorption in the proximal small intestine by absorbing substantial amounts of fats, carbohydrates, and amino acids. Similarly, the colon, which normally absorbs water, sodium, chloride, and bicarbonate, can compensate for small intestine malabsorption by absorbing additional end-products of bacterial carbohydrate metabolism.

The conditions that impair one or more steps involved in digestion and absorption of nutrients can be divided into three broad categories: intraluminal maldigestion, mucosal malabsorption, and lymphatic obstruction. Intraluminal maldigestion involves a defect in processing of nutrients in the intestinal lumen. The most common causes are pancreatic insuf-

ficiency, hepatobiliary disease, and intraluminal bacterial growth. Mucosal malabsorption is caused by mucosal lesions that impair uptake and transport of available intraluminal nutrients across the mucosal surface of the intestine. They include disorders such as celiac disease, and Crohn's disease. Lymphatic obstruction interferes with the transport of the products of fat digestion to the systemic circulation after they have been absorbed by the intestinal mucosa. The process can be interrupted by congenital defects, neoplasms, trauma, and selected infectious diseases.

Malabsorption Syndrome

The term syndrome implies a common constellation of symptoms arising from multiple causes. Persons with conditions that diffusely affect the small intestine and reduce its absorptive functions share certain common features referred to as *mal*absorption syndrome. Among the causes of malabsorption syndrome are celiac sprue, Crohn's disease, and resection of large segments of the small bowel.

Celiac sprue is a relatively rare chronic disease in which there is a characteristic mucosal lesion of the small intestine and impaired nutrient absorption, which improves when gluten is removed from the diet.^{6,11} There is convincing evidence that the disorder is caused by an immunologic response to the gliadin fraction of gluten. The condition results in loss of absorptive villi from the small intestine. When the resulting lesions are extensive, they may impair absorption of virtually all nutrients. In approximately one third of the cases, symptoms begin in childhood. The effects of celiac sprue usually are reversed after removal of all wheat, rye, barley, and oat gluten from the diet. Corn and rice products are not toxic and can be used as substitutes.

Persons with intestinal malabsorption usually have symptoms directly referable to the gastrointestinal tract that include diarrhea, steatorrhea (fatty stools), flatulence, bloating, abdominal pain, and cramps. Weakness, muscle wasting, weight loss, and abdominal distention often are present. Weight loss often occurs despite normal or excessive caloric intake. In a person consuming a diet containing 80 to 100 g of fat each day, excretion of 7 to 9 g of fat indicates steatorrhea.

Along with loss of fat in the stools, there is failure to absorb the fat-soluble vitamins. This can lead to easy bruising and bleeding (*i.e.*, vitamin K deficiency), bone pain, a predisposition to the development of fractures and tetany (*i.e.*, vitamin D and calcium deficiency), macrocytic anemia, and glossitis (*i.e.*, folic acid deficiency). Neuropathy, atrophy of the skin, and peripheral edema may be present. Table 27-2 describes the signs and symptoms of impaired absorption of dietary constituents.

Neoplasms

Epithelial cell tumors of the intestines are a major cause of morbidity and mortality worldwide. Although the small intestine accounts for approximately 75% of the length of the gastrointestinal tract, its tumors account for only 3% to 6% of gastrointestinal tumors.11

Adenomatous Polyps

By far the most common types of neoplasms of the intestine are adenomatous polyps. A gastrointestinal polyp can be described as a mass that protrudes into the lumen of gut.^{11,16} Polyps can be subdivided according to their attachment to the bowel wall (sessile [raised mucosal nodules] or pedunculated [attached by a stalk]); their histopathologic appearance (hyperplastic or adenomatous); and their neoplastic potential (benign or malignant).16

Adenomatous polyps (adenomas) are benign neoplasms that arise from the mucosal epithelium of the intestine (Fig. 27-9). They are composed of neoplastic cells that have proliferated in excess of those needed to replace the cells that normally are shed from the mucosal surface. The pathogenesis of adenoma formation involves neoplastic alteration in the replication of the crypt epithelial cells in the crypts of Lieberkühn. There may be diminished apoptosis (see Chapter 2), persistence of cell replication, and failure of cell maturation and differentiation of the cells that migrate to the surface of the crypts.¹⁶ Normally, DNA synthesis ceases as the cells reach the upper two thirds of the crypts, after which they mature, migrate to the surface, and become senescent. They then become apoptotic and are shed from the surface.¹⁶ Adenomas arise from a disruption in this sequence, such that the epithelial cells retain their proliferative ability throughout the entire length of the crypt. Alterations in cell differentiation can lead to dysplasia and progression to the development of invasive carcinoma.

Most cases of colorectal cancer begin as benign adenomatous colonic polyps (Fig. 27-10). The frequency of polyps increases with age, and the prevalence of adenomatous polyps, which is approximately 20% to 30% before 40 years of age, rises to 40% to 50% after age 60 years.¹¹ Men and women are equally affected. The peak incidence of adenomatous polyps precedes by some years the peak for colorectal cancer. Programs that provide careful follow-up for persons with adenomatous polyps and removal of all suspect lesions have substantially reduced the incidence of colorectal cancer.11

Colorectal Cancer

Colorectal cancer is the second leading cause of cancer death in the United States. In 2003, there will be an estimated 105,500 new cases of colorectal cancer and 57,100 deaths associated with the disease.²¹ The death rate for colorectal cancer has been steadily declining since the early 1980s. This may attributable to a decreased number of cases, because more of the cases are found earlier, and because treatments have improved.

The cause of cancer of the colon and rectum is largely unknown. Its incidence increases with age, as evidenced by the fact that approximately 80% of persons who have this form of cancer are older than 50 years.11 The incidence of colorectal cancer is increased among persons with a family history of cancer, persons with Crohn's disease or ulcerative colitis, and those with familial adenomatous polyposis of the colon. Diet also is thought to play a role. Attention has focused on dietary fat intake, refined sugar intake, fiber intake, and the adequacy of such protective micronutrients as vitamins A, C, and E in the diet. Reports indicate that aspirin may protect against colorectal cancer.^{40,41} Although the mechanism of aspirin's action is unknown, it may be related to its effect on the synthesis of prostaglandins, one or more of which may be involved in signal systems that influence cell proliferation or tumor growth.

Usually, cancer of the colon and rectum is present for a long time before it produces symptoms. Bleeding is a highly significant early symptom, and it usually is the one that causes persons to seek medical care. Other symptoms include a change in

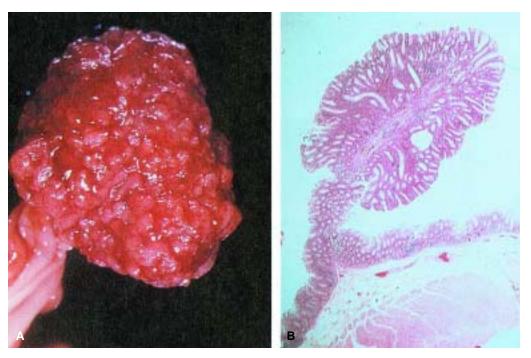
TABLE 27-2 Sites of and Requirements for Absorption of Dietary Constituents and Manifestations of Malabsorption

Dietary Constituent	Site of Absorption	Requirements	Manifestations
Water and electrolytes	Mainly small bowel	Osmotic gradient	Diarrhea
			Dehydration
			Cramps
Fat	Upper jejunum	Pancreatic lipase	Weight loss
		Bile salts	Steatorrhea
		Functioning lymphatic channels	Fat-soluble vitamin deficiency
Carbohydrates			
Starch	Small intestine	Amylase	Diarrhea
		Maltase	Flatulence
		Isomaltase α -dextrins	Abdominal discomfort
Sucrose	Small intestine	Sucrase	
Lactose	Small intestine	Lactase	
Maltose	Small intestine	Maltase	
Fructose	Small intestine		
Protein	Small intestine	Pancreatic enzymes (e.g., trypsin,	Loss of muscle mass
		chymotrypsin, elastin)	Weakness
			Edema
Vitamins			
A	Upper jejunum	Bile salts	Night blindness
			Dry eyes
			Corneal irritation
Folic acid	Duodenum and jejunum	Absorptive; may be impaired by some	Cheilosis
		drugs (<i>i.e.,</i> anticonvulsants)	Glossitis
			Megaloblastic anemia
B ₁₂	lleum	Intrinsic factor	Glossitis
			Neuropathy
			Megaloblastic anemia
D	Upper jejunum	Bile salts	Bone pain
			Fractures
			Tetany
E	Upper jejunum	Bile salts	Uncertain
К	Upper jejunum	Bile salts	Easy bruising and bleeding
Calcium	Duodenum	Vitamin D and parathyroid hormone	Bone pain
			Fractures
			Tetany
Iron	Duodenum and jejunum	Normal pH (hydrochloric acid secretion)	Iron-deficiency anemia
			Glossitis

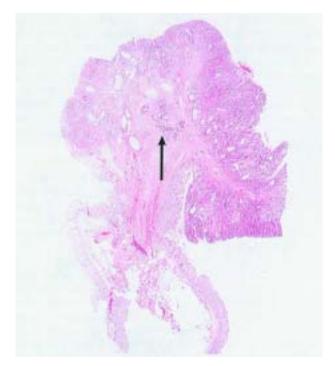
bowel habits, diarrhea or constipation, and sometimes a sense of urgency or incomplete emptying of the bowel. Pain usually is a late symptom.

The prognosis for persons with colorectal cancer depends largely on the extent of bowel involvement and on the presence of metastasis at the time of diagnosis. Colorectal cancer commonly is divided into four categories according to the Dukes classification or its variants.^{15,20} A stage A tumor is limited to invasion of the mucosal and submucosal layers of the colon and is associated with a 5-year survival rate of almost 100%.¹¹ A stage B tumor involves the entire wall of the colon, but without lymph node involvement, and is associated with a 5-year survival rate of 43% to 67%.¹¹ With a stage C tumor, there is invasion of the serosal layer and involvement of the regional lymph nodes. The 5-year survival rate is approximately 23%.¹¹ Stage D colorectal cancers involve far-advanced metastasis and have a much poorer prognosis.

Screening, Diagnosis, and Treatment. Among the methods used for the detection of colorectal cancers are stool occult blood tests and digital rectal examination, usually done during routine physical examinations; x-ray studies using barium (*e.g.*, barium enema); and flexible sigmoidoscopy and colonoscopy. Digital rectal examinations are most helpful in detecting neoplasms of the rectum. Rectal examination should be considered a routine part of a good physical examination. The American Cancer Society recommends that all asymptomatic men and women older than 40 years should have a digital rectal examination, and that those older than 50 years should have



■ FIGURE 27-9 ■ Tubular adenoma of the colon. (A) A pedunculated tubular adenoma. (B) A low-power photomicrograph of a tubular adenoma shows closely packed epithelial tubules. The fibrous stalk is vascular and covered with normal colonic epithelium. (From Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 737]. Philadelphia: Lippincott Williams & Wilkins)



■ FIGURE 27-10 ■ Adenocarcinoma arising from a pedunculated adenomatous polyp. A low-grade micrograph shows irregular neoplastic glands (*arrow*) invading the stalk. (From Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 738]. Philadelphia: Lippincott Williams & Wilkins)

an annual stool test for occult blood and a flexible sigmoidoscopy examination every 5 years and colonoscopy every 10 years or double contrast barium enema every 5 years, as recommended by their physician.³⁹

Carcinoembryonic antigen (CEA) can be used as a marker for colorectal cancer. However, blood levels are of little screening or diagnostic value because they become elevated only after the tumor has reached considerable size. Moreover, CEA is produced by other types of cancers and noncancerous conditions, such as alcoholic cirrhosis, pancreatitis, and ulcerative colitis. This marker is of greatest value for monitoring tumor recurrence in persons after resection of the primary tumor.^{11,39}

The only recognized treatment for cancer of the colon and rectum is surgical removal. Preoperative radiation therapy may be used and has in some cases demonstrated increased 5-year survival rates. Postoperative adjuvant chemotherapy may be used. Radiation therapy and chemotherapy are used as palliative treatment methods.

In summary, disorders of the small and large intestines include irritable bowel syndrome, inflammatory bowel disease, diverticular disease, disorders of motility (*i.e.*, diarrhea, constipation, infectious colitis, and intestinal obstruction), peritonitis alterations in intestinal absorption, and colorectal cancer.

Irritable bowel syndrome is a functional disorder characterized by a variable combination of chronic and recurrent intestinal symptoms not explained by structural or biochemical

492 Unit Seven: Alterations in the Gastrointestinal System

abnormalities. The term *inflammatory bowel disease* is used to designate two inflammatory conditions: Crohn's disease, which affects the small and large bowel, and ulcerative colitis, which affects the colon and rectum. Both are chronic diseases characterized by remissions and exacerbations of diarrhea, weight loss, fluid and electrolyte disorders, and systemic signs of inflammation.

Infectious forms of colitis include those caused by *C. difficile*, which is associated with antibiotic therapy, and *E. coli* O157:H7, which is found in undercooked hamburger and unpasteurized milk. Diverticular disease includes diverticulosis, which is a condition in which the mucosal layer of the colon herniates through the muscularis layer, and diverticulitis, in which there is inflammation and gross or microscopic perforation of the diverticulum.

Diarrhea and constipation represent disorders of intestinal motility. Diarrhea, characterized by excessively frequent passage of stools, can be divided into large-volume diarrhea, characterized by an increased water content in the feces, and small-volume diarrhea, associated with intrinsic bowel disease and frequent passage of small stools. Constipation can be defined as the infrequent passage of stools; it commonly is caused by failure to respond to the urge to defecate, inadequate fiber or fluid intake, weakness of the abdominal muscles, inactivity and bed rest, pregnancy, hemorrhoids, and gastrointestinal disease. Intestinal obstruction designates an impairment of movement of intestinal contents in a cephalocaudal direction as the result of mechanical or paralytic mechanisms. Peritonitis is an inflammatory response of the serous membrane that lines the abdominal cavity and covers the visceral organs. It can be caused by bacterial invasion or chemical irritation resulting from perforation of the viscera or abdominal organs.

Malabsorption results from the impaired absorption of nutrients and other dietary constituents from the intestine. It can involve a single dietary constituent, such as vitamin B_{12} , or extend to involve all of the substances absorbed in a particular part of the small intestine. Malabsorption can result from disease of the small bowel and disorders that impair digestion and in some cases obstruct the lymph flow by which fats are transported to the general circulation.

Colorectal cancer, the second most common fatal cancer, is seen most commonly in persons older than 50 years. Most, if not all, cancers of the colon and rectum arise in pre-existing adenomatous polyps. Programs that provide careful follow-up for persons with adenomatous polyps and removal of all suspect lesions have substantially reduced the incidence of colorectal cancer.

REVIEW QUESTIONS

Relate the pathophysiology of gastroesophageal reflux to measures used in diagnosis and treatment of the disorder in adults and children.

State the reason for the poor prognosis associated with esophageal cancer.

Describe the factors that contribute to the gastric mucosal barrier.

Characterize the proposed role of *Helicobacter pylori* in the development of chronic gastritis and peptic ulcer and cite methods for diagnosing the infection.

Describe the predisposing factors in the development of peptic ulcer and cite the three complications of peptic ulcer.

State the diagnostic criteria for irritable bowel syndrome.

Compare the characteristics of Crohn's disease and ulcerative colitis.

Compare the causes and manifestations of small-volume diarrhea and large-volume diarrhea.

Differentiate between mechanical and paralytic intestinal obstruction in terms of cause and manifestations.

List conditions that cause malabsorption by impaired intraluminal malabsorption, mucosal malabsorption, and lymphatic obstruction.

List the risk factors associated with colorectal cancer and cite the screening methods for detection.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- National Institute of Diabetes and Digestive and Kidney Diseases (National Digestive Diseases Clearinghouse). (1999). Digestive diseases statistics. [On-line]. Available: http://www.niddk.nih.gov/health/ digest/pubs/ddstats/ddstats.htm#all.
- 2. Spieker M.R. (2000). Evaluating dysphagia. American Family Physician 61, 3639–3648.
- 3. Mittal R.K. (1998). The spectrum of diaphragmatic hernia. *Hospital Practice* November 15, 65–79.
- Mittal R.K., Balaban D.H. (1997). The esophagogastric junction. New England Journal of Medicine 336, 924–931.
- Katzka D.A., Rustgi A.K. (2000). Gastroesophageal reflux disease and Barrett's esophagus. *Medical Clinics of North America* 84, 1137–1161.
- McQuaid K.R. (2001). Alimentary tract. In Tierney L.M., McPhee S.J., Papadakis M. (Eds.), *Current medical diagnosis and treatment 2001* (40th ed., pp. 585–589, 604–615, 618–621, 624–665). New York: Lange Medical Books/McGraw-Hill.
- 7. Alexander J.A., Hunt L.W., Patel A.M. (1999). Prevalence, pathophysiology, and treatment of patients with asthma and gastroesophageal reflux disease. *Mayo Clinic Proceedings* 75, 1055–1063.
- DeVault K.R., Castell D.G., and the Practice Parameters Committee of the American College of Gastroenterology (1999). Updated guidelines for the diagnosis and treatment of gastroesophageal reflux disease. American Journal of Gastroenterology 94(6), 1434–1442.
- 9. Mason D.B. (2000). Gastroesophageal reflux in children. Nursing Clinics of North America 35, 15–36.
- Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.). (2000). Nelson textbook of pediatrics (16th ed., pp. 1125–1126). Philadelphia: W.B. Saunders.
- Cotran R.S., Kumar V.S., Collins, T. (1999). Robbins pathologic basis of disease (6th ed., pp. 776–787, 813–814, 826–836, 927). Philadelphia: W.B. Saunders.
- 12. National Cancer Institute. (2001). Esophageal cancer. [On-line]. Available: http://www.cancernet.nci.nih.gov/cancer_Types/Esophageal_ Cancer.shtml.
- 13. Fromm D. (1987). Mechanisms involved in gastric mucosal resistance to injury. *Annual Review of Medicine* 38, 119.

Chapter 27: Alterations in Gastrointestinal Function

- 14. Wolfe M.M., Lichtenstein D.R., Singh G. (1999). Gastrointestinal toxicity of nonsteroidal antiinflammatory drugs. New England Journal of Medicine 340, 1888-1899.
- 15. Cryer B., Lee E., Feldman M. (1992). Factors influencing mucosal prostaglandin concentrations: Role of smoking and aging. Annals of Internal Medicine 116, 636-640.
- 16. Hamilton S.R., Farber J.L., Rubin E. (1999). The gastrointestinal tract. In Rubin E., Farber J.L. (Eds.), Pathology (3rd ed., pp. 688-695, 727-746). Philadelphia: Lippincott Williams & Wilkins.
- 17. Shiotani A., Nurgalieva Z.Z., Yamaoka Y., et al. (2000). Helicobacter pylori. Medical Clinics of North America 84, 1125–1136.
- 18. Soll A.H. (1996). Medical treatment of peptic ulcer disease: Practice guidelines. Journal of the American Medical Association 275, 622-629.
- 19. Fass R. (1995). Zollinger-Ellison syndrome: Diagnosis and management. Hospital Practice, November 15, 73-80.
- 20. Zuckerman G.R., Cort D., Schuman R.B. (1988). Stress ulcer syndrome. Journal of Intensive Care Medicine 3, 21.
- 21. American Cancer Society. (2003). Cancer facts & figures 2003. [Online]. Available: http://www.cancer.org.
- 22. Rothenstein R.D. (2000). Irritable bowel syndrome. Medical Clinics of North America 85, 1247-1257.
- 23. Dalton C.B., Drossman D.A. (1997). Diagnosis and treatment of irritable bowel syndrome. American Family Physician 55, 875-880, 883-885
- 24. Thompson W.G., Doteval G., Drossman D.A., et al. (1989). Irritable bowel syndrome: Guidelines for diagnosis. Gastrointestinal International 2, 92-95.
- 25. Stotland B.R., Stein R.B., Lichtenstein G.R. (2000). Advances in inflammatory bowel disease. Medical Clinics of North America 84, 1107–1123.
- 26. Botoman V.A., Bonner G.F., Botoman D.A. (1998). Management of inflammatory bowel disease. American Family Physician 57, 57-68, 71-72.
- 27. Lewis C. (1999). Crohn's disease: New drug may help when others fail. FDA Consumer Magazine. [On-line]. Available: http://www.fda. gov/fdac/features/1999/599_crohn.html.

- 28. Kelly C.P., Pothoulakis C., LaMont J.T. (1994). Clostridium difficile colitis. New England Journal of Medicine 330, 257-261.
- 29. Mylonakis E., Ryan E.T., Claderswood S.B. (2001). Clostridium difficile-associated diarrhea: A review. Archives of Internal Medicine 161, 525-533.
- 30. Greenwald D.A., Brandt L.J. (1997). Recognizing E. coli O157:H7 infection. Hospital Practice 32, 123-140.
- 31. Ferzoco L.B., Raptopoulos V., Silen W. (1998). Acute diverticulitis. New England Journal of Medicine 338, 1521-1526.
- 32. Gishan F.K. (2000). Chronic diarrhea. In Bierman R.E., Kliegman R.M., Jenson H.B. (Eds.), Nelson textbook of pediatrics (16th ed., pp. 1171-1176). Philadelphia: W.B. Saunders.
- 33. Limbos M.A., Lieberman J.M. (1995). Management of acute diarrhea in children. Contemporary Pediatrics 12 (12), 68-88.
- 34. Schiller L.R. (2000). Diarrhea. Medical Clinics of North America 84, 1259-1275.
- 35. Field M., Rao M.C., Chang E.B. (1989). Intestinal electrolyte transport and diarrheal disease (part 2). New England Journal of Medicine 321, 879-883
- 36. American Academy of Pediatrics, Subcommittee on Acute Gastroenteritis. (1996). Practice parameter: The management of acute gastroenteritis in young children. Pediatrics 97 (3), 424-435.
- 37. Wald A. (2000). Constipation. Medical Clinics of North America 84, 1231-1246
- 38. Wyllie R. (2000). Ileus, adhesions, intussusception, and closed-loop obstruction. In Bierman R.E., Kliegman R.M., Jenson H.B. (Eds.), Nelson textbook of pediatrics (16th ed., pp. 1142-1143). Philadelphia: W.B. Saunders.
- 39. American Cancer Society. (2003). Can colon and rectum cancer be found early? [On-line]. Available: http://www.cancer.org.
- 40. Marcus A.J. (1995). Aspirin as prophylaxis against colorectal cancer. New England Journal of Medicine 333, 656–657.
- 41. Bond J.H. (2000). Colorectal cancer update. Medical Clinics of North America 84, 1163-1182.

28

Alterations in Hepatobiliary Function

The Liver and Hepatobiliary System Metabolic Functions of the Liver Carbohydrate, Protein, and Lipid Metabolism Drug and Hormone Metabolism **Bile Production and Cholestasis** Cholestasis **Bilirubin Elimination Tests of Hepatobiliary Function Disorders of the Liver Hepatitis Acute Viral Hepatitis Chronic Hepatitis** Intrahepatic Biliary Disorders **Primary Biliary Cirrhosis** Primary Sclerosing Cholangitis Secondary Biliary Cirrhosis Alcohol-Induced Liver Disease Cirrhosis, Portal Hypertension, and Liver Failure Cirrhosis **Portal Hypertension** Liver Failure Cancer of the Liver **Disorders of the Gallbladder and Extrahepatic Biliary Ducts Cholelithiasis and Cholecystitis** Cholelithiasis Acute and Chronic Cholecystitis Choledocholithiasis and Cholangitis Cancer of the Gallbladder **Disorders of the Exocrine Pancreas Acute Pancreatitis Chronic Pancreatitis Cancer of the Pancreas**

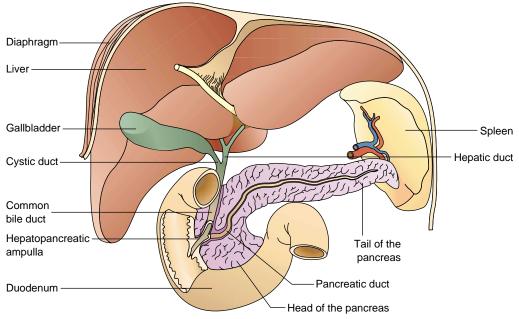
he liver, the gallbladder, and the exocrine pancreas are classified as accessory organs of the gastrointestinal tract. In addition to producing digestive secretions, the liver and the pancreas have other important functions. For example, the endocrine pancreas supplies the insulin and glucagon needed in cell metabolism, whereas the liver synthesizes glucose, plasma proteins, and blood clotting factors and is responsible for the degradation and elimination of drugs and hormones, among other functions. This chapter focuses on functions and disorders of the liver, the biliary tract and gallbladder, and the exocrine pancreas.

THE LIVER AND HEPATOBILIARY SYSTEM

The liver is the largest visceral organ in the body, weighing approximately 1.3 kg (3 lb) in the adult (Fig. 28-1). It lies below and on the right side of the diaphragm. Except for the portion that is in the epigastric area, the liver is contained within the rib cage and in healthy persons cannot normally be palpated. The liver is surrounded by a tough fibroelastic capsule called *Glisson's capsule*.

The liver is unique among the abdominal organs in having a dual blood supply—the hepatic artery and the portal vein. Approximately 300 mL of blood per minute enters the liver through the hepatic artery; another 1050 mL/minute enters by way of the valveless portal vein, which carries blood from the stomach, the small and the large intestines, the pancreas, and the spleen¹ (Fig. 28-2). Although the blood from the portal vein is incompletely saturated with oxygen, it supplies approximately 60% to 70% of the oxygen needs of the liver. The venous outflow from the liver is carried by the valveless hepatic veins, which empty into the inferior vena cava just below the level of the diaphragm. The pressure difference between the hepatic vein and the portal vein normally is such that the liver stores approximately 450 mL of blood.1 This blood can be shifted back into the general circulation during periods of hypovolemia and shock. In congestive heart failure, in which the pressure in the vena cava increases, blood backs up and accumulates in the liver.

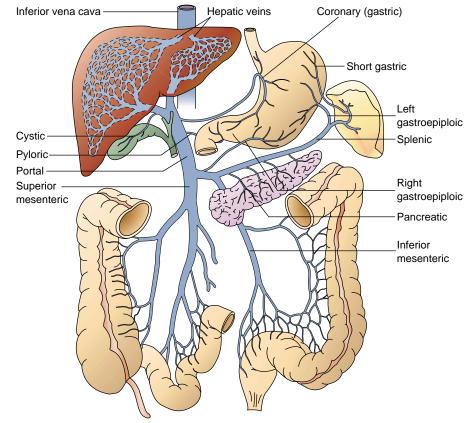
The *lobules* are the functional units of the liver. Each lobule is a cylindrical structure that measures approximately 0.8 to

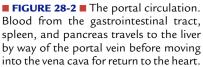


■ FIGURE 28-1 ■ The liver and biliary system, including the gallbladder and bile ducts.

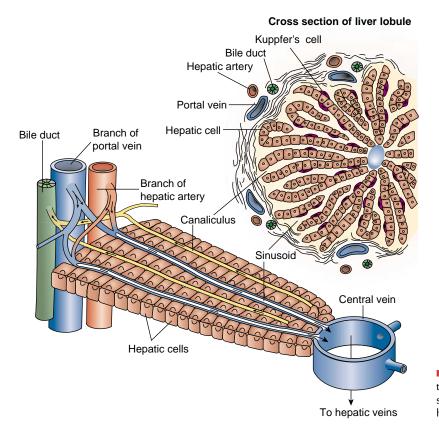
2 mm in diameter and several millimeters long. There are approximately 50,000 to 100,000 lobules in the liver.¹ Each lobule is organized around a central vein that empties into the hepatic veins and from there into the vena cava. The terminal bile ducts and small branches of the portal vein and hepatic

artery are located at the periphery of the lobule. Plates of hepatic cells radiate centrifugally from the central vein like spokes on a wheel (Fig. 28-3). These hepatic plates are separated by wide, thin-walled channels, called *sinusoids*, that extend from the periphery of the lobule to its central vein. The sinusoids are





496 Unit Seven: Alterations in the Gastrointestinal System



■ FIGURE 28-3 ■ A section of liver lobule showing the location of the hepatic veins, hepatic cells, liver sinusoids, and branches of the portal vein and hepatic artery.

supplied by blood from the portal vein and hepatic artery. Because the plates of hepatic cells are no more than two layers thick, every cell is exposed to the blood that travels through the sinusoids. Thus, the hepatic cells can remove substances from the blood or can release substances into the blood as it moves through the sinusoids.

The venous sinusoids are lined with two types of cells: endothelial cells and Kupffer's cells. *Kupffer's cells* are phagocytic cells that are capable of removing old and defective blood cells, bacteria, and other foreign material from the portal blood as it flows through the sinusoid. This phagocytic action removes the enteric bacilli and other harmful substances that filter into the blood from the intestine.

The lobules also are supplied by small tubular channels, called bile canaliculi, that lie between the cell membranes of adjacent hepatocytes. The bile produced by the hepatocytes flows into the canaliculi and then to the periphery of the lobules, which drain into progressively larger ducts, until it reaches the right and left hepatic ducts. The intrahepatic and extrahepatic bile ducts often are collectively referred to as the *hepatobiliary* tree. These ducts unite to form the common duct (see Fig. 28-1). The common duct, which is approximately 10 to 15 cm long, descends and passes behind the pancreas and enters the descending duodenum. The pancreatic duct joins the common duct at a short dilated tube called the hepatopancreatic ampulla (ampulla of Vater), which empties into the duodenum through the duodenal papilla. Muscle tissue at the junction of the papilla, called the sphincter of the bile duct, regulates the flow of bile into the duodenum. When this sphincter is closed, bile moves back into the common duct and gallbladder.

Metabolic Functions of the Liver

The liver is one of the most versatile and active organs in the body. It produces bile; metabolizes hormones and drugs; synthesizes proteins, glucose, and clotting factors; stores vitamins and minerals; changes ammonia produced by deamination of amino acids to urea; and converts fatty acids to ketones. In its capacity for metabolizing drugs and hormones, the liver serves as an excretory organ. In this respect, the bile, which carries the end-products of substances metabolized by the liver, is much like the urine, which carries the body wastes filtered by the kidneys.

Carbohydrate, Protein, and Lipid Metabolism

The liver plays an essential role in carbohydrate, fat, and proteins metabolism. It degrades excess nutrients and converts them into substances essential to the body. It builds carbohydrates from proteins, converts sugars to fats that can be stored, and interchanges chemical groups on amino acids so that they can be used for a number of purposes.

Carbohydrate Metabolism. The liver is especially important in maintaining glucose homeostasis. It stores excess glucose as glycogen and releases it into the circulation when blood glucose levels fall. The liver converts galactose and fructose to glucose and it synthesizes glucose from amino acids, glycerol, and lactic acid as a means of maintaining blood glucose during periods of fasting or increased need. The liver also converts excess carbohydrates to triglycerides for storage in adipose tissue. **Protein Synthesis and Conversion of Ammonia to Urea.** Even though the muscle contains the greatest amount of protein, the liver has the greatest rate of protein synthesis per gram of tissue. It produces the proteins for its own cellular needs and secretory proteins that are released into the circulation. The most important of these secretory proteins is albumin. Albumin contributes significantly to the plasma colloidal osmotic pressure (see Chapter 6) and to the binding and transport of numerous substances, including some hormones, fatty acids, bilirubin, and other anions. The liver also produces other important proteins, such as fibrinogen and the blood clotting factors.

Proteins are made up of amino acids. Protein synthesis and degradation involves two major reactions: transamination and deamination. In transamination, the amino group (NH₂) from an amino acid is transferred to α -ketoglutaric acid (a Krebs cycle keto acid) to form glutamic acid. The transferring amino acid becomes a keto acid and α -ketoglutaric acid becomes an amino acid (glutamic acid). The reaction is fully reversible. The process of transamination is catalyzed by aminotransferases, enzymes that are found in high amounts in the liver. Oxidative *deamination* involves the removal of an amino group from an amino acid. This occurs mainly by transamination, in which the amino group of glutamic acid is removed as ammonia, and α -ketoglutaric is regenerated. Because ammonia is very toxic to body tissues, particularly neurons, it is converted to urea in the liver and then excreted by the kidneys.² The goal of amino acid degradation is to produce molecules that can be used to produce energy or be converted to glucose.

Pathways of Lipid Metabolism. Although most body cells can metabolize fat, certain aspects of lipid metabolism occur mainly in the liver. These include the oxidation of fatty acids to supply energy for other body functions; the synthesis of large quantities of cholesterol, phospholipids, and most lipoproteins; and

KEY CONCEPTS

DISEASES OF THE LIVER

- Diseases of the liver can affect the hepatocytes or the biliary drainage system.
- Disorders of hepatocyte function impair the metabolic and synthetic functions of the liver, causing disorders in carbohydrate, protein, and fat metabolism; metabolism and removal of drugs, hormones, toxins, ammonia, and bilirubin from the blood; and the interconversion of amino acids and synthesis of proteins. Elevations in serum aminotransferase levels signal the presence of hepatocyte damage.
- Disorders of the biliary drainage system obstruct the flow of bile and interfere with the elimination of bile salts and bilirubin, producing cholestatic liver damage because of the backup of bile into the lobules of the liver. Elevations in bilirubin and alkaline phosphatase signal the presence of cholestatic liver damage.

the formation of triglycerides from carbohydrates and proteins. To derive energy from neutral fats (triglycerides), the fat must first be split into glycerol and fatty acids, and then the fatty acids split into acetyl-coenzyme A (acetyl-CoA). Acetyl-CoA can be used by the liver to produce adenosine triphosphate (ATP) or it can be converted to acetoacetic acid and released into the bloodstream and transported to other tissues, where it is used for energy. The acetyl-CoA units from fat metabolism also are used to synthesize cholesterol and bile acids. Cholesterol has several fates in the liver. It can be esterified and stored; it can be exported bound to lipoproteins; or it can be converted to bile acids.

Drug and Hormone Metabolism

By virtue of its many enzyme systems that are involved in biochemical transformations and modifications, the liver has an important role in the metabolism of many drugs and chemical substances. The liver is particularly important in terms of metabolizing lipid-soluble substances that cannot be directly excreted by the kidneys. Two major types of reactions are involved in the hepatic detoxification and metabolism of drugs and other chemicals: phase 1 reactions, which involve chemical modification or inactivation of a substance, and phase 2 reactions, which involve conversion of lipid-soluble substances to water-soluble derivatives.^{3,4} Often, the two types of reactions are linked. Many phase 1 reactants are not soluble and must therefore undergo a subsequent phase 2 reaction to be eliminated. These reactions, which are called biotransformations, are important considerations in drug therapy. Because the liver is central to metabolic disposition of virtually all drugs and foreign substances, drug-induced liver toxicity is a potential complication of many medications.

In addition to its role in metabolism of drugs and chemicals, the liver also is responsible for hormone inactivation or modification. Insulin and glucagon are inactivated by proteolysis or deamination. Thyroxine and triiodothyronine are metabolized by reactions involving deiodination. Steroid hormones such as the glucocorticoids are first inactivated by a phase 1 reaction and then converted to a more water-soluble product by a phase 2 reaction.

Bile Production and Cholestasis

The secretion of bile, approximately 600 to 1200 mL daily, is one of the many functions of the liver.¹ Bile functions in the digestion and absorption of fats and fat-soluble vitamins from the intestine, and it serves as a vehicle for excretion of bilirubin, excess cholesterol, and metabolic end-products that cannot be eliminated in the urine.

Bile contains water, electrolytes, bile salts, bilirubin, cholesterol, and certain products of organic metabolism. Of these, only the bile salts, which are formed from cholesterol, are important in digestion. Bile salts aid in emulsifying dietary fats, and they are necessary for the formation of the micelles that transport fatty acids and fat-soluble vitamins to the surface of the intestinal mucosa for absorption. Approximately 94% of bile salts that enter the intestine are reabsorbed into the portal circulation by an active transport process that takes place in the distal ileum. From the portal circulation, the bile salts pass into the liver, where they are recycled. Normally, bile salts travel this entire circuit approximately 18 times before being expelled in the feces.¹ This system for recirculation of bile is called the *enterohepatic circulation*.

Cholestasis

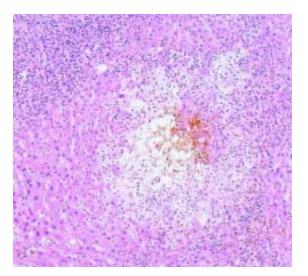
Cholestasis represents a decrease in bile flow through the intrahepatic canaliculi and a reduction in secretion of water, bilirubin, and bile acids by the hepatocytes. As a result, the materials normally transferred to the bile, including bilirubin, cholesterol, and bile acids, accumulate in the blood.^{5,6} The condition may be caused by intrinsic liver disease, in which case it is referred to as *intrahepatic cholestasis*, or by obstruction of the larger bile ducts located outside the liver, a condition known as *extrahepatic cholestasis*.

Common to all types of obstructive and hepatocellular cholestasis is the accumulation of bile pigment within the bile canaliculi and hepatocytes. Prolonged obstructive cholestasis leads not only to fatty changes in the hepatocytes but to destruction of the supporting connective tissue, giving rise to bile lakes filled with cellular debris and pigment (Fig. 28-4).⁵ Unrelieved obstruction leads to portal tract fibrosis and ultimately to end-stage biliary cirrhosis.

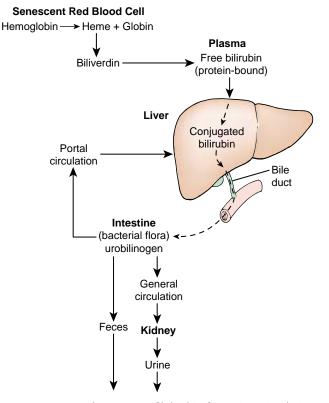
Pruritus is the most common presenting symptom in persons with cholestasis, probably related to an elevation in plasma bile acids. Skin xanthomas (focal accumulations of cholesterol) may occur, the result of hyperlipidemia and impaired excretion of cholesterol. A characteristic laboratory finding is an elevated serum alkaline phosphatase level. Alkaline phosphatase is present in the membranes between liver cells and the bile duct and is released by disorders affecting the bile duct. Other manifestations of reduced bile flow relate to intestinal absorption, including nutritional deficiencies of the fat-soluble vitamins A, D, and K.

Bilirubin Elimination

Bilirubin is the substance that gives bile its color. It is formed during the breakdown of senescent red blood cells. In the process of degradation, the heme portion of the hemoglobin mol-



■ FIGURE 28-4 ■ Bile infarct (bile lake). A photomicrograph of the liver in a patient with extrahepatic biliary obstruction shows an area of necrosis and the accumulation of extravasated bile (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 768]. Philadelphia: Lippincott Williams & Wilkins)



■ FIGURE 28-5 ■ The process of bilirubin formation, circulation, and elimination.

ecule is oxidized to form biliverdin, which is then converted to free bilirubin (Fig. 28-5). Free bilirubin, which is insoluble in plasma, is transported in the blood attached to plasma albumin. Even when it is bound to albumin, this bilirubin is still called free bilirubin. As it passes through the liver, free bilirubin is released from the albumin carrier molecule and moved into the hepatocytes. Inside the hepatocytes, free bilirubin is converted to conjugated bilirubin, making it soluble in bile. Conjugated bilirubin is secreted as a constituent of bile, and in this form it passes through the bile ducts into the small intestine. In the intestine, approximately one half of the bilirubin is converted into a highly soluble substance called *urobilinogen* by the intestinal flora. Urobilinogen is either absorbed into the portal circulation or excreted in the feces. Most of the urobilinogen that is absorbed is returned to the liver to be re-excreted into the bile. A small amount of urobilinogen, approximately 5%, is absorbed into the general circulation and then excreted by the kidneys.

Usually, only a small amount of bilirubin (0.1 to 1.2 mg/dL) is found in the blood. Laboratory measurements of bilirubin usually measure the free and the conjugated bilirubin as well as the total bilirubin. These are reported as the direct (conjugated) bilirubin and the indirect (unconjugated or free) bilirubin.

Jaundice. Jaundice (*i.e.*, icterus), which results from an abnormally high accumulation of bilirubin in the blood, is a yellowish discoloration to the skin and deep tissues. Jaundice becomes evident when the serum bilirubin levels rise above 2.0 to 2.5 mg/dL.^{5,6} Because normal skin has a yellow cast, the early signs of jaundice often are difficult to detect, especially in per-

Chapter 28: Alterations in Hepatobiliary Function 49



■ FIGURE 28-6 ■ Jaundice. A patient with hepatic failure displays a yellow sclera. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 762]. Philadelphia: Lippincott Williams & Wilkins)

sons with dark skin. Bilirubin has a special affinity for elastic tissue. The sclera of the eye, which contains considerable elastic fibers, usually is one of the first structures in which jaundice can be detected (Fig 28-6).

The four major causes of jaundice are excessive destruction of red blood cells, impaired uptake of bilirubin by the liver cells, decreased conjugation of bilirubin, and obstruction of bile flow in the canaliculi of the hepatic lobules or in the intrahepatic or extrahepatic bile ducts. From an anatomic standpoint, jaundice can be categorized as prehepatic, intrahepatic, and posthepatic. Chart 28-1 lists the common causes of prehepatic, hepatic, and posthepatic jaundice.

CHART 28-1 Causes of Jaundice

Prehepatic (Excessive Red Blood Cell Destruction)

Hemolytic blood transfusion reaction Hereditary disorders of the red blood cell Sickle cell anemia Thalassemia Spherocytosis Acquired hemolytic disorders Hemolytic disease of the newborn

Autoimmune hemolytic anemias

Intrahepatic

Decreased bilirubin uptake by the liver Decreased conjugation of bilirubin Hepatocellular liver damage Hepatitis Cirrhosis Cancer of the liver Drug-induced cholestasis

Posthepatic (Obstruction of Bile Flow)

Structural disorders of the bile duct Cholelithiasis Congenital atresia of the extrahepatic bile ducts Bile duct obstruction caused by tumors The major cause of prehepatic jaundice is excessive hemolysis of red blood cells. Hemolytic jaundice occurs when red blood cells are destroyed at a rate in excess of the liver's ability to remove the bilirubin from the blood. It may follow a hemolytic blood transfusion reaction or may occur in diseases such as hereditary spherocytosis, in which the red cell membranes are defective, or in hemolytic disease of the newborn (see Chapter 13). Neonatal hyperbilirubinemia results in an increased production of bilirubin in newborn infants and their limited ability to excrete it.⁷ Premature infants are at particular risk because their red cells have a shorter life span and higher turnover rate. In prehepatic jaundice, there is mild jaundice, the unconjugated bilirubin is elevated, the stools are of normal color, and there is no bilirubin in the urine.

Intrahepatic or hepatocellular jaundice is caused by disorders that directly affect the ability of the liver to remove bilirubin from the blood or conjugate it so it can be eliminated in the bile. Liver diseases such as hepatitis and cirrhosis are the most common causes of intrahepatic jaundice. Intrahepatic jaundice usually interferes with all phases of bilirubin metabolism—uptake, conjugation, and excretion. Both conjugated and unconjugated bilirubin are elevated and the urine often is dark because of the presence of bilirubin.

Posthepatic or obstructive jaundice, also called *cholestatic jaundice*, occurs when bile flow is obstructed between the liver and the intestine. Among the causes of posthepatic jaundice are strictures of the bile duct, gallstones, and tumors of the bile duct or the pancreas. Conjugated bilirubin levels usually are elevated; the stools are clay colored because of the lack of bilirubin in the bile; the urine is dark; the levels of serum alkaline phosphatase are markedly elevated; and the aminotransferase levels are slightly increased. Blood levels of bile acids often are elevated in obstructive jaundice. As the bile acids accumulate in the blood, pruritus develops. A history of pruritus preceding jaundice is common in obstructive jaundice.

Tests of Hepatobiliary Function

The history and physical examination, in most instances, provide clues about liver function. Diagnostic tests help to evaluate liver function and the extent of liver damage. Laboratory tests commonly are used to assess liver function and confirm the diagnosis of liver disease.

Liver function tests, including serum levels of liver enzymes, are used to assess injury to liver cells, the liver's ability to synthesize proteins, and the excretory functions of the liver.^{8,9} Elevated serum enzyme tests usually indicate liver injury earlier than do other indicators of liver function. The key enzymes are alanine aminotransferase (ALT) and aspartate aminotransferase (AST), which are present in liver cells. ALT is liver specific, whereas AST is derived from organs other than the liver. In most cases of liver damage, there are parallel increases in ALT and AST. The most dramatic rise is seen in cases of acute hepatocellular injury, such as occurs with viral hepatitis, hypoxic or ischemic injury, and acute toxic injury.

The liver's synthetic capacity is reflected in measures of serum protein levels and prothrombin time (*i.e.*, synthesis of coagulation factors). Hypoalbuminemia caused by depressed synthesis may complicate severe liver disease. Deficiencies of coagulation factor V and vitamin K-dependent factors (II, VII, IX, and X) may occur.

500 Unit Seven: Alterations in the Gastrointestinal System

Serum bilirubin, γ -glutamyltransferase (GGT), and alkaline phosphatase measure hepatic excretory function. Alkaline phosphatase is present in the membranes of cells that line the bile duct and is released by disorders affecting the bile duct.⁸ GGT is thought to function in the transport of amino acids and peptides into liver cells; it is a sensitive indicator of hepatobiliary disease. Measurement of GGT may be helpful in diagnosing alcohol abuse.⁹

Ultrasonography and computed tomography (CT) scanning provide information about the size, structure, and composition of the liver. Magnetic resonance imaging (MRI) has proved to be useful in some disorders. Selective angiography of the celiac, superior mesenteric, or hepatic artery may be used to visualize the hepatic or portal circulation. A liver biopsy affords a means of examining liver tissue without surgery.

In summary, the hepatobiliary system consists of the liver, gallbladder, and bile ducts. The liver is the largest and, in functions, one of the most versatile organs in the body. It is located between the gastrointestinal tract and the systemic circulation; venous blood from the intestine flows through the liver before it is returned to the heart. In this way, nutrients can be removed for processing and storage, and bacteria and other foreign matter can be removed by Kupffer's cells before the blood is returned to the systemic circulation.

The liver synthesizes fats, glucose, and plasma proteins. Other important functions of the liver include the deamination of amino acids, conversion of ammonia to urea, and the interconversion of amino acids and other compounds that are important to the metabolic processes of the body. The liver produces approximately 600 to 1200 mL of bile daily. Bile contains bile salts that are essential for digestion of fats and absorption of fat-soluble vitamins, and it serves as an excretory vehicle for bilirubin, cholesterol, and certain products of organic metabolism. Cholestasis represents a decrease in bile flow through the intrahepatic bile ducts. It results in destructive liver changes and an accumulation of bile components in the blood. The liver also removes, conjugates, and secretes bilirubin into the bile. Jaundice occurs when bilirubin accumulates in the blood. It can occur because of excessive red blood cell destruction, failure of the liver to remove and conjugate the bilirubin, or obstructed biliary flow.

Liver function tests, including serum aminotransferase levels, are used to assess injury to liver cells. Serum bilirubin, GGT, and alkaline phosphatase measure hepatic excretory function. Ultrasonography, CT scans, and MRI are used to evaluate liver structures. Angiography may be used to visualize the hepatic or portal circulation, and a liver biopsy may be used to obtain tissue specimens for microscopic examination.

DISORDERS OF THE LIVER

The liver is subject to many of the same pathologic conditions that affect other body organs: infection, inflammation, and immune responses; metabolic disorders; and neoplasms. This section focuses on alterations in liver function caused by viral and autoimmune hepatitis; intrahepatic biliary tract disorders; alcohol-induced liver disease; cirrhosis, portal hypertension, and liver failure; and cancer of the liver.

Hepatitis

Hepatitis refers to inflammation of the liver. It can be caused by reactions to chemical agents, drugs, and toxins; disorders such as autoimmune diseases and infectious mononucleosis that cause secondary hepatitis; and by hepatotropic viruses that primarily affect liver cells or hepatocytes.

Acute Viral Hepatitis

The known hepatotropic viruses include hepatitis A virus (HAV), hepatitis B virus (HBV), the hepatitis B-associated delta virus (HDV), hepatitis C virus (HCV), and hepatitis E virus (HEV). Although all of these viruses cause acute hepatitis, they differ in the mode of transmission and incubation period; mechanism, degree, and chronicity of liver damage; and ability to evolve to a carrier state.

There are two mechanisms of liver injury in viral hepatitis: direct cellular injury and immune responses against viral antigens in infected hepatocytes. The immune-mediated mechanisms of injury have been most closely studied in HBV. It is thought that the extent of inflammation and necrosis depends on the person's immune response. Accordingly, a prompt immune response during the acute phase of the infection would be expected to cause cell injury but at the same time eliminate the virus. Thus, people who respond with fewer symptoms and a marginal immune response are less likely to eliminate the virus, and hepatocytes expressing the viral antigens persist, leading to the chronic or carrier state. Fulminant hepatitis would be explained in terms of an accelerated immune response with severe liver necrosis.

The clinical course of viral hepatitis involves a number of syndromes, including asymptomatic infection with only sero-logic evidence of disease; acute symptomatic hepatitis; the carrier state without clinically apparent disease or with chronic hepatitis; chronic hepatitis with or without progression to cirrhosis; or fulminating disease (>1% to 3%) with rapid onset of liver failure. Not all hepatotropic viruses provoke each of the clinical syndromes.

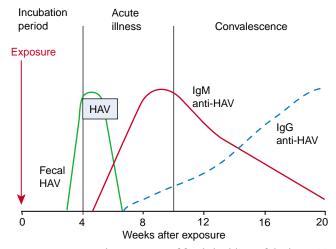
The manifestations of acute hepatitis can be divided into three phases: the prodromal or preicterus period, the icterus period, and the convalescent period. The *prodromal period* is marked by nonspecific symptoms, which vary from abrupt to insidious. There are usually complaints of malaise, easy fatigability, nausea, and loss of appetite. Weight loss, low-grade fever, headaches, muscle aches and pains, vomiting, and diarrhea are less constant symptoms. In some persons, the nonspecific symptoms are more severe with higher fever, chills, and headache, sometimes accompanied by right upper quadrant abdominal pain and liver enlargement and tenderness. Serum levels of AST and ALT show variable increases during the preicterus phase and precede a rise in serum bilirubin that accompanies the onset of the icterus or jaundice phase of infection. The icterus or jaundice phase, if it occurs, usually follows the prodromal phase by 5 to 10 days. Jaundice is less likely to occur with HCV infection. The symptoms may become worse with the onset of jaundice, followed by progressive clinical improvement. Severe pruritus and liver tenderness are common during the icterus period. The *convalescent phase* is characterized by an increased sense of well-being, return of appetite, and disappearance of jaundice. The acute illness usually subsides gradually during a 2- to 3-week period, with complete clinical recovery by approximately 9 weeks in hepatitis A and 16 weeks in uncomplicated hepatitis B.

Infection with HBV and HCV can produce a *carrier state*, in which the person does not have symptoms but harbors the virus and can transmit the disease. Evidence also indicates a carrier state for HDV infection. There is no carrier state for HAV infection. There are two types of carriers: healthy carriers who have few or no ill effects, and those with chronic disease who may or may not have symptoms. Factors that increase the risk of becoming a carrier are age at time of infection and immune status. Persons at high risk for becoming carriers are infants of HBV-infected mothers, persons with impaired immunity, those who have received multiple transfusions or blood products, those who are on hemodialysis, and drug addicts.

Hepatitis A. Hepatitis A, formerly called *infectious hepatitis*, is caused by the small, unenveloped, RNA-containing HAV. It usually is a benign, self-limited disease, although in rare cases it can cause acute fulminant hepatitis and liver failure, leading to death. The onset of symptoms usually is abrupt and includes fever, malaise, nausea, anorexia, abdominal discomfort, dark urine, and jaundice. The likelihood of having symptoms is related to age.¹⁰ Children younger than 6 years often are asymptomatic. The illness in older children and adults usually is symptomatic, and jaundice occurs in approximately 90% of cases. Symptoms usually last approximately 2 months but can last longer. HAV does not cause chronic hepatitis or induce a carrier state.

Hepatitis A has a brief incubation period (15 to 45 days) and usually is transmitted by the fecal-oral route.^{5,6,11} The virus replicates in the liver, is excreted in the bile, and shed in the stool. The fecal shedding of HAV occurs as much as 2 weeks before the development of symptoms and ends as the immunoglobulin M (IgM) levels rise.⁵ The disease often occurs sporadically or in epidemics. Drinking contaminated milk or water and eating shellfish from infected waters are fairly common routes of transmission. At special risk are persons traveling abroad who have not previously been exposed to the virus. Because young children are asymptomatic, they play an important role in the spread of the disease. Institutions housing large numbers of persons (usually children) sometimes are stricken with an epidemic of hepatitis A. Oral behavior and lack of toilet training promote viral infection among children attending preschool day care centers, who then carry the virus home to older siblings and parents. Hepatitis A usually is not transmitted by transfusion of blood or plasma derivatives, presumably because its short period of viremia usually coincides with clinical illness, so that the disease is apparent and blood donations are not accepted.

Antibodies to HAV (anti-HAV) appear early in the disease and tend to persist in the serum (Fig. 28-7). The IgM antibodies (see Chapter 8) usually appear during the first week of symptomatic disease and begin to decline in a few months. Their presence coincides with a decline in fecal shedding of the virus. Peak levels of IgG antibodies occur after 1 month of illness and may persist for years; they provide long-term protective immunity against reinfection. The presence of IgM anti-



■ FIGURE 28-7 ■ The sequence of fecal shedding of the hepatitis A virus (HAV), HAV viremia, and HAV antibody (IgM and IgG anti-HAV) changes in hepatitis A.

HAV is indicative of acute hepatitis A, whereas IgG anti-HAV merely documents past exposure.

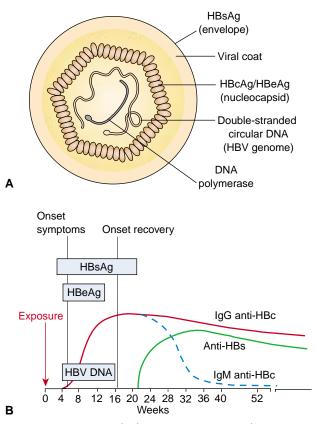
Two commercially prepared hepatitis A vaccines are available; vaccination is recommended for persons at risk for HAV exposure. It is recommended for international travelers to regions where sanitation is poor and endemic HAV infections are high, children living in communities with high rates of HAV infection, homosexually active men, and users of illicit drugs.¹⁰ A public health benefit also may be derived from vaccinating persons with increased potential for transmitting the disease (*e.g.*, food handlers). An immune globulin is available for persons with known HAV exposure who have not been immunized.

Hepatitis B. The Centers for Disease Control (CDC) estimates that there are 200,000 to 300,000 new cases of hepatitis B each year and 1 to 1.25 million chronic carriers in the United States.¹² The CDC also estimates that each year in the United States there are 4000 to 5000 deaths associated with hepatitis B-related cirrhosis and hepatocellular carcinoma. These figures are dwarfed by a much higher frequency of hepatitis B on a global scale. For example, the infection is endemic in regions of Africa and Southeast Asia.

Hepatitis B is caused by a double-stranded DNA virus (HBV).^{5,6,13,14} The complete virion, also called a *Dane particle*, consists of an outer envelope and an inner core that contains HBV DNA and DNA polymerase (Fig. 28-8). Hepatitis B has a longer incubation period and represents a more serious health problem than does hepatitis A. It can produce acute hepatitis, chronic hepatitis, progression of chronic hepatitis to cirrhosis, fulminant hepatitis with massive hepatic necrosis, and the carrier state. It also participates in the development of hepatitis D (delta hepatitis).

The HBV usually is transmitted through inoculation with infected blood or serum. However, the viral antigen can be found in most body secretions and can be spread by oral or sexual contact. In the United States, most persons with hepatitis B acquire the infection as adults or adolescents. The disease is highly prevalent among injecting drug users, persons with multiple sex

502 Unit Seven: Alterations in the Gastrointestinal System



■ FIGURE 28-8 ■ (A) The hepatitis B virus. (B) The sequence of hepatitis B virus (HBV) viral antigens (HBsAg, HBeAg), HBV DNA, and HBV antibody (IgM, IgG, anti-HBc, and anti-HBs) changes in acute resolving hepatitis B.

partners, and men who have sex with men.¹⁵ Health care workers are at risk because of blood exposure and accidental needle injuries. Although the virus can be spread through transfusion or administration of blood products, routine screening methods have appreciably reduced transmission through this route. The risk that infants born to infected mothers will have hepatitis B ranges from 10% to 85%, depending on the mother's HBV core antigen (HBeAg) status. Infants who become infected have a 90% risk of becoming chronic carriers, and as many as 25% will die of chronic liver disease as adults.¹²

Three well-defined antigens are associated with the virus: two nucleocapsid "core" antigens, HBcAg and HBeAg, and a third, HbsAg, surface antigen (see Fig. 28-8). These HBV antigens evoke specific antibodies: anti-HBs, anti-HBc, and anti-HBe. These antigens and their antibodies serve as serologic markers for following the course of the disease.

The *HBsAg* is the viral antigen measured most routinely in blood. It is produced in abundance by infected liver cells and released into the serum. HBsAg is the earliest serologic marker to appear; it appears before the onset of symptoms and is an indicator of acute or chronic infection. The HBsAg level begins to decline after the onset of the illness and usually is undetectable in 3 to 6 months. Persistence beyond 6 months indicates continued viral replication, infectivity, and the risk of chronic hepatitis. *Anti-HBs*, a specific antibody to HBsAg, occurs in most individuals after clearance of HBsAg and after successful immunization for hepatitis B. There often is a delay in appearance of anti-HBs after clearance of HBsAg. During this period of serologic gap, called the *window period*, infectivity has been demonstrated. Development of anti-HBs signals recovery from HBV infection, noninfectivity, and protection from future HBV infection.

The *HBeAg* is thought to be a cleavage product of the viral core antigen; it may be found in the serum and is an active marker for the disease and shedding of complete virions into the bloodstream. It appears during the incubation period, shortly after the appearance of HBsAg, and is found only in the presence of HBsAg. HBeAg usually disappears before HBsAg. The antibody to HBeAg, *anti-HBe*, begins to appear in the serum at about the time that HBeAg disappears, and its appearance signals the onset of resolution of the acute illness. The clinical usefulness of the antigen and its antibody lies in their predictive value as markers for infectivity.

The *HBcAg* does not circulate in the blood; therefore, it is not a useful marker for the disease. Although the antigen is not found in the blood, its antibodies (anti-HBc) are the first to be detected. They appear toward the end of the incubation period and persist during the acute illness and for several months to years after that. The initial HBcAg antibody is IgM; it serves as a marker for recent infection and is followed in 6 to 18 months by IgG antibodies. These antibodies are not protective and are detectable in the presence of chronic disease.

The presence of viral DNA (HBV DNA) in the serum is the most certain indicator of hepatitis B infection. It is transiently present during the presymptomatic period and for a brief time during the acute illness. The presence of DNA polymerase, the enzyme used in viral replication, usually is transient but may persist for years in persons who are chronic carriers and is an indication of continued infectivity.

Hepatitis B vaccine provides long-term protection against HBV infection.¹⁶ Vaccination is recommended for all children ages 0 to 18 years as a means of preventing HBV transmission.¹⁵ The vaccine also is recommended for all persons who are at high risk for exposure to the virus. All pregnant women should be tested for HBsAg during an early prenatal visit, and infants born to HBsAg-positive mothers should receive appropriate doses of hepatitis immune globulin and hepatitis B vaccine.¹²

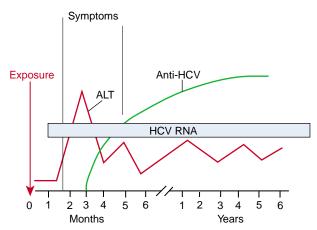
Hepatitis **C**. Hepatitis C is caused by a single-stranded RNA virus (HCV) that is distantly related to the viruses that cause yellow fever and dengue fever. There are at least six genotypes of the hepatitis C virus and multiple subtypes of the virus.¹⁷ It is likely that the wide diversity of genotypes contributes to the pathogenicity of the virus, allowing it to escape the actions of host immune mechanisms and antiviral medications, and to the difficulties in developing a preventative vaccine.^{18,19}

Hepatitis C is the most common cause of chronic hepatitis, cirrhosis, and hepatocellular cancer in the world.^{17,18} An estimated 2.7 million people in the United States have active HCV infection.¹⁷ Most of these people are chronically infected and unaware of their infection because they are not clinically ill. Infected persons serve as a source of infection to others and are at risk for chronic liver disease during the first two or more decades after initial infection.

Before 1990, the main route of transmission was through contaminated blood transfusions or blood products. With implementation of HCV testing in blood banks, the risk of HCV infection from blood transfusion is less than 1 in 103,000.¹⁴ Currently, injecting drug use is thought to be the single most important risk factor for HCV infection. There also is concern that transmission of small amounts of blood during tattooing, acupuncture, and body piercing may facilitate the transmission of HCV.19 The incidence of sexual and vertical transmission from mother to child is uncertain. Occupational exposure through incidents such as unintentional needle sticks can result in infection. However, the prevalence of HCV among health care, emergency medical, and public safety workers who are exposed to blood in the workplace is reported to be no greater than in the general public.^{17,18} Sporadic cases of hepatitis C of unknown source account for approximately 40% of cases.

The incubation period for HCV infection ranges from 15 to 150 days (average, 50 days). Clinical symptoms with acute hepatitis C tend to be milder than those seen in persons with other types of viral hepatitis. Children and adults who acquire the infection usually are asymptomatic, or have a nonspecific clinical disease characterized by fatigue, malaise, anorexia, and weight loss. Jaundice is uncommon, and only 25% to 30% of symptomatic adults have jaundice.²⁰ Unlike hepatitis A and B viral infections, fulminant hepatic failure is rare, and only a few cases have been reported. The most alarming aspects of HCV infection are its high rate of persistence and ability to induce chronic hepatitis and cirrhosis. HCV also increases the risk for development of hepatocellular cancer.

Both antibody and viral tests are available for detecting the presence of hepatitis C infection (Fig 28-9). Antibody testing has the advantage of being readily available and having a relatively low cost. False-negative results can occur in immunocompromised people and early in the course of the disease before antibodies develop. Direct measurement of HCV in the serum remains the most accurate test for infection. The viral tests are highly sensitive and specific but more costly than antibody tests. With newer antibody testing methods, infection often can be detected as early as 6 to 8 weeks after exposure, and as early as 1 to 2 weeks with viral tests that use the polymerase chain reaction (PCR) testing methods. Unlike hepatitis A and B, antibodies to HCV are not protective, but they serve as markers for the disease. At present, there is no vaccine that protects against HCV infection.



■ FIGURE 28-9 ■ The sequence of serologic changes in chronic hepatitis C with persistence of hepatitis C virus (HCV) RNA and exacerbations and remissions of clinical symptoms designated by changes in serum alanine aminotransferase (ALT) levels.

Hepatitis D. Also called the hepatitis delta virus, HDV is a defective RNA virus that can cause acute or chronic hepatitis. Infection depends on concomitant infection with hepatitis B, specifically the presence of HBsAg. Acute hepatitis D occurs in two forms: coinfection that occurs simultaneously with acute hepatitis B, and as a super-infection in which hepatitis D is imposed on chronic hepatitis B or hepatitis B carrier state.²¹ The HDV often increases the severity of HBV infection. It can convert mild HBV infection into severe, fulminating hepatitis, cause acute hepatitis in asymptomatic carriers, or increase the tendency for progression to chronic hepatitis and cirrhosis.

The routes of transmission of HDV are similar to those for HBV. In the United States, infection is restricted largely to persons at high risk for HBV infection, particularly injecting drug users and persons receiving clotting factor concentrates. The greatest risk is in HBV carriers; these persons should be informed about the dangers of HDV superinfection. Hepatitis D is diagnosed by detection of antibody to HDV (anti-HDV) in the serum or HDV RNA in the serum. There is no specific treatment for hepatitis D. Because the infection is linked to hepatitis B, prevention of hepatitis D should begin with prevention of hepatitis B through vaccination.

Hepatitis E. The hepatitis E virus is an unenveloped, singlestranded RNA virus. It is transmitted by the fecal-oral route and causes manifestations of acute hepatitis that are similar to hepatitis A. It does not cause chronic hepatitis or the carrier state. Its distinguishing feature is the high mortality rate (approximately 20%) among pregnant women attributable to the development of fulminant hepatitis. The infection occurs primarily in developing areas, such as India, other Southeast Asian countries, parts of Africa, and Mexico. The only reported cases in the United States have been in persons who have recently been in an endemic area.

Chronic Hepatitis

Chronic hepatitis is defined as inflammatory reaction of the liver of more than 3 to 6 months' duration. It is characterized by persistently elevated serum aminotransferase levels and characteristic histologic findings on liver biopsy. The causes of chronic hepatitis include HBV, HCV, HDV, autoimmune hepatitis, and hepatitis associated with certain medications.²²

Chronic Viral Hepatitis. Chronic viral hepatitis is the principal cause of chronic liver disease, cirrhosis, and hepatocellular cancer in the world and now ranks as the chief reason for liver transplantation in adults.²³

The clinical features of chronic viral hepatitis are highly variable and not predictive of outcome. The most common symptoms are fatigue, malaise, loss of appetite, and occasional bouts of jaundice. Elevation of serum aminotransferase levels depends on the degree of disease activity.

Chronic hepatitis B accounts for 5% to 10% of chronic liver disease and cirrhosis in the United States.²³ Hepatitis B is less likely than hepatitis C to progress to chronic infection. Chronic hepatitis B is characterized by the persistence of HBV DNA and usually by HBeAg in the serum, indicating active viral replication. Many persons are asymptomatic at the time of diagnosis, and elevated serum aminotransferase levels are the first sign of infection. Chronic hepatitis D depends on concurrent infection with HBV.

Chronic hepatitis C accounts for most cases of chronic viral hepatitis. HCV infection becomes chronic in 75% to 80% of cases.¹⁸ Chronic HCV infection often smolders during a period of years, silently destroying liver cells. Most persons with chronic hepatitis C are asymptomatic, and diagnosis usually follows a finding of elevated serum aminotransferase levels, a tender liver, or complaints of fatigue or nonspecific weakness. Because the course of acute hepatitis C often is mild, many persons do not recall the events of the acute infection.

There are no simple and effective treatment methods for chronic viral hepatitis.^{23,24} Persons with chronic hepatitis B who have evidence of active viral replication may be treated with a course of recombinant interferon α -2b. The antiviral drug lamivudine may be used as a substitute for interferon α .²³ Liver transplantation is a treatment option for end-stage liver disease caused by viral hepatitis. Liver transplantation has been more successful in persons with hepatitis C than in those with hepatitis B. Although the graft often is reinfected, the disease seems to progress more slowly.

Autoimmune Hepatitis. Chronic autoimmune hepatitis is a chronic inflammatory liver disease of unknown origin, but it is associated with circulating autoantibodies and high serum gamma globulin levels. Autoimmune hepatitis accounts for only approximately 10% of chronic hepatitis cases in the United States, a decrease from previously reported rates that probably reflects not a true change in incidence but better methods of detecting viral pathogens. The pathogenesis of the disorder is one of a genetically predisposed person exposed to an environmental agent that triggers an autoimmune response directed at liver cell antigens.²⁵ The resulting immune response produces a necrotizing inflammatory response that eventually leads to destruction of liver cells and development of cirrhosis. The factors surrounding the genetic predisposition and the triggering events that lead to the autoimmune response are unclear. Autoimmune hepatitis is mainly a disease of young women, although it can occur at any age and in men or women.

Clinical manifestations of the disorder cover a spectrum that extends from no apparent symptoms to the signs accompanying liver failure. In asymptomatic cases, the disorder may be discovered when abnormal serum enzyme levels are discovered during performance of routine screening tests.

The differential diagnosis includes measures to exclude other causes of liver disease, including hepatitis B and C. A characteristic laboratory finding is that of a marked elevation in serum gamma globulins. A biopsy is used to confirm the diagnosis. Corticosteroid drugs and immunosuppressant drugs are the treatment of choice for this type of hepatitis. Liver transplantation may be the only treatment for end-stage disease.

Intrahepatic Biliary Disorders

Intrahepatic biliary diseases disrupt the flow of bile through the liver, causing cholestasis and biliary cirrhosis. Among the causes of intrahepatic biliary disease are primary biliary cirrhosis, primary sclerosing cholangitis, and secondary biliary cirrhosis.

Primary Biliary Cirrhosis

Primary biliary cirrhosis involves inflammation and scarring of small intrahepatic bile ducts, portal inflammation, and progressive scarring of liver tissue.²⁶ The disease is seen most com-

monly in women 30 to 65 years of age and accounts for 2% to 5% of cases of cirrhosis. Familial occurrences of the disease are found between parents and children and among siblings. Abnormalities of cell-mediated and humoral immunity suggest an autoimmune mechanism. Antimitochondrial antibodies are found in 98% of persons with the disease, but their role in the pathogenesis of the disease is unclear.^{26,27} As many as 84% of persons with primary biliary cirrhosis have at least one other autoimmune disorder, such as scleroderma, Hashimoto's thyroiditis, rheumatoid arthritis, or the sicca complex of dry eyes and mouth (Sjögren's syndrome).

The disorder is characterized by an insidious onset and progressive scarring and destruction of liver tissue. The liver becomes enlarged and takes on a green hue because of the accumulated bile. The earliest symptoms are unexplained pruritus or itching, weight loss, and fatigue, followed by dark urine and pale stools. Jaundice is a late manifestation of the disorder, as are other signs of liver failure. Serum alkaline phosphatase levels are elevated in persons with primary biliary cirrhosis.

Treatment is largely symptomatic. Liver transplantation remains the only treatment for advanced disease.

Primary Sclerosing Cholangitis

Primary sclerosing cholangitis is a chronic cholestatic disease of unknown origin that causes destruction and fibrosis of intrahepatic and extrahepatic bile ducts.²⁸ Bile flow is obstructed (*i.e.*, cholestasis), and the bile retention destroys hepatic structures. The disease commonly is associated with inflammatory bowel disease, occurs more often in men than women, and is seen most commonly in the third to fifth decades of life. Primary sclerosing cholangitis, although much less common than alcoholic cirrhosis, is the fourth leading indication for liver transplantation in adults in the United States.²⁸

Most persons with the disorder are initially asymptomatic, with the disorder being detected during routine liver function tests that reveal elevated levels of serum alkaline phosphatase or GGT. Alternatively, some persons may present with progressive fatigue, jaundice, and pruritus. The later stages of the disease are characterized by cirrhosis, portal hypertension, and liver failure. Ten-year survival rates range from 50% to 75%.²⁸ Other than measures aimed at symptom relief, the only treatment is liver transplantation.

Secondary Biliary Cirrhosis

Secondary biliary cirrhosis results from prolonged obstruction of the extrabiliary tree. The most common cause is cholelithiasis. Other causes of secondary biliary cirrhosis are malignant neoplasms of the biliary tree or head of the pancreas and strictures of the common duct caused by previous surgical procedures. Extrahepatic biliary cirrhosis may benefit from surgical procedures designed to relieve the obstruction.

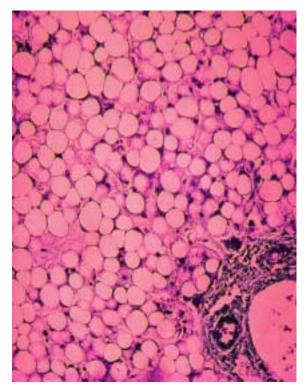
Alcohol-Induced Liver Disease

The spectrum of alcoholic liver disease includes fatty liver disease, alcoholic hepatitis, and cirrhosis. Alcoholic cirrhosis causes 200,000 deaths annually and is the fifth leading cause of death in the United States.⁵ Most deaths associated with alcoholic cirrhosis are attributable to liver failure, bleeding esophageal varices, or kidney failure. It has been estimated that there are 10 million alcoholics in the United States. However, only approximately 10% to 15% of alcoholics have cirrhosis, suggesting that other conditions such as genetic and environmental factors contribute to its occurrence.

Although the mechanism by which alcohol exerts its toxic effects on liver structures is somewhat uncertain, the changes that develop can be divided into three stages: fatty changes, alcoholic hepatitis, and cirrhosis.^{5,6} Alcoholic cirrhosis is the end result of repeated bouts of drinking-related liver injury and designates the onset of end-stage alcoholic liver disease.

Fatty liver is characterized by the accumulation of fat in hepatocytes, a condition called steatosis (Fig. 28-10). The liver becomes yellow and enlarged because of excessive fat accumulation. The pathogenesis of fatty liver is not completely understood and can depend on the amount of alcohol consumed, dietary fat content, body stores of fat, hormonal status, and other factors. There is evidence that ingestion of large amounts of alcohol can cause fatty liver changes even with an adequate diet. For example, young, nonalcoholic volunteers had fatty liver changes after 2 days of consuming 18 to 24 oz of alcohol, even though adequate carbohydrates, fats, and proteins were included in the diet.29 The fatty changes that occur with ingestion of alcohol usually do not produce symptoms and are reversible after the alcohol intake has been discontinued.

Alcoholic hepatitis is the intermediate stage between fatty changes and cirrhosis. It often is seen after an abrupt increase in alcohol intake and is common in "spree" drinkers. Alcoholic hepatitis is characterized by inflammation and necrosis of liver



■ FIGURE 28-10 ■ Alcoholic fatty liver. A photomicrograph shows the cytoplasm of almost all the hepatocytes to be distended by fat, which displaces the nucleus to the periphery. Note the absence of inflammation and fibrosis. (Rubin E., Farber J.L. [1999]. Pathology [3rd ed., p. 791]. Philadelphia: Lippincott Williams & Wilkins)

cells. This stage usually is characterized by hepatic tenderness, pain, anorexia, nausea, fever, jaundice, ascites, and liver failure, but some individuals may be asymptomatic. The condition is always serious and sometimes fatal. The immediate prognosis correlates with the severity of liver cell injury. In some cases, the disease progresses rapidly to liver failure and death. The mortality rate in the acute stage ranges from 10% to 30%.6 In persons who survive and continue to drink, the acute phase often is followed by persistent alcoholic hepatitis, with progression to cirrhosis in a matter of 1 to 2 years.⁶

Cirrhosis, Portal Hypertension, and Liver Failure

Cirrhosis

Cirrhosis represents the end stage of chronic liver disease in which the normal architecture of the liver is replaced by fibrous septa that encompass regenerative nodules of hepatic tissue.^{5,6} The gross appearance of the liver in early alcoholic cirrhosis is one of fine, uniform nodules on its surface. The condition has traditionally been called micronodular or Laennec cirrhosis. Although the initial nodules that develop are similar in size to the lobules in normal liver tissue, they lack all the landmarks of normal lobular architecture in terms of portal tracts and central venules. With more advanced cirrhosis, regenerative processes cause the nodules to become larger and more irregular in size and shape. The nodules may compress the hepatic veins and bile ducts, producing portal hypertension, extrahepatic portosystemic shunts, and cholestasis.

Although cirrhosis usually is associated with alcoholism, it can develop in the course of other disorders, including viral hepatitis, toxic reactions to drugs and chemicals, and biliary obstruction. Cirrhosis also accompanies metabolic disorders that cause the deposition of minerals in the liver. Two of these disorders are hemochromatosis (*i.e.*, iron deposition) and Wilson's disease (*i.e.*, copper deposition).

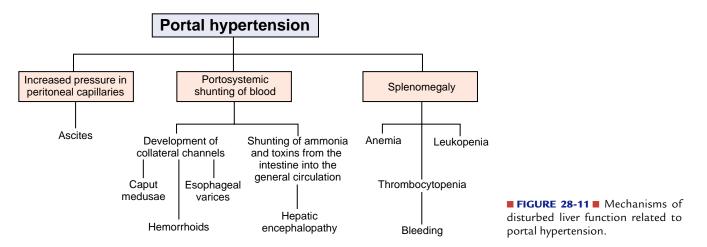
The manifestations of cirrhosis are variable, ranging from asymptomatic hepatomegaly to hepatic failure. Often there are no symptoms until the disease is far advanced. The most common signs and symptoms of cirrhosis are weight loss (sometimes masked by ascites), weakness, and anorexia. Diarrhea frequently is present, although some persons may report constipation. Hepatomegaly and jaundice also are common signs of cirrhosis. There may be abdominal pain because of liver enlargement or stretching of Glisson's capsule.

The late manifestations of cirrhosis are related to portal hypertension and liver failure (Fig. 28-11, Fig. 28-12). Splenomegaly, ascites, and portosystemic shunts (i.e., esophageal varices, anorectal varices, and caput medusae) result from portal hypertension. Other complications include bleeding caused by decreased clotting factors; thrombocytopenia caused by splenomegaly; gynecomastia, a feminizing pattern of pubic hair distribution, and testicular atrophy in men because of altered testosterone and estrogen metabolism; spider angiomas and palmar erythema; and encephalopathy with asterixis and neurologic signs.

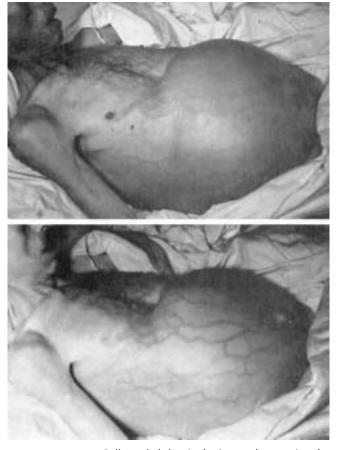
Portal Hypertension

Portal hypertension is characterized by increased resistance to flow in the portal venous system and sustained portal vein pressure above 12 mm Hg (normal, 5–10 mm Hg).^{6,30} Normally,





venous blood returning to the heart from the abdominal organs collects in the portal vein and travels through the liver before entering the vena cava. Portal hypertension can be caused by a variety of conditions that increase resistance to hepatic blood flow, including prehepatic, posthepatic, and intrahepatic



■ FIGURE 28-12 ■ Collateral abdominal veins on the anterior abdominal wall in a patient with alcoholic liver disease as recorded by black and white photography (top) and infrared photography (bottom). (Schiff L. [1982]. *Diseases of the liver*. Philadelphia: J.B. Lippincott)

obstructions (with *hepatic* referring to the liver lobules, rather than the entire liver).⁶

Prehepatic causes of portal hypertension include portal vein thrombosis and external compression caused by cancer or enlarged lymph nodes that produce obstruction of the portal vein before it enters the liver. *Posthepatic* obstruction refers to any obstruction to blood flow through the hepatic veins beyond the liver lobules, either within or distal to the liver. It is caused by conditions such as thrombosis of the hepatic veins, venoocclusive disease, and severe right-sided heart failure that impede the outflow of venous blood from the liver. *Intrahepatic* causes of portal hypertension include conditions that cause obstruction of blood flow within the liver. In alcoholic cirrhosis, which is the major cause of portal hypertension, bands of fibrous tissue and fibrous nodules distort the architecture of the liver and increase the resistance to portal blood flow, which leads to portal hypertension.

Complications of portal hypertension arise from the increased pressure and dilatation of the venous channels behind the obstruction. In addition, collateral channels open that connect the portal circulation with the systemic circulation. The major complications of the increased portal vein pressure and the opening of collateral channels are ascites, splenomegaly,

KEY CONCEPTS PORTAL HYPERTENSION

- Venous blood from the gastrointestinal tract empties into the portal vein and travels through the liver before moving into the general venous circulation.
- Obstruction of blood flow and development of portal hypertension produces an increase in the hydrostatic pressure within the peritoneal capillaries, contributing to the development of ascites, splenic engorgement with sequestration and destruction of blood cells and platelets, and shunting of blood to collateral venous channels causing varicosities of the hemorrhoidal and esophageal veins.

Chapter 28: Alterations in Hepatobiliary Function

507

and the formation of portosystemic shunts with bleeding from esophageal varices (Fig. 28-11).

Ascites. Ascites occurs when the amount of fluid in the peritoneal cavity is increased and is a late-stage manifestation of cirrhosis and portal hypertension.³¹ It is not uncommon for persons with advanced cirrhosis to present with an accumulation of 15 L or more of ascitic fluid. Those who gain this much fluid often experience abdominal discomfort, dyspnea, and insomnia.

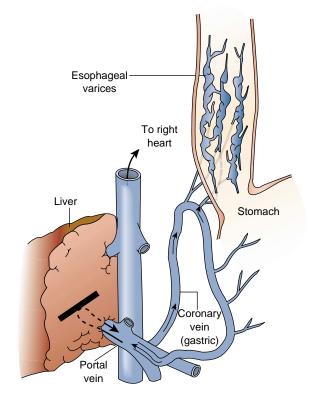
Although the mechanisms responsible for the development of ascites are not completely understood, several factors seem to contribute to fluid accumulation, including an increase in capillary pressure caused by portal hypertension and obstruction of venous flow through the liver, salt and water retention by the kidney, and decreased colloidal osmotic pressure caused by impaired synthesis of albumin by the liver (see Chapter 6).

Treatment of ascites usually focuses on dietary restriction of sodium and administration of diuretics. Water intake also may need to be restricted. Because of the many limitations in sodium restriction, the use of diuretics has become the mainstay of treatment for ascites. Oral potassium supplements often are given to prevent hypokalemia. The upright position is associated with the activation of the renin-angiotensin-aldosterone system; therefore, bed rest may be recommended for persons with a large amount of ascites.³² Large-volume paracentesis (removal of 5 L or more of ascitic fluid) may be done in persons with massive ascites and pulmonary compromise.

Splenomegaly. The spleen enlarges progressively in portal hypertension because of shunting of blood into the splenic vein. The enlarged spleen often gives rise to sequestering of significant numbers of blood elements and development of a syndrome known as hypersplenism. Hypersplenism is characterized by a decrease in the life span and a subsequent decrease in all the formed elements of the blood, leading to anemia, thrombocytopenia, and leukopenia. The person with thrombocytopenia is subject to purpura, easy bruising, hematuria, and abnormal menstrual bleeding, and is vulnerable to bleeding from the esophagus and other segments of the gastrointestinal tract.

Portosystemic Shunts. With the gradual obstruction of venous blood flow in the liver, the pressure in the portal vein increases, and large collateral channels develop between the portal and systemic veins that supply the lower rectum and esophagus and the umbilical veins of the falciform ligament that attaches to the anterior wall of the abdomen. The collaterals between the inferior and internal iliac veins may give rise to hemorrhoids. In some persons, the fetal umbilical vein is not totally obliterated; it forms a channel on the anterior abdominal wall (Fig. 28-12). Dilated veins around the umbilicus are called *caput medusae*. Portopulmonary shunts also may develop and cause blood to bypass the pulmonary capillaries, interfering with blood oxygenation and producing cyanosis.

Clinically, the most important collateral channels are those connecting the portal and coronary veins that lead to reversal of flow and formation of thin-walled varicosities in the submucosa of the esophagus (Fig. 28-13). These thin-walled esophageal varices are subject to rupture, producing massive and sometimes fatal hemorrhage. Impaired hepatic synthesis of co-



■ FIGURE 28-13 ■ Obstruction of blood flow in the portal circulation, with portal hypertension and diversion of blood flow to other venous channels, including the gastric and esophageal veins.

agulation factors and decreased platelet levels (i.e., thrombocytopenia) caused by splenomegaly may further complicate the control of esophageal bleeding. Esophageal varices develop in approximately 65% of persons with advanced cirrhosis and cause massive hemorrhage and death in approximately half of them.5

Treatment of portal hypertension and esophageal varices is directed at prevention of initial hemorrhage, management of acute hemorrhage, and prevention of recurrent variceal hemorrhage. Mechanical compression of the bleeding vessel may be accomplished through the use of an esophageal balloon, designed to compress the bleeding vessel. Prevention of recurrent hemorrhage focuses on lowering portal venous pressure and diverting blood flow away from the easily ruptured collateral channels. Pharmacologic therapy is used to lower portal venous pressure and prevent initial hemorrhage. β -Adrenergic-blocking drugs (e.g., propranolol) commonly are used for this purpose. Endoscopic sclerotherapy and band ligation is used to prevent rebleeding. Portosystemic shunt procedures involve the creation of an opening between the portal vein and a systemic vein as a means of bypassing the liver. The transjugular intrahepatic portosystemic shunt (TIPS) procedure involves the placement of an expandable metal stent between a branch of the hepatic and portal vein inserted via the internal jugular vein.

Liver Failure

The most severe clinical consequence of liver disease is hepatic failure. It may result from sudden and massive hepatic destruction, as in fulminant hepatitis, or it may be the result of progressive damage to the liver, such as occurs in alcoholic cirrhosis. Whatever the cause, 80% to 90% of hepatic functional capacity must be lost before hepatic failure occurs.⁵ In many cases, the progressive decompensating aspects of the disease are hastened by intercurrent conditions, such as gastrointestinal bleeding, systemic infection, electrolyte disturbances, or superimposed diseases such as heart failure.

The manifestations of liver failure reflect the various synthesis, storage, metabolic, and excretory functions of the liver, (Fig. 28-14). *Fetor hepaticus* refers to a characteristic musty, sweetish odor of the breath in the patient with advanced liver failure, resulting from the metabolic by-products of the intestinal bacteria.

Hematologic Disorders. Liver failure can cause anemia, thrombocytopenia, coagulation defects, and leukopenia. Anemia may be caused by blood loss, excessive red blood cell destruction, and impaired formation of red blood cells. A folic acid deficiency may lead to severe megaloblastic anemia. Changes in the lipid composition of the red cell membrane increase hemolysis. Because factors V, VII, IX, and X, prothrombin, and fibrinogen are synthesized by the liver, their decline in liver disease contributes to bleeding disorders. Malabsorption of the fat-soluble vitamin K contributes further to the impaired synthesis of these clotting factors.

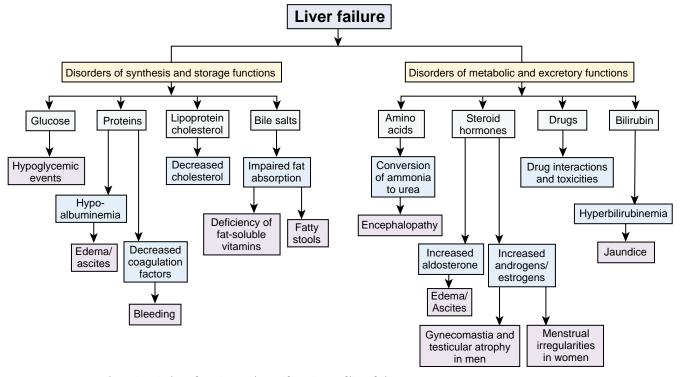
Endocrine Disorders. The liver metabolizes the steroid hormones. Endocrine disorders, particularly disturbances in gonadal function, are common accompaniments of cirrhosis and liver failure. Women may have menstrual irregularities (usually amenorrhea), loss of libido, and sterility. In men, testosterone

levels usually fall, the testes atrophy, and loss of libido, impotence, and gynecomastia occur. A decrease in aldosterone metabolism may contribute to salt and water retention by the kidney, along with a lowering of serum potassium resulting from increased elimination of potassium.

Skin Disorders. Liver failure brings on numerous skin disorders. These lesions, called variously *vascular spiders, telangiectases, spider angiomas,* and *spider nevi*, are seen most often in the upper half of the body. They consist of a central pulsating arteriole from which smaller vessels radiate. Palmar erythema is redness of the palms, probably caused by increased blood flow from higher cardiac output. Clubbing of the fingers may be seen in persons with cirrhosis. Jaundice usually is a late manifestation of liver failure.

Hepatorenal Syndrome. The hepatorenal syndrome refers to a functional state of renal failure sometimes seen during the terminal stages of liver failure with ascites. It is characterized by progressive azotemia, increased serum creatinine levels, and oliguria. Although the basic cause is unknown, a decrease in renal blood flow is believed to play a part. Ultimately, when renal failure is superimposed on liver failure, azotemia and elevated levels of blood ammonia occur; this condition is thought to contribute to hepatic encephalopathy and coma.

Hepatic Encephalopathy. Hepatic encephalopathy refers to the totality of central nervous system manifestations of liver failure. It is characterized by neural disturbances ranging from a lack of mental alertness to confusion, coma, and convulsions. A very early sign of hepatic encephalopathy is a flapping tremor



■ FIGURE 28-14 ■ Alterations in liver function and manifestations of liver failure.

Although the cause of hepatic encephalopathy is unknown, the accumulation of neurotoxins, which appear in the blood because the liver has lost its detoxifying capacity, is believed to be a factor. Hepatic encephalopathy develops in approximately 10% of persons with portosystemic shunts. One of the suspected neurotoxins is ammonia. A particularly important function of the liver is the conversion of ammonia, a by-product of protein and amino acid metabolism, to urea. The ammonium ion is produced in abundance in the intestinal tract, particularly in the colon, by the bacterial degradation of luminal proteins and amino acids. Normally, these ammonium ions diffuse into the portal blood and are transported to the liver, where they are converted to urea before entering the general circulation. When the blood from the intestine bypasses the liver or the liver is unable to convert ammonia to urea, ammonia moves directly into the general circulation and then into the cerebral circulation. Hepatic encephalopathy may become worse after a large protein meal or gastrointestinal tract bleeding.

Treatment. The treatment of liver failure is directed toward eliminating alcohol intake when the condition is caused by alcoholic cirrhosis; preventing infections; providing sufficient carbohydrates and calories to prevent protein breakdown; correcting fluid and electrolyte imbalances, particularly hypokalemia; and decreasing ammonia production in the gastrointestinal tract by controlling protein intake. In many cases, liver transplantation remains the only effective treatment.

Cancer of the Liver

Primary liver tumors are relatively rare in the United States, accounting for approximately 0.5% to 2% of all cancers.³ The American Cancer Society estimates that more than 17,300 new cases of primary liver and intrahepatic cancer will be diagnosed during 2003, and more than 14,400 people will die of the disease during the same period.³³ In contrast to many other cancers, the number of people who have liver cancer and die of it is increasing. Liver cancer is approximately 10 times more common in developing countries in Southeast Asia and Africa. In many of these countries, it is the most common type of cancer.

There are two major types of primary liver cancer: hepatocellular carcinoma, which arises from the liver cells (Fig. 28-15), and cholangiocarcinoma, which is a primary cancer of bile duct cells.⁵ Hepatocellular cancer is one of the few cancers for which an underlying etiology can be identified in most cases and is unique because it usually occurs in a background of chronic liver disease.³⁴ Among the factors identified as etiologic agents in hepatocellular cancer are chronic viral hepatitis (HBV, HCV, HDV), cirrhosis, long-term exposure to aflatoxin, and drinking water contaminated with arsenic. With HBV and HCV, both of which become integrated into the host DNA, repeated cycles of cell death and regeneration afford the potential for develop-



■ FIGURE 28-15 ■ Hepatocellular carcinoma. Cross-section of a cirrhotic liver showing a poorly circumscribed, nodular area ofyellow, partially hemorrhagic carcinoma. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 826]. Philadelphia: Lippincott Williams & Wilkins)

ment of cancer-producing mutations. Aflatoxins, produced by food spoilage molds in certain areas endemic for hepatocellular carcinoma, are particularly potent carcinogenic agents.⁵ These carcinogenic agents are activated by hepatocytes and their products incorporated into the host DNA with the potential for producing cancer-producing mutations.

The manifestations of hepatocellular cancer often are insidious in onset and masked by those related to cirrhosis or chronic hepatitis. The initial symptoms include weakness, anorexia, weight loss, fatigue, bloating, a sensation of abdominal fullness, and a dull, aching abdominal pain. Ascites, which often obscures weight loss, is common. Jaundice, if present, usually is mild. There may be a rapid increase in liver size and worsening of ascites in persons with pre-existing cirrhosis. Usually, the liver is enlarged when these symptoms appear, and there is a low fever without apparent cause. Serum α -fetoprotein, a serum protein present during fetal life, normally is barely detectable in the serum after the age of 2 years, but it is present in 60% to 75% of cases of hepatocellular carcinoma.⁵

Cholangiocarcinoma occurs much less frequently than hepatocellular carcinoma. The etiology, clinical features, and prognosis vary considerably with the part of the biliary tree that is the site of origin. Cholangiocarcinoma is not associated with the same risk factors as hepatocellular carcinoma. Instead, most of the risk factors revolve around longstanding inflammation and injury of the bile duct epithelium. Cholangiocarcinoma often presents with pain, weight loss, anorexia, and abdominal swelling or awareness of a mass in the right hypochondrium. Tumors affecting the central or distal bile ducts may present with jaundice.

Primary cancers of the liver usually are far advanced at the time of diagnosis; the 5-year survival rate is approximately 1%, and most patients die within 6 months. The treatment of choice is subtotal hepatectomy, if conditions permit. Chemo-therapy and radiation therapy are largely palliative. Although

liver transplantation may be an option for people with wellcompensated cirrhosis and small tumors, it often is impractical because of the shortage of donor organs.

Metastatic tumors of the liver are much more common than primary tumors. Common sources include colorectal cancer and spread from the breast, lung, or urogenital cancers. In addition, tumors of neuroendocrine origin spread to the liver. It often is difficult to distinguish primary from metastatic tumors with the use of CT scans, MRI, or ultrasonography. Usually the diagnosis is confirmed by biopsy.

In summary, the liver is subject to most of the disease processes that affect other body structures, such as vascular disorders, inflammation, metabolic diseases, toxic injury, and neoplasms.

Hepatitis is characterized by inflammation of the liver. Acute viral hepatitis is caused by hepatitis viruses A, B, C, D, and E. Although all these viruses cause acute hepatitis, they differ in terms of mode of transmission, incubation period, mechanism, degree and chronicity of liver damage, and the ability to evolve to a carrier state. HBV, HCV, and HDV have the potential for progression to the carrier state, chronic hepatitis, and hepatocellular carcinoma.

Intrahepatic biliary diseases disrupt the flow of bile through the liver, causing cholestasis and biliary cirrhosis. Among the causes of intrahepatic biliary diseases are primary biliary cirrhosis, primary sclerosing cholangitis, and secondary biliary cirrhosis.

Because alcohol competes for use of intracellular cofactors normally needed by the liver for other metabolic processes, it tends to disrupt the metabolic functions of the liver. The spectrum of alcoholic liver disease includes fatty liver disease, alcoholic hepatitis, and cirrhosis.

Cirrhosis represents the end stage of chronic liver disease in which much of the functional liver tissue has been replaced by fibrous tissue. The fibrous tissue replaces normally functioning liver tissue and forms constrictive bands that disrupt flow in the vascular channels and biliary duct systems of the liver. The disruption of vascular channels predisposes to portal hypertension and its complications, loss of liver cells, and eventual liver failure. Portal hypertension is characterized by increased resistance to flow and increased pressure in the portal venous system; the pathologic consequences of the disorder include ascites, the formation of collateral bypass channels (e.g., esophageal varices) from the portosystemic circulation, and splenomegaly. Liver failure represents the end stage of a number of liver diseases and occurs when less than 10% of liver tissue is functional. The manifestations of liver failure reflect the various functions of the liver, including hematologic disorders, disruption of endocrine function, skin disorders, hepatorenal syndrome, and hepatic encephalopathy.

Cancers of the liver include metastatic and primary neoplasms. Primary hepatic neoplasms are rare, accounting for less than 2% of cancers, and those involving the hepatocytes or liver cells are commonly associated with underlying diseases of the liver such as cirrhosis and chronic hepatitis. Liver cancer usually is far advanced at the time of diagnosis; the 5-year survival rate is approximately 1%.

DISORDERS OF THE GALLBLADDER AND EXTRAHEPATIC BILE DUCTS

The so-called hepatobiliary system consists of the gallbladder, the left and right hepatic ducts, which come together to form the common hepatic duct, the cystic duct, which extends to the gallbladder, and the common bile duct, which is formed by the union of the common hepatic duct and the cystic duct (see Fig. 28-16). The common bile duct descends posterior to the first part of the duodenum, where it comes in contact with the main pancreatic duct. These ducts unite to form the hepatopancreatic ampulla (ampulla of Vater). The circular muscle around the distal end of the bile duct is thickened to form the sphincter of the bile duct (Fig. 28-16).

The gallbladder is a distensible, pear-shaped, muscular sac located on the ventral surface of the liver. It has a outer serous peritoneal layer, a middle smooth muscle layer, and an inner mucosal layer that is continuous with the linings of the bile duct. The function of the gallbladder is to store and concentrate bile. Bile contains bile salts, cholesterol, bilirubin, lecithin, fatty acids, water, and the electrolytes normally found in the plasma. The cholesterol found in bile has no known function; it is assumed to be a by-product of bile salt formation, and its presence is linked to the excretory function of bile. Normally insoluble in water, cholesterol is rendered soluble by the action of bile salts and lecithin, which combine with it to form micelles. In the gallbladder, water and electrolytes are absorbed from the liver bile, causing the bile to become more concentrated. Because neither lecithin nor bile salts are absorbed in the gallbladder, their concentration increases along with that of cholesterol; in this way, the solubility of cholesterol is maintained.

Entrance of food into the intestine causes the gallbladder to contract and the sphincter of the bile duct to relax, such that bile stored in the gallbladder moves into the duodenum. The stimulus for gallbladder contraction is primarily hormonal. Products of food digestion, particularly lipids, stimulate the release of a gastrointestinal hormone called cholecystokinin from the mucosa of the duodenum. Cholecystokinin provides a strong stimulus for gallbladder contraction. The role of other gastrointestinal hormones in bile release is less clearly understood.

Cholelithiasis and Cholecystitis

Two common disorders of the gallbladder system are cholelithiasis (*i.e.*, gallstones) and inflammation of the gallbladder (cholecystitis) or common bile duct (cholangitis). At least 10% of adults have gallstones. Approximately twice as many women as men have gallstones, and there is an increased prevalence with age—after 60 years of age, 10% to 15% among men and 20% to 40% among women.³⁵

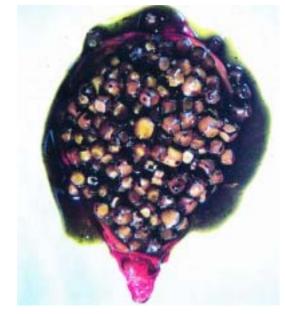
Cholelithiasis

Gallstones are caused by precipitation of substances contained in bile, mainly cholesterol and bilirubin. The bile of which gallstones are formed usually is supersaturated with cholesterol or bilirubin. Approximately 75% of gallstones are composed primarily of cholesterol; the other 25% are black or brown pigment stones consisting of calcium salts with bilirubin.³⁵ Many

FIGURE 28-16 (A) Extrahepatic bile passages, gallbladder, and pancreatic ducts. (B) Entry of bile duct and pancreatic duct into the hepatopancreatic ampulla, which opens into the duodenum.

stones have a mixed composition. Figure 28-17 shows a gallbladder with numerous cholesterol gallstones.

Three factors contribute to the formation of gallstones: (1) abnormalities in the composition of bile, (2) stasis of bile, and (3) inflammation of the gallbladder. The formation of cholesterol stones is associated with obesity and occurs more frequently in women, especially women who have had multiple pregnancies or who are taking oral contraceptives. All of these factors cause the liver to excrete more cholesterol into the bile. Estrogen reduces the synthesis of bile acid in women. Drugs that lower serum cholesterol levels, such as clofibrate, also cause increased cholesterol excretion into the bile. Malabsorption disorders stemming from ileal disease or intestinal bypass surgery interfere with intestinal reabsorption of bile salts, which are needed to maintain cholesterol solubility. Gallbladder sludge (thickened gallbladder mucoprotein with tiny trapped cholesterol crystals) causes stasis of bile flow and is thought to be a precursor of gallstones. Sludge frequently occurs with pregnancy, starvation, and rapid weight loss.³⁵ Inflammation of the gallbladder alters the absorptive characteristics of the mucosal layer, allowing excessive absorption of water and bile salts. Cholesterol gallstones are extremely common among Native Americans, which suggests that a genetic component may have a role in gallstone formation. Pigment



■ FIGURE 28-17 ■ Cholesterol gallstones. The gallbladder has been opened to reveal numerous yellow cholesterol gallstones (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 791]. Philadelphia: Lippincott Williams & Wilkins)

stones containing bilirubin are seen in persons with hemolytic disease (*e.g.*, sickle cell disease) and hepatic cirrhosis.

Many persons with gallstones have no symptoms. Gallstones cause symptoms when they obstruct bile flow. Small stones (*e.g.*, <8 mm in diameter) pass into the common duct, producing symptoms of indigestion and biliary colic. Larger stones are more likely to obstruct flow and cause jaundice. The pain of biliary colic usually is abrupt in onset and increases steadily in intensity until it reaches a climax in 30 to 60 minutes. It is usually located in the upper right quadrant or epigastric area and is often referred to the back, above the waist or to the right shoulder and scapula region. The pain usually persists for 2 to 8 hours and is followed by soreness in the upper right quadrant.

Acute and Chronic Cholecystitis

The term *cholecystitis* refers to inflammation of the gallbladder. Both acute and chronic cholecystitis are associated with cholelithiasis. Acute cholecystitis may be superimposed on chronic cholecystitis.

Acute cholecystitis almost always is associated with complete or partial obstruction. It is believed that the inflammation is caused by chemical irritation from the concentrated bile, along with mucosal swelling and ischemia resulting from venous congestion and lymphatic stasis. The gallbladder usually is markedly distended. Bacterial infections may arise secondary to the ischemia and chemical irritation. The bacteria reach the injured gallbladder through the blood, lymphatics, or bile ducts or from adjacent organs. Among the common pathogens are staphylococci and enterococci. The wall of the gallbladder is most vulnerable to the effects of ischemia, as a result of which mucosal necrosis and sloughing occur. The process may lead to gangrenous changes and perforation of the gallbladder.

The signs and symptoms of acute cholecystitis vary with the severity of obstruction and inflammation. Pain, initially similar to that of biliary colic, is characteristic of acute cholecystitis. It often is precipitated by a fatty meal and may initiate with complaints of indigestion. However, it does not subside spontaneously and responds poorly or only temporarily to potent analgesics. When the inflammation progresses to involve the peritoneum, the pain becomes more pronounced in the right upper quadrant. The right subcostal region is tender, and the muscles that surround the area spasm. Approximately 75% of patients have vomiting, and approximately 25% have jaundice.²² Fever and an abnormally high white blood cell count attest to inflammation. Total serum bilirubin, aminotransferase, and alkaline phosphatase levels usually are elevated.

Chronic cholecystitis results from repeated episodes of acute cholecystitis or chronic irritation of the gallbladder by stones. It is characterized by varying degrees of chronic inflammation. Gallstones almost always are present. Cholelithiasis with chronic cholecystitis may be associated with acute exacerbations of gallbladder inflammation, common duct stones, pancreatitis, and rarely, carcinoma of the gallbladder.

The manifestations of chronic cholecystitis are more vague than those of acute cholecystitis. There may be intolerance to fatty foods, belching, and other indications of discomfort. Often, there are episodes of colicky pain with obstruction of biliary flow caused by gallstones. The gallbladder, which in chronic cholecystitis usually contains stones, may be enlarged, shrunken, or of normal size. *Diagnosis and Treatment.* The most commonly used methods for diagnosis of gallbladder disease include ultrasonography and nuclear scanning (cholescintigraphy).³⁵ Ultrasonography is widely used in diagnosing gallbladder disease. It can detect stones as small as 1 to 2 cm, and its overall accuracy in detecting gallbladder disease is high. In addition to stones, ultrasonography can detect wall thickening, which indicates inflammation. Cholescintigraphy, also called a *gallbladder scan*, relies on the ability of the liver to extract a rapidly injected radionuclide, technetium-99m, bound to one of several iminodiacetic acids, that is excreted into the bile ducts. The gallbladder scan is highly accurate in detecting acute cholecystitis.

Gallbladder disease usually is treated by removing the gallbladder or by dissolving the stones or fragmenting them. Laparoscopic cholecystectomy has largely become the treatment of choice for symptomatic gallbladder disease.

Choledocholithiasis and Cholangitis

Choledocholithiasis refers to stones in the common duct and cholangitis to inflammation of the common duct. Common duct stones usually originate in the gallbladder but can form spontaneously in the common duct. The stones frequently are clinically silent unless there is obstruction.

The manifestations of choledocholithiasis are similar to those of gallstones and acute cholecystitis. There is a history of acute biliary colic and right upper abdominal pain, with chills, fever, and jaundice associated with episodes of abdominal pain. Bilirubinuria and an elevated serum bilirubin are present if the common duct is obstructed.

Complications include acute suppurative cholangitis accompanied by pus in the common duct. It is characterized by the presence of an altered sensorium, lethargy, and septic shock.²² Acute suppurative cholangitis represents an endoscopic or surgical emergency. Common duct stones also can obstruct the outflow of the pancreatic duct, causing a secondary pancreatitis.

Diagnostic measures include the use of ultrasonography, CT scans, and radionuclide to detect dilatation of the bile ducts and impaired blood flow. Endoscopic ultrasonography and magnetic resonance cholangiography may be used for detecting common duct stones. Both percutaneous transhepatic cholangiography (PTC) and endoscopic retrograde cholangiopancreatography (ERCP) provide a direct means for determining the cause, location, and extent of obstruction. PTC involves the injection of dye directly into the biliary tree. It requires the insertion of a thin, flexible needle through a small incision in the skin with advancement into the biliary tree. ERCP involves the passage of an endoscope into the duodenum and the passage of a catheter into the hepatopancreatic ampulla. ERCP can be used to enlarge the opening of the bile duct sphincter so that a lodged stone can pass into the intestine or an instrument may be inserted into the common duct to remove the stone.

Common duct stones in persons with cholelithiasis usually are treated by stone extraction followed by laparoscopic cholecystectomy. Antibiotic therapy, with an agent that penetrates the bile, is used to treat the infection. Emergency decompression of the common duct, usually by ERCP, may be necessary for persons who are septic or fail to experience improvement with antibiotic treatment.

Cancer of the Gallbladder

Cancer of the gallbladder is found in approximately 2% of persons operated on for biliary tract disease. The onset of symptoms usually is insidious, and they resemble those of cholecystitis; the diagnosis often is made unexpectedly at the time of gallbladder surgery. Because of their ability to produce chronic irritation of the gallbladder mucosa, it is believed that gallstones play a role in the development of gallbladder cancer. The 5-year survival rate is only approximately 3%.5

In summary, the biliary tract serves as a passageway for the delivery of bile from the liver to the intestine. This tract consists of the bile ducts and gallbladder. The most common causes of biliary tract disease are cholelithiasis and cholecystitis. Three factors contribute to the development of cholelithiasis: abnormalities in the composition of bile, stasis of bile, and inflammation of the gallbladder. Cholelithiasis predisposes to obstruction of bile flow, causing biliary colic and acute or chronic cholecystitis. Cancer of the gallbladder, which has a poor 5-year survival rate, occurs in 2% of persons with biliary tract disease.

DISORDERS OF THE EXOCRINE PANCREAS

The pancreas lies transversely in the posterior part of the upper abdomen. The head of the pancreas is at the right of the abdomen; it rests against the curve of the duodenum in the area of the hepatopancreatic ampulla and its entrance into the duodenum (Fig. 28-1). The body of the pancreas lies beneath the stomach. The tail touches the spleen. The pancreas is virtually hidden because of its posterior position; unlike many other organs, it cannot be palpated. Because of the position of the pancreas and its large functional reserve, symptoms from conditions such as cancer of the pancreas do not usually appear until the disorder is far advanced.

The pancreas is both an endocrine and exocrine organ. Its function as an endocrine organ is discussed in Chapter 31. The exocrine pancreas is made up of lobules that consist of acinar cells, which secrete digestive enzymes into a system of microscopic ducts. These ducts empty into the main pancreatic duct, which extends from left to right through the substance of the pancreas. The main pancreatic duct and the bile duct unite to form the hepatopancreatic ampulla, which empties into the duodenum. The sphincter of the pancreatic duct controls the flow of pancreatic secretion into duodenum (Fig. 28-16).

The pancreatic secretions contain proteolytic enzymes that break down dietary proteins, including trypsin, chymotrypsin, carboxypolypeptidase, ribonuclease, and deoxyribonuclease. The pancreas also secretes pancreatic amylase, which breaks down starch, and lipases, which hydrolyze neutral fats into glycerol and fatty acids. The pancreatic enzymes are secreted in the inactive form and become activated in the intestine. This is important because the enzymes would digest the tissue of the pancreas itself if they were secreted in the active form. The acinar cells secrete a trypsin inhibitor, which prevents trypsin activation. Because trypsin activates other proteolytic enzymes, the trypsin inhibitor prevents subsequent activation of those other enzymes. The smaller pancreatic ducts are lined with epithelial cells that secrete water and bicarbonate and thereby modify the fluid and electrolyte composition of the pancreatic secretions. Two types of pancreatic disease are discussed in this chapter: acute and chronic pancreatitis and cancer of the pancreas.

Acute Pancreatitis

Acute pancreatitis represents an inflammation of the pancreas that ranges from a mild self-limited disease, consisting of inflammation and interstitial edema, to an acute hemorrhagic pancreatitis that is associated with massive necrosis of tissue.³⁶ Acute hemorrhagic pancreatitis is a severe, life-threatening disorder associated with the escape of activated pancreatic enzymes into the pancreas and surrounding tissues. These enzymes cause fat necrosis, or autodigestion, of the pancreas and produce fatty deposits in the abdominal cavity with hemorrhage from the necrotic vessels.

Although a number of factors are associated with the development of acute pancreatitis, most cases result from gallstones (stones in the common duct) or alcohol abuse.^{37,38} In the case of biliary tract obstruction caused by gallstones, pancreatic duct obstruction or biliary reflux is believed to activate the enzymes in the pancreatic duct system. The precise mechanisms whereby alcohol exerts its action are largely unknown. Alcohol is known to be a potent stimulator of pancreatic secretions, and it also is known to cause constriction of the sphincter of the pancreatic duct. Acute pancreatitis also is associated with hyperlipidemia, hyperparathyroidism, infections (particularly viral), abdominal and surgical trauma, and drugs such as steroids and thiazide diuretics.

The onset of acute pancreatitis usually is abrupt and dramatic, and it may follow a heavy meal or an alcoholic binge. The most common initial symptom is severe epigastric and abdominal pain that radiates to the back. The pain is aggravated when the person is lying supine; it is less severe when the person is sitting and leaning forward. Abdominal distention accompanied by hypoactive bowel sounds is common. An important disturbance related to acute necrotizing pancreatitis is the loss of a large volume of fluid into the retroperitoneal and peripancreatic spaces and the abdominal cavity. Tachycardia, hypotension, cool and clammy skin, and fever often are evident. Signs of hypocalcemia may develop, probably as a result of the precipitation of serum calcium in the areas of fat necrosis. Mild jaundice may appear after the first 24 hours because of biliary obstruction. Complications include acute respiratory distress syndrome and acute tubular necrosis. Hypocalcemia occurs in approximately 25% of patients.

Total serum amylase is the test used most frequently in the diagnosis of acute pancreatitis. Serum amylase levels increase within the first 24 hours after onset of symptoms and remain elevated for 48 to 72 hours. The serum lipase level also increases during the first 24 to 48 hours and remains elevated for 5 to 14 days. Urinary clearance of amylase is increased. Because the serum amylase level may be increased as a result of other serious illnesses, the urinary level of amylase is often measured. The white blood cell count may be increased, and blood glucose and serum bilirubin levels may be elevated. Plain radiographs of the abdomen may be used for detecting gallstones

or abdominal complications. CT scans and dynamic contrastenhanced CT of the pancreas are used to detect necrosis and fluid accumulation.

The treatment consists of measures directed at pain relief, "putting the pancreas to rest," and restoration of lost plasma volume. Antibiotic prophylaxis is used to prevent infection of necrotic pancreatic tissue. Oral foods and fluids are withheld, and gastric suction is instituted to treat distention of the bowel and prevent further stimulation of the secretion of pancreatic enzymes. Intravenous fluids and electrolytes are administered to replace those lost from the circulation and to combat hypotension and shock. Intravenous colloid solutions are given to replace the fluid that has become sequestered in the abdomen and retroperitoneal space.

Persons who survive acute necrotizing pancreatitis are at risk for development of pancreatic abscesses and pseudocysts. A pseudocyst is a collection of degraded blood, debris, and necrotic pancreatic tissue enclosed in a layer of connective tissue (Fig. 28-18). The pseudocyst most often is connected to a pancreatic duct, so that it continues to increase in mass. The symptoms depend on its location; for example, jaundice may occur when a cyst develops near the head of the pancreas, close to the common duct. Pseudocysts may become secondarily infected and form an abscess. Pseudocysts may resolve or, if they persist, may require surgical intervention.

Chronic Pancreatitis

Chronic pancreatitis is characterized by progressive destruction of the pancreas. It can be divided into two types: chronic calcifying pancreatitis and chronic obstructive pancreatitis.³⁸ In chronic calcifying pancreatitis, calcified protein plugs (*i.e.*, calculi) form in the pancreatic ducts. This form is seen most often in alcoholics. Alcohol damages pancreatic cells directly and also increases the concentration of proteins in the pancreatic secretions, which eventually leads to formation of protein plugs.³⁹ Other causes of chronic pancreatitis are cystic fibrosis and chronic obstructive pancreatitis caused by stenosis of the sphincter of the pancreatic duct. In obstructive pancreatitis, lesions are more prominent in the head of the pancreas. The disease usually is caused by cholelithiasis and sometimes is relieved by removal of the stones.

Chronic pancreatitis is manifested in episodes that are similar, albeit of lesser severity, to those of acute pancreatitis. Patients have persistent, recurring episodes of epigastric and



■ FIGURE 28-18 ■ Pancreatic pseudocyst. A cystic cavity arises from the head of the pancreas. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 847]. Philadelphia: Lippincott Williams & Wilkins)

upper left quadrant pain; the attacks often are precipitated by alcohol abuse or overeating. Anorexia, nausea, vomiting, constipation, and flatulence are common. Eventually the disease progresses to the extent that endocrine and exocrine pancreatic functions become deficient. At this point, signs of diabetes mellitus and the malabsorption syndrome (*e.g.*, weight loss, fatty stools [steatorrhea]) become apparent.³⁸

Treatment consists of measures to treat coexisting biliary tract disease. A low-fat diet usually is prescribed. The signs of malabsorption may be treated with pancreatic enzymes. When diabetes is present, it is treated with insulin. Alcohol is forbidden because it frequently precipitates attacks. Because of the frequent episodes of pain, narcotic addiction is a potential problem in persons with chronic pancreatitis. Surgical intervention sometimes is needed to relieve the pain and usually focuses on relieving any obstruction that may be present. In advanced cases, a subtotal or total pancreatectomy may be necessary.²²

Cancer of the Pancreas

Pancreatic cancer is now the fourth leading cause of cancer death in the United States, with more than 28,000 deaths attributed to the neoplasm each year.⁴⁰ Considered to be one of the most deadly malignancies, pancreatic cancer is associated with a mortality:incidence ratio of approximately 0.99. The risk of pancreatic cancer increases after the age of 50 years, with most cases occurring between the ages of 60 and 80 years. The incidence and mortality rates for both male and female African Americans are higher than those for whites.

The cause of pancreatic cancer is unknown. Smoking appears to be a major risk factor.^{41,42} The incidence of pancreatic cancer is twice as high among smokers than nonsmokers. The second most important factor appears to be diet. There appears to be an association of pancreatic cancer with an increasing total calorie intake and a high intake of fat, meat, salt, dehydrated foods, fried foods, refined sugars, soy beans, and nitrosamines. Data from animal studies indicate that nitrosamines and tobacco smoke are carcinogenic in the pancreas. A protective effect has been ascribed to a diet containing dietary fiber, vitamin C, fresh fruits and vegetables, and no preservatives. Diabetes and chronic pancreatitis also are associated with pancreatic cancer, although neither the nature nor the sequence of the possible cause-and-effect relation has been established.^{41,42} Genetic alterations appear to play a role. There appears to be an association between pancreatic cancer and certain genetic disorders, including nonpolyposis colon cancer, familial breast cancer with the BRCA2 gene mutation (see Chapter 5), ataxia-telangiectasia syndrome, familial atypical multiple molemelanoma syndrome, and hereditary pancreatitis.⁴¹

Cancer of the pancreas usually has an insidious onset. Pain, jaundice, and weight loss constitute the classic presentation of the disease. The most common pain is a dull epigastric pain often accompanied by back pain, often worse in the supine position, and relieved by sitting forward. Duodenal obstruction with nausea and vomiting is a late sign.

Because of the proximity of the pancreas to the common duct and the hepatopancreatic ampulla, cancer of the head of the pancreas tends to obstruct bile flow; this causes distention of the gallbladder and jaundice. Jaundice frequently is the presenting symptom of a person with cancer of the head of the pancreas, and it usually is accompanied by complaints of pain and pruritus. Cancer of the body of the pancreas usually impinges on the celiac ganglion, causing pain. The pain usually worsens with ingestion of food or with assumption of the supine position. Cancer of the tail of the pancreas usually has metastasized before symptoms appear.

Ultrasonography and CT scanning are the most frequently used diagnostic methods to confirm the disease. Percutaneous fine-needle aspiration cytology of the pancreas has been one of the major advances in the diagnosis of pancreatic cancer. Endoscopic retrograde cholangiopancreatography may be used for evaluation of persons with suspected pancreatic cancer and obstructive jaundice.

Most cancers of the pancreas have metastasized by the time of diagnosis. Surgical resection of the tumor is done when the tumor is localized or as a palliative measure. Radiation therapy may be useful when the disease is not resectable but appears to be localized. The use of irradiation and chemotherapy for pancreatic cancer continues to be investigated. Pain control is one of the most important aspects in the management of persons with end-stage pancreatic cancer.

In summary, the pancreas is an endocrine and exocrine organ. The exocrine pancreas produces digestive enzymes that are secreted in an inactive form and transported to the small intestine through the main pancreatic duct, which usually empties into the ampulla of Vater and then into the duodenum through the sphincter of Oddi. Acute and chronic types of pancreatitis are associated with biliary reflux and chronic alcoholism. Acute pancreatitis is a dramatic and lifethreatening disorder in which there is autodigestion of pancreatic tissue. Chronic pancreatitis causes progressive destruction of the endocrine and exocrine pancreas. It is characterized by episodes of pain and epigastric distress that are similar to but less severe than those that occur with acute pancreatitis. Cancer of the pancreas is the fourth leading cause of death in the United States. It usually is far advanced at the time of diagnosis, and the 5-year survival rate is less than 3%.

REVIEW QUESTIONS

Characterize the function of the liver in terms of bilirubin elimination and describe the pathogenesis of unconjugated and conjugated hyperbilirubinemia.

Relate the mechanism of bile formation and elimination to the development of cholestasis.

Compare hepatitis A, B, C, D, and E in terms of source of infection, incubation period, acute disease manifestations, development of chronic disease, and the carrier state.

Define the term *chronic hepatitis* and compare the pathogenesis of chronic autoimmune and chronic viral hepatitis.

Explain the metabolism of alcohol by the liver and state the metabolic mechanisms that can be used to explain liver injury.

Describe the pathogenesis of intrahepatic biliary tract disease.

Characterize the liver changes that occur with cirrhosis.

Describe the physiologic basis for portal hypertension and relate it to the development of ascites, esophageal varices, and splenomegaly.

Relate the functions of the liver to the manifestations of liver failure.

Relate the function of the gallbladder to the development of gallstones.

Describe the clinical manifestations of acute and chronic cholecystitis.

Characterize the effects of choledocholithiasis and cholangitis on bile flow and the potential for hepatic and pancreatic complications.

Compare the causes and manifestations of acute and chronic pancreatitis.

State the reason for the poor prognosis in pancreatic cancer.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- Guyton A., Hall J.E. (2000). *Textbook of medical physiology* (10th ed., pp. 781–802). Philadelphia: W.B. Saunders.
- Rose S. (1998). Gastrointestinal and hepatic pathophysiology. Madison, CT: Fence Creek Publishing.
- Katzung B.G. (2001). Basic and clinical pharmacology (8th ed., pp. 51–63). New York: Lange Medical Books/McGraw-Hill.
- Lee W.M. (1995). Drug-induced hepatotoxicity. New England Journal of Medicine 333, 1118–1127.
- Crawford J.M. (1999). The liver and biliary tract. In Kumar V., Cotran R.S., Collins T. (Eds.), *Robbins pathologic basis of disease* (6th ed., pp. 845–901). Philadelphia: W.B. Saunders.
- Rubin E., Farber J.L. (1999). The liver and biliary system. In Rubin E., Farber J.L. (Eds.), *Pathophysiology* (3rd ed., pp. 757–838). Philadelphia: Lippincott Williams & Wilkins.
- Denery P.A., Seidman D.S., Stevenson D.K. (2001). Neonatal hyperbilirubinemia. *New England Journal of Medicine* 344, 581–590.
- Herlong H.F. (1994). Approach to the patient with abnormal liver enzymes. *Hospital Practice* 29 (11), 32–38.
- Pratt D.S., Kaplan M.M. (2000). Evaluation of abnormal liver-enzyme results in asymptomatic patients. *New England Journal of Medicine* 342, 1266–1271.
- Advisory Committee on Immunization Practices. (1999). Prevention of hepatitis A through active and passive immunization. *Morbidity and Mortality Weekly Report* 48 (RR-12), 1–25.
- 11. Kemmer N.M., Miskovsky E.P. (2000). Hepatitis A. Infectious Disease Clinics of North America 14, 605–615.
- Centers for Disease Control. (1991). Hepatitis B virus: A comprehensive strategy for eliminating transmission in the United States through universal childhood vaccination. *Morbidity and Mortality Weekly Report* 40 (RR-13), 1–25.
- Lee W.M. (1997). Hepatitis B virus infection. New England Journal of Medicine 337, 1733–1745.
- Befeler A.S., DiBisceglie A.M. (2000). Hepatitis B. Infectious Disease Clinics of North America 14, 617–632.
- Advisory Committee on Immunization Practices. (1999). Notice to readers update: Recommendations to prevent hepatitis B transmission—United States. *Morbidity and Mortality Weekly Report* 48 (2), 33–34.

516 Unit Seven: Alterations in the Gastrointestinal System

- Lemon S.M., Thomas D.L. (1997). Vaccines to prevent viral hepatitis. New England Journal of Medicine 336, 196–203.
- Lauer G.M., Walker R.D. (2001). Hepatitis C virus infection. New England Journal of Medicine 345 (1), 41–52.
- Alter M.J., Kruszon-Moran D., Nainam O.V., et al. (1999). The prevalence of hepatitis C virus infection in the United States 1988 through 1994. *New England Journal of Medicine* 341, 556–562.
- Centers for Disease Control and Prevention. (1998). Recommendations for prevention and control of hepatitis C virus (HCV) infection and HCV-related chronic disease. *Morbidity and Mortality Weekly Report* 47 (RR-19), 1–25.
- Cheney C.P., Chopra S., Graham C. (2000). Hepatitis C. Infectious Disease Clinics of North America 14, 633–659.
- 21. Hoffnagle J.H. (1989). Type D (delta) hepatitis. *Journal of the American Medical Association* 261, 1321–1325.
- Friedman S. (2001). Liver, biliary tract and pancreas. In Tierney L.M., McPhee S.J., Papadakis M.A. (Eds.), *Current medical diagnosis* and treatment (40th ed., 662–705). New York: Lange Medical Books/ McGraw-Hill.
- Hoffnagle J.H., DiBisceglie A.M. (1997). The treatment of chronic hepatitis. New England Journal of Medicine 336, 347–355.
- Lin O.S., Keeffe E.B. (2001). Current treatment strategies for chronic hepatitis B and C. Annual Review of Medicine 52, 29–49.
- Krawitt E.L. (1996). Autoimmune hepatitis. New England Journal of Medicine 334, 897–902.
- Kaplan M.M. (1996). Primary biliary cirrhosis. New England Journal of Medicine 335, 1570–1580.
- Gershwin E., Mackay J.R. (1995). New knowledge in primary biliary cirrhosis. *Hospital Practice* 30 (8), 29–36.
- Lee Y.-M., Kaplan M.M. (1995). Primary sclerosing cholangitis. New England Journal of Medicine 332, 924–932.

- Rubin E., Lieber C.S. (1968). Alcohol-induced hepatic injury in nonalcoholic volunteers. New England Journal of Medicine 278, 869–876.
- Trevillyan J., Carroll P.J. (1997). Management of portal hypertension and esophageal varices in alcoholic cirrhosis. *American Family Phy*sician 55, 1851–1858.
- Roberts L.R., Kamath P.S. (1996). Ascites and hepatorenal syndrome: Pathophysiology and management. *Mayo Clinic Proceedings* 71, 874–881.
- Garcia N., Sanyal A.J. (2001). Minimizing ascites: Complications of cirrhosis signals clinical deterioration. *Postgraduate Medicine* 109 (2), 91–103.
- American Cancer Society. (2003). Cancer facts & figures 2003. [Online]. Available: http://www.cancer.org.
- DiBisceglie A.M. (1999). Malignant neoplasms of the liver. In Schiff E.R., Sorrell M.F., Maddrey W.C. (Eds.), Schiff's diseases of the liver (8th ed., pp. 1281–1300). Philadelphia: Lippincott Williams & Wilkins.
- 35. Johnston D.E., Kaplan M.M. (1993). Pathogenesis and treatment of gallstones. *New England Journal of Medicine* 328, 412–421.
- Baron T.H., Morgan D.E. (1999). Acute necrotizing pancreatitis. New England Journal of Medicine 340, 1412–1417.
- Cartmell M.T., Kingsnorth A.N. (2000). Acute pancreatitis. *Hospital Medicine* 61, 382–385.
- Steer M.L., Waxman L., Freeman S. (1995). Chronic pancreatitis. New England Journal of Medicine 332, 1482–1490.
- 39. Isla A.M. (2000). Chronic pancreatitis. Hospital Medicine 61, 386-389.
- Lillemoe K.D. (2000). Pancreatic cancer: State-of-the-art care. CA: A Cancer Journal for Clinicians 50, 241–268.
- 41. Warshaw A.I., Castillo C.F. (1992). Pancreatic carcinoma. New England Journal of Medicine 326, 455–465.
- Wanebo H.J., Vezeridis M.P. (1996). Pancreatic cancer in perspective. Cancer 76, 580–587.

Alterations in Body Nutrition

Regulation of Food Intake and Energy Metabolism **Energy Storage Adipose Tissue** Nutritional Needs **Recommended Dietary Allowances and Dietary Reference Intakes** Calories Proteins, Fats, and Carbohydrates Vitamins and Minerals **Regulation of Food Intake and Energy Storage** Hunger, Appetite, and Food Intake Assessment of Energy Stores and Nutritional Status **Overnutrition and Obesity Causes of Obesity Types of Obesity** Health Risks Associated With Obesity **Prevention and Treatment Childhood Obesity** Undernutrition Protein and Calorie Malnutrition Malnutrition and Wasting in Illness Treatment **Eating Disorders** Anorexia Nervosa Bulimia Nervosa and Binge Eating

CHAPTER

Nutritional status describes the condition of the body related to the availability and use of nutrients. Nutrients provide the energy and materials necessary for performing the activities of daily living; for maintaining healthy skin, muscles, and other body tissues; for replacing and healing tissues; and for the effective functioning of all body systems, including the immune and respiratory systems. Nutrients are derived from the digestive tract through the ingestion of foods. Once inside the body, nutrients are used for energy or as the building blocks for tissue growth and repair. When excess nutrients are available, they frequently are stored for future use. If the required nutrients are unavailable, the body adapts by conserving and using its nutrient stores.

REGULATION OF FOOD INTAKE AND ENERGY METABOLISM

Energy is measured in heat units called *calories*. A calorie, spelled with a small c and also called a *gram calorie*, is the amount of heat or energy required to raise the temperature of 1 g of water by 1°C. A *kilocalorie* (kcal), or *large calorie*, is the amount of energy needed to raise the temperature of 1 kg of water by 1°C. Because a calorie is so small, kilocalories often are used in nutritional and physiologic studies.

Metabolism is the organized process through which nutrients such as carbohydrates, fats, and proteins are broken down, transformed, or otherwise converted into cellular energy. The oxidation of proteins provides 4 kcal/g; fats, 9 kcal/g; carbohydrates, 4 kcal/g; and alcohol, 7 kcal/g.

The process of metabolism is unique in that it enables the continual release of energy, and it couples this energy with physiologic functioning. For example, the energy used for muscle contraction is derived largely from energy sources that are stored in muscle cells and then released as the muscle contracts. Because most of our energy sources come from the nutrients in the food that is eaten, the ability to store energy and control its release is important. Normally, energy utilization is balanced with energy expenditure. When a person is overfed and intake of food consistently exceeds energy expenditure, the excess energy is stored as fat, and the person becomes overweight. Conversely, when food intake is less than energy expenditure, fat stores and other body tissues are broken down, and the person loses weight.

Energy Storage

Adipose Tissue

More than 90% of body energy is stored in the adipose tissues of the body. *Adipocytes*, or fat cells, occur singly or in small groups in loose connective tissue. In many parts of the body, they cushion body organs such as the kidneys. In addition to

KEY CONCEPTS

ENERGY METABOLISM

- Energy is required for all the body's activities. Food is the source of the body's energy, which is measured in kilocalories (kcal).
- Fats, which are a concentrated water-free energy source, contain 9 kcal/g. They are stored in fat cells as triglycerides, which are the main storage sites for energy.
- Carbohydrates are hydrated fuels, which supply 4 kcal/g. They are stored in limited quantities as glycogen and can be converted to fatty acids and stored in fat cells as triglycerides.
- Amino acids, which supply 4 kcal/g, are used in building body proteins. Amino acids in excess of those needed for protein synthesis are converted to fatty acids, ketones, or glucose and are stored or used as metabolic fuel.

isolated groups of fat cells, entire regions of fat tissue are committed to fat storage. Collectively, fat cells constitute a large body organ that is metabolically active in the uptake, synthesis, storage, and mobilization of lipids, which are the main source of fuel storage for the body. Some tissues, such as liver cells, are able to store small amounts of lipids, but when these lipids accumulate, they begin to interfere with cell function. Adipose tissue not only serves as a storage site for body fuels, it provides insulation for the body, fills body crevices, and protects body organs.

Studies of adipocytes in the laboratory have shown that fully differentiated cells do not divide. However, such cells have a long life span, and anyone born with large numbers of adipocytes runs the risk of becoming obese. Some immature adipocytes capable of division are present in postnatal life; these cells respond to estrogen stimulation and are the potential source of additional fat cells during postnatal life.¹ Fat deposition results from proliferation of these existing immature adipocytes and can occur as a consequence of excessive caloric intake when a woman is breast-feeding or during estrogen stimulation around the time of puberty. An increase in fat cells also may occur during late adolescence and in middle-aged persons who already are overweight.

There are two types of adipose tissue: white fat and brown fat. White fat, which despite its name is cream colored or yellow, is the prevalent form of adipose tissue in postnatal life. It constitutes 10% to 20% of body weight in adult males and 15% to 25% in adult females. At body temperature, the lipid content of fat cells exists as an oil. It consists of triglycerides, which are three molecules of fatty acids esterified to a glycerol molecule. Triglycerides, which contain no water, have the highest caloric content of all nutrients and are an efficient form of energy storage. Fat cells synthesize triglycerides, the major fat storage form, from dietary fats and carbohydrates. Insulin is required for transport of glucose into fat cells. When calorie intake is restricted for any reason, fat cell triglycerides are broken down, and the resultant fatty acids and glycerol are released as energy sources.

Brown fat differs from white fat in terms of its thermogenic capacity or ability to produce heat. Brown fat, the site of diet-induced thermogenesis and nonshivering thermogenesis, is found primarily in early neonatal life in humans and in animals that hibernate. In humans, brown fat decreases with age but is still detectable in the sixth decade. This small amount of brown fat has a minimal effect on energy expenditure.

Nutritional Needs

Recommended Dietary Allowances and Dietary Reference Intakes

The *Recommended Daily Allowances* (RDAs) define the average daily intakes that meet the nutrient needs of almost all healthy persons in a specific age and sex group.² The RDAs, which are periodically updated, have been published since 1941 by the National Academy of Sciences. The RDA is used in advising persons about the level of nutrient intake they need to decrease the risk of chronic disease.

The Dietary Reference Intake (DRI) includes a set of at least four nutrient-based reference values, each of which has specific uses: the RDA, the Adequate Intake, the Estimated Average Requirement, and the Tolerable Upper Intake Level.³ The United States RDA (USRDA) was established for the purpose of labeling foods. It takes the highest recommended daily intake of each nutrient for children older than 4 years of age and for adults (excluding those pregnant or lactating); therefore, the USRDA sometimes provides a margin of nutritional safety higher than the RDA.

The Adequate Intake (AI) is set when there is not enough scientific evidence to estimate an average requirement. An *Estimated Average Requirement* is the intake that meets the estimated nutrient need of half of the persons in a specific group. The *Tolerable Upper Intake Level* is the maximum intake that is judged unlikely to pose a health risk in almost all healthy persons in a specific group. The DRIs are regularly reviewed and updated by the Food and Nutrition Board of the Institute of Medicine and the National Academy of Science.

Proteins, fats, carbohydrates, vitamins, and minerals each have their own function in providing the body with what it needs to maintain life and health. Recommended allowances have not been established for every nutrient; some are given as a safe and adequate intake, but others, such as carbohydrates and fats, are expressed as a percentage of the calorie intake.

Calories

Energy requirements are greater during growth periods. Infants require approximately 115 kcal/kg at birth, 105 kcal/kg at 1 year, and 80 kcal/kg of body weight between 1 to 10 years of age. During adolescence, boys require 45 kcal/kg of body weight and girls require 38 kcal/kg of body weight. During pregnancy, a woman needs an extra 300 kcal/day above her usual requirement, and during the first 3 months of breastfeeding, she requires an additional 500 kcal.² Table 29-1 can be used to predict the caloric requirements of healthy adults.

TABLE 29-1Caloric RequirementsBased on Body Weight andActivity Level						
	Sedentary	Moderate	Active			
Overweight	20–25 kcal/kg	30 kcal/kg	35 kcal/kg			
Normal	30 kcal/kg	35 kcal/kg	40 kcal/kg			
Underweight	30 kcal/kg	40 kcal/kg	45–50 kcal/kg			

(Adapted from Goodhart R.S., Shils M.E. [1980]. Modern nutrition in health and disease [6th ed.]. Philadelphia: Lea and Febiger)

Proteins, Fats, and Carbohydrates

Proteins are required for growth and maintenance of body tissues, formation of enzymes and antibodies, fluid and electrolyte balance, and nutrient transport. Proteins are composed of amino acids, nine of which are essential to the body. These are leucine, isoleucine, methionine, phenylalanine, threonine, tryptophan, valine, lysine, and histidine. The foods that provide these essential amino acids in adequate amounts are milk, eggs, meat, fish, and poultry. Dried peas and beans, nuts, seeds, and grains contain all the essential amino acids but in less than adequate proportions. The proteins in these foods need to be combined with each other or with complete proteins to meet the amino acid requirements for protein synthesis.

Unlike carbohydrates and fats, which are composed of hydrogen, carbon, and oxygen, proteins contain 16% nitrogen; therefore, nitrogen excretion is an indicator of protein intake. If the amount of nitrogen taken in by way of protein is equivalent to the nitrogen excreted, the person is said to be in nitrogen balance. A person is in positive nitrogen balance when the nitrogen consumed by way of protein is greater than the amount excreted. This occurs during growth, pregnancy, or healing after surgery or injury. A negative nitrogen balance often occurs with fever, illness, infection, trauma, or burns, when more nitrogen is excreted than is consumed. It represents a state of tissue breakdown.

Dietary fats are composed primarily of triglycerides (*i.e.*, a mixture of fatty acids and glycerol). The fatty acids are saturated (*i.e.*, no double bonds), monounsaturated (*i.e.*, one double bond), or polyunsaturated (*i.e.*, two or more double bonds). The saturated fatty acids elevate blood cholesterol, whereas the monounsaturated and polyunsaturated fats lower blood cholesterol. Saturated fats usually are from animal sources and remain solid at room temperature. With the exception of coconut and palm oils (which are saturated), unsaturated fats are found in plant oils and usually are liquid at room temperature.

Dietary fats provide energy, serve as carriers for the fatsoluble vitamins, are precursors of prostaglandins, and are a source of fatty acids. The polyunsaturated fatty acid linoleic acid is the only fatty acid that is required. A deficiency of linoleic acid results in dermatitis. The daily requirement is 5 g or 1% to 2% of the total daily calories. Because vegetable oils are rich sources of linoleic acid, this level can be met by including two teaspoons of oil.

Other than the requirement for linoleic acid, there is no specific requirement for dietary fat, provided there is adequate nutrition available for energy. Fat is the most concentrated source of energy. It is recommended that 30% or less of the calories in the diet should come from fats.

Cholesterol is the major constituent of cell membranes and is synthesized by the body. Cholesterol metabolism and transport are discussed in Chapter 15. The daily dietary recommendation for cholesterol is less than 300 mg.

Dietary carbohydrates are composed of simple sugars, complex carbohydrates, and undigested carbohydrates (*i.e.*, fiber). Within the body, carbohydrate is transformed into glucose, a six-carbon molecule. Excess glucose is stored as glycogen or converted to triglycerides for storage in fat cells. Because of their vitamin, mineral, and fiber content, it is recommended that the bulk of the carbohydrate content in the diet be in the complex form, rather than as simple sugars that contain few nutrients. Sucrose (*i.e.*, table sugar) is implicated in the development of dental caries.

There is no specific dietary requirement for carbohydrates. All of the energy requirements can be met by dietary fats and proteins. Although some tissues, such as the nervous system, require glucose as an energy source, this need can be met through the conversion of amino acids and the glycerol part of the triglyceride molecule to glucose. A carbohydrate-deficient diet usually results in the loss of tissue proteins and the development of ketosis. Because protein and fat metabolism increases the production of osmotically active metabolic wastes that must be eliminated through the kidneys, there is a danger of dehydration and electrolyte imbalances. The amount of carbohydrate needed to prevent tissue wasting and ketosis is 50 to 100 g/day. In practice, most of the daily energy requirement should be from carbohydrate. This is because protein is an expensive source of calories and because it is recommended that no more than 30% of the calories in the diet be derived from fat. The current recommendation is that the diet should provide 50% to 60% of the calories as carbohydrates.

Vitamins and Minerals

Vitamins are a group of organic compounds that act as catalysts in various chemical reactions. A compound cannot be classified as a vitamin unless it is shown that a deficiency of it causes disease. Contrary to popular belief, vitamins do not provide energy directly. As catalysts, they are part of the enzyme systems required for the release of energy from protein, fat, and carbohydrates. Vitamins also are necessary for the formation of red blood cells, hormones, genetic materials, and the nervous system. They are essential for normal growth and development.

There are two types of vitamins: fat soluble and water soluble. The four fat-soluble vitamins are vitamins A, D, E, and K. The nine required water-soluble vitamins are thiamine, ribo-flavin, niacin, pyridoxine (Vitamin B_6), pantothenic acid, vitamin B_{12} , folic acid, biotin, and vitamin C. Because the water-soluble vitamins are excreted in the urine, it is less likely that they may become toxic to the body, but the fat-soluble vitamins are stored in the body, and they may reach toxic levels.

Minerals serve many functions. They are involved in acidbase balance and in the maintenance of osmotic pressure in body compartments. Minerals are components of vitamins, hormones, and enzymes. They maintain normal hemoglobin levels, play a role in nervous system function, and are involved in muscle contraction and skeletal development and maintenance. Minerals that are present in relatively large amounts in the body are called *macrominerals*. These include calcium, phosphorus, sodium, chloride, potassium, magnesium, and sulfur. The remainder are classified as *trace minerals*; they include iron, manganese, copper, iodine, zinc, cobalt, fluorine, and selenium.

Regulation of Food Intake and Energy Storage

Stability of the weight and body composition over time requires that energy intake matches energy utilization. Environmental, cultural, genetic, and psychological factors all influence food intake and energy expenditure. However, powerful physiological control systems also regulate hunger and food intake.¹

Hunger, Appetite, and Food Intake

The sensation of *hunger* is associated with several sensory perceptions, such as the rhythmic contractions of the stomach and that "empty feeling" in the stomach that stimulates a person to seek food. A person's *appetite* is the desire for a particular type of food. It is useful in helping the person determine the type of food that is eaten. Satiety is the feeling of fullness or decreased desire for food.

The hypothalamus contains centers for hunger and satiety (Fig 29-1). It receives neural input from the gastrointestinal tract that provides information about stomach filling, chemical signals from the blood about the nutrients in food, and input from the cerebral cortex regarding the smell, sight, and taste of the food. Centers in the hypothalamus also control the secretion of several hormones (*e.g.*, thyroid and adrenocortical hormones) that regulate energy balance and metabolism.

The control of food intake can be divided into short-term regulation, which is concerned with the amount of food that is consumed at a meal or snack, and intermediate and long-term regulation, which is concerned with the maintenance of energy stores over time.¹

The short-term regulation of food intake provides a person with the feeling of satiety and turns off the desire for eating when adequate food has been consumed. It requires rapid feedback mechanisms that signal the adequacy of food intake be-



fore digestion has taken place and nutrients have been absorbed into the blood. These mechanisms include receptors that monitor filling of the gastrointestinal tract, gastrointestinal tract hormones, and oral receptors that monitor food intake. Stretch receptors in the gastrointestinal tract monitor gastrointestinal filling and send inhibitory impulses by way of the vagus nerve to the feeding center to suppress the desire for food. The gastrointestinal hormone cholecystokinin, which is released in response to fat in the duodenum, has a strong suppressant effect on the feeding center. The presence of food in the stomach increases the release of insulin and glucagon, both of which suppress the neurogenic feeding signals from the brain.¹ The act of tasting, chewing, and swallowing also appears to suppress the feeling of hunger.

The intermediate and long-term regulation of food intake is determined by the amount of nutrients that are in the blood and in storage sites. It has long been known that a decrease in blood glucose causes hunger. In contrast, an increase in breakdown products of lipids such as ketoacids produces a decrease in appetite. The ketogenic weight-loss diet relies on the appetite suppressant effects of ketones in the blood. Recent evidence suggests that the hypothalamus also senses the amount of energy that is stored in fat cells through a hormone called *leptin*. Increased amounts of leptin are released from the adipocytes when fat stores are increased. The stimulation of leptin receptors in the hypothalamus produces a decrease in appetite and food intake as well as an increase in metabolic rate and energy consumption. It also produces a decrease in insulin release from the beta cells, which decreases energy storage in fat cells.

Assessment of Energy Stores and Nutritional Status

Anthropometric measurements provide a means for assessing body composition, particularly fat stores and skeletal muscle mass. This is done by measuring height, weight, body circumferences, and thickness of various skinfolds. These measurements commonly are used to determine growth patterns in children and appropriateness of current weight in adults.

Body weight is the most frequently used method of assessing nutritional status; it should be used in combination with measurements of body height to establish whether a person is underweight or overweight.

Relative weight is the actual weight divided by the desirable weight and multiplied by 100. A relative weight greater than 120% is indicative of obesity. Recent changes in weight are probably a better indication of undernutrition than a low relative weight. An unintentional loss of 10% of body weight or more within the past 6 months usually is considered predictive of a poor clinical outcome, especially if weight loss is continuing.⁴ The body mass index (BMI) uses height and weight to determine healthy weight (Table 29-2). It is calculated by dividing the weight in kilograms by the height in meters squared (BMI = weight [kg]/height [m²]). A BMI between 18.5 and 25 has the lowest statistical health risk.⁵ A BMI of 25 to 29.9 is considered overweight; a BMI of 30 or greater as obese; and a BMI greater than 40 as very or morbidly obese.⁶

Body weight reflects both lean body mass and adipose tissue and cannot be used as a method for describing body composition or the percentage of fat tissue present. Statistically, the best percentage of body fat for men is between 12% and

TABLE 29-2 Classification of Overweight and Obesity by BMI, Waist Circumference, and Associated Disease Risk*						
			Disease Risk* Relative to Normal Weight and Waist Circumference			
	BMI (kg/m²)	Obesity Class	Men ≤102 cm (≤40 in) Women ≤88 cm (≤35 in)	Men >102 cm (>40 in) Women >88 cm (>35 in)		
Underweight	<18.5		_	_		
Normal [†]	18.5–24.9		_	_		
Overweight	25.0-29.9		Increased	High		
Obesity	30.0-34.9		High	Very high		
	35.0-39.9	11	Very high	Very high		
Extreme obesity	≥40	Ш	Extremely high	Extremely high		

BMI, body mass index.

*Disease risk for type 2 diabetes, hypertension, and cardiovascular disease.

[†]Increased waist circumference also can be a marker for increased risk, even in persons of normal weight.

(Expert Panel. [1998]. Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in

adults. National Institutes of Health. [On-line.] Available: http://nhlbi.nih.gov/guidelines/ob_gdlns.htm.

20%, and for women, it is between 20% and 30%.⁷ During physical training, body fat usually decreases and lean body mass increases.

Among the methods used to estimate body fat are skinfold thickness, body circumferences, bioelectrical impedance, computed tomography (CT), and magnetic resonance imaging (MRI). The measurement of *body circumferences*, most commonly waist and hip, provides an objective measurement of body fat and supplies the information needed to calculate the waist circumference to hip circumference. The measurement of body circumference has received attention because of an interest in excess visceral or intra-abdominal fat. The waist circumference is commonly used for this purpose.⁶

Bioelectrical impedance involves the use of electrodes attached to the wrist and ankles to send a harmless current through the body. The flow of the current is affected by the amount of water in the body. Because fat-free tissue contains virtually all the water and current-conducting electrolytes, measurements of the resistance (*i.e.*, impedance) can be used to estimate the percentage of body fat present.⁸

Computed tomography and *MRI* can be used to provide quantitative pictures from which the thickness of fat can be determined. CT scans also can be used to provide quantitative estimates of regional fat and give a ratio of intra-abdominal to extra-abdominal fat. Because these methods are costly, they usually are reserved for research studies.

Various laboratory tests can aid in evaluating nutritional status. Some of the most commonly performed tests are serum albumin to assess the protein status, total lymphocyte count and delayed hypersensitivity reaction to assess cellular immunity, and creatinine-height index to assess skeletal muscle protein.

In summary, nutritional status describes the condition of the body related to the availability and use of nutrients. Metabolism is the organized process whereby nutrients such as carbohydrates, fats, and proteins are broken down, transformed, or otherwise converted to cellular energy. Glucose, fats, and amino acids from proteins serve as fuel sources for cellular metabolism. These fuel sources are ingested during meals and stored for future use. Glucose is stored as glycogen or converted to triglycerides in fat cells for storage. Fats are stored in adipose tissue as triglycerides. Amino acids are the building blocks of proteins, and most of the stored amino acids are contained in body proteins and as fuel sources for cellular metabolism. Energy is measured in heat units called *kilocalories*. The RDA is the recommended daily allowance needed to meet the known nutritional needs of healthy persons.

Nutritional status reflects the continued daily intake of nutrients over time and the deposition and use of these nutrients in the body. When a person is consistently overfed, the excess energy is stored as fat, and the person gains weight. When energy expenditure exceeds food intake, body fat and other tissues are broken down, and the person loses weight. The nutritional status of a person can be assessed by evaluation of dietary intake, anthropometric measurements, health assessment, and laboratory tests. Anthropometric measurements are used for assessing body composition; they include height and weight measurements and measurements to determine the composition of the body in relation to lean body mass and fat tissue (*e.g.*, skinfold thickness, body circumferences, bioelectrical impedance, and CT scans).

OVERNUTRITION AND OBESITY

Obesity is defined as a condition characterized by excess body fat. Clinically, obesity and overweight have been defined in terms of the BMI. Historically, various world organizations have used different BMI cutoff points to define obesity. In 1997, the World Health Organization defined the various classifications of overweight (BMI \ge 25) and obesity (BMI \ge 30). This classification system was subsequently adopted by the National Institutes of Health.⁶ The use of a BMI cutoff of 25 as a measure of overweight raised some concern that the BMI in some men might be attributable to muscle, rather than fat, weight. However, it has been shown that a BMI cutoff of 25 can sensitively detect most overweight people and does not erroneously detect overlean people.⁹

Overweight and obesity have become national health problems, increasing the risk of hypertension, hyperlipidemia, type 2 diabetes, coronary heart disease, and other health problems. Fifty-five percent of the U.S. population is estimated to be overweight (BMI \ge 25). Obesity is particularly prevalent among some minority groups, lower income groups, and people with less education. The prevalence of obesity (BMI \ge 30) in the United States has increased from 12.0% in 1991 to 17.9% in 1998.¹⁰

Causes of Obesity

The excess body fat of obesity often significantly impairs health. This excess body fat is generated when the calories consumed exceed those expended through exercise and activity.¹¹ The physiologic mechanisms that lead to this imbalance are poorly understood. They probably exist in different combinations among obese persons.

Although factors that lead to the development of obesity are not understood, they are thought to involve the interaction of the person's genotype with environmental influences such as social, behavioral, and cultural factors.¹² Obesity is known to run in families, suggesting a hereditary component. The question that surrounds this observation is whether the disorder arises because of genetic endowment or environmental influences. Studies of twin and adopted children have provided evidence that heredity contributes to the disorder.¹³ It is now believed that the heritability of the BMI is approximately 33%.¹⁴

Although genetic factors may explain some of the individual variations in terms of excess weight, environmental influences also must be taken into account. These influences include family dietary patterns, decreased level of activity because of labor-saving devices and time spent using the computer, re-

KEY CONCEPTS

OBESITY

- Obesity results from an imbalance between energy intake and energy consumption. Because fat is the main storage form of energy, obesity represents an excess of body fat.
- Overweight and obesity are determined by measurements of body mass index (BMI; weight [kg]/height [m²]) and waist circumference. A BMI of 25 to 29.9 is considered overweight; a BMI of 30 or greater, obese; and a BMI greater than 40, very or morbidly obese.
- Waist circumference is used to determine the distribution of body fat. Central, or abdominal, obesity is an independent predictor of morbidity and mortality associated with obesity.

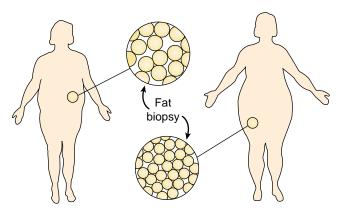
liance on the automobile for transportation, easy access to food, energy density of food, and large food portions. The obese may be greatly influenced by the availability of food, the flavor of food, time of day, and other cues. The composition of the diet also may be a causal factor, and the percentage of dietary fat independent of total calorie intake may play a part in the development of obesity. Psychological factors include using food as a reward, comfort, or means of getting attention. Eating may be a way to cope with tension, anxiety, and mental fatigue. Some persons may overeat and use obesity as a means of avoiding emotionally threatening situations.

Types of Obesity

Two types of obesity based on distribution of fat have been described: upper body and lower body obesity. *Upper body obesity* is also referred to as *central, abdominal,* or *male* obesity. Lower body obesity is known as *peripheral, gluteal-femoral,* or *female* obesity. The obesity type is determined by dividing the waist by the hip circumference. A waist-hip ratio greater than 1.0 in men and 0.8 in women indicates upper body obesity (Fig. 29-2).¹⁵ Research suggests that fat distribution may be a more important factor for morbidity and mortality than overweight or obesity.

The presence of excess fat in the abdomen out of proportion to total body fat is an independent predictor of risk factors and mortality. Waist circumference is positively correlated with abdominal fat content. Waist circumference 35 inches or greater in women and 40 inches or greater in men has been associated with increased health risk⁶ (see Table 29-2). Central obesity can be further differentiated into intra-abdominal, or visceral, fat and subcutaneous fat by the use of CT or MRI scans. However, intra-abdominal fat usually is synonymous with central fat distribution. One of the characteristics of intra-abdominal fat is that fatty acids released from the viscera go directly to the liver before entering the systemic circulation, having a potentially greater impact on hepatic function. Higher levels of circulating free fatty acids in obese persons, particularly those with upper body obesity, are thought to be associated with many of the adverse effects of obesity.11

In general, men have more intra-abdominal fat and women more subcutaneous fat. As men age, the proportion of intra-



■ FIGURE 29-2 ■ Distribution of body fat and size of fat cells in persons with upper and lower body obesity. (Courtesy of Ahmed Kissebah, M.D., Ph.D., Medical College of Wisconsin, Milwaukee)

abdominal fat to subcutaneous fat increases. After menopause, women tend to acquire more central fat distribution. Increasing weight gain, alcohol, and low levels of activity are associated with upper body obesity. These changes place persons with upper body obesity at greater risk for ischemic heart disease, stroke, and death independent of total body fat. They also tend to exhibit hypertension, elevated levels of triglycerides and decreased levels of high-density lipoproteins, hyperinsulinemia and diabetes mellitus, breast and endometrial cancer, gallbladder disease, menstrual irregularities, and infertility. Visceral fat also is associated with abnormalities of metabolic and sex hormone levels.15

Weight reduction causes a loss of intra-abdominal fat and has resulted in improvements in metabolic and hormonal abnormalities.¹⁶ In terms of weight reduction, some studies have shown that persons with upper body obesity are easier to treat than those with lower body obesity. Other studies have shown no difference in terms of success with weight reduction programs between the two types of obesity.

Weight cycling (the losing and gaining of weight) has been found to have little or no effect on metabolic variables, central obesity, or cardiovascular risk factors or future amount of weight loss.¹⁷ More research is needed to determine its effect on dietary preference for fat, psychological adjustment, disordered eating, and mortality.

Health Risks Associated With Obesity

Obese persons are more likely to have high blood pressure, hyperlipidemia, cardiovascular disease, glucose intolerance, insulin resistance, type 2 diabetes, stroke, gallbladder disease, infertility, and cancer of the endometrium, prostate, colon, and in postmenopausal women, the breast.6 The increased weight associated with obesity stresses the bones and joints, increasing the likelihood of arthritis. Other conditions associated with obesity include sleep apnea and pulmonary dysfunction, nonalcoholic steatohepatitis, carpal tunnel syndrome, venous insufficiency and deep vein thrombosis, and poor wound healing.⁶ Because some drugs are lipophilic and exhibit increased distribution in fat tissue, the administration of these drugs, including some anesthetic agents, can be more dangerous in obese persons. If surgery is required, the obese person tends to heal more slowly than a nonobese person of the same age.

Massive obesity, because of its close association with so many health problems, can be regarded as a disease in its own right.18 It is the second leading cause of preventable death. In men who have never smoked, the risk of death increases from 1.06 at a BMI of 24.5 to 1.67 at a BMI higher than 26.19

Prevention and Treatment

It has been theorized that obesity is preventable because the effect of hereditary factors is no more than moderate. A more active lifestyle together with a low-fat diet (<30% of calories) is seen as the strategy for prevention. The target audience should be young children, adolescents, and young adults.²⁰ Tools needed to achieve this goal include promotion of regular lowfat meals, avoidance of snacking, substitution of water for calorie-containing beverages, reduction of time spent in sedentary activities, such as television viewing, and an increase in activity. Other high-risk periods are from 25 to 35 years, at menopause,²⁰ and the year after successful weight loss.

The goals for weight loss are, at a minimum, prevention of further weight gain, reduction of current weight, and maintenance of a lowered body weight indefinitely. An algorithm has been designed for use in treating overweight and obesity. The initial goal in treatment is to lower body weight by 10% from baseline during a 6-month period. This degree of weight loss requires a calorie reduction of 300 to 500 kcal/day in individuals with a BMI of 27 to 35. For those persons with a BMI greater than 35, the calorie intake needs to be reduced by 500 to 1000 kcal/day. After 6 months, the person should be given strategies for maintaining the new weight. The person who is unable to achieve significant weight loss should be enrolled in a weight management program to prevent further weight gain.

There are many ways to treat obesity. It is currently recommended that treatment should focus on lifestyle modification through a combination of a low-calorie diet, increased physical activity, and behavior therapy.⁶

Dietary therapy should be individually prescribed based on the person's overweight status and risk profile. The diet should be a personalized plan that is 300 to 1000 kcal/day less than the current dietary intake. If the patient's risk status warrants it, the diet also should be decreased in saturated fat and contain 30% or less of total calories from fat. Reduction of dietary fat without a calorie deficit will not result in weight loss.

Physical activity is important in the prevention of weight gain. In addition, it reduces cardiovascular and diabetes risk beyond that achieved by weight loss alone. Although physical activity is an important part of weight loss therapy, it does not lead to a significant weight loss. Exercise should be started slowly, with the duration and intensity increased independent of each other. The goal should be 30 minutes or more of moderate-intensity activity on most days of the week. The activity can be performed at one time or intermittently over the day.

Techniques for changing behavior include self-monitoring of eating habits and physical activity, stress management, stimulus control, problem solving, contingency management, cognitive restructuring, and social support.

Pharmacotherapy and surgery are available as an adjunct to lifestyle changes in individuals who meet specific criteria.²¹ Pharmacotherapy is usually considered only after combined diet, exercise, and behavioral therapy has been in effect for a minimum of 6 months. Weight loss surgery is limited to persons with a BMI greater than 40; those with a BMI greater than 35 who have comorbid conditions and in whom efforts at medical therapy have failed; and those who have complications of extreme obesity.

Ē Childhood Obesity

Obesity is the most prevalent nutritional disorder affecting the pediatric population in the United States. The findings from the third National Health and Nutrition Examination Survey (NHANES III), conducted between 1988 and 1994, showed that 14% of children and 12% of adolescents were overweight.²² The definition for overweight for the NHANES III study was a BMI at or above the sex- and age-specific 95th percentile. Another definition of overweight in children is a weight for height greater than 120% of ideal when controlled for age and sex.23 Children who are 120% of weight expected for their height are overweight, but they are only considered to have excess fat when the triceps skinfold thickness is greater than the 85th percentile. This distinction is important in preventing misdiagnosis of obesity and creating anxiety for the parents and the child.

The major concern of childhood obesity is that obese children will grow up to become obese adults. Health care providers are beginning to see hypertension, dyslipidemia, and type 2 diabetes in obese children and adolescents.²⁴ In addition, there is a growing concern that childhood and adolescent obesity may be associated with negative psychosocial consequences, such as low self-esteem and discrimination by adults and peers.²⁵

Childhood obesity is determined by a combination of hereditary and environmental factors. It is associated with obese parents, higher socioeconomic status, increased parental education, small family size, and sedentary lifestyle.²⁶ Children with overweight parents are at highest risk; the risk for those with two overweight parents is much higher than for children in families in which neither parent is obese. One of the trends leading to childhood obesity is the increase in inactivity. Increasing perceptions that neighborhoods are unsafe has resulted in less time spent outside playing and walking and more time spent indoors engaging in sedentary activities, such as television viewing. Television viewing is associated with consumption of calorie-dense snacks and decreased indoor activity. Obese children also may have a deficit in recognizing hunger sensations, stemming perhaps from parents who use food as gratification.

Because adolescent obesity is predictive of adult obesity, treatment of childhood obesity is desirable. Weight loss without adverse health effects and maintenance of that loss are the goals. Each child should be assessed and treated individually. In young children who have mild to moderate weight problems, weight maintenance or a reduced rate of weight gain is sufficient. Studies indicate that physical activity in combination with diet therapy is more effective than diet therapy alone.

When weight loss is required, a loss of 1 pound per month is reasonable together with permanent changes in food consumption and activity. However, the focus should be on normalizing food intake, particularly fat intake, and increasing physical activity. If weight gain can be slowed or maintained during growth, lean body mass increases, and some of the abnormal metabolic effects of obesity may be reversed. Family members need to be involved so they can learn to provide appropriate support and assist the child in taking responsibility for his or her own actions. Highly restrictive diets should be limited to the rare child or adolescent who has morbid complications. They should not be used for children or adolescents with renal, liver, or cardiac disease. These diets should contain a minimum of 2 g protein/kg body weight.²² There should be close monitoring for sustained nitrogen losses, cardiac dysrhythmias, and cholelithiasis. Commercial diets are not recommended.

In summary, obesity is defined as excess body fat resulting from the consumption of calories in excess of those expended for exercise and activities. Heredity, socioeconomic, cultural, and environmental factors, psychological influences, and activity levels have been implicated as causative factors in the development of obesity. The health risks associated with obesity include hypertension and cardiovascular disease, hyperlipidemia, insulin resistance and type 2 diabetes mellitus, menstrual irregularities and infertility, cancer of the endometrium, breast, prostate, and colon, and gallbladder disease. There are two types of obesity—upper body and lower body obesity. Upper body obesity is associated with a higher incidence of complications. The treatment of obesity focuses on nutritionally adequate weight-loss diets, behavior modification, exercise, social support, and in situations of marked obesity, surgical methods. Obesity is the most prevalent nutritional disorder affecting the pediatric population in the United States.

UNDERNUTRITION

Undernutrition ranges from the selective deficiency of a single nutrient to starvation in which there is deprivation of all ingested nutrients. Undernutrition can result from eating disorders; lack of food availability; or health problems that impair food intake and decrease its absorption and use. Weight loss and malnutrition are common during illness, recovery from trauma, and hospitalization.

The prevalence of malnutrition in children is substantial. Globally, nearly 195 million children younger than 5 years of age are undernourished.²⁷ Malnutrition is most obvious in developing countries of the world, where the condition takes severe forms. Even in developed nations, malnutrition remains a problem. In 1992, it was estimated that 12 million American children consumed diets that were significantly below the recommended allowances of the National Academy of Sciences.²⁷

Protein and Calorie Malnutrition

Malnutrition and starvation are conditions in which a person does not receive or is unable to use an adequate amount of calories and nutrients for body function. Among the many causes of starvation, some are willful, such as the person with anorexia nervosa who does not consume enough food to maintain weight and health, and some are medical, such as persons with Crohn's disease who are unable to absorb their food. Most cases of food deprivation result in semistarvation with protein and calorie malnutrition.

Protein-calorie malnutrition, also referred to as *marasmus*, is characterized by loss of muscle and fat stores. Marasmus is characterized by progressive wasting from inadequate food intake that is equally deficient in calories and protein. The person appears emaciated, with sparse, dry, and dull hair and depressed heart rate, blood pressure, and body temperature. The female experiences anovulation and amenorrhea, and the male experiences decreased testicular function. The child with marasmus has a wasted appearance, with stunted growth and loss of subcutaneous fat, but with relatively normal skin, hair, liver function, and affect.

Kwashiorkor is caused by protein deficiency. The term *kwashiorkor* comes from the African word meaning "the disease suffered by the displaced child," because the condition devel-

ops soon after a child is displaced from the breast after the arrival of a new infant and placed on a starchy gruel feeding. The child with kwashiorkor is characterized by edema, desquamating skin, discolored hair, enlarged abdomen, anorexia, and extreme apathy. The serum albumin level is less that 3.0 g/dL, and there is pitting edema of the extremities. There is less weight loss and wasting of skeletal muscles than in marasmus. Other manifestations include skin lesions, easily pluckable hair, enlarged liver and distended abdomen, cold extremities, and decreased cardiac output and tachycardia.

Marasmus-kwashiorkor is an advanced protein-calorie deficit together with increased protein requirement or loss. This results in a rapid decrease in anthropometric measurements with obvious edema and wasting and loss of organ mass.

Malnutrition and Wasting in Illness

Malnutrition and wasting are common in persons with trauma, sepsis, and serious illnesses such as cancer and acquired immunodeficiency syndrome. Approximately half of all persons with cancer experience tissue wasting in which the tumor induces metabolic changes, leading to a loss of adipose tissue and muscle mass.²⁸ In healthy adults, body protein homeostasis is maintained by a cycle in which the net loss of protein in the postabsorptive state is matched by a net postprandial gain of protein.²⁹ In persons with severe injury or illness, net protein breakdown is accelerated and protein rebuilding disrupted. Consequently, these persons may lose up to 20% of body protein, much of which originates in skeletal muscle.²⁹ Protein mass is lost from the liver, gastrointestinal tract, kidneys, and heart. As protein is lost from the liver, hepatic synthesis of serum proteins decreases and decreased levels of serum proteins are observed. There is a decrease in immune cells and those needed for wound healing. The lungs are affected primarily by weakness and atrophy of the respiratory muscles. The gastrointestinal tract undergoes mucosal atrophy with loss of villi in the small intestine, resulting in malabsorption. The loss of protein from cardiac muscle leads to a decrease in myocardial contractility and cardiac output.

In hospitalized patients, malnutrition increases morbidity and mortality rates, incidence of complications, and length of stay. Malnutrition may present at the time of admission or develop during hospitalization. The hospitalized patient often finds eating a healthful diet difficult and commonly has restrictions on food and water intake in preparation for tests and surgery. Pain, medications, special diets, and stress can decrease appetite. Even when the patient is well enough to eat, being alone in a room where unpleasant treatments may be given is not conducive to eating. Although hospitalized patients may appear to need fewer calories because they are on bed rest, their actual need for caloric intake may be higher because of other energy expenditures. For example, more calories are expended during fever, when the metabolic rate is increased. There also may be an increased need for protein to support tissue repair after trauma or surgery.

Treatment

The treatment of severe protein-calorie malnutrition involves the use of measures to correct fluid and electrolyte abnormalities and replenish proteins, calories, and micronutrients.³⁰ Treatment is started with modest quantities of proteins and calories based on the person's actual weight. Concurrent administration of vitamins and minerals is needed. Either the enteral or parenteral route can be used. The treatment should be undertaken slowly to avoid complications. The administration of water and sodium with carbohydrates can overload a heart that has been weakened by malnutrition and result in congestive failure. Enteral feedings can result in malabsorptive symptoms caused by abnormalities in the gastrointestinal tract. Refeeding edema is benign dependent edema that results from renal sodium reabsorption and poor skin and blood vessel integrity. It is treated by elevation of the dependent area and modest sodium restrictions. Diuretics are ineffective and may aggravate electrolyte deficiencies.

Eating Disorders

Eating disorders affect an estimated 5 million Americans each year.³¹These illnesses, which include anorexia nervosa, bulimia nervosa, and binge-eating disorder and their variants, incorporate serious disturbances in eating, such as restriction of intake and binging, with an excessive concern over body shape or body weight. Eating disorders typically occur in adolescent girls and young women, although 5% to 15% of cases of anorexia nervosa and 40% of cases of binge-eating disorder occur in boys and men.³¹ The mortality rate associated with anorexia nervosa, 0.56% per year, is more than 12 times the mortality rate among young women in the general population.³¹

Eating disorders are more prevalent in industrialized societies and occur in all socioeconomic and major ethnic groups. A combination of genetic, neurochemical, developmental, and sociocultural factors is thought to contribute to the development of the disorders. The American Psychiatric Society's

KEY CONCEPTS

- Eating disorders are serious disturbances in eating, such as willful restriction of intake and binge eating, as well as excessive concern over body weight and shape.
- Anorexia nervosa is characterized by a refusal to maintain a minimally normal body weight (*e.g.*, at least 85% of minimal expected weight); an excessive concern over gaining weight and how the body is perceived in terms of size and shape; and amenorrhea (in girls and women after menarche).
- Bulimia nervosa is characterized by recurrent binge eating; inappropriate compensatory behaviors, such as self-induced vomiting, fasting, or excessive exercise that follow the binge-eating episode; and extreme concern over body shape and weight.
- Binge eating consists of consuming unusually large quantities of food during a discrete period (*e.g.*, within any 2-hour period) along with a lack of control over the binge-eating episode.

Diagnostic and Statistical Manual of Mental Disorders, Text Revision (DSM-IV-TR) has established criteria for the diagnosis of anorexia nervosa and bulimia nervosa.32 Although these criteria allow clinicians to make a diagnosis in persons with a specific eating disorder, the symptoms often occur along a continuum between those of anorexia nervosa and bulimia nervosa. Preoccupation with weight and excessive self-evaluation of weight and shape are common to both disorders, and persons with eating disorders may demonstrate a mixture of both disorders.³² The female athlete triad, which includes disordered eating, amenorrhea, and osteoporosis, does not meet the strict DSM-IV-TR criteria for anorexia nervosa or bulimia nervosa but shares many of the characteristics and therapeutic concerns of the two disorders (see Chapter 43). Persons with eating disorders may require concomitant evaluation for psychiatric illness because eating disorders often are accompanied by mood, anxiety, and personality disorders. Suicidal behavior may accompany anorexia nervosa and bulimia nervosa and should be excluded.31

Anorexia Nervosa

Anorexia nervosa was first described in the scientific literature more than 100 years ago by Sir William Gull.33 The DSM-IV-TR diagnostic criteria for anorexia nervosa are (1) a refusal to maintain a minimally normal body weight for age and height (e.g., $\leq 85\%$ of minimal expected weight or BMI ≥ 17.5); (2) an intense fear of gaining weight or becoming fat; (3) a disturbance in the way one's body size, weight, or shape is perceived; and (4) amenorrhea (in girls and women after menarche).³² Anorexia nervosa is more prevalent among young women than men. The disorder typically begins in teenage girls who are obese or perceive themselves as being obese. An interest in weight reduction becomes an obsession, with severely restricted caloric intake and frequently with excessive physical exercise. The term *anorexia*, meaning "loss of appetite," is a misnomer because hunger is felt but in this case is denied.

Many organ systems are affected by the malnutrition that occurs in persons with anorexia nervosa. The severity of the abnormalities tends to be related to the degree of malnutrition and is reversed by refeeding. The most frequently occurring complication of anorexia is amenorrhea and loss of secondary sex characteristics with decreased levels of estrogen, which can eventually lead to osteoporosis. Bone loss can occur in young women after as short a period of illness as 6 months.³¹ Symptomatic compression fractures and kyphosis have been reported. Constipation, cold intolerance and failure to shiver in cold, bradycardia, hypotension, decreased heart size, electrocardiographic changes, blood and electrolyte abnormalities, and skin with lanugo (*i.e.*, increased amounts of fine hair) are common. Unexpected sudden deaths have been reported; the risk appears to increase as weight drops to less than 35% to 40% of ideal weight. It is believed that these deaths are caused by myocardial degeneration and heart failure, rather than dysrhythmias.

The most exasperating aspect of the treatment of anorexia is the inability of the person with anorexia to recognize there is a problem. Because anorexia is a form of starvation, it can lead to death if left untreated. A multidisciplinary approach appears to be the most effective method of treating persons with the disorder.^{34,35} The goals of treatment are eating and weight gain, resolution of issues with the family, healing of pain from the past, and efforts to work on psychological, relationship, and emotional issues.

Bulimia Nervosa and Binge Eating

Bulimia nervosa and binge eating are eating disorders that encompass an array of distinctive behaviors, feelings, and thoughts. Binge eating is characterized by the consumption of an unusually large quantity of food during a discrete time along with lack of control over the binge-eating episode. Bulimia nervosa is a condition of recurrent binge eating that is accompanied by purging, excessive exercise, or fasting.

Bulimia Nervosa. Bulimia nervosa is 10 times more common in women than men; it usually begins between 13 and 20 years of age, and affects as many as 3% of young women.³⁶ The DSM-IV-TR criteria for bulimia nervosa are (1) recurrent binge eating (at least two times per week for 3 months); (2) inappropriate compensatory behaviors, such as self-induced vomiting, abuse of laxatives or diuretics, fasting, or excessive exercise that follow the binge-eating episode; (3) self-evaluation that is unduly influenced by body shape and weight; and (4) a determination that the eating disorder does not occur exclusively during episodes of anorexia nervosa.³² The diagnostic criteria for bulimia nervosa now include subtypes to distinguish patients who compensate by purging (e.g., vomiting or abuse of laxatives or diuretics) and those who use nonpurging behaviors (e.g., fasting or excessive exercise). The disorder may be associated with other psychiatric disorders, such as substance abuse.31,36

The complications of bulimia nervosa include those resulting from overeating, self-induced vomiting, and cathartic and diuretic abuse. Among the complications of self-induced vomiting are dental disorders, parotitis, and fluid and electrolyte disorders. Dental abnormalities, such as sensitive teeth, increased dental caries, and periodontal disease, occur with frequent vomiting because the high acid content of the vomitus causes tooth enamel to dissolve. Esophagitis, dysphagia, and esophageal stricture are common. With frequent vomiting, there often is reflux of gastric contents into the lower esophagus because of relaxation of the lower esophageal sphincter. Vomiting may lead to aspiration pneumonia, especially in intoxicated or debilitated persons. Potassium, chloride, and hydrogen are lost in the vomitus, and frequent vomiting predisposes to metabolic alkalosis with hypokalemia (see Chapter 6). An unexplained physical response to vomiting is the development of benign, painless parotid gland enlargement.

The weights of persons with bulimia nervosa may fluctuate, although not to the dangerously low levels seen in anorexia nervosa. Their thoughts and feelings range from fear of not being able to stop eating to a concern about gaining too much weight. They also experience feelings of sadness, anger, guilt, shame, and low self-esteem.

Treatment strategies include psychological and pharmacologic treatments. Unlike persons with anorexia nervosa, persons with bulimia nervosa or binge eating are upset by the behaviors practiced and the thoughts and feelings experienced, and they are more willing to accept help. Pharmacotherapeutic agents include the tricyclic antidepressants (*e.g.*, desipramine, imipramine), the selective serotonin reuptake inhibitors (*e.g.*, fluoxetine), and other antidepressant medications.³⁶

Binge Eating. Binge eating is characterized by recurrent episodes of binge eating at least 2 days per week for 6 months and at least three of the following: (1) eating rapidly; (2) eating until becoming uncomfortably full; (3) eating large amounts when not hungry; (4) eating alone because of embarrassment; and (5) disgust, depression, or guilt because of eating episodes.^{31,34,36}

The primary goal of therapy for binge-eating disorders is to establish a regular, healthful eating pattern. Persons with binge-eating disorders who have been successfully treated for their eating disorder have reported that making meal plans, eating a balanced diet at three regular meals a day, avoiding high-sugar foods and other binge foods, recording food intake and binge-eating episodes, exercising regularly, finding alternative activities, and avoiding alcohol and drugs are helpful in maintaining their more healthful eating behaviors after treatment.

In summary, undernutrition can range from a selective deficiency of a single nutrient to starvation in which there is deprivation of all ingested nutrients. Malnutrition and starvation are among the most widespread causes of morbidity and mortality in the world. Malnutrition is common during illness, recovery from trauma, and hospitalization. The effects of malnutrition and starvation on body function are widespread. They include loss of muscle mass, impaired wound healing, impaired immunologic function, decreased appetite, loss of calcium and phosphate from bone, anovulation and amenorrhea in women, and decreased testicular function in men.

Anorexia nervosa, bulimia nervosa, and binge eating are eating disorders that result in malnutrition. In anorexia nervosa, distorted attitudes about eating lead to serious weight loss and malnutrition. Bulimia nervosa is characterized by secretive episodes or binges of eating large quantities of easily consumed, high-caloric foods, followed by compensatory behaviors such as fasting, self-induced vomiting, or abuse of laxatives or diuretics. Binge-eating disorder is characterized by eating large quantities of food but is not accompanied by purging and other inappropriate compensatory behaviors seen in persons with bulimia nervosa.

REVIEW QUESTIONS

Define *calorie* and state the number of calories derived from the oxidation of 1 g of protein, fat, or carbohydrate.

State the purpose of the Recommended Dietary Allowance of calories, proteins, fats, carbohydrates, vitamins, and minerals.

State the factors used in determining body mass index and explain its use in evaluating body weight in terms of undernutrition and overnutrition.

Define and discuss the causes of obesity and health risks associated with obesity.

Explain the effect of malnutrition on muscle mass, respiratory function, acid-base balance, wound healing, immune function,

bone mineralization, the menstrual cycle, and testicular function.

Compare the eating disorders and complications associated with anorexia nervosa and the binge-purge syndrome.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- 1. Guyton A.C., Hall J.E. (1996). *Textbook of medical physiology* (9th ed., p. 805, 806, 909). Philadelphia: W.B. Saunders.
- Subcommittee on the Tenth Edition of the RDAs. (1989). Recommended dietary allowances (10th ed.). Commission on Life Sciences-National Council. Washington, DC: National Academy Press.
- National Academic Press. (2000). Introduction to dietary references intakes. In Dietary reference intakes for vitamin C, vitamin E, selenium, and carotenoids. Washington, DC: National Academy Press. [On-line.] Available: http://www.nap.edu/books/0309069351/html.
- 4. Detsky A.S., Smalley P.S., Chang J. (1994). Is this patient malnourished? *Journal of the American Medical Association* 271, 54–58.
- World Health Organization. (1989). Measuring obesity: Classification and description of anthropometric data. Copenhagen: World Health Organization.
- North American Association for Study of Obesity. (1998). Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults. NIH publication no. 98-4083. Obesity Research 6 (Suppl. 2), 51S–209S. Also available on-line: http://www. nhlbi.nih.gov/guidelines/ob_gdlns.htm.
- Abernathy R.P., Black D.R. (1996). Healthy body weight: An alternative perspective. American Journal of Clinical Nutrition 63 (Suppl.), 448S-451S.
- Willett W.C., Dietz W.H., Colditz G.A. (1999). Guidelines to healthy weight. New England Journal of Medicine 341, 427–434.
- Mokdad A.H., Serdula M.K., Dietz W.H., et al. (1999). The spread of the obesity epidemic in the United States, 1991–1998. *Journal of the American Medical Association* 282, 1519–1522.
- Allison D.B., Saunders S.E. (2000). Obesity in North America: An overview. Medical Clinics of North America 84, 305–328.
- Goran M.I. (2000). Energy metabolism and obesity. Medical Clinics of North America 84, 347–362.
- Hill J.O., Wyatt H.R., Melanson E.L. (2000). Genetic and environmental contributions to obesity. *Medical Clinics of North America* 84, 333–346.
- Soreneson T.J., Holst C., Stunkard A.J., et al. (1992). Correlations of body mass index of adult adoptees and their biological and adoptive relatives. *International Journal of Obesity and Related Metabolic Disorders* 16, 227–236.
- 14. Bouchard C. (1994). The genetics of obesity. Boca Raton, FL: CRC Press.
- 15. Kissebah A.H., Krakower G.R. (1994). Regional adiposity and morbidity. *Physiological Reviews* 74, 761–811.
- Pleuss J.A., Hoffman R.G., Sonnentag G.E., et al. (1993). Effects of abdominal fat on insulin and androgen levels. *Obesity Research* 1 (Suppl. 1), 25F.
- Jeffery R.W. (1996). Does weight cycling present a health risk? American Journal of Clinical Nutrition 63 (Suppl.), 452S-455S.
- Dwyer J. (1996). Policy and healthy weight. American Journal of Nutrition 63 (Suppl. 3), 4155–418S.
- Lee I., Manson J.E., Hennekens C.H., et al. (1993). Body weight and mortality: A 27-year follow-up of middle-aged men. *Journal of the American Medical Association* 270, 2823–2828.

Unit Seven: Alterations in the Gastrointestinal System

- 20. Task Force on Prevention and Treatment of Obesity. (1994). Towards prevention of obesity: Research directives. Obesity Research 2, 571.
- 21. Berke E.M., Morden N.E. (2000). Medical management of obesity. American Family Physician 62, 419–426.
- 22. CDC. (1997). Update: Prevalence of overweight among children, adolescents, and adults-United States, 1988-1994. Morbidity and Mortality Weekly Report 46, 198-202.
- 23. Dietz W.H., Robinson T.N. (1993). Assessment and treatment of childhood obesity. Pediatrics in Review 14, 337-343.
- 24. Dietz W.H. (1998). Health consequences of obesity in youth: Childhood predictors of adult disease. Pediatrics 101 (Suppl. 3), 518-525.
- 25. Hill J.O., Trowbridge F.L. (1998). Childhood obesity: Future directions and research priorities. Pediatrics 101 (Suppl. 3), 570-574.
- 26. Birch L.L., Fisher J.O. (1998). Development of eating behaviors among children and adolescents. Pediatrics 101 (Suppl. 3), 539-554.
- 27. Brown J.L., Pollitt E. (1996). Malnutrition, poverty, intellectual development. Scientific American 274(2), 38-43.
- 28. Chiolero R., Revelly J., Tappy L. (1999). Energy metabolism in sepsis and injury. Journal of Nutrition 129 (IS Suppl.), 45S-51S.

- 29. Biolo G., Gabriele T., Ciccchi B., et al. (1999). Metabolic response to injury and sepsis: Changes in protein metabolism. Journal of Nutrition 129 (IS Suppl.), 53S-57S.
- 30. Baron R.B. (2001). Nutrition. In Tierney L.M., McPhee S.J., Papadakis M.A. (Eds.), Current diagnosis and treatment (40th ed., pp. 1232-1233). New York: Lange Medical Books/McGraw-Hill.
- 31. Becker A., Grinspoon S.K., Klibanski A., et al. (1999). Eating disorders. New England Journal of Medicine 340, 1092-1098.
- 32. American Psychiatric Society. (2000). Practice guideline for treatment of patients with eating disorders (Revision). American Journal of Psychiatry 157, 1-38.
- 33. Gull W.W. (1868). Anorexia nervosa. Transactions of the Clinical Society of London 7, 22-27.
- 34. Kreipe R.E., Birndorf S.A. (2000). Eating disorders in adolescents and young adults. Medical Clinics of North America 84, 1027-1049.
- 35. Gordon A. (2001). Eating disorders: Anorexia nervosa. Hospital Practice 36(2), 36-38.
- 36. McGilley B.M., Pryor T.L. (1998). Assessment and treatment of bulimia nervosa. American Family Physician 57 (6), 27-43.

Eight Alterations in the Endocrine System

Organization and Control of the Endocrine System

The Endocrine System

CHAPTER

Hormones Paracrine and Autocrine Actions **Eicosanoids and Retinoids** Structural Classification Synthesis and Transport Metabolism and Elimination Mechanisms of Action **Control of Hormone Levels** Hypothalamic-Pituitary Regulation **Feedback Regulation Diagnostic Tests Blood Tests Urine Tests** Stimulation and Suppression Tests **Genetic Tests** Imaging

he endocrine system is involved in all of the integrative aspects of life, including growth, sex differentiation, metabolism, and adaptation to an ever-changing environment. This chapter focuses on general aspects of endocrine function, organization of the endocrine system, hormone receptors and hormone actions, and regulation of hormone levels.

THE ENDOCRINE SYSTEM

The endocrine system uses chemical substances called hormones as a means of regulating and integrating body functions. The endocrine system participates in the regulation of digestion, use, and storage of nutrients; growth and development; electrolyte and water metabolism; and reproductive functions. Although the endocrine system once was thought to consist solely of discrete endocrine glands, it is now known that a number of other tissues release chemical messengers that modulate body processes. The functions of the endocrine system are closely linked with those of the nervous system and the immune system. For example, neurotransmitters such as epinephrine can act as neurotransmitters or as hormones. The functions of the immune system also are closely linked with those of the endocrine system. The immune system responds to foreign agents by means of chemical messengers (cytokines, e.g., interleukins, interferons) and complex receptor mechanisms (see Chapter 8). The immune system also is extensively regulated by hormones such as the adrenal corticosteroid hormones.

Hormones

Hormones generally are thought of as chemical messengers that are transported in body fluids. They are highly specialized organic molecules produced by endocrine organs that exert their action on specific target cells. Hormones do not initiate reactions; they are modulators of systemic and cellular responses. Most hormones are present in body fluids at all times but in greater or lesser amounts, depending on the needs of the body.

A characteristic of hormones is that a single hormone can exert various effects in different tissues or, conversely, a single function can be regulated by several hormones. For example, estradiol, which is produced by the ovary, can act on the ovarian follicles to promote their maturation, on the uterus to stimulate its growth and maintain the cyclic changes in the uterine mucosa, on the mammary gland to stimulate ductal growth, on the hypothalamic-pituitary system to regulate the secretion of gonadotropins and prolactin, and on general metabolic processes to affect adipose tissue distribution. Lipolysis, which is the release of free fatty acids from adipose tissue, is an example of a single function that is regulated by several hormones, including the catecholamines, glucagon, and secretin. Table 30-1 lists the major functions and sources of body hormones.

Paracrine and Autocrine Actions

In the past, hormones were described as chemical substances that were released into the bloodstream and transported to distant target sites, where they exerted their action (Fig. 30-1). Although many hormones travel by this mechanism, some hormones and hormone-like substances never enter the bloodstream but instead act locally in the vicinity in which they are released. When they act locally on cells other than those that produced the hormone, the action is called *paracrine*. The action of sex steroids on the ovary is a paracrine action. Hormones also can exert an *autocrine* action on the cells from which they were produced. For example, the release of insulin from

KEY CONCEPTS

HORMONES

- Hormones function as chemical messengers, moving through the blood to distant target sites of action, or acting more locally as paracrine or autocrine messengers that incite more local effects.
- Most hormones are present in body fluids at all times but in greater or lesser amounts, depending on the needs of the body.
- Hormones exert their actions by interacting with high-affinity receptors, which in turn are linked to one or more effector systems in the cell. Some hormone receptors are located on the surface of the cell and act through second messenger mechanisms, and others are located in the cell, where they modulate the synthesis of enzymes, transport proteins, or structural proteins.

pancreatic beta cells can inhibit its release from the same cells. *Juxtacrine* refers to a mechanism whereby a cytokine that is embedded in, bound to, or associated with the plasma membrane of one cell interacts with a specific receptor in a juxtaposed cell.

Eicosanoids and Retinoids

A group of compounds that have a hormone-like action are the eicosanoids, which are derived from polyunsaturated fatty acids in the cell membrane. Among these, *arachidonic acid* is the most important and abundant precursor of the various eicosanoids. The most important of the eicosanoids are the prostaglandins, leukotrienes, and thromboxanes. These fatty acid derivatives are produced by most body cells, are rapidly cleared from the circulation, and are thought to act mainly by paracrine and autocrine mechanisms. Eicosanoid synthesis often is stimulated in response to hormones, and they serve as mediators of hormone action. Retinoids (*e.g.*, retinoic acid) also are derived from fatty acids and have an important role in regulating nuclear receptor action.

Structural Classification

Hormones have diverse structures, ranging from single amino acids to complex proteins and lipids. Hormones usually are divided into four categories according to their structures: (1) amines and amino acids; (2) peptides, polypeptides, glycoproteins, and proteins; (3) steroids; and (4) fatty acid derivatives (Table 30-2). The first category, the amines, includes norepinephrine and epinephrine, which are derived from a single amino acid (*i.e.*, tyrosine), and the thyroid hormones, which are derived from two iodinated tyrosine amino acid residues. The second category, the peptides, polypeptides, glycoproteins, and proteins, can be as small as thyrotropinreleasing hormone (TRH), which contains three amino acids, and as large and complex as growth hormone (GH) and follicle-stimulating hormone (FSH), which have approximately 200 amino acids. Glycoproteins are large peptide hormones associated with a carbohydrate (e.g., FSH). The third category comprises the steroid hormones, which are derivatives of cholesterol. The fourth category, the fatty acid derivatives, includes the eicosanoids and retinoids.

Synthesis and Transport

The mechanisms for hormone synthesis vary with hormone structure. Protein and peptide hormones are synthesized and stored in granules or vesicles in the cytoplasm of the cell until secretion is required. The lipid-soluble steroid hormones are released as they are synthesized.

Protein and peptide hormones are synthesized in the rough endoplasmic reticulum in a manner similar to the synthesis of other proteins (see Chapter 1). The appropriate amino acid sequence is dictated by messenger RNAs from the nucleus. Usually, synthesis involves the production of a precursor hormone, which is modified by the addition of peptides or sugar units. These precursor hormones often contain extra peptide units that ensure proper folding of the molecule and insertion of essential linkages. If extra amino acids are present, as in insulin, the precursor hormone is called a prohormone. After synthesis and sequestration in the endoplasmic reticulum, the protein and peptide hormones move into the Golgi complex, where they are packaged in granules or vesicles. It is in the Golgi complex that prohormones are converted into hormones.

TABLE 30-1	0-1 Major Action and Source of Selected Hormones		
Source	Hormone	Major Action	
Hypothalamus	Releasing and inhibiting hormones Corticotropin-releasing hormone (CRH) Thyrotropin-releasing hormone (TRH) Growth hormone-releasing hormone (GHRH)	Controls the release of pituitary hormones	
Antonion pituiton (Gonadotropin-releasing hormone (GnRH)	Ctimulates growth of home and muscle promotes protein arm	
Anterior pituitary	Growth hormone (GH) Adrenocorticotropic hormone (ACTH) Thyroid-stimulating hormone (TSH) Follicle-stimulating hormone (FSH)	Stimulates growth of bone and muscle, promotes protein syn- thesis and fat metabolism, decreases carbohydrate metabolism Stimulates synthesis and secretion of adrenal cortical hormones Stimulates synthesis and secretion of thyroid hormone Female: stimulates growth of ovarian follicle, ovulation Male: stimulates sperm production	
	Luteinizing hormone (LH)	Female: stimulates development of corpus luteum, release of oocyte, production of estrogen and progesterone Male: stimulates secretion of testosterone, development of inter- stitial tissue of testes	
Posterior pituitary	Antidiuretic hormone (ADH) Oxytocin	Increases water reabsorption by kidney Stimulates contraction of pregnant uterus, milk ejection from breasts after childbirth	
Adrenal cortex	Mineralocorticosteroids, mainly aldosterone Glucocorticoids, mainly cortisol	Increases sodium absorption, potassium loss by kidney Affects metabolism of all nutrients; regulates blood glucose levels, affects growth, has anti-inflammatory action, and decreases effects of stress	
	Adrenal androgens, mainly dehydroepiandros- terone (DHEA) and androstenedione	Have minimal intrinsic androgenic activity; they are converted to testosterone and dihydrotestosterone in the periphery	
Adrenal medulla	Epinephrine Norepinephrine	Serve as neurotransmitters for the sympathetic nervous system	
Thyroid (follicular cells)	Thyroid hormones: triiodothyronine (T ₃), thyroxine (T ₄)	Increase the metabolic rate; increase protein and bone turnover; increase responsiveness to catecholamines; necessary for fetal and infant growth and development	
Thyroid C cells	Calcitonin	Lowers blood calcium and phosphate levels	
Parathyroid glands	Parathyroid hormone	Regulates serum calcium	
Pancreatic islet cells	Insulin	Lowers blood glucose by facilitating glucose transport across cell membranes of muscle, liver, and adipose tissue	
	Glucagon	Increases blood glucose concentration by stimulation of glycogenolysis and glyconeogenesis	
	Somatostatin	Delays intestinal absorption of glucose	
Kidney	1,25-Dihydroxyvitamin D	Stimulates calcium absorption from the intestine	
Ovaries	Estrogen	Affects development of female sex organs and secondary sex characteristics	
	Progesterone	Influences menstrual cycle; stimulates growth of uterine wall; maintains pregnancy	
Testes	Androgens, mainly testosterone	Affect development of male sex organs and secondary sex characteristics; aid in sperm production	

Steroid hormones are synthesized in the smooth endoplasmic reticulum, and steroid-secreting cells can be identified by their large amounts of smooth endoplasmic reticulum. Certain steroids serve as precursors for the production of other hormones. For example, in the adrenal cortex, progesterone and other steroid intermediates are enzymatically converted into aldosterone, cortisol, or androgens (see Chapter 31).

Hormones that are released into the bloodstream circulate as either free, unbound molecules or as hormones attached to transport carriers (Fig. 30-2). Peptide hormones and protein hormones usually circulate unbound in the blood. Steroid hormones and thyroid hormone are carried by specific carrier proteins synthesized in the liver. The extent of carrier binding influences the rate at which hormones leave the blood and enter the cells. The half-life of a hormone—the time it takes for the body to reduce the concentration of the hormone by one half—is positively correlated with its percentage of protein binding. Thyroxine, which is more than 99% protein bound, has a half-life of 6 days. Aldosterone, which is only 15% bound, has a half-life of only 25 minutes. Drugs that compete with a hormone for binding with transport carrier molecules increase hormone action by increasing the availability of the active unbound hormone. For example, aspirin competes with thyroid hormone for binding to transport proteins; when the



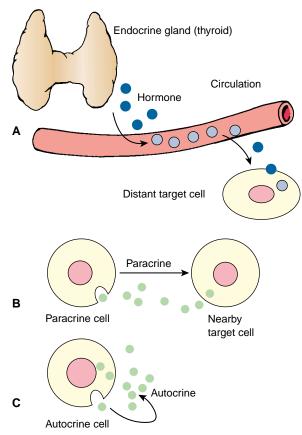
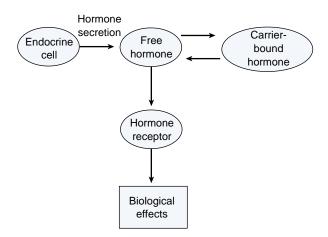


FIGURE 30-1 Examples of endocrine (A), paracrine (B), and autocrine (C) secretions.

drug is administered to persons with excessive levels of circulating thyroid hormone, such as during thyroid crisis, serious effects may occur.

Metabolism and Elimination

Metabolism of hormones and their precursors can generate more or less active products or it can degrade them to inactive forms. In some cases, hormones are eliminated in the intact



■ FIGURE 30-2 ■ Relationship of free and carrier-bound hormone.

form. Hormones secreted by endocrine cells must be inactivated continuously to prevent their accumulation. Intracellular and extracellular mechanisms participate in the termination of hormone function. Some hormones are enzymatically inactivated at receptor sites where they exert their action. The catecholamines, which have a very short half-life, are degraded by catechol-O-methyl transferase (COMT) and monoamine oxidase (MAO). Because of their short half-life, their production is measured by some of their metabolites. In general, peptide hormones also have a short life span in the circulation. Their major mechanism of degradation is through binding to cell surface receptors, with subsequent uptake and degradation by enzymes in the cell membrane or inside the cell. Peptide hormones have a short life span and are inactivated by enzymes that split peptide bonds. Steroid hormones are bound to protein carriers for transport and are inactive in the bound state. Their activity depends on the availability of transport carriers. Unbound adrenal and gonadal steroid hormones are conjugated in the liver, which renders them inactive, and then excreted in the bile or urine. Thyroid hormones also are transported by carrier molecules. The free hormone is rendered inactive by the removal of amino acids (*i.e.*, deamination) in the tissues, and the hormone is conjugated in the liver and eliminated in the bile.

TABLE 30-2 Classes of Hormones Based on Structure				
Amines and Amin	o Acids	Peptides, Polypeptides, and Proteins	Steroids	Fatty Acid Compounds
Dopamine Epinephrine Norepinephrine Thyroid hormone		Corticotropin-releasing hormone (CRH) Growth hormone–releasing hormone (GHRH) Thyrotropin-releasing hormone (TRH) Adrenocorticotropic hormone (ACTH) Follicle-stimulating hormone (FSH) Luteinizing hormone (LH) Thyroid-stimulating hormone (TSH) Growth hormone (GH) Antidiuretic hormone (ADH) Oxytocin Insulin Glucagon Somatostatin Calcitonin Parathyroid hormone	Aldosterone Glucocorticoids Estrogens Testosterone Progesterone Androstenedione 1,25-Dihydroxyvitamin D Dihydrotestosterone (DHT) Dehydroepiandrosterone (DHEA)	Eicosanoids Retinoids

Mechanisms of Action

Hormones produce their effects through interaction with highaffinity receptors, which in turn are linked to one or more effector systems within the cell. These mechanisms involve many of the cell's metabolic activities, ranging from ion transport at the cell surface to stimulation of nuclear transcription of complex molecules. The rate at which hormones react depends on their mechanism of action. The neurotransmitters, which control the opening of ion channels, have a reaction time of milliseconds. Thyroid hormone, which functions in the control of cell metabolism and synthesis of intracellular signaling molecules, requires days for its full effect to occur.

Receptors. Hormones exert their action by binding to highaffinity receptors located either on the surface or inside the target cells. The function of these receptors is to recognize a specific hormone and translate the hormonal signal into a cellular response. The structure of these receptors varies in a manner that allows target cells to respond to one hormone and not to others. For example, receptors in the thyroid are specific for thyroid-stimulating hormone, and receptors on the gonads respond to the gonadotropic hormones.

The response of a target cell to a hormone varies with the *number* of receptors present and with the *affinity* of these receptors for hormone binding. A variety of factors influence the number of receptors that are present on target cells and their affinity for hormone binding.

There are approximately 2000 to 100,000 hormone receptor molecules per cell. The number of hormone receptors on a cell may be altered for any of several reasons. Antibodies may destroy or block the receptor proteins. Increased or decreased hormone levels often induce changes in the activity of the genes that regulator receptor synthesis. For example, decreased hormone levels often produce an increase in receptor numbers by means of a process called *up-regulation*; this increases the sensitivity of the body to existing hormone levels. Likewise, sustained levels of excess hormone often bring about a decrease in receptor numbers by down-regulation, producing a decrease in hormone sensitivity. In some instances, the reverse effect occurs, and an increase in hormone levels appears to recruit its own receptors, thereby increasing the sensitivity of the cell to the hormone. The process of up-regulation and down-regulation of receptors is regulated largely by inducing or repressing the transcription of receptor genes.

The affinity of receptors for binding hormones also is affected by a number of conditions. For example, the pH of the body fluids plays an important role in the affinity of insulin receptors. In ketoacidosis, a lower pH reduces insulin binding.

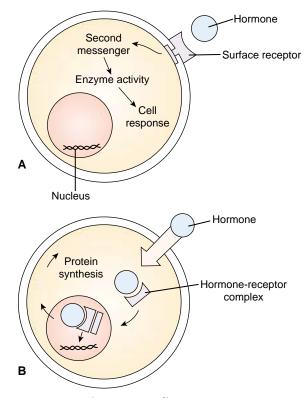
Some hormone receptors are located on the surface of the cell and act through second messenger mechanisms, and others are located within the cell, where they modulate the synthesis of enzymes, transport proteins, or structural proteins. The receptors for thyroid hormones, which are found in the nucleus, are thought to be directly associated with controlling the activity of genes located on one or more of the chromosomes. Chart 30-1 lists hormones that act through the two types of receptors.

Surface Receptors. Because of their low solubility in the lipid layer of cell membranes, peptide hormones and catecholamines cannot readily cross the cell membrane. Instead, these hormones interact with surface receptors in a manner that in-

CHART 30-1 Hormone–Receptor Interactions					
Surface (Second Messenger) Receptors					
Glucagon Insulin Epinephrine Parathyroid hormone Thyroid-stimulating hormone (TSH) Adrenocorticotropic hormone (ACTH) Follicle-stimulating hormone (FSH) Luteinizing hormone (LH) Antidiuretic hormone (ADH) Secretin					
Intracellular Interactions					
Estrogens Testosterone Progesterone Adrenal cortical hormones Thyroid hormones					

cites the generation of an intracellular signal or message. The intracellular signal system is termed the *second messenger*, and the hormone is considered to be the first messenger (Fig. 30-3). For example, the first messenger glucagon binds to surface receptors on liver cells to incite glycogen breakdown by way of the second messenger system.

The most widely distributed second messenger is cyclic adenosine monophosphate (cAMP). cAMP is formed from



■ FIGURE 30-3 ■ The two types of hormone-receptor interactions: the surface receptor (A) and the intracellular receptor (B).

534 Unit Eight: Alterations in the Endocrine System

cellular adenosine triphosphate (ATP) by the enzyme adenylate cyclase, a membrane-bound enzyme that is located on the inner aspect of the cell membrane. Adenylate cyclase is functionally coupled to various cell surface receptors by the regulatory actions of G proteins (see Chapter 1 and Fig. 1-12). A second messenger similar to cAMP is cyclic GMP, derived from guanine triphosphate (GTP). As a result of binding to specific cell receptors, many peptide hormones incite a series of enzymatic reactions that produce an almost immediate increase in cAMP. Some hormones act to decrease cAMP levels and have an opposite effect.

In some cells, the binding of hormones or neurotransmitters to surface receptors acts directly, rather than through a second messenger, to open ion channels in the cell membrane. The influx of ions serves as an intracellular signal to convey the hormonal message to the cell interior. In many instances, the activation of hormone receptors results in the opening of calcium channels. The increasing cytoplasmic concentration of calcium may result in direct activation of calcium-dependent enzymes or calcium-calmodulin complexes with their attendant effects.

Intracellular Receptors. A second type of receptor mechanism is involved in mediating the action of hormones such as the steroid and thyroid hormones (see Fig. 30-3). These hormones are lipid soluble and pass freely through the cell membrane. They then attach to intracellular receptors and form a hormone-

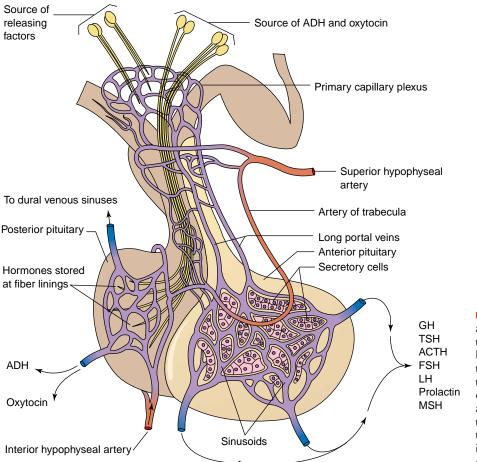
receptor complex that travels to the cell nucleus. The hormonemessenger complex then activates or suppresses intracellular mechanisms such as gene activity, with subsequent production or inhibition of messenger RNA and protein synthesis.

Control of Hormone Levels

Hormone secretion varies widely during a 24-hour period. Some hormones, such as growth hormone (GH) and adrenocorticotropic hormone (ACTH), have diurnal fluctuations that vary with the sleep-wake cycle. Others, such as the female sex hormones, are secreted in a complicated cyclic manner. The levels of hormones such as insulin and antidiuretic hormone (ADH) are regulated by feedback mechanisms that monitor substances such as glucose (insulin) and water (ADH) in the body. The levels of many of the hormones are regulated by feedback mechanisms that involve the hypothalamic-pituitary-target cell system.

Hypothalamic-Pituitary Regulation

The hypothalamus and pituitary (*i.e.*, hypophysis) form a unit that exerts control over many functions of several endocrine glands as well as a wide range of other physiologic functions. These two structures are connected by blood flow in the hypophyseal portal system, which begins in the hypothalamus and drains into the anterior pituitary gland, and by the nerve axons that connect the supraoptic and paraventricular nuclei of the hypothalamus with the posterior pituitary gland (Fig. 30-4).



■ FIGURE 30-4 ■ The hypothalamus and the anterior and posterior pituitary. The hypothalamic releasing or inhibiting hormones are transported to the anterior pituitary by way of the portal vessels. ADH and oxytocin are produced by nerve cells in the supraoptic and paraventricular nuclei of the hypothalamus and then transported through the nerve axon to the posterior pituitary, where they are released into the circulation. The pituitary is enclosed in the bony sella turcica ("Turkish saddle") and is bridged by the diaphragma sellae. Embryologically, the anterior pituitary gland developed from glandular tissue and the posterior pituitary developed from neural tissue.

Hypothalamic Hormones. The synthesis and release of anterior pituitary hormones are largely regulated by the action of releasing or inhibiting hormones from the hypothalamus, which is the coordinating center of the brain for endocrine, behavioral, and autonomic nervous system function. It is at the level of the hypothalamus that emotion, pain, body temperature, and other neural input are communicated to the endocrine system (Fig. 30-5). The posterior pituitary hormones, ADH and oxytocin, are synthesized in the cell bodies of neurons in the hypothalamus that have axons that travel to the posterior pituitary. The release and function of ADH are discussed in Chapter 31.

The hypothalamic hormones that regulate the secretion of anterior pituitary hormones include GH-releasing hormone (GHRH), somatostatin, dopamine, thyrotropin-releasing hormone (TRH), corticotropin-releasing hormone (CRH), and gonadotropin-releasing hormone (GnRH). With the exception of GH and prolactin, most of the pituitary hormones are regulated by hypothalamic stimulatory hormones. GH secretion is stimulated by GHRH; thyroid-stimulating hormone (TSH) by TRH; ACTH by CRH; and luteinizing hormone (LH) and FSH

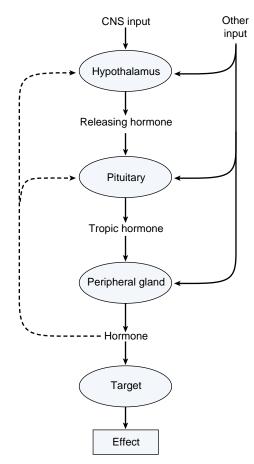


FIGURE 30-5 Hypothalamic-pituitary control of hormone levels. The dashed line represents feedback control.

by GnRH. Somatostatin functions as an inhibitory hormone for GH and TSH. Prolactin secretion is inhibited by dopamine; thus, persons receiving antipsychotic drugs that block dopamine often have increased prolactin levels.

The activity of the hypothalamus is regulated by both hormonally mediated signals (e.g., negative feedback signals) and by neuronal input from a number of sources. Neuronal signals are mediated by neurotransmitters such as acetylcholine, dopamine, norepinephrine, serotonin, γ -aminobutyric acid, and opioids. Cytokines that are involved in immune and inflammatory responses, such as the interleukins, also are involved in the regulation of hypothalamic function. This is particularly true of the hormones involved in the hypothalamic-pituitaryadrenal axis. Thus, the hypothalamus can be viewed as a bridge by which signals from multiple systems are relayed to the pituitary gland.

Pituitary Hormones. The pituitary gland has been called the master gland because its hormones control the functions of many target glands and cells. Hormones produced by the anterior pituitary control body growth and metabolism (GH), function of the thyroid gland (TSH), glucocorticoid hormone levels (ACTH), function of the gonads (FSH and LH), and breast growth and milk production (prolactin). Melanocyte-stimulating hormone, which is involved in the control of pigmentation of the skin, is produced by the pars intermedia of the pituitary gland. The functions of many of these hormones are discussed in other parts of this book (e.g., thyroid hormone, GH, and the corticosteroids in Chapter 31, the sex hormones in Chapters 33 and 34, and ADH from the posterior pituitary in Chapter 6).

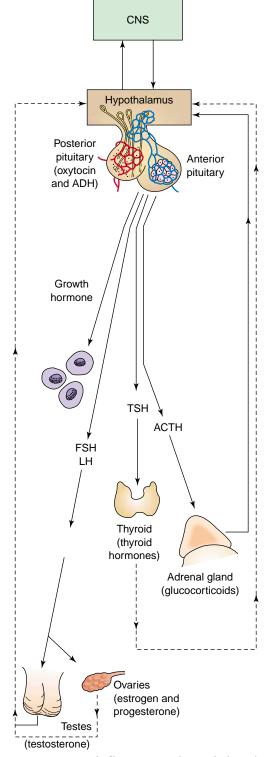
Feedback Regulation

The level of many of the hormones in the body is regulated by negative feedback mechanisms. The function of this type of system is similar to that of the thermostat in a heating system. In the endocrine system, sensors detect a change in the hormone level and adjust hormone secretion so that body levels are maintained within an appropriate range. When the sensors detect a decrease in hormone levels, they initiate changes that cause an increase in hormone production; when hormone levels rise above the set point of the system, the sensors cause hormone production and release to decrease. For example, an increase in thyroid hormone is detected by sensors in the hypothalamus or anterior pituitary gland, and this causes a reduction in the secretion of TSH, with a subsequent decrease in the output of thyroid hormone from the thyroid gland. The feedback loops for the hypothalamic-pituitary feedback mechanisms are illustrated in Figures 30-5 and 30-6.

Exogenous forms of hormones (given as drug preparations) can influence the normal feedback control of hormone production and release. One of the most common examples of this influence occurs with the administration of the corticosteroid hormones, which causes suppression of the hypothalamicpituitary-target cell system that regulates the production of these hormones.

Although the levels of most hormones are regulated by negative feedback mechanisms, a small number are under positive feedback control, in which increasing levels of a hormone cause another gland to release a hormone that is stimulating to the first. However, there must be a mechanism for shutting off the release of the first hormone, or its production would





■ FIGURE 30-6 ■ Control of hormone production by hypothalamicpituitary-target cell feedback mechanism. Hormone levels from the target glands regulate the release of hormones from the anterior pituitary by means of a negative feedback system. The *dashed line* represents feedback control.

continue unabated. An example of such a system is that of the female ovarian hormone estradiol. Increased estradiol production during the follicular stage of the menstrual cycle causes increased gonadotropin (FSH) production by the anterior pituitary gland. This stimulates further increases in estradiol levels until the demise of the follicle, which is the source of estradiol, results in a decrease in gonadotropin levels.

In addition to positive and negative feedback mechanisms that monitor changes in hormone levels, some hormones are regulated by the level of the substance they regulate. For example, insulin levels normally are regulated in response to blood glucose levels and those of aldosterone in response to body levels of sodium and potassium. Other factors such as stress, environmental temperature, and nutritional status can alter feedback regulation of hormone levels.

Diagnostic Tests

Several techniques are available for assessing endocrine function and hormone levels. One technique measures the effect of a hormone on body function. For example, measurement of blood glucose reflects insulin levels and is an indirect method of assessing insulin availability. Another method is to measure hormone levels.

Blood Tests

Hormones circulating in the plasma were first detected by bioassays using the intact animal or a portion of tissue from the animal. At one time, female rats or male frogs were used to test women's urine for the presence of human chorionic gonadotropin, which is produced by the placenta during pregnancy. Unfortunately, most bioassays lack the precision, sensitivity, and specificity to measure low concentrations of hormones in plasma, and they are inconvenient to perform.

Blood hormone levels provide information about hormone levels at a specific time. For example, blood insulin levels can be measured along with blood glucose after administration of a challenge dose of glucose to measure the time course of change in blood insulin levels.

Real progress in measuring plasma hormone levels came more than 40 years ago with the use of competitive binding and the development of radioimmunoassay (RIA) methods. This method uses a radiolabeled form of the hormone and a hormone antibody that has been prepared by injecting an appropriate animal with a purified form of the hormone. The unlabeled hormone in the sample being tested competes with the radiolabeled hormone for attachment to the binding sites of the antibody. Measurement of the radiolabeled hormoneantibody complex then provides a means of arriving at a measure of the hormone level in the sample. Because hormone binding is competitive, the amount of radiolabeled hormoneantibody complex that is formed decreases as the amount of unlabeled hormone in the sample is increased. Newer techniques of RIA have been introduced, including the immunoradiometric assay (IRMA). IRMA uses two antibodies instead of one. These two antibodies are directed against two different parts of the molecule, so IRMA assays are more specific. RIA has several disadvantages, including limited shelf-life of the radiolabeled hormone and the cost for the disposal of radioactive waste.

Nonradiolabeled methods have been developed in which the antigen of the hormone being measured is linked to an enzyme-activated label (*e.g.*, fluorescent label, chemiluminescent label) or latex particles that can be agglutinated with an antigen and measured. The enzyme-linked immunosorbent assays (ELISA) use antibody-coated plates and an enzyme-labeled reporter antibody. Binding of the hormone to the enzymelabeled reporter antibody produces a colored reaction that can be measured using a spectrophotometer.

Urine Tests

Measurements of urinary hormone or hormone metabolite excretion often are done on a 24-hour urine sample and provide a better measure of hormone levels during that period than hormones measured in an isolated blood sample. The advantages of a urine test include the relative ease of obtaining urine samples and the fact that blood sampling is not required. The disadvantage is that reliably timed urine collections often are difficult to obtain. For example, a person may be unable to urinate at specific timed intervals, and urine samples may be accidentally discarded or inaccurately preserved. Because many urine tests involve the measure of a hormone metabolite. rather than the hormone itself, drugs or disease states that alter hormone metabolism may interfere with the test result. Some urinary hormone metabolite measurements include hormones from more than one source and are of little value in measuring hormone secretion from a specific source. For example, urinary 17-ketosteroids are a measure of both adrenal and gonadal androgens.

Stimulation and Suppression Tests

Stimulation tests are used when hypofunction of an endocrine organ is suspected. A tropic or stimulating hormone can be administered to test the capacity of an endocrine organ to increase hormone production. The capacity of the target gland to respond is measured by an increase in the appropriate hormone. For example, the function of the hypothalamic-pituitarythyroid system can be evaluated through stimulation tests using TRH and measuring TSH response. Failure to effect an increase in TSH after a TRH stimulation test suggests inadequate production of TSH by the pituitary.

Suppression tests are used to determine if negative feedback control mechanisms are intact. For example, a glucocorticoid hormone can be administered to persons suspected of having hypercortisolism to assess the capacity to inhibit CRH.

Genetic Tests

Deoxyribonucleic acid (DNA) analysis is being increasingly used for the identification of affected family members in a kindred harboring a known mutation (*e.g.*, looking for the RET protooncogene in certain multiple endocrine neoplasia syndromes).

Imaging

Imaging studies are gaining increasing importance in the diagnosis and follow-up of endocrine diseases. Magnetic resonance imaging (MRI) and computed tomography scans are especially useful for imaging endocrine glands and endocrine tumors. Nuclear scanning also is widely used for assessing thyroid, parathyroid, and adrenal disorders. Ultrasound scanning is recommended for managing thyroid nodules. Positron emission tomography scanning is being used more widely for parathyroid detection after failed parathyroid surgery for hyperparathyroidism.

In summary, the endocrine system acts as a communication system that uses chemical messengers, or hormones, for the transmission of information from cell to cell and from organ to organ. Hormones act by binding to receptors that are specific for the different types of hormones. Many of the endocrine glands are under the regulatory control of other parts of the endocrine system. The hypothalamus and the pituitary gland form a complex integrative network that joins the nervous system and the endocrine system; this central network controls the output from many of the other glands in the body.

Endocrine function can be assessed directly by measuring hormone levels or indirectly by assessing the effects that a hormone has on the body (*e.g.*, assessment of insulin function through blood glucose). Imaging techniques are increasingly used to visualize endocrine structures, and genetic techniques are used to determine the presence of genes that contribute to the development of endocrine disorders.

REVIEW QUESTIONS

Characterize a hormone.

- State a difference between the synthesis of protein hormones and that of steroid hormones.
- Describe mechanisms of hormone transport and inactivation.

State the function of a hormone receptor, and state the difference between cell surface hormone receptors and intracellular hormone receptors.

Describe the role of the hypothalamus in regulating pituitary control of endocrine function.

- **State** the major difference between positive and negative feedback control mechanisms.
- Describe methods used in the diagnosis of endocrine disorders.

connection-J

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

BIBLIOGRAPHY

- DeGroot L.J., Jameson J.L. (Eds.). (2001). *Endocrinology*. Philadelphia: W.B. Saunders.
- Greenspan F.S., Gardner D.G. (2001). Basic and clinical endocrinology (6th ed.). Norwalk, CT: Appleton & Lange.
- Griffin J.E., Sergio R.O. (Eds.). (2000). Textbook of endocrine physiology (4th ed.). New York: Oxford University Press.
- Kacsoh B. (2000). Endocrine physiology. New York: McGraw-Hill.
- Neal J.M. (2000). Basic endocrinology. Malden, MA: Blackwell Science.
- Nussey S.S., Whitehead S.A. (2001). Endocrinology—an integrated approach. London: Bios.

CHAPTER 31

Alterations in Pituitary, Thyroid, Parathyroid, and Adrenal Function

General Aspects of Altered Endocrine Function Hypofunction and Hyperfunction Primary, Secondary, and Tertiary Disorders **Pituitary and Growth Hormone Disorders Hypopituitarism** Growth and Growth Hormone Disorders **Growth Hormone** Short Stature and Growth Hormone Deficiency in Children Growth Hormone Deficiency in Adults Tall Stature and Growth Hormone Excess in Children Growth Hormone Excess in Adults **Isosexual Precocious Puberty Thyroid Disorders Control of Thyroid Function** Actions of Thyroid Hormone **Tests of Thyroid Function Alterations in Thyroid Function Hypothyroidism** Congenital Hypothyroidism Acquired Hypothyroidism and Myxedema **Hyperthyroidism** Graves' Disease **Thyroid Storm Parathyroid Hormone Disorders Hypoparathyroidism** Hyperparathyroidism **Disorders of Adrenal Cortical Function Control of Adrenal Cortical Function** Biosynthesis, Transport, and Metabolism **Adrenal Sex Hormones** Mineralocorticoids

Glucocorticoids Suppression of Adrenal Function Tests of Adrenal Function Congenital Adrenal Hyperplasia Adrenal Cortical Insufficiency Primary Adrenal Cortical Insufficiency Secondary Adrenal Cortical Insufficiency Acute Adrenal Crisis Glucocorticoid Hormone Excess (Cushing's Syndrome)

he endocrine system affects all aspects of body function, including growth and development, energy metabolism, muscle and adipose tissue distribution, sexual development, fluid and electrolyte balance, and inflammation and immune responses. This chapter focuses on disorders of pituitary function, growth and growth hormone levels, thyroid function, and adrenocortical function.

GENERAL ASPECTS OF ALTERED ENDOCRINE FUNCTION

Hypofunction and Hyperfunction

Disturbances of endocrine function usually can be divided into two categories: hypofunction and hyperfunction. Hypofunction of an endocrine gland can occur for a variety of reasons. Congenital defects can result in the absence or impaired development of the gland or the absence of an enzyme needed for hormone synthesis. The gland may be destroyed by a disruption in blood flow, infection, inflammation, autoimmune responses, or neoplastic growth. There may be a decline in function with aging, or the gland may atrophy as the result of drug therapy or for unknown reasons. Some endocrine-deficient states are associated with receptor defects: hormone receptors may be absent, the receptor binding of hormones may be defective, or the cellular response to the hormone may be impaired. It is suspected that in some cases a gland may produce a biologically inactive hormone or that an active hormone may be destroyed by circulating antibodies before it can exert its action.

Hyperfunction usually is associated with excessive hormone production. This can result from excessive stimulation and hyperplasia of the endocrine gland or from a hormoneproducing tumor of the gland. An ectopic tumor can produce hormones; for example, certain bronchogenic tumors produce hormones such as antidiuretic hormone (ADH) and adrenocorticotropic hormone (ACTH).

Primary, Secondary, and Tertiary Disorders

Endocrine disorders in general can be divided into primary, secondary, and tertiary groups. *Primary defects* in endocrine function originate in the target gland responsible for producing the hormone. In *secondary disorders* of endocrine function, the target gland is essentially normal, but its function is altered by defective levels of stimulating hormones or releasing factors from the pituitary system. For example, adrenalectomy produces a primary deficiency of adrenal corticosteroid hormones. Removal or destruction of the pituitary gland eliminates ACTH stimulation of the adrenal cortex and brings about a secondary deficiency. A *tertiary disorder* results from hypothalamic dysfunction (as may occur with craniopharyngiomas or cerebral irradiation); thus, both the pituitary and target organ are understimulated.

In summary, endocrine disorders are the result of hypofunction or hyperfunction of an endocrine gland. They can occur as a primary defect in hormone production by a target gland or as a secondary or tertiary disorder resulting from a defect in the hypothalamic-pituitary system that controls a target gland's function.

PITUITARY AND GROWTH HORMONE DISORDERS

The anterior lobe of the pituitary gland produces ACTH, thyroid stimulating hormone (TSH), growth hormone (GH), the gonadotrophic hormones (follicle stimulating hormone [FSH] and luteinizing hormone [LH]), and prolactin (see Chapter 30, Fig. 30-4). Four of these, ACTH, TSH, LH, and FSH control the secretion of hormones from other endocrine glands. ACTH controls the release of cortisol from the adrenal gland and TSH the secretion of thyroid hormone from the thyroid gland; LH regulates sex hormones, and FSH regulates fertility.

Hypopituitarism

Hypopituitarism, which is characterized by a decreased secretion of pituitary hormones, is a condition that affects many of the other endocrine systems. Typically, 70% to 90% of the anterior pituitary must be destroyed before hypopituitarism becomes clinically evident.¹ The cause may be congenital or result from a variety of acquired abnormalities (Chart 31-1). The manifestations of hypopituitarism usually occur gradually, but it can present as an acute and life-threatening condition. Patients usually report being chronically unfit, with weakness, fatigue, loss of appetite, impairment of sexual function, and

CHART 31-1 Causes of Hypopituitarism

- Tumors and mass lesions—pituitary adenomas, cysts, metastatic cancer, and other lesions
- Pituitary surgery or radiation
- Infiltrative lesions and infections—hemochromatosis, lymphocytic hypophysitis
- Pituitary infarction—infarction of the pituitary gland after substantial blood loss during childbirth (Sheehan's syndrome)
- Pituitary apoplexy—sudden hemorrhage into the pituitary gland
- Genetic diseases—rare congenital defects of one or more pituitary hormones
- Empty sella syndrome—an enlarged sella turcica that is not entirely filled with pituitary tissue
- Hypothalamic disorders—tumors and mass lesions (*e.g.*, craniopharyngiomas and metastatic malignancies), hypothalamic radiation, infiltrative lesions (*e.g.*, sarcoidosis), trauma, infections

cold intolerance. However, ACTH deficiency (secondary adrenal failure) is the most serious endocrine deficiency, leading to weakness, nausea, anorexia, fever, and postural hypotension. Hypopituitarism is associated with increased morbidity and mortality.

Anterior pituitary hormone loss tends to follow a typical sequence, especially with progressive loss of pituitary reserve caused by tumors or previous pituitary radiation therapy (which may take 10 to 20 years to produce hypopituitarism). Usually GH secretion is lost first, then LH and FSH, followed by TSH deficiency. ACTH is usually the last to become deficient.

Treatment of hypopituitarism includes treating any identified underlying cause. Hormone deficiencies are treated with replacement of the target gland hormone. Cortisol replacement is started when ACTH deficiency is present; thyroid replacement when TSH deficiency is detected; and sex hormone replacement when LH and FSH are deficient. GH replacement is being used increasingly to treat GH deficiency.

Growth and Growth Hormone Disorders

Several hormones are essential for normal body growth and maturation, including growth hormone (GH), insulin, thyroid hormone, and androgens. In addition to its actions on carbohydrate and fat metabolism, insulin plays an essential role in growth processes. Children with diabetes, particularly those with poor control, often fail to grow normally even though GH levels are normal. When levels of thyroid hormone are lower than normal, bone growth and epiphyseal closure are delayed. Androgens such as testosterone and dihydrotestosterone exert anabolic growth effects through their actions on protein synthesis. Glucocorticoids at excessive levels inhibit growth, apparently because of their antagonistic effect on GH secretion.

Growth Hormone

Growth hormone, also called *somatotropin*, is a 191-aminoacid polypeptide hormone synthesized and secreted by special cells in the anterior pituitary called *somatotropes*. For many

540 Unit Eight: Alterations in the Endocrine System

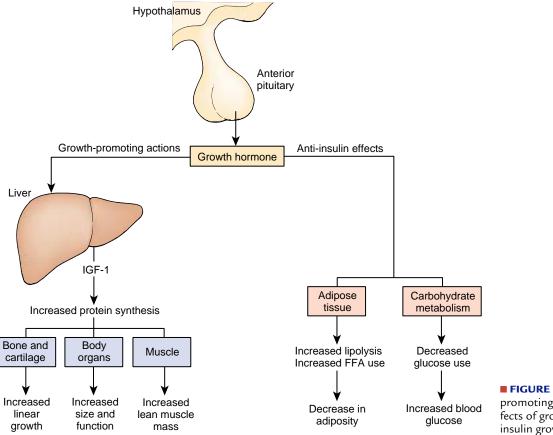
years, it was thought that GH was produced primarily during periods of growth. However, this has proved to be incorrect because the rate of GH production in adults is almost as great as in children. GH is necessary for growth and contributes to the regulation of metabolic functions (Fig. 31-1).^{2,3} All aspects of cartilage growth are stimulated by GH; one of the most striking effects of GH is on linear bone growth, resulting from its action on the epiphyseal growth plates of long bones. The width of bone increases because of enhanced periosteal growth; visceral and endocrine organs, skeletal and cardiac muscle, skin, and connective tissue all undergo increased growth in response to GH. In many instances, the increased growth of visceral and endocrine organs is accompanied by enhanced functional capacity. For example, increased growth of cardiac muscle is accompanied by an increase in cardiac output.

In addition to its effects on growth, GH facilitates the rate of protein synthesis by all of the cells of the body; it enhances fatty acid mobilization and increases the use of fatty acids for fuel; and it maintains or increases blood glucose levels by decreasing the use of glucose for fuel. GH has an initial effect of increasing insulin levels. However, the predominant effect of prolonged GH excess is to increase glucose levels despite an insulin increase. This is because GH induces a resistance to insulin in the peripheral tissues, inhibiting the uptake of glucose by muscle and adipose tissues.

Many of the effects of GH depend on a family of peptides called *insulin-like growth factors* (IGF), also called *somatomedins*, which are produced mainly by the liver. GH cannot directly produce bone growth; instead, it acts indirectly by causing the liver to produce IGF. These peptides act on cartilage and bone to promote their growth. At least four IGFs have been identified; of these, IGF-1 appears to be the more important in terms of growth, and it is the one that usually is measured in laboratory tests. The IGFs have been sequenced and have structures that are similar to that of proinsulin. This undoubtedly explains the insulin-like activity of the IGFs and the weak action of insulin on growth. IGF levels are themselves influenced by a family of at least six binding factors called *IGF-binding proteins* (IGFBPs).

GH is carried unbound in the plasma and has a half-life of approximately 20 to 50 minutes. The secretion of GH is regulated by two hypothalamic hormones: GH-releasing hormone (GHRH), which increases GH release, and somatostatin, which inhibits GH release. A third hormone, the recently identified ghrelin, also may be important. These hypothalamic influences (*i.e.*, GHRH and somatostatin) are tightly regulated by neural, metabolic, and hormonal factors. The secretion of GH fluctuates during a 24-hour period, with peak levels occurring 1 to 4 hours after onset of sleep (*i.e.*, during sleep stages 3 and 4). The nocturnal sleep bursts, which account for 70% of daily GH secretion, are greater in children than in adults.⁴

GH secretion is stimulated by hypoglycemia, fasting, starvation, increased blood levels of amino acids (particularly arginine), and stress conditions such as trauma, excitement, emotional stress, and heavy exercise. GH is inhibited by increased glucose levels, free fatty acid release, cortisol, and obesity. Impairment of secretion, leading to growth retardation, is not uncommon in children with severe emotional deprivation.



■ FIGURE 31-1 ■ Growthpromoting and anti-insulin effects of growth hormone. IGF-1, insulin growth factor 1.

KEY CONCEPTS

GROWTH HORMONE

Growth hormone (GH), which is produced by somatotropes in the anterior pituitary, is necessary for linear bone growth in children. It also stimulates cells to increase in size and divide more rapidly; it enhances amino acid transport across cell membranes and increases protein synthesis; and it increases the rate at which cells use fatty acids and decreases the rate at which they use carbohydrates.

The effects of GH on cartilage growth require insulin-like growth factors (IGFs), also called somatomedins, which are produced mainly by the liver.

In children, GH deficiency interferes with linear bone growth, resulting in short stature or dwarfism. In a rare condition called Laron-type dwarfism, GH levels are normal or elevated, but there is a hereditary defect in IGF production.

GH excess in children results in increased linear bone growth, or gigantism. In adults, GH excess results in overgrowth of the cartilaginous parts of the skeleton, enlargement of the heart and other organs of the body, and metabolic disturbances resulting in altered fat metabolism and impaired glucose tolerance.

Short Stature and Growth Hormone Deficiency in Children

Short stature is a condition in which the attained height is well below the fifth percentile or linear growth is below normal for age and gender. Short stature, or growth retardation, has a variety of causes, including chromosomal abnormalities such as Turner's syndrome (see Chapter 4), GH deficiency, hypothyroidism, and panhypopituitarism. Other conditions known to cause short stature include protein-calorie malnutrition, chronic diseases such as renal failure and poorly controlled diabetes mellitus, malabsorption syndromes, and certain therapies such as corticosteroid administration. Emotional disturbances can lead to functional endocrine disorders, causing psychosocial dwarfism. The causes of short stature are summarized in Chart 31-2.

Accurate measurement of height is an extremely important part of the physical examination of children. Completion of the developmental history and growth charts is essential. Growth curves and growth velocity studies also are needed. Diagnosis of short stature is not made on a single measurement but is based on actual height and on velocity of growth and parental height. The diagnostic procedures for short stature include tests to exclude nonendocrine causes. Extensive hormonal testing procedures are initiated in cases where an endocrine cause is suspected. Usually, tests to determine GH and IGF-1 levels are done. Radiologic films are used to assess bone age, which most

CHART 31-2 Causes of Short Stature

Variants of Normal

Genetic or "familial" short stature Constitutional short stature

Low Birth Weight (e.g., intrauterine growth retardation)

Endocrine Disorders

Growth hormone (GH) deficiency Primary GH deficiency Idiopathic GH deficiency **Pituitary** agenesis Secondary GH deficiency (panhypopituitarism) Biologically inactive GH production Deficient IGF-1 production in response to normal or elevated GH (Laron-type dwarfism) Hypothyroidism Diabetes mellitus in poor control Glucocorticoid excess Endogenous (Cushing's disease) Exogenous (glucocorticoid drug treatment) Abnormal mineral metabolism (e.g., pseudohypoparathyroidism)

Chronic Illness and Malnutrition

Chronic organic or systemic disease (e.g., asthma, especially when treated with glucocorticoids; heart or renal disease) Nutritional deprivation Malabsorption syndrome

Functional Endocrine Disorders (Psychosocial Dwarfism)

Chromosomal Disorders (e.g., Turner's Syndrome)

Skeletal Abnormalities (e.g., achondroplasia)

often is delayed. Lateral skull x-rays may be used to evaluate the size and shape of the sella turcica (*i.e.*, depression in the sphenoid bone that contains the pituitary gland) and determine if a pituitary tumor exists. Magnetic resonance imaging (MRI) or computed axial tomography (CT) scans of the hypothalamicpituitary area may be done if a pituitary lesion is suspected. After the cause of short stature has been determined, treatment can be initiated.

Genetic and Constitutional Short Stature. Two forms of short stature, genetic short stature and constitutional short stature, are not disease states but variations from population norms.⁵ Genetically short children tend to be well proportioned and to have a height close to the midparental height of their parents. The midparental height for boys can be calculated by adding 13 cm (5 inches) to the height of the mother, adding the father's height, and dividing the total by two. For girls, 13 cm (5 inches) is subtracted from the father's height, the result is added to the mother's height, and the total is divided by two. Ninety-five percent of normal children are within 8.5 cm of the midparental height.

Constitutional short stature is a term used to describe children (particularly boys) who have moderately short stature, thin

542 Unit Eight: Alterations in the Endocrine System

build, delayed skeletal and sexual maturation, and absence of other causes of decreased growth. *Catch-up growth* is a term used to describe an abnormally high growth rate that occurs as a child approaches normal height for age. It occurs after the initiation of therapy for GH deficiency and hypothyroidism and the correction of chronic diseases.

Psychosocial Dwarfism. Psychosocial dwarfism involves a functional hypopituitarism and is seen in some emotionally deprived children. These children usually present with poor growth, potbelly, and poor eating and drinking habits.^{5,6} Typically, there is a history of disturbed family relationships in which the child has been severely neglected or disciplined. Often, the neglect is confined to one child in the family. GH function usually returns to normal after the child is removed from the constraining environment. The prognosis depends on improvement in behavior and catch-up growth. Family therapy usually is indicated, and foster care may be necessary.

Growth Hormone Deficiency in Children. There are seseveral forms of GH deficiency that present in childhood. Children with idiopathic GH deficiency lack the hypothalamic GHRH but have adequate somatotropes, whereas children with pituitary tumors or agenesis of the pituitary lack somatotropes. The term *panhypopituitarism* refers to conditions that cause a deficiency of all of the anterior pituitary hormones. In a rare condition called *Laron-type dwarfism*, GH levels are normal or elevated, but there is a hereditary defect in IGF production that can be treated directly with IGF-1 replacement.⁷

Congenital GH deficiency is associated with normal birth length, followed by a decrease in growth rate that can be identified by careful measurement during the first year and that becomes obvious by 1 to 2 years of age. Persons with classic GH deficiency have normal intelligence, short stature, obesity with immature facial features, and some delay in skeletal maturation (Fig. 31-2). Puberty often is delayed, and males with the disorder have microphallus (abnormally small penis), especially if the condition is accompanied by gonadotropinreleasing hormone (GnRH) deficiency. In the neonate, GH deficiency can lead to hypoglycemia and seizures; if adrenocorticotropic hormone (ACTH) deficiency also is present, the hypoglycemia often is more severe. Acquired GH deficiency develops in later childhood; it may be caused by a hypothalamicpituitary tumor, particularly if it is accompanied by other pituitary hormone deficiencies.

When short stature is due to a GH deficiency, GH replacement therapy is the treatment of choice. GH is species specific, and only human GH (hGH) is effective in humans. Human GH, which is synthesized by recombinant DNA techniques, is now available in adequate supply. It is administered subcutaneously in multiple weekly doses during the period of active growth, and its use can be continued into adulthood.

Children with short stature due to Turner's syndrome and chronic renal insufficiency also are treated with hGH.⁸ hGH therapy may be considered for children with short stature but without GH deficiency. Several studies suggest that short-term treatment with GH increases the rate of growth in these children. Although the effect of GH on adult height is not great, it can result in improved psychological well-being. There are concerns about misuse of the drug to produce additional growth in children with normal GH function who are of near-



normal height. Guidelines for use of the hormone continue to be established.

Growth Hormone Deficiency in Adults

There are two categories of GH deficiency in adults: (1) GH deficiency that was present in childhood, and (2) GH deficiency that developed during adulthood, mainly as the result of hypopituitarism resulting from a pituitary tumor or its treatment. GH levels also can decline with aging, and there has been interest in the effects of declining GH levels in the elderly (described as the *somatopause*).

Several studies have shown that cardiovascular mortality is increased in GH-deficient adults.⁹ Increased arterial intimamedia thickness and a higher prevalence of atherosclerotic plaques and endothelial dysfunction have been reported in both childhood and adult GH deficiency. The GH deficiency syndrome is associated with a cluster of cardiovascular risk factors, including central adiposity (increased waist-hip ratio), increased visceral fat, insulin resistance, and dyslipidemia. These features also are associated with the *metabolic syndrome* (syndrome X; see Chapter 32). In addition to these so-called traditional cardiovascular risk factors, nontraditional cardiovascular risk factors (*e.g.*, C-reactive protein and interleukin-6, which are markers of the inflammatory pathway) also are elevated.

The diagnosis of GH deficiency in adults is made by finding subnormal serum GH responses to two stimulation tests that measure GH reserve. Measurements of the serum IGF-1 or basal GH do not distinguish reliably between normal and subnormal GH secretion in adults. Insulin-induced hypoglycemia is the gold standard test for GH reserve. The L-dopa test probably is the next best test. Other stimulation tests involve the use of arginine or arginine plus GHRH, clonidine (an α -adrenergic agonist), glucagon, or GHRH.

GH replacement obviously is important in the growing child; however, the role in adults (especially for the somato-

pause) is being assessed. Adults with documented GH deficiency may be treated with hGH. In the United States, persons with GH deficiency acquired as an adult must meet at least two criteria for therapy: a poor GH response to at least two standard stimuli, and hypopituitarism caused by pituitary or hypothalamic damage.⁸

Tall Stature and Growth Hormone Excess in Children

Tall Stature. Just as there are children who are short for their age and sex, there also are children who are tall for their age and sex. Normal variants of tall stature include genetic tall stature and constitutional tall stature. Children with exceptionally tall parents tend to be taller than children with shorter parents. The term *constitutional tall stature* is used to describe a child who is taller than his or her peers and is growing at a velocity that is within the normal range for bone age.^{5,10} Other causes of tall stature are genetic or chromosomal disorders, such as Marfan's syndrome or XYY syndrome (see Chapter 4). Endocrine causes of tall stature include sexual precocity because of early onset of estrogen and androgen secretion and excessive GH.

Exceptionally tall children (*i.e.*, genetic tall stature and constitutional tall stature) can be treated with sex hormones estrogens in girls and testosterone in boys—to effect early epiphyseal closure. Such treatment is undertaken only after full consideration of the risks involved. To be effective, such treatment must be instituted 3 to 4 years before expected epiphyseal fusion.

Gigantism. Growth hormone excess occurring before puberty and the fusion of the epiphyses of the long bones results in gigantism (Fig. 31-3). Excessive secretion of GH by somatotrope adenomas causes gigantism in the prepubertal child. It occurs when the epiphyses are not fused and high levels of IGF stimulate excessive skeletal growth. Fortunately, the condition is rare because of early recognition and treatment of the adenoma.

Growth Hormone Excess in Adults

When GH excess occurs in adulthood or after the epiphyses of the long bones have fused, the condition is referred to as *acromegaly*.¹¹ Acromegaly results from excess levels of GH that stimulate the hepatic secretion of IGF-1, which causes most of the clinical manifestations of acromegaly. The annual incidence of acromegaly is 3 to 4 cases per 1 million people, with a mean age at the time of diagnosis of 40 to 45 years.

The most common cause of acromegaly is a somatotrope adenoma. Approximately 75% of persons with acromegaly have large expansive tumors that erode the sella turcica and impinge on adjacent cranial structures, and most of the remainder have small tumors contained within the pituitary gland. Other causes of acromegaly (<5%) are excess secretion of GHRH by hypothalamic tumors, ectopic GHRH secretion by nonendocrine tumors such as carcinoid tumors or small cell lung cancers, and ectopic secretion of GH by nonendocrine tumors.

The disorder usually has an insidious onset, and symptoms often are present for a considerable period before a diagnosis is made. When the production of excessive GH occurs after the epiphyses of the long bones have closed, as in the adult, the person cannot grow taller, but the soft tissues continue to grow. Enlargement of the small bones of the hands and feet and of the membranous bones of the face and skull results in a pro-



nounced enlargement of the hands and feet, a broad and bulbous nose, a protruding lower jaw, and a slanting forehead. The teeth become splayed, causing a disturbed bite and difficulty in chewing. The cartilaginous structures in the larynx and respiratory tract also become enlarged, resulting in a deepening of the voice and tendency to develop bronchitis. Vertebral changes often lead to kyphosis, or hunchback. Bone overgrowth often leads to arthralgias and degenerative arthritis of the spine, hips, and knees. Virtually every organ of the body is increased in size. Enlargement of the heart and accelerated atherosclerosis may lead to an early death.

The metabolic effects of excess levels of GH include alterations in fat and carbohydrate metabolism. GH causes increased release of free fatty acids from adipose tissue, leading to increased concentration of free fatty acids in body fluids. In addition, GH enhances the formation of ketones and the utilization of free fatty acids for energy in preference to use of carbohydrates and proteins. GH exerts multiple effects on carbohydrate metabolism, including decreased glucose uptake by tissues such as skeletal muscle and adipose tissue, increased glucose production by the liver, and increased insulin secretion. Each of these changes results in GH-induced insulin resistance (see Chapter 32). This leads to glucose intolerance, which stimulates the beta cells of the pancreas to produce additional insulin. Long-term elevation of GH results in overstimulation of the beta cells, causing them literally to "burn out." Impaired glucose tolerance occurs in as many as 50% to 70% of persons with acromegaly; overt diabetes mellitus subsequently can result.

The pituitary gland is located in the pituitary fossa of the sphenoid bone (i.e., sella turcica), which lies directly below the optic nerve. Enlargement of the pituitary gland eventually causes erosion of the surrounding bone, and because of its location, this can lead to headaches, visual field defects resulting from compression of the optic nerve, and palsies of cranial nerves III, IV, and VI. Compression of other pituitary structures can cause secondary hypothyroidism, hypogonadism, and adrenal insufficiency. Other manifestations include excessive sweating with an unpleasant odor, oily skin, heat intolerance, moderate weight gain, muscle weakness and fatigue, menstrual irregularities, and decreased libido. Hypertension is relatively common. Paresthesias may develop because of nerve entrapment and compression caused by excess soft tissue and accumulation of subcutaneous fluid (especially carpal tunnel syndrome). Acromegaly also is associated with an increased risk of colonic polyps and colorectal cancer. The mortality rate of patients with acromegaly is two to three times the expected rate, mostly from cardiovascular diseases and cancer. The cardiovascular disease results from the combination of cardiomyopathy, hypertension, insulin resistance and hyperinsulinemia, and hyperlipidemia.

Acromegaly often develops insidiously, and only a small number of persons seek medical care because of changes in appearance. The diagnosis of acromegaly is facilitated by the typical features of the disorder—enlargement of the hands and feet and coarsening of facial features. Laboratory tests to detect elevated levels of GH not suppressed by a glucose load are used to confirm the diagnosis. CT and MRI scans can detect and localize the pituitary lesions. Because most of the effects of GH are mediated by IGF-1, IGF-1 levels may provide information about disease activity.

The treatment goals for acromegaly focus on the correction of metabolic abnormalities and include normalization of the GH response to an oral glucose load; normalization of IGF-1 levels to age- and gender-matched control levels; removal or reduction of the tumor mass; relieving the central pressure effects; improvement of adverse clinical features; and normalization of the mortality rate.¹² Pituitary tumors can be removed surgically using the transsphenoidal approach or, if that is not possible, a transfrontal craniotomy. Radiation therapy may be used, but remission (reduction in GH levels) may not occur for several years after therapy. Radiation therapy also significantly increases the risk of hypopituitarism, hypothyroidism, hypoadrenalism, and hypogonadism.⁵

Medications used for the treatment of acromegaly include the somatostatin analogs, which produce feedback inhibition of GH. Sustained-release preparations, which effectively inhibit GH secretion for 30 days after a single intramuscular injection, are now available.⁴

Isosexual Precocious Puberty

Sexual development is considered precocious and warrants investigation when it occurs before 8 years of age for girls and before 9 years of age for boys. Precocious sexual development may be idiopathic or may be caused by gonadal, adrenal, or hypothalamic tumors. True isosexual precocious puberty is caused by early activation of the hypothalamic-pituitary-gonadal axis. The gonadotropin-mediated increase in size and activity of the gonads leads to increasing sex hormone production with early development of sexual characteristics and fertility.^{13,14} Benign and malignant tumors of the central nervous system (CNS) can cause precocious puberty. These tumors are thought to remove the inhibitory influences normally exerted on the hypothalamus during childhood. Gonadotropin-independent causes of precocious puberty include functioning ovarian tumors and feminizing adrenal tumors in girls and congenital adrenal hyperplasia and Leydig cell tumors in boys.¹³

Diagnosis of precocious puberty is based on physical findings of early thelarche (*i.e.*, beginning of breast development), adrenarche (*i.e.*, beginning of augmented adrenal androgen production), and menarche (*i.e.*, beginning of menstrual function) in girls. The most common sign in boys is early genital enlargement. Radiologic findings may indicate advanced bone age. Persons with precocious puberty usually are tall for their age as children but short as adults because of the early closure of the epiphyses. CT or MRI should be used to exclude intracranial lesions.

Depending on the cause of precocious puberty, the treatment may involve surgery, medication, or no treatment. The treatment of choice for gonadotropin-dependent precocious puberty is administration of a long-acting GnRH agonist. Constant levels of the hormone cause a decrease in pituitary responsiveness to GnRH, leading to decreased secretion of gonadotropic hormones and sex steroids. Parents often need education, support, and anticipatory guidance in dealing with their feelings and the child's physical needs and in relating to a child who appears older than his or her years.

In summary, hypopituitarism, which is characterized by a decreased secretion of pituitary hormones, is a condition that affects many of the other endocrine systems. Depending on the extent of the disorder, it can result in decreased levels of GH, thyroid hormones, adrenal corticosteroid hormones, and testosterone in the male and of estrogens and progesterone in the female.

A number of hormones are essential for normal body growth and maturation, including GH, insulin, thyroid hormone, and androgens. GH exerts its growth effects through a group of IGFs. GH also exerts an effect on metabolism and is produced in the adult and in the child. Its metabolic effects include a decrease in peripheral use of carbohydrates and an increased mobilization and use of fatty acids.

In children, alterations in growth include short stature, tall stature, and isosexual precocious puberty. Short stature is a condition in which the attained height is well below the fifth percentile or the linear growth velocity is below normal for a child's age or gender. Short stature can occur as a variant of normal growth (*i.e.*, genetic short stature or constitutional short stature) or as the result of endocrine disorders, chronic illness, malnutrition, emotional disturbances, or chromosomal disorders. Short stature resulting from GH deficiency can be treated with human GH preparations. In adults, GH deficiency represents a deficiency carried over from childhood or one that develops during adulthood as the result of a pituitary tumor or its treatment. GH levels also can decline with aging,

and there has been interest in the effects of declining GH levels in the elderly (described as the *somatopause*).

Tall stature refers to the condition in which children are tall for their age and gender. It can occur as a variant of normal growth (*i.e.*, genetic tall stature or constitutional tall stature) or as the result of a chromosomal abnormality or GH excess. GH excess in adults results in acromegaly, which involves proliferation of bone, cartilage, and soft tissue along with the metabolic effects of excessive hormone levels.

Isosexual precocious puberty defines a condition of early activation of the hypothalamic-pituitary-gonadal axis (*i.e.*, before 8 years of age in girls and 9 years of age in boys), resulting in the development of appropriate sexual characteristics and fertility. It causes tall stature during childhood but results in short stature in adulthood because of the early closure of the epiphyses.

THYROID DISORDERS

Control of Thyroid Function

The thyroid gland is a shield-shaped structure located immediately below the larynx in the anterior middle portion of the neck. It is composed of a large number of tiny, saclike structures called *follicles* (Fig. 31-4). These are the functional units of the thyroid. Each follicle is formed by a single layer of epithelial (follicular) cells and is filled with a secretory substance called *colloid*, which consists largely of a glycoprotein-iodine complex called *thyroglobulin*.

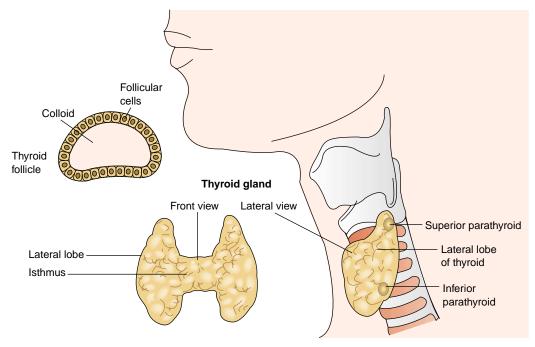
The thyroglobulin that fills the thyroid follicles is a large glycoprotein molecule that contains 140 tyrosine amino acids. In the process of thyroid synthesis, iodine is attached to these tyrosines. Both thyroglobulin and iodide are secreted into the colloid of the follicle by the follicular cells.

The thyroid is remarkably efficient in its use of iodide. A daily absorption of 100 to 200 μ g of dietary iodide is sufficient to form normal quantities of thyroid hormone. In the process of removing it from the blood and storing it for future use, iodide is pumped into the follicular cells against a concentration gradient. As a result, the concentration of iodide in the normal thyroid gland is approximately 30 times that in the blood.²

Once inside the follicle, most of the iodide is oxidized by the enzyme peroxidase in a reaction that facilitates combination with a tyrosine molecule to form monoiodotyrosine and then diiodotyrosine (Fig. 31-5). Two diiodotyrosine residues are coupled to form thyroxine (T_4), or a monoiodotyrosine and a diiodotyrosine are coupled to form triiodothyronine (T_3). Only T_4 (93%) and T_3 (7%) are released into the circulation.² There is evidence that T_3 is the active form of the hormone and that T_4 is converted to T_3 before it can act physiologically.

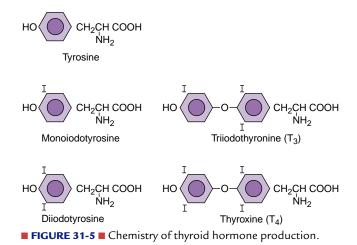
Thyroid hormones are bound to thyroid-binding globulin and other plasma proteins for transport in the blood. Only the free hormone enters cells and regulates the pituitary feedback mechanism. Protein-bound thyroid hormone forms a large reservoir that is slowly drawn on as free thyroid hormone is needed. There are three major thyroid-binding proteins: thyroid hormone-binding globulin (TBG), thyroxine-binding prealbumin (TBPA), and albumin. More than 99% of T₄ and T₃ is carried in the bound form. TBG carries approximately 70% of T₄ and T₃; TBPA binds approximately 10% of circulating T₄ and lesser amounts of T₃; and albumin binds approximately 15% of circulating T₄ and T₃.

A number of disease conditions and pharmacologic agents can decrease the amount of binding protein in the plasma or influence the binding of hormone. Congenital TBG deficiency



■ FIGURE 31-4 ■ The thyroid gland and the follicular structure.

546 Unit Eight: Alterations in the Endocrine System



is an X-linked trait that occurs in 1 of every 2500 live births. Corticosteroid medications and systemic disease conditions such as protein malnutrition, nephrotic syndrome, and cirrhosis decrease TBG concentrations. Medications such as phenytoin, salicylates, and diazepam may bind to TBG, displacing T_4 and T_3 , effectively producing a low TBG state.

The secretion of thyroid hormone is regulated by the hypothalamic-pituitary-thyroid feedback system (Fig. 31-6). In this system, thyrotropin-releasing hormone (TRH), which is produced by the hypothalamus, controls the release of thyroidstimulating hormone (TSH) from the anterior pituitary gland. TSH, in turn, increases the overall activity of the thyroid gland by promoting the release of thyroid hormone into the bloodstream and increasing T_4 and T_3 synthesis. The effect of TSH on the release of thyroid hormones occurs within approximately 30 minutes, but the other effects require days or weeks.

Increased levels of thyroid hormone act in the feedback inhibition of TRH or TSH. High levels of iodide (e.g., from

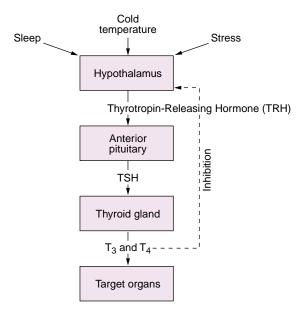


FIGURE 31-6 The hypothalamic-pituitary-thyroid feedback system, which regulates the body levels of thyroid hormone.

iodide-containing cough syrup or kelp tablets) also cause a temporary decrease in thyroid activity that lasts for several weeks, probably through a direct inhibition of TSH on the thyroid. Cold exposure is one of the strongest stimuli for increased thyroid hormone production and probably is mediated through TRH from the hypothalamus. Various emotional reactions also can affect the output of TRH and TSH and therefore indirectly affect secretion of thyroid hormones.

Actions of Thyroid Hormone

All the major organs in the body are affected by altered levels of thyroid hormone. Thyroid hormone has two major functions: it increases metabolism and protein synthesis, and it is necessary for growth and development in children, including mental development and attainment of sexual maturity.

Metabolic Rate. Thyroid hormone increases the metabolism of all body tissues except the retina, spleen, testes, and lungs. The basal metabolic rate can increase by 60% to 100% above normal when large amounts of T_4 are present.² As a result of this higher metabolism, the rate of glucose, fat, and protein use increases. Lipids are mobilized from adipose tissue, and the catabolism of cholesterol by the liver is increased. Blood levels of cholesterol are decreased in hyperthyroidism and increased in hypothyroidism. Muscle proteins are broken down and used as fuel, probably accounting for some of the muscle fatigue that occurs with hyperthyroidism. The absorption of glucose from the gastrointestinal tract is increased. Because vitamins are essential parts of metabolic enzymes and coenzymes, an increase in the metabolic rate "speeds up" the use of vitamins and tends to cause vitamin deficiency.

Cardiovascular Function. Cardiovascular and respiratory functions are strongly affected by thyroid function. With an increase in metabolism, there is an increase in oxygen consumption and production of metabolic end-products, with an accompanying increase in vasodilatation. Blood flow to the skin, in particular, is augmented as a means of dissipating the body heat that results from the higher metabolism. Blood volume, cardiac output, and ventilation all are increased as a means of maintain-

KEY CONCEPTS

THYROID HORMONE

- Thyroid hormone increases the metabolism and protein synthesis in nearly all of the tissues of the body.
- Hypothyroidism produces a decrease in metabolic rate, an accumulation of a hydrophilic mucopolysaccharide substance (myxedema) in the connective tissues throughout the body, and an elevation in serum cholesterol.
- Hyperthyroidism has an effect opposite that of hypothyroidism. It produces an increase in metabolic rate and oxygen consumption, increased use of metabolic fuels, and increased sympathetic nervous ystem responsiveness.

ing blood flow and oxygen delivery to body tissues. Heart rate and cardiac contractility are enhanced as a means of maintaining the needed cardiac output. However, blood pressure is likely to change little because the increase in vasodilatation tends to offset the increase in cardiac output.

Gastrointestinal Function. Thyroid hormone enhances gastrointestinal function, causing an increase in motility and production of gastrointestinal secretions that often results in diarrhea. An increase in appetite and food intake accompanies the higher metabolic rate that occurs with increased thyroid hormone levels. At the same time, weight loss occurs because of the increased use of calories.

Neuromuscular Effects. Thyroid hormone has marked effects on neural control of muscle function and tone. Slight elevations in hormone levels cause skeletal muscles to react more vigorously, and a drop in hormone levels causes muscles to react more sluggishly. In the hyperthyroid state, a fine muscle tremor is present. The cause of this tremor is unknown, but it may represent an increased sensitivity of the neural synapses in the spinal cord that control muscle tone.

In the infant, thyroid hormone is necessary for normal brain development. The hormone enhances cerebration; in the hyperthyroid state, it causes extreme nervousness, anxiety, and difficulty in sleeping.

Evidence suggests a strong interaction between thyroid hormone and the sympathetic nervous system. Many of the signs and symptoms of hyperthyroidism suggest overactivity of the sympathetic division of the autonomic nervous system, such as tachycardia, palpitations, and sweating. Tremor, restlessness, anxiety, and diarrhea also may reflect autonomic nervous system imbalances. Drugs that block sympathetic activity have proved to be valuable adjuncts in the treatment of hyperthyroidism because of their ability to relieve some of these undesirable symptoms.

Tests of Thyroid Function

The diagnosis of thyroid disorders is based on tests of $T_{3_{4}}$ $T_{4_{4}}$ and TSH levels, radioiodine uptake studies, thyroid scans, ultrasound, and CT or MRI scans.¹⁵ Measures of T₃, T₄, and TSH have been made available through immunoassay methods. The free T_4 test measures the unbound portion of T_4 that is free to enter cells to produce its effects. Ultrasensitive TSH measurements are used to diagnose hyperthyroidism and hypothyroidism, as well differentiate between primary and secondary thyroid disorders. T₃, T₄, and free T₄ levels are low in primary hypothyroidism, and the TSH level (via negative feedback) is elevated.

The radioiodine (123I) uptake test measures the ability of the thyroid gland to remove and concentrate iodine from the blood. Thyroid scans (i.e., 123I, 99mTc-pertechnetate) can be used to detect thyroid nodules and determine the functional activity of the thyroid gland. Ultrasonography can be used to differentiate cystic from solid thyroid lesions, and CT and MRI scans are used to demonstrate tracheal compression or impingement on other neighboring structures.

Alterations in Thyroid Function

An alteration in thyroid function can represent a hypofunctional or a hyperfunctional state. The manifestations of these two altered states are summarized in Table 31-1. Disorders of the thyroid may be caused by a congenital defect in thyroid development, or they may develop later in life, with a gradual or sudden onset.

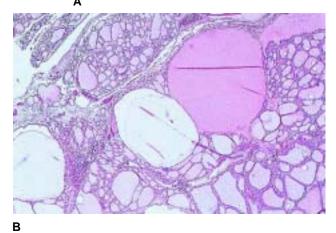
Goiter is an increase in the size of the thyroid gland. It can occur in hypothyroid, euthyroid, and hyperthyroid states.

TABLE 31-1 Manifestations of Hypothyroid and Hyperthyroid States				
Level of Organization	Hypothyroidism	Hyperthyroidism		
Basal metabolic rate	Decreased	Increased		
Sensitivity to catecholamines	Decreased	Increased		
General features	Myxedematous features Deep voice Impaired growth (child)	Exophthalmos Lid lag Decreased blinking		
Blood cholesterol levels	Increased	Decreased		
General behavior	Mental retardation (infant) Mental and physical sluggishness Somnolence	Restlessness, irritability, anxiety Hyperkinesis Wakefulness		
Cardiovascular function	Decreased cardiac output Bradycardia	Increased cardiac output Tachycardia and palpitations		
Gastrointestinal function	Constipation Decreased appetite	Diarrhea Increased appetite		
Respiratory function	Hypoventilation	Dyspnea		
Muscle tone and reflexes	Decreased	Increased, with tremor and fibrillatory twitching		
Temperature tolerance	Cold intolerance	Heat intolerance		
Skin and hair	Decreased sweating Coarse and dry skin and hair	Increased sweating Thin and silky skin and hair		
Weight	Gain	Loss		

Goiters may be diffuse, involving the entire gland without evidence of nodularity, or they may contain nodules. Diffuse goiters usually become nodular. Goiters may be toxic, producing signs of extreme hyperthyroidism, or thyrotoxicosis, or they may be nontoxic. Diffuse nontoxic and multinodular goiters are the result of compensatory hypertrophy and hyperplasia of follicular epithelium from some derangement that impairs thyroid hormone output (Fig. 31-7).

The degree of thyroid enlargement usually is proportional to the extent and duration of thyroid deficiency. Multinodular goiters produce the largest thyroid enlargements and often are associated with thyrotoxicosis. When sufficiently enlarged, they





■ FIGURE 31-7 ■ (A) Middle-aged woman with a nontoxic (nodular) goiter that has enlarged to produce a conspicuous neck mass. (B) Microscopic view of one of the macroscopic nodules shows marked variation in size of the follicles. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 1164]. Philadelphia: Lippincott Williams & Wilkins) may compress the esophagus and trachea, causing difficulty in swallowing, a choking sensation, and inspiratory stridor. Such lesions also may compress the superior vena cava, producing distention of the veins of the neck and upper extremities, edema of the eyelids and conjunctiva, and syncope with coughing.

Hypothyroidism

Hypothyroidism can occur as a congenital or an acquired defect. Congenital hypothyroidism develops prenatally and is present at birth. Acquired hypothyroidism develops later in life because of primary disease of the thyroid gland or secondary to disorders of hypothalamic or pituitary origin.

Congenital Hypothyroidism

Congenital hypothyroidism is a common cause of preventable mental retardation. It affects approximately 1 of 4000 infants.¹⁶ Hypothyroidism in the infant may result from a congenital lack of the thyroid gland, abnormal biosynthesis of thyroid hormone, or deficient TSH secretion. With congenital lack of the thyroid gland, the infant usually appears normal and functions normally at birth because of the transplacental passage of moderate amounts of maternal T₄. The manifestations of untreated congenital hypothyroidism are referred to as *cretinism*. However, the term does not apply to the normally developing infant in whom replacement thyroid hormone therapy was instituted shortly after birth.

Thyroid hormone is essential for normal brain development and growth, almost half of which occurs during the first 6 months of life. If untreated, congenital hypothyroidism causes mental retardation and impairs growth. Fortunately, neonatal screening tests have been instituted to detect congenital hypothyroidism during early infancy. Screening usually is done in the hospital nursery. In this test, a drop of blood is taken from the infant's heel and analyzed for T_4 and TSH.

Transient congenital hypothyroidism has been recognized more frequently since the introduction of neonatal screening. It is characterized by high TSH levels and low thyroid hormone levels. The fetal and infant thyroids are sensitive to iodine excess. Iodine crosses the placenta and mammary glands and is readily absorbed by infant skin. Transient hypothyroidism may be caused by maternal or infant exposure to substances such as topical iodine-containing disinfectants (*e.g.*, iodine-containing douches or nursery disinfectants). Antithyroid drugs such as propylthiouracil, methimazole, and carbimazole also cross the placenta and block fetal thyroid function.

Congenital hypothyroidism is treated by hormone replacement. Evidence indicates that it is important to normalize T_4 levels as rapidly as possible (in the first 6 weeks of life) because a delay is accompanied by poorer psychomotor and mental development. Dosage levels are adjusted as the child grows. Infants with transient hypothyroidism usually can have the replacement therapy withdrawn at 6 to 12 months. When early and adequate treatment regimens are followed, the risk of mental retardation in infants detected by screening programs essentially is nonexistent.

Acquired Hypothyroidism and Myxedema

Acquired hypothyroidism in older children and adults represents a decrease in thyroid function resulting from destruction or dysfunction of the thyroid gland (*i.e.*, primary hypothyroidism) or impaired hypothalamic or pituitary function (*i.e.*, secondary hypothyroidism).

Primary hypothyroidism is much more common than secondary hypothyroidism. It may result from thyroidectomy (*i.e.*, surgical removal) or ablation of the gland with radiation. Certain goitrogenic agents, such as lithium carbonate (i.e., used in the treatment of manic-depressive states) and the antithyroid drugs propylthiouracil and methimazole in continuous dosage can block hormone synthesis and produce hypothyroidism with goiter. Iodine-containing drugs (e.g., kelp tablets, iodide-containing cough syrups, radiographic contrast media) also can block thyroid hormone production, particularly in persons with autoimmune thyroid disease. Amiodarone (an antiarrhythmic drug), which contains 75 mg of iodine per 200-mg tablet, is being increasingly implicated in causing thyroid problems. Iodine deficiency, which can cause goiter and hypothyroidism, is rare in the United States because of the widespread use of iodized salt and other iodide sources.

Hashimoto's thyroiditis is an autoimmune disorder in which the thyroid gland may be totally destroyed by an immunologic process.¹ It is the major cause of goiter and hypothyroidism in children and adolescents.¹⁶ Hashimoto's thyroiditis is predominantly a disease of women, with a female-to-male ratio of 10:1 to 20:1.¹ The course of the disease varies. At the onset, only a goiter may be present. In time, hypothyroidism usually becomes evident. Although the disorder usually causes hypothyroidism, a hyperthyroid state may develop midcourse in the disease. The transient hyperthyroid state is caused by leakage of preformed thyroid hormone from damaged cells of the gland.

Clinical Manifestations. The clinical manifestations of hypothyroidism can range from mild nonspecific complaints associated with subclinical hypothyroidism to those associated with overt hypothyroidism.¹⁷ The manifestations of the overt hypothyroidism are related largely to two factors: the hypometabolic state resulting from thyroid hormone deficiency, and myxedematous involvement of body tissues.

The hypometabolic state is characterized by a gradual onset of weakness and fatigue, a tendency to gain weight despite a loss of appetite, and cold intolerance. As the condition progresses, the skin becomes dry and rough and acquires a pale yellowish cast, which primarily results from carotene deposition, and the hair becomes coarse and brittle. There can be loss of the lateral one third of the eyebrows. Gastrointestinal motility is decreased, producing constipation, flatulence, and abdominal distention. Nervous system involvement is manifested in mental dullness, lethargy, and impaired memory.

Myxedema represents the presence of a nonpitting mucous type of edema caused by an accumulation of a hydrophilic mucopolysaccharide substance in the connective tissues throughout the body. As a result of myxedematous fluid accumulation, the face takes on a characteristic puffy look, especially around the eyes, the tongue becomes enlarged, and the voice hoarse and husky (Fig. 31-8). Mucopolysaccharide deposits in the heart can cause generalized cardiac dilatation, bradycardia, and other signs of altered cardiac function. The signs and symptoms of hypothyroidism are summarized in Table 31-1.

Diagnosis and Treatment. Diagnosis of hypothyroidism is based on history and physical examination and laboratory mea-



surement of TSH and free T_4 levels. The condition is treated with replacement therapy using synthetic preparations of T_3 or T_4 . Most people are treated with T_4 . Serum TSH levels are used to estimate the adequacy of T_4 replacement therapy.

Myxedematous Coma. Myxedematous coma is a lifethreatening, end-stage expression of hypothyroidism. It is characterized by coma, hypothermia, cardiovascular collapse, hypoventilation, and severe metabolic disorders that include hyponatremia, hypoglycemia, and lactic acidosis.¹⁸ It occurs most often in elderly women who have chronic hypothyroidism from a spectrum of causes. It occurs more frequently in the winter months, which suggests that cold exposure may be a precipitating factor. The severely hypothyroid person is unable to metabolize sedatives, analgesics, and anesthetic drugs, and buildup of these agents may precipitate coma.

Treatment includes aggressive management of precipitating factors; supportive therapy such as management of cardiorespiratory status, hyponatremia, and hypoglycemia; and thyroid replacement therapy. Prevention is preferable to treatment and entails special attention to high-risk populations, such as women with a history of Hashimoto's thyroiditis. These persons should be informed about the signs and symptoms of severe hypothyroidism and the need for early medical treatment.

Hyperthyroidism

Hyperthyroidism, or thyrotoxicosis, results from excessive delivery of thyroid hormone to the peripheral tissues. The most common cause of hyperthyroidism is Graves' disease, which is accompanied by ophthalmopathy (exophthalmos, *i.e.*, bulging of the eyeballs) and goiter. Other causes of hyperthyroidism are multinodular goiter, adenoma of the thyroid, and occasionally, ingestion of excessive thyroid hormone. Thyroid crisis, or storm, is an acutely exaggerated manifestation of the hyperthyroid state.

Many of the manifestations of hyperthyroidism are related to the increase in oxygen consumption and use of metabolic fuels associated with the hypermetabolic state, as well as to the increase in sympathetic nervous system activity. With the hypermetabolic state, there are frequent complaints of nervousness, irritability, and fatigability. Weight loss is common, despite a large appetite. Other manifestations include tachycardia, palpitations, shortness of breath, excessive sweating, muscle cramps, and heat intolerance. The person appears restless and has a fine muscle tremor. Even in persons without exophthalmos, there is an abnormal retraction of the eyelids and infrequent blinking such that they appear to be staring. The hair and skin usually are thin and have a silky appearance. The signs and symptoms of hyperthyroidism are summarized in Table 31-1.

The treatment of hyperthyroidism is directed toward reducing the level of thyroid hormone. This can be accomplished with eradication of the thyroid gland with radioactive iodine, through surgical removal of part or all of the gland, or the use of drugs that decrease thyroid function and thereby the effect of thyroid hormone on the peripheral tissues. Destruction of thyroid tissue with radioactive iodine is used more frequently than surgery. The β -adrenergic–blocking drug propranolol may be used to block the effects of the hyperthyroid state on sympathetic nervous system function. It is given in conjunction with other antithyroid drugs such as propylthiouracil and methimazole. These drugs prevent the thyroid gland from converting iodine to its organic (hormonal) form and block the conversion of T₄ to T₃ in the tissues.

Graves' Disease

Graves' disease is a state of hyperthyroidism, goiter, and ophthalmopathy (or, less commonly, dermopathy).^{19,20} The onset usually is between the ages of 20 and 40 years, and women are five times more likely to experience the disease than men. Graves' disease is an autoimmune disorder characterized by abnormal stimulation of the thyroid gland by thyroid-stimulating antibodies (thyroid-stimulating immunoglobulins [TSI]) that act through the normal TSH receptors. It may be associated with other autoimmune disorders, such as myasthenia gravis and pernicious anemia. The disease is associated with human leukocyte antigen (HLA)-DR3 and HLA-B8, and a familial tendency is evident.

The exophthalmos, which occurs in as many as one third of persons with Graves' disease, is thought to result from thyroidstimulating antibodies sensitized to interact with antigens found in fibroblasts in orbital tissue behind the eyeball and in the extraocular muscles that move the eyeball.^{21,22} The ophthalmopathy of Graves' disease can cause severe eye problems, including paralysis of the extraocular muscles; involvement of the optic nerve, with some visual loss; and corneal ulceration because the lids do not close over the protruding eyeball. The exophthalmos usually tends to stabilize after treatment of the hyperthyroidism. Unfortunately, not all of the ocular changes are reversible with treatment. Smoking tends to aggravate the condition. Figure 31-9 depicts a woman with Graves' disease.



■ FIGURE 31-9 ■ Graves' disease. A young woman with hyperthyroidism presented with a mass in the neck and exophthalmos. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 1167]. Philadelphia: Lippincott Williams & Wilkins)

Thyroid Storm

Thyroid storm, or crisis, is an extreme and life-threatening form of thyrotoxicosis, rarely seen today because of improved diagnosis and treatment methods. When it does occur, it is seen most often in undiagnosed cases or in persons with hyperthyroidism who have not been adequately treated. It often is precipitated by stress such as an infection (usually respiratory), diabetic ketoacidosis, physical or emotional trauma, or manipulation of a hyperactive thyroid gland during thyroidectomy. Thyroid storm is manifested by a very high fever, extreme cardiovascular effects (*i.e.*, tachycardia, congestive failure, and angina), and severe CNS effects (*i.e.*, agitation, restlessness, and delirium). Thyroid storm requires rapid diagnosis and implementation of treatment.

In summary, thyroid hormones play a role in the metabolic process of almost all body cells and are necessary for normal physical and mental growth in the infant and young child. Alterations in thyroid function can manifest as a hypothyroid or a hyperthyroid state. Hypothyroidism can occur as a congenital or an acquired defect. Congenital hypothyroidism leads to mental retardation and impaired physical growth unless treatment is initiated during the first months of life. Acquired hypothyroidism leads to a decrease in metabolic rate and an accumulation of a mucopolysaccharide substance in the intercellular spaces; this substance attracts water and causes a mucous type of edema called myxedema. Hyperthyroidism causes an increase in metabolic rate and alterations in body function similar to those produced by enhanced sympathetic nervous system activity. Graves' disease, which is caused by thyroid-stimulating antibodies, is characterized by hyperthyroidism, goiter, and ophthalmopathy.

PARATHYROID HORMONE DISORDERS

Parathyroid hormone (PTH), a major regulator of serum calcium and phosphate, is secreted by the parathyroid glands. There are four parathyroid glands located on the dorsal surface of the thyroid gland.¹ The dominant regulator of PTH is a decrease in serum calcium concentration. A unique calcium receptor within the parathyroid cell membrane responds rapidly to changes in serum calcium levels. The response to a decrease in serum calcium is prompt, occurring within seconds. Phosphate does not exert a direct effect on PTH secretion. Instead, it acts indirectly by complexing with calcium and decreasing serum calcium concentration.

The secretion, synthesis, and action of PTH are also influenced by magnesium. Magnesium serves as a cofactor in the generation of cellular energy and is important in the function of second messenger systems. Magnesium's effects on the synthesis and release of PTH are thought to be mediated through these mechanisms.²³ Because of its function in regulating PTH release, severe and prolonged hypomagnesemia can markedly inhibit PTH levels.

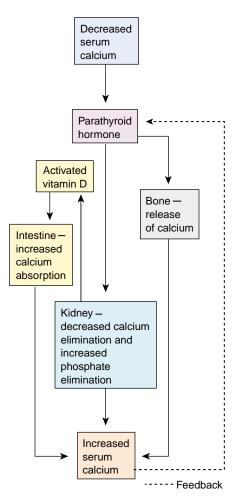
The main function of PTH is to maintain the calcium concentration of the extracellular fluids. It performs this function by promoting the release of calcium from bone, increasing the activation of vitamin D as a means of enhancing intestinal absorption of calcium, and stimulating calcium conservation by the kidney while increasing phosphate excretion (Fig. 31-10).

Parathyroid hormone acts on bone to accelerate the mobilization and transfer of calcium to the extracellular fluid. The skeletal response to PTH is a two-step process. There is an immediate response in which calcium that is present in bone fluid is released into the extracellular fluid and a second more slowly developing response in which completely mineralized bone is resorbed resulting in the release of both calcium and phosphate The actions of PTH in terms of bone resorption require normal levels of both vitamin D and magnesium.

The activation of vitamin D by the kidney is enhanced by the presence of PTH; it is through the activation of vitamin D that PTH increases intestinal absorption of calcium and phosphate. PTH acts directly on the kidney to increase tubular reabsorption of calcium and magnesium while increasing phosphate elimination. The accompanying increase in phosphate elimination ensures that calcium released from bone does not produce hyperphosphatemia and increase the risk of soft tissue deposition of calcium/phosphate crystals.

Hypoparathyroidism

Hypoparathyroidism reflects deficient PTH secretion resulting in hypocalcemia. PTH deficiency may be caused by a congenital absence of all of the parathyroid glands, such as occurs in DiGeorge syndrome.²⁴ An acquired deficiency of PTH may occur after neck surgery, particularly if the surgery involves removal of a parathyroid adenoma, thyroidectomy, or bilateral neck resection for cancer. Hypoparathyroidism may also have an autoimmune origin. Antiparathyroid antibodies have been detected in some persons with hypoparathyroidism, particularly those with multiple endocrine disorders. Other causes of hypoparathyroidism include heavy metal damage, such as occurs with Wilson's disease, metastatic tumors, and infection.



■ FIGURE 31-10 ■ Regulation of serum calcium concentration by parathyroid hormone.

Functional impairment of parathyroid function occurs with magnesium deficiency. Correction of the hypomagnesemia results in rapid disappearance of the condition.

Manifestations of acute hypoparathyroidism that result from a decrease in serum calcium include tetany with muscle cramps, carpopedal spasm, and convulsions (see discussion of hypocalcemia in Chapter 6). Paresthesias, such as tingling of the circumoral area and in the hands and feet, are almost always present. There may be prolongation of the QT interval caused by low calcium levels, resistance to digitalis, hypotension, and refractory heart failure. Symptoms of chronic deficiency include lethargy, anxiety state, and personality changes. There may be blurring of vision caused by cataracts, which develop during an extended period of time. Extrapyramidal signs, such as those seen with Parkinson's disease, may occur because of calcification of the basal ganglia. Successful treatment of the hypocalcemia may improve the disorder and is sometimes associated with decreases in basal ganglia calcification on x-ray. Teeth may be defective if the disorder occurs during childhood.

Diagnosis of hypoparathyroidism is based on low serum calcium levels, high serum phosphate levels, and low serum PTH levels. Serum magnesium levels are usually measured to exclude hypomagnesemia as a cause of the disorder. Acute hypoparathyroid tetany is treated with intravenous calcium gluconate followed by oral administration of calcium salts and vitamin D. Magnesium supplementation is used when the disorder is caused by magnesium deficiency. Persons with chronic hypoparathyroidism are treated with oral calcium and vitamin D. Serum calcium levels are monitored at regular intervals (at least every 3 months) as a means of maintaining serum calcium within a slightly low but asymptomatic range. Maintaining serum calcium within this range helps to prevent hypercalciuria and kidney damage.

Pseudohypoparathyroidism is a rare familial disorder characterized by target tissue resistance to PTH. It is characterized by hypocalcemia, increased parathyroid function, and a variety of congenital defects in the growth and development of the skeleton, including short stature and short metacarpal and metatarsal bones. There are variants in the disorders, with some persons having the pseudohypoparathyroidism with the congenital defects and others having the congenital defects with normal calcium and phosphate levels. The manifestations of the disorder are primarily attributable to hypocalcemia. Treatment is similar to that for hypoparathyroidism.

Hyperparathyroidism

Hyperparathyroidism is caused by hypersecretion of parathyroid hormone. Hyperparathyroidism can manifest as a primary disorder caused by hyperplasia, an adenoma, or carcinoma of the parathyroid glands or as a secondary disorder seen in persons with renal failure.²⁵

Primary hyperparathyroidism is seen more commonly after age 50 years and is more common in women than men. Primary hyperparathyroidism causes hypercalcemia and an increase in calcium in the urine filtrate, resulting in hypercalcuria and the potential for development of kidney stones. Chronic bone resorption may produce diffuse demineralization, pathologic fractures, and cystic bone lesions. Signs and symptoms of the disorder are related to skeletal abnormalities, exposure of the kidney to high calcium levels, and elevated serum calcium levels (see hypercalcemia). Diagnostic procedures include serum calcium and parathyroid hormone levels. Imaging studies of the parathyroid area may be used to identify a parathyroid adenoma.

Secondary hyperparathyroidism involves hyperplasia of the parathyroid glands and occurs primarily in persons with renal failure (see Chapter 24).²⁵ In early renal failure, an increase in PTH results from decreased serum calcium and activated vitamin D levels. As the disease progresses, there is a decrease in vitamin D and calcium receptors, making the parathyroid glands more resistant to vitamin D and calcium. At this point, elevated phosphate levels induce hyperplasia of the parathyroid glands independent of calcium and activated vitamin D.

The bone disease seen in persons with secondary hyperparathyroidism caused by renal failure is known as renal osteodystrophy. Treatment includes resolving the hypercalcemia with large fluid intake. Persons with mild disease are advised to keep active and drink adequate fluids. They also are advised to avoid calcium-containing antacids, vitamin D, and thiazide diuretics, which increase reabsorption of calcium by the kidney. Bisphosphonates (*e.g.*, pamidronate and alendronate), which are potent inhibitors of bone resorption, may be used temporarily to treat the hypercalcemia of hyperparathyroidism. Parathyroidectomy may be indicated in persons with symptomatic hyperthyroidism, kidney stones, or bone disease.

DISORDERS OF ADRENAL CORTICAL FUNCTION

Control of Adrenal Cortical Function

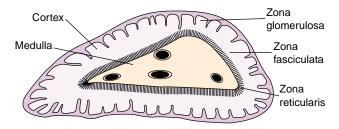
The adrenal glands are small, bilateral structures that weigh approximately 5 g each and lie retroperitoneally at the apex of each kidney (Fig. 31-11). The medulla or inner portion of the gland secretes epinephrine and norepinephrine and is part of the sympathetic nervous system. The cortex forms the bulk of the adrenal gland and is responsible for secreting three types of hormones: the glucocorticoids, the mineralocorticoids, and the adrenal sex hormones. Because the sympathetic nervous system also secretes epinephrine and norepinephrine, adrenal medullary function is not essential for life, but adrenal cortical function is.

Biosynthesis, Transport, and Metabolism

More than 30 hormones are produced by the adrenal cortex. Of these hormones, aldosterone is the principal mineralocorticoid, cortisol (hydrocortisone) is the major glucocorticoid, and androgens are the chief sex hormones.¹ All of the adrenal cortical hormones have a similar structure in that all are steroids and are synthesized from acetate and cholesterol. Each of the steps involved in the synthesis of the various hormones requires a specific enzyme (Fig. 40-11). The secretion of the glucocorticoids and the adrenal androgens is controlled by the ACTH secreted by the anterior pituitary gland.

Cortisol and the adrenal androgens are secreted in an unbound state and bind to plasma proteins for transport in the circulatory system. Cortisol binds largely to corticosteroidbinding globulin and to a lesser extent to albumin. Aldosterone circulates mostly bound to albumin. It has been suggested that the pool of protein-bound hormones may extend the duration of their action by delaying metabolic clearance.

The main site for metabolism of the adrenal cortical hormones is the liver, where they undergo a number of metabolic conversions before being conjugated and made water soluble. They are then eliminated in either urine or bile.



■ FIGURE 31-11 ■ The adrenal gland, showing the medulla and the three layers of the cortex. The zona glomerulosa is the outer layer of the cortex and is primarily responsible for mineralocorticoid production. The middle layer, the zona fasciculata, and the inner layer, the zona reticularis, produce the glucocorticoids and the adrenal sex hormones.

KEY CONCEPTS

ADRENAL CORTICAL HORMONES

- The adrenal cortex produces three types of steroid hormones: the mineralocorticoids (principally aldosterone), which function in sodium, potassium, and water balance; the glucocorticoids (principally cortisol), which aid in regulating the metabolic functions of the body and in controlling the inflammatory response, and are essential for survival in stress situations; and the adrenal sex hormones (principally androgens), which serve mainly as a source of androgens for women.
- The manifestations of adrenal cortical insufficiency are related mainly to mineralocorticoid deficiency and glucocorticoid deficiency.
- The manifestations of adrenal cortical excess are related to mineralocorticoid excess, glucocorticoid excess with derangements in glucose metabolism and impaired ability to respond to stress because of inhibition of inflammatory and immune responses, and signs of increased androgen levels, such as hirsutism in women.

Adrenal Sex Hormones

The adrenal sex hormones are synthesized primarily by the zona reticularis and the zona fasciculata of the cortex (see Fig. 31-12). These sex hormones probably exert little effect on normal sexual function. However, there is evidence that the adrenal sex hormones (the most important of which is dehydroepiandrosterone) contribute to the pubertal growth of body hair, particularly pubic and axillary hair in women. They also may play a role in the steroid hormone economy of the pregnant woman and the fetal-placental unit.

Mineralocorticoids

The mineralocorticoids play an essential role in regulating potassium and sodium levels and water balance. They are produced in the zona glomerulosa, the outer layer of cells of the adrenal cortex. Aldosterone secretion is regulated by the reninangiotensin mechanism and by blood levels of potassium. Increased levels of aldosterone promote sodium retention by the distal tubules of the kidney while increasing urinary losses of potassium. The influence of aldosterone on fluid and electrolyte balance is discussed in Chapter 6, and its effect on blood pressure is discussed in Chapter 16.

Glucocorticoids

The glucocorticoid hormones, mainly cortisol, are synthesized in the zona fasciculata and the zona reticularis of the adrenal gland. The blood levels of these hormones are regulated by negative feedback mechanisms of the hypothalamic-pituitaryadrenal (HPA) system (Fig. 31-13). The hypothalamus releases corticotropin-releasing hormone, which is important in controlling the release of ACTH from the pituitary.¹ Cortisol levels increase as ACTH levels rise and decrease as ACTH levels fall. There is considerable diurnal variation in ACTH levels, which reach their peak in the early morning (around 6 to 8 AM) and decline as the day progresses (Fig. 31-14). This appears to be attributable to rhythmic activity in the CNS, which causes bursts of corticotropin-releasing hormone (CRH) secretion and, in turn, ACTH secretion. This diurnal pattern is reversed in people who work during the night and sleep during the day. The rhythm also may be changed by physical and psychological stresses, endogenous depression, manic-depressive psychosis, and liver disease or other conditions that affect cortisol metabolism. One of the earliest signs of Cushing's syndrome, a disorder of cortisol excess, is the loss of diurnal variation in CRH and ACTH secretion.

The glucocorticoids perform a necessary function in response to stress and are essential for survival. When produced as part of the stress response, these hormones aid in regulating the metabolic functions of the body and in controlling the inflammatory response. The actions of cortisol are summarized in Table 31-2.

Metabolic Effects. Cortisol stimulates glucose production by the liver, promotes protein breakdown, and causes mobilization of fatty acids. As body proteins are broken down, amino acids are mobilized and transported to the liver, where they are used in the production of glucose (*i.e.*, gluconeogenesis). Mobilization of fatty acids converts cell metabolism from the use of glucose for energy to the use of fatty acids. As glucose production by the liver increases and peripheral glucose use decreases, a moderate resistance to insulin develops. In persons with diabetes and those who are diabetes prone, this has the effect of raising the blood glucose level.

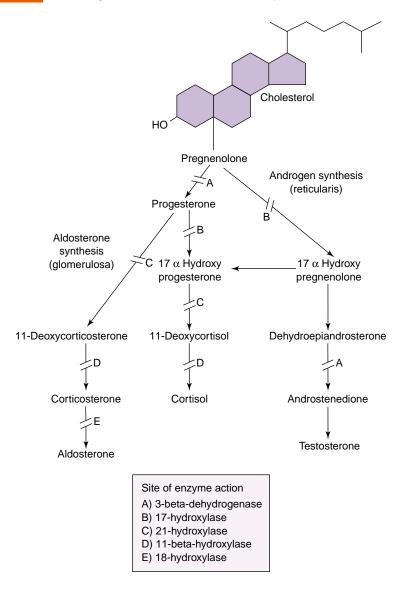
Psychological Effects. The glucocorticoid hormones appear to be involved directly or indirectly in emotional behavior. Receptors for these hormones have been identified in brain tissue, which suggests that they play a role in the regulation of behavior. Persons treated with adrenal cortical hormones have been known to display behavior ranging from mildly aberrant to psychotic.

Immunologic and Inflammatory Effects. Cortisol influences multiple aspects of immunologic function and inflammatory responsiveness. It blocks inflammation at an early stage by decreasing capillary permeability and stabilizing the lysosomal membranes so that inflammatory mediators are not released. Cortisol suppresses the immune response by reducing humoral and cell-mediated immunity. With this lessened inflammatory response comes a reduction in fever. During the healing phase, cortisol suppresses fibroblast activity and thereby lessens scar formation. Cortisol also inhibits prostaglandin synthesis, which may account in large part for its antiinflammatory actions. Large quantities of cortisol are required for an effective anti-inflammatory action. This is achieved by the administration of pharmacologic, rather than physiologic, doses of synthetic cortisol.

Suppression of Adrenal Function

A highly significant aspect of long-term therapy with pharmacologic preparations of the adrenal cortical hormones is adrenal insufficiency on withdrawal of the drugs. The deficiency

554 Unit Eight: Alterations in the Endocrine System



results from suppression of the HPA system. Chronic suppression causes atrophy of the adrenal gland, and the abrupt withdrawal of drugs can cause acute adrenal insufficiency. Recovery to a state of normal adrenal function may be prolonged, requiring 12 months or more.

Tests of Adrenal Function

Several diagnostic tests can be used to evaluate adrenal cortical function and the HPA system. Blood levels of cortisol, aldosterone, and ACTH can be measured using immunoassay methods. A 24-hour urine specimen measures the excretion of 17-ketosteroids, 17-ketogenic steroids, and 17-hydroxycorticosteroids. These metabolic end-products of the adrenal hormones and the male androgens provide information about alterations in the biosynthesis of the adrenal cortical hormones. The 24-hour urinary free cortisol is an excellent screening test for Cushing's syndrome.

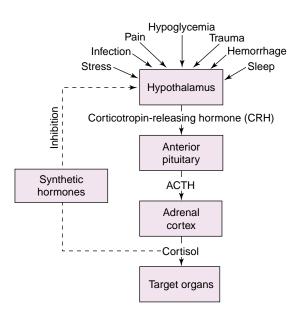
Suppression and stimulation tests afford a means of assessing the state of the HPA feedback system. For example, a test dose of ACTH can be given to assess the response of the adrenal

■ FIGURE 31-12 ■ Predominant biosynthetic pathways of the adrenal cortex. Critical enzymes in the biosynthetic process include 11-beta-hydroxylase and 21-hydroxylase. A deficiency in one of these enzymes blocks the synthesis of hormones dependent on that enzyme and routes the precursors into alternative pathways.

cortex to stimulation. Similarly, administration of dexamethasone, a synthetic glucocorticoid drug, provides a means of measuring negative feedback suppression of ACTH. Adrenal tumors and ectopic ACTH-producing tumors usually are unresponsive to ACTH suppression by dexamethasone. CRH tests can be used to diagnose a pituitary ACTH-secreting tumor (*i.e.*, Cushing's disease), especially when combined with inferior petrosal venous sampling (this allows the blood drainage of the pituitary to be sampled directly). Metyrapone (Metopirone) blocks the final step in cortisol synthesis, resulting in the production of 11-dehydroxycortisol, which does not inhibit ACTH. This test measures the ability of the pituitary to release ACTH. The gold standard test for assessing the HPA axis is the insulin hypoglycemic stress test.

Congenital Adrenal Hyperplasia

Congenital adrenal hyperplasia (CAH), or the adrenogenital syndrome, describes a congenital disorder caused by an autosomal recessive trait in which a deficiency exists in any of the



■ FIGURE 31-13 ■ The hypothalamic-pituitary-adrenal (HPA) feedback system that regulates glucocorticoid (cortisol) levels. Cortisol release is regulated by ACTH. Stress exerts its effects on cortisol release through the HPA system and the corticotropin-releasing hormone (CRH), which controls the release of ACTH from the anterior pituitary gland. Increased cortisol levels incite a negative feedback inhibition of ACTH release. Pharmacologic doses of synthetic steroids inhibit ACTH release by way of the hypothalamic CRH.

enzymes necessary for the synthesis of cortisol.^{26,27} A common characteristic of all types of CAH is a defect in the synthesis of cortisol that results in increased levels of ACTH and adrenal hyperplasia. The increased levels of ACTH overstimulate the pathways for production of adrenal androgens. Mineralocorticoids may be produced in excessive or insufficient amounts, depending on the precise enzyme deficiency. Infants of both genders are affected. The condition is seldom diagnosed in males at birth unless they have enlarged genitalia or lose salt

and manifest adrenal crisis. In female infants, an increase in androgens is responsible for creating the virilization syndrome of ambiguous genitalia, with an enlarged clitoris, fused labia, and urogenital sinus (Fig. 31-15). In male and female children, other secondary sex characteristics are normal, and fertility is unaffected if appropriate therapy is instituted.

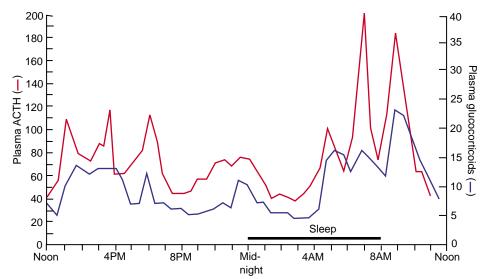
The two most common enzyme deficiencies are 21hydroxylase (accounting for >90% of cases) and 11- β hydroxylase deficiency. The clinical manifestations of both deficiencies are largely determined by the functional properties of the steroid intermediates and the completeness of the block in the cortisol pathway.

A spectrum of 21-hydroxylase deficiency states exists, ranging from simple virilizing CAH to a complete salt-losing enzyme deficiency. Simple virilizing CAH impairs the synthesis of cortisol, and steroid synthesis is shunted to androgen production. Persons with these deficiencies usually produce sufficient aldosterone or aldosterone intermediates to prevent signs and symptoms of mineralocorticoid deficiency. The salt-losing form is accompanied by deficient production of aldosterone and its intermediates. This results in fluid and electrolyte disorders after the fifth day of life, including hyponatremia, hyperkalemia, vomiting, dehydration, and shock.

The 11- β -hydroxylase deficiency is rare and manifests a spectrum of severity. Affected persons have excessive androgen production and impaired conversion of 11-deoxycorticosterone to corticosterone. The overproduction of 11-deoxycorticosterone, which has mineralocorticoid activity, is responsible for the hypertension that accompanies this deficiency. Diagnosis of adrenogenital syndrome depends on the precise biochemical evaluation of metabolites in the cortisol pathway and on clinical signs and symptoms.

Medical treatment of adrenogenital syndrome includes oral or parenteral cortisol replacement. Fludrocortisone acetate, a mineralocorticoid, also may be given to children who are salt losers. Depending on the degree of virilization, reconstructive surgery during the first 2 years of life is indicated to reduce the size of the clitoris, separate the labia, and exteriorize the vagina. Surgery has provided excellent results and does not impair sexual function.

■ FIGURE 31-14 ■ Pulsatile changes in the concentration of adrenocorticotropic hormone (ACTH) and glucocorticoids over a 24-hour period. The amplitude of the pulses of ACTH and glucocorticoids is lower in the evening hours and then increases greatly during the early morning hours. This is due to the diurnal oscillation of the hypothalamic-pituitary axis. (Modified from Krieger D.T. [1979]. Rhythms of CRF, ACTH and corticosteroids. In Krieger D.T. [Ed.], *Endocrine rhythms* [pp. 123–142]. New York: Raven)



556 Unit Eight: Alterations in the Endocrine System

TABLE 31-2 Actions of Cortisol		
Major Influence	Effect on Body	
Glucose metabolism	Stimulates gluconeogenesis Decreases glucose use by the tissues	
Protein metabolism	Increases breakdown of proteins Increases plasma protein levels	
Fat metabolism	Increases mobilization of fatty acids Increases use of fatty acids	
Anti-inflammatory action (pharmacologic levels)	Stabilizes lysosomal membranes of the inflammatory cells, preventing the release of inflammatory mediators Decreases capillary permeability to prevent inflammatory edema Depresses phagocytosis by white blood cells to reduce the release of inflammatory mediators Suppresses the immune response Causes atrophy of lymphoid tissue Decreases eosinophils Decreases antibody formation Decreases the development of cell-mediated immunity Reduces fever Inhibits fibroblast activity	
Psychic effect Permissive effect	May contribute to emotional instability Facilitates the response of the tissues to humoral and neural influences, such as that of the catecholamines, during trauma and extreme stress	

Adrenal Cortical Insufficiency

There are two forms of adrenal insufficiency: primary and secondary. Primary adrenal insufficiency, or Addison's disease, is caused by destruction of the adrenal gland. Secondary adrenal insufficiency results from a disorder of the HPA system.

Primary Adrenal Cortical Insufficiency

In 1855, Thomas Addison, an English physician, provided the first detailed clinical description of primary adrenal insufficiency, now called *Addison's disease*. The use of this term is reserved for primary adrenal insufficiency in which adrenal cortical hormones are deficient and ACTH levels are elevated because of lack of feedback inhibition.²⁸⁻³⁰



■ FIGURE 31-15 ■ A female infant with congenital adrenal hyperplasia demonstrating virilization of the genitalia with hypertrophy of the clitoris and partial fusion of labioscrotal folds. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 1186]. Philadelphia: Lippincott Williams & Wilkins)

Addison's disease is a relatively rare disorder in which all the layers of the adrenal cortex are destroyed. Autoimmune destruction is the most common cause of Addison's disease in the United States. Before 1950, tuberculosis was the major cause of Addison's disease in the United States, and it continues to be a major cause of the disease in countries where it is more prevalent. Rare causes include metastatic carcinoma, fungal infection (particularly histoplasmosis), cytomegalovirus infection, amyloid disease, and hemochromatosis. Bilateral adrenal hemorrhage may occur in persons taking anticoagulants, during open heart surgery, and during birth or major trauma. Adrenal insufficiency can be caused by acquired immunodeficiency syndrome (AIDS), in which the adrenal gland is destroyed by a variety of opportunistic infectious agents.

Addison's disease, like type 1 diabetes mellitus, is a chronic metabolic disorder that requires lifetime hormone replacement therapy. The adrenal cortex has a large reserve capacity, and the manifestations of adrenal insufficiency usually do not become apparent until approximately 90% of the gland has been destroyed. These manifestations are related primarily to mineralocorticoid deficiency, glucocorticoid deficiency, and hyperpigmentation resulting from elevated ACTH levels. Although lack of the adrenal androgens (*i.e.*, DHEAS) exerts few effects in men because the testes produce these hormones, women have sparse axillary and public hair.³¹

Mineralocorticoid deficiency causes increased urinary losses of sodium, chloride, and water, along with decreased excretion of potassium. The result is hyponatremia, loss of extracellular fluid, decreased cardiac output, and hyperkalemia. There may be an abnormal appetite for salt. Orthostatic hypotension is common. Dehydration, weakness, and fatigue are common early symptoms. If loss of sodium and water is extreme, cardiovascular collapse and shock ensue. Because of a lack of glucocorticoids, the person with Addison's disease has poor tolerance to stress. This deficiency causes hypoglycemia, lethargy, weakness, fever, and gastrointestinal symptoms, such as anorexia, nausea, vomiting, and weight loss.

Hyperpigmentation results from elevated levels of ACTH. The skin looks bronzed or suntanned in exposed and unexposed areas, and the normal creases and pressure points tend to become especially dark. The gums and oral mucous membranes may become bluish-black. The amino acid sequence of ACTH is strikingly similar to that of melanocyte-stimulating hormone; hyperpigmentation occurs in more than 90% of persons with Addison's disease and is helpful in distinguishing the primary and secondary forms of adrenal insufficiency.

The daily regulation of the chronic phase of Addison's disease usually is accomplished with oral replacement therapy, with higher doses being given during periods of stress. The pharmacologic agent that is used should have both glucocorticoid and mineralocorticoid activity. Mineralocorticoids are needed only in primary adrenal insufficiency. Hydrocortisone usually is the drug of choice. In mild cases, hydrocortisone alone may be adequate. Fludrocortisone (a mineralocorticoid) is used for persons who do not obtain a sufficient salt-retaining effect from hydrocortisone. DHEAS replacement also may be helpful in the female patient.

Because persons with the disorder are likely to have episodes of hyponatremia and hypoglycemia, they need to have a regular schedule for meals and exercise. Persons with Addison's disease also have limited ability to respond to infections, trauma, and other stresses. Such situations require immediate medical attention and treatment. All persons with Addison's disease should be advised to wear a medical alert bracelet or medal.

Secondary Adrenal Cortical Insufficiency

Secondary adrenal insufficiency can occur as the result of hypopituitarism or because the pituitary gland has been surgically removed. Tertiary adrenal insufficiency results from a hypothalamic defect. However, a far more common cause than either of these is the rapid withdrawal of glucocorticoids that have been administered therapeutically. These drugs suppress the HPA system, with resulting adrenal cortical atrophy and loss of cortisol production. This suppression continues long after drug therapy has been discontinued and can be critical during periods of stress or when surgery is performed.

Acute Adrenal Crisis

Acute adrenal crisis is a life-threatening situation. If Addison's disease is the underlying problem, exposure to even a minor illness or stress can precipitate nausea, vomiting, muscular weakness, hypotension, dehydration, and vascular collapse. The onset of adrenal crisis may be sudden, or it may progress during a period of several days. The symptoms may occur suddenly in children with salt-losing forms of the adrenogenital syndrome. Massive bilateral adrenal hemorrhage causes an acute fulminating form of adrenal insufficiency. Hemorrhage can be caused by meningococcal septicemia (*i.e.*, Waterhouse-Friderichsen syndrome), adrenal trauma, anticoagulant therapy, adrenal vein thrombosis, or adrenal metastases.

Acute adrenal insufficiency is treated with intravenous fluids and corticosteroid replacement therapy. Corticosteroid replacement is accomplished through the intravenous administration of either dexamethasone or hydrocortisone.

Glucocorticoid Hormone Excess (Cushing's Syndrome)

The term *Cushing's syndrome* refers to the manifestations of hypercortisolism from any cause. Three important forms of Cushing's syndrome result from excess glucocorticoid production by the body. One is a pituitary form, which results from excessive production of ACTH by a tumor of the pituitary gland. This form of the disease was the one originally described by Cushing; therefore, it is called *Cushing's disease*.^{32,33} The second form is the adrenal form, caused by a benign or malignant adrenal tumor. The third form is ectopic Cushing's, caused by a nonpituitary ACTH-secreting tumor.³⁴ Certain extrapituitary malignant tumors such as small cell carcinoma of the lung may secrete ACTH or rarely CRH and produce Cushing's syndrome. Cushing's syndrome also can result from long-term therapy with one of the potent pharmacologic preparations of glucocorticoids; this form is called *iatrogenic Cushing's syndrome*.

The major manifestations of Cushing's syndrome represent an exaggeration of the many actions of cortisol (see Table 31-2). Altered fat metabolism causes a peculiar deposition of fat characterized by a protruding abdomen; subclavicular fat pads or "buffalo hump" on the back; and a round, plethoric "moon face" (Fig. 31-16). There is muscle weakness, and the extremities are thin because of protein breakdown and muscle wasting. In advanced cases, the skin over the forearms and legs becomes thin, having the appearance of parchment. Purple striae, or stretch marks, from stretching of the catabolically weakened skin and subcutaneous tissues are distributed over the breast, thighs, and abdomen. Osteoporosis may develop because of destruction of bone proteins and alterations in calcium metabolism, resulting in back pain, compression fractures of the vertebrae, and rib fractures. As calcium is mobilized from bone, renal calculi may develop.

Derangements in glucose metabolism are found in approximately 75% of patients, with clinically overt diabetes mellitus occurring in approximately 20%. The glucocorticoids possess mineralocorticoid properties; this causes hypokalemia as a result of excessive potassium excretion and hypertension resulting



■ FIGURE 31-16 ■ Cushing's syndrome. A woman who suffered from a pituitary adenoma that produced ACTH exhibits a moon face, buffalo hump, increased facial hair, and thinning of the scalp hair. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 1193]. Philadelphia: Lippincott Williams & Wilkins)

from sodium retention. Inflammatory and immune responses are inhibited, resulting in increased susceptibility to infection. Cortisol increases gastric acid secretion, which may provoke gastric ulceration and bleeding. An accompanying increase in androgen levels causes hirsutism, mild acne, and menstrual irregularities in women. Excess levels of the glucocorticoids may give rise to extreme emotional lability, ranging from mild euphoria and absence of normal fatigue to grossly psychotic behavior.

Diagnosis of Cushing's syndrome depends on the finding of cortisol hypersecretion. The determination of 24-hour excretion of cortisol in urine provides a reliable and practical index of cortisol secretions. One of the prominent features of Cushing's syndrome is loss of the diurnal pattern of cortisol secretion. Cortisol determinations often are made on three blood samples: one taken in the morning, one in late afternoon or early evening, and a third drawn the following morning after a midnight dose of dexamethasone. Measurement of the plasma levels of ACTH, measurement of 24-hour urinary 17-ketosteroids, 17-ketogenic steroids, and 17-hydroxycorticosteroids, and suppression or stimulation tests of the HPA system often are made. MRI or CT scans afford a means for locating adrenal or pituitary tumors.

Untreated, Cushing's syndrome produces serious morbidity and even death. The choice of surgery, irradiation, or pharmacologic treatment is determined largely by the cause of the hypercortisolism. The goal of treatment for Cushing's syndrome is to remove or correct the source of hypercortisolism without causing any permanent pituitary or adrenal damage. Transsphenoidal removal of a pituitary adenoma or a hemihypophysectomy is the preferred method of treatment for Cushing's disease. This allows removal of only the tumor, rather than the entire pituitary gland. After successful removal, the person must receive cortisol replacement therapy for 6 to 12 months or until adrenal function returns. Patients also may receive pituitary radiation therapy, but the full effects of treatment may not be realized for 3 to 12 months. Unilateral or bilateral adrenalectomy may be done in the case of adrenal adenoma. When possible, ectopic ACTH-producing tumors are removed. Pharmacologic agents that block steroid synthesis (i.e., etomidate, mitotane, ketoconazole, metyrapone, and aminoglutethimide) may be used to treat persons with ectopic tumors that cannot be resected. Many of these patients also require Pneumocystis carinii pneumonia prophylaxis because of the profound immunosuppression caused by the excessive glucocorticoid.

In summary, the adrenal cortex produces three types of hormones: mineralocorticoids, glucocorticoids, and adrenal sex hormones. The mineralocorticoids along with the reninangiotensin mechanism contribute to the control of body levels of sodium and potassium. The glucocorticoids have antiinflammatory actions and aid in regulating glucose, protein, and fat metabolism during periods of stress. These hormones are under the control of the HPA system. The adrenal sex hormones exert little effect on daily control of body function, but they probably contribute to the development of body hair in women. The adrenogenital syndrome describes a genetic defect in the cortisol pathway resulting from a deficiency of one of the enzymes needed for its synthesis. Depending on the enzyme involved, the disorder causes virilization of female infants and, in some instances, fluid and electrolyte disturbances because of impaired mineralocorticoid synthesis.

Chronic adrenal insufficiency (Addison's disease) can be caused by destruction of the adrenal gland or by dysfunction of the HPA system. Adrenal insufficiency requires replacement therapy with cortical hormones. Acute adrenal insufficiency is a life-threatening situation. Cushing's syndrome refers to the manifestations of excessive cortisol levels. This syndrome may be a result of pharmacologic doses of cortisol, a pituitary or adrenal tumor, or an ectopic tumor that produces ACTH. The clinical manifestations of Cushing's syndrome reflect the very high level of cortisol that is present.

REVIEW QUESTIONS

Use thyroid hormone as an example for describing the etiology of primary, secondary, and tertiary hypothyroidism.

Differentiate genetic short stature from constitutional short stature.

State the mechanisms of short stature in hypothyroidism, poorly controlled diabetes mellitus, treatment with adrenal glucocorticosteroid hormones, malnutrition, and psychosocial dwarfism.

List three causes of tall stature.

Relate the functions of growth hormone to the manifestations of acromegaly.

Explain why children with isosexual precocious puberty are tall-statured children but short-statured adults.

Diagram the hypothalamic-pituitary-thyroid feedback system.

Relate the functions of thyroid hormone to hypothyroidism and hyperthyroidism.

Explain how a defect in a single step of corticosteroid hormone synthesis produces the manifestations of the adrenogenital syndrome.

Relate the functions of the adrenal cortical hormones to Addison's disease (*i.e.*, adrenal insufficiency) and Cushing's syndrome (*i.e.*, cortisol excess).

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- Cotran R.S., Kumar V., Collins T. (1999). Robbins pathologic basis of disease (6th ed., pp. 1127–1169). Philadelphia: W.B. Saunders.
- Guyton A.C., Hall J.E. (2000). *Textbook of medical physiology* (10th ed., pp. 836–856, 858–864, 869–883, 906–908). Philadelphia: W.B. Saunders.
- Ganong W.R. (1999). Review of medical physiology (19th ed., pp. 378– 389). Stamford, CT: Appleton & Lange.

- Aron D.C., Findling J.W., Tyrell B.J. (2001). Hypothalamus and pituitary. In Greenspan F.S., Gardner D.G. (Eds.), *Basic and clinical endocrinology* (6th ed., pp. 100–158). New York: Lange Medical Books/ McGraw-Hill.
- Styne D. (2001). Growth. In Greenspan F.S., Gardner D.G. (Eds.), Basic and clinical endocrinology (6th ed., pp. 100–158). New York: Lange Medical Books/McGraw-Hill.
- Parks J.S. (2000). Hypopituitarism. In Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.), *Nelson textbook of pediatrics* (16th ed., pp. 1675– 1680). Philadelphia: W.B. Saunders.
- Laron Z. (1995). Laron syndrome (primary GH resistance) from patient to laboratory to patient. *Journal of Clinical Endocrinology and Metabolism* 80, 1526–1531.
- 8. Vance M.L., Mauras N. (1999). Growth hormone therapy in adults and children. *New England Journal of Medicine* 341, 1206–1216.
- 9. Khan A.S., Sane D.C., Wanneburg T., et al. (2002). Growth hormone, insulin-like growth factor-1 and the aging cardiovascular system. *Cardiovascular Research* 54, 25–35.
- Cohen P. (2000). Hyperpituitarism, tall stature, and overgrowth syndromes. In Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.), *Nelson textbook of pediatrics* (16th ed., pp. 1685–1687). Philadelphia: W.B. Saunders.
- Ezzat S. (1997). Acromegaly. Endocrinology and Metabolism Clinics of North America 26, 703–723.
- Utiger R.D. (2000). Treatment of acromegaly. New England Journal of Medicine 342, 1210–1211.
- Garibaldi L. (2000). Disorders of pubertal development. In Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.), *Nelson textbook of pediatrics* (16th ed., pp. 1688–1695). Philadelphia: W.B. Saunders.
- Lebrethon M.C., Bourguignon J.P. (2001). Central and peripheral isosexual precocious puberty. *Current Opinion in Endocrinology and Dia*betes 8, 17–22.
- 15. Dayan C.M. (2001). Interpretation of thyroid function tests. *The Lancet* 357, 619–624.
- LaFranchi S. (2000). Thyroid development and physiology. In Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.), *Nelson textbook of pediatrics* (16th ed., pp. 1696–1704). Philadelphia: W.B. Saunders.
- 17. Cooper D.S. (2001). Subclinical hypothyroidism. *New England Journal of Medicine* 345, 260–264.
- Wall C.R. (2000). Myxedematous coma. American Family Physician 62, 2485–2490.
- 19. McKenna T.J. (2001). Graves' disease. The Lancet 357, 1793-1796.

- Weetman A.P. (2000). Graves' disease. New England Journal of Medicine 343, 1236–1248.
- Bahn R.S., Heufelder A.E. (1993). Pathogenesis of Graves' ophthalmopathy. New England Journal of Medicine 329, 1468–1474.
- Bartalena L., Pinchera A., Marocci C. (2000). Management of Graves' ophthalmopathy: Reality and perspectives. *Endocrine Reviews* 21(2), 168–199.
- 23. Korbin S.M., Goldfarb S. (1990). Magnesium deficiency. Seminars in Nephrology 10, 525–535.
- Marx S.J. (2000). Hyperparathyroid and hypoparathyroid disorders. New England Journal of Medicine 343, 1863–1875.
- Slaptopolsky E., Brown A., Dusso A. (1999). Pathogenesis of secondary hyperparathyroidism. *Kidney International* 56 (Suppl. 73), S14–S19.
- Levine L.S., DiGeorge A.M. (2000). Adrenal disorders and genital abnormalities. In Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.), *Nelson textbook of pediatrics* (16th ed., pp. 1729–1737). Philadelphia: W.B. Saunders.
- Lack E.A., Farber J.L., Rubin E. (1999). The endocrine system. In Rubin E., Farber J.L. (Eds.), *Pathology* (3rd ed., pp. 1184–1186). Philadelphia: Lippincott Williams & Wilkins.
- Oelker W. (1996). Adrenal insufficiency. New England Journal of Medicine 335, 1206–1212.
- 29. Ten S., New M., Maclaren N. (2001). Addison's disease 2001. Journal of Endocrinology and Metabolism 86, 2909–2922.
- Aron D.C., Findling J.W., Tyrrell J.B. (2001). Glucocorticoids and adrenal androgens. In Greenspan F.S., Gardner D.G. (Eds.), *Basic and clinical endocrinology* (6th ed. pp. 334–376). New York: Lange Medical Books/McGraw-Hill.
- Arlt W., Callies F., Van Vlymen J.C., et al. (1999). Dehydroepiandrosterone replacement in women with adrenal insufficiency. *New England Journal of Medicine* 341, 1013–1020.
- 32. Boscaro M. (2001). Cushing's syndrome. The Lancet 357, 783-791.
- Newell-Price J., Trainer P., Besser M., et al. (1998). The diagnosis and differential diagnosis of Cushing's syndrome and pseudo-Cushing's states. *Endocrine Reviews* 19(5), 647–672.
- Orth D.N. (1995). Cushing's syndrome. New England Journal of Medicine 332, 791–803.
- 35. Utiger R.D. (1997). Treatment and retreatment of Cushing's disease. *New England Journal of Medicine* 336, 215–217.
- White P.C., Speiser P.W. (2000). Congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Endocrine Reviews* 21, 245–291.

Nine

Alterations in the Male and Female Reproductive Systems

CHAPTER 33

Alterations in the Male Reproductive System

Physiologic Basis of Male Reproductive Function **Spermatogenesis** Hormonal Control of Male Reproductive Function Action of the Hypothalamic and Anterior **Pituitary Hormones Neural Control of Sexual Function Erectile Dysfunction** Disorders of the Penis, the Scrotum and Testes, and the Prostate **Disorders of the Penis** Inflammation and Infection Peyronie's Disease **Priapism** Cancer of the Penis **Disorders of the Scrotum and Testes** Disorders of the Testicular Tunica **Testicular Torsion** Inflammation and Infection Cancer of the Scrotum and Testes **Disorders of the Prostate Prostatitis Benign Prostatic Hyperplasia** Cancer of the Prostate

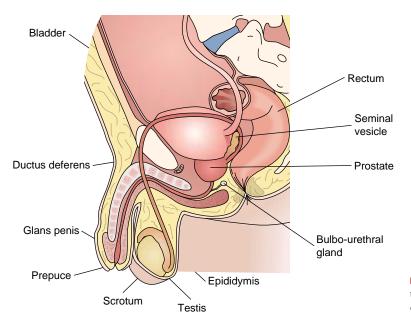
Disorders in Childhood and Aging Changes Disorders of Childhood Hypospadias and Epispadias Phimosis and Paraphimosis Cryptorchidism Aging Changes

he male genitourinary system is subject to structural defects, inflammation, and neoplasms, all of which can affect urine elimination, sexual function, and fertility. This chapter focuses on spermatogenesis and hormonal control of male reproductive function; neural control of sexual function and erectile dysfunction; disorders of the penis, scrotum, testes, and prostate; disorders of the male reproductive system in children; and changes in function as a result of the aging process.

PHYSIOLOGIC BASIS OF MALE REPRODUCTIVE FUNCTION

The male genitourinary system is composed of the paired gonads, or testes, genital ducts, accessory organs, and penis (Fig. 33-1). The dual function of the testes is to produce male sex androgens (*i.e.*, male sex hormones), mainly testosterone, and spermatozoa (*i.e.*, male germ cells). The internal accessory organs produce the fluid constituents of semen, and the ductile system aids in the storage and transport of spermatozoa. The penis functions in urine elimination and sexual function.

582 Unit Nine: Alterations in the Male and Female Reproductive Systems



■ FIGURE 33-1 ■ The structures of the male reproductive system, including the testes, the scrotum, and the excretory ducts.

Spermatogenesis

Spermatogenesis refers to the generation of spermatozoa or sperm. It begins at an average age of 13 years and continues throughout the reproductive years of a man's life. Spermatogenesis occurs in the seminiferous tubules of the testes (Fig. 33-2). Internally, the testes are composed of several hundred compartments or lobules. Each lobule contains one or more coiled seminiferous tubules. The outer layer of the seminiferous tubules is made up of connective tissue and smooth muscle; the inner lining is composed of Sertoli's cells, which are embedded with sperm in

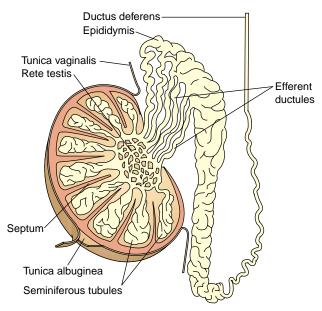
KEY CONCEPTS

MALE REPRODUCTIVE SYSTEM

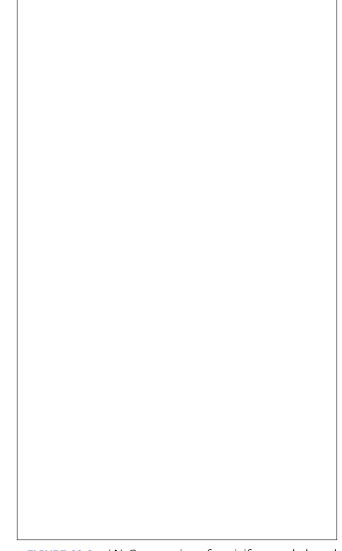
- The male genitourinary system functions in both urine elimination and reproduction.
- The testes function in both production of male germ cells (spermatogenesis) and secretion of the male sex hormone, testosterone.
- The ductile system (epididymides, vas deferens, and ejaculatory ducts) transports and stores sperm and assists in their maturation, and the accessory glands (seminal vesicles, prostate gland, and bulbourethral glands) prepare the sperm for ejaculation.
- Sperm production requires temperatures that are 2° to 3°C below body temperature. The position of the testes in the scrotum and the unique blood flow-cooling mechanisms provide this environment.
- The urethra, which is enclosed in the penis, is the terminal portion of the male genitourinary system.
 Because it conveys both urine and semen, it serves both urinary and reproductive functions.

various stages of development (Fig. 33-3). Sertoli's cells secrete a special fluid that contains nutrients to bathe and nourish the immature germ cells; they provide digestive enzymes that play a role in spermiation (*i.e.*, converting the spermatocytes to sperm); and they are thought to play a role in shaping the head and tail of the sperm. Sertoli's cells also secrete several hormones, including the principal feminizing sex hormone, estradiol, which seems to be required in the male for spermatogenesis, and inhibin, which controls the function of Sertoli's cells through feedback inhibition of follicle-stimulating hormone (FSH) from the anterior pituitary gland.

After the spermatozoa are formed in the seminiferous tubules, they travel through the efferent tubules to the epididymis, which is the final site for sperm maturation. Because the spermatozoa are not motile at this stage of development,



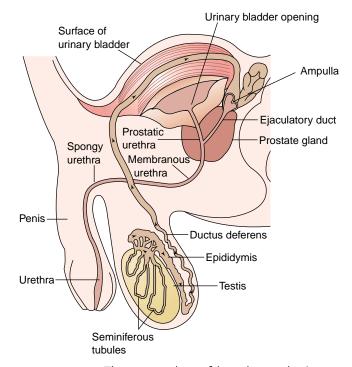
■ FIGURE 33-2 ■ The parts of the testes and epididymis.



■ FIGURE 33-3 ■ (A) Cross-section of seminiferous tubule and (B) stages of development of spermatozoa.

peristaltic movements of the ductal walls of the epididymis aid in their movement.¹ The spermatozoa continue their migration through the ductus deferens, also called the *vas deferens*. The ampulla of the vas deferens serves as a storage reservoir for sperm. Sperm are stored in the ampulla until they are released through the penis during ejaculation (Fig. 33-4). Spermatozoa can be stored in the genital ducts for as long as 42 days and still maintain their fertility. Surgical disconnection of the vas deferens in the scrotal area (*i.e.*, vasectomy) serves as an effective method of male contraception. Because sperm are stored in the ampulla, men can remain fertile for 4 to 5 weeks after performance of a vasectomy.

The seminal vesicles, the prostate gland, and the bulbourethral glands form the accessory reproductive structures of the male genitourinary system. The spermatozoa plus the secretions from the genital ducts and accessory organs make up



■ FIGURE 33-4 ■ The excretory ducts of the male reproductive system and the path that sperm follows as it leaves the testis and travels to the urethra.

the semen (from the Latin word meaning seed). The seminal vesicles consist of two highly tortuous tubes that secrete fluid for the semen. Each of the paired seminal vesicles is lined with secretory epithelium containing an abundance of fructose, prostaglandins, and several other proteins. The fructose secreted by the seminal vesicles provides the energy for sperm motility. The prostaglandins are thought to assist in fertilization by making the cervical mucus more receptive to sperm and by causing reverse peristaltic contractions in the uterus and fallopian tubes to move the sperm toward the ovaries. Each seminal vesicle joins its corresponding vas deferens to form the ejaculatory duct, which enters the posterior part of the prostate and continues through until it ends in the prostatic portion of the urethra. During the emission phase of coitus, each vesicle empties fluid into the ejaculatory duct, adding bulk to the semen. Approximately 70% of the ejaculate originates in the seminal vesicles.

Hormonal Control of Male Reproductive Function

The male sex hormones are called *androgens*. The testes secrete several male sex hormones, including *testosterone*, *dihydrotes*-*tosterone*, and *androstenedione*.² Testosterone, which is the most abundant of these hormones, is considered the main testicular hormone. The adrenal cortex also produces androgens, although in much smaller quantities (<5% of the total male androgens) than those produced in the testes. The testes also secrete small quantities of estradiol and estrone.

Testosterone is produced and secreted by the interstitial Leydig's cells in the testes. It is metabolized in the liver and excreted by the kidneys. In the bloodstream, testosterone exists in a free (unbound) or a bound form. The bound form is attached to plasma proteins, including albumin and the sex hormonebinding protein produced by the liver. Only approximately 2% of circulating testosterone is unbound and therefore able to enter the cell and exert its metabolic effects. Much of the testosterone that becomes fixed to the tissues is converted to dihydrotestosterone, especially in certain target tissues such as the prostate gland. Some of the actions of testosterone depend on this conversion, whereas others do not. Testosterone also can be aromatized or converted to estradiol in the peripheral tissues.

Testosterone exerts a variety of biologic effects in the male (Chart 33-1). In the male embryo, testosterone is essential for the appropriate differentiation of the internal and external genitalia, and it is necessary for descent of the testes in the fetus. Testosterone is essential to the development of primary and secondary male sex characteristics during puberty and for the maintenance of these characteristics during adult life. It causes growth of pubic, chest, and facial hair; it produces changes in the larynx that result in the male bass voice; and it increases the thickness of the skin and the activity of the sebaceous glands, predisposing to acne.

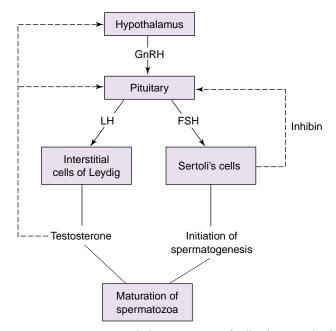
All or almost all of the actions of testosterone and other androgens result from increased protein synthesis in target tissues. Androgens function as anabolic agents in males and females to promote metabolism and musculoskeletal growth. Testosterone and the androgens have a great effect on the development of increasing musculature during puberty, with boys averaging approximately 50% more of an increase in muscle mass than do girls.

Action of the Hypothalamic and Anterior Pituitary Hormones

The hypothalamus and the anterior pituitary gland play an essential role in promoting spermatogenic activity in the testes and maintaining the endocrine function of the testes by means of the gonadotropic hormones. The synthesis and release of the gonadotropic hormones from the pituitary gland are regulated by gonadotropin-releasing hormone (GnRH), which is synthesized by the hypothalamus and secreted into the hypothalamohypophysial portal circulation (Fig. 33-5).

CHART 33-1 Main Actions of Testosterone

- Induces differentiation of the male genital tract during fetal development
- Induces development of primary and secondary sex
 - characteristics
 - Gonadal function
 - External genitalia and accessory organs
- Male voice timbre
- Male skin characteristics
- Male hair distribution
- Anabolic effects
 - Promotes protein metabolism
 - Promotes musculoskeletal growth
- Influences subcutaneous fat distribution
- Promotes spermatogenesis (in FSH-primed tubules) and maturation of sperm



■ **FIGURE 33-5** ■ Hypothalamic-pituitary feedback control of spermatogenesis and testosterone levels in the male. The *dashed line* represents negative feedback.

Two gonadotropic hormones are secreted by the pituitary gland: FSH and luteinizing hormone (LH). In the male, LH also is called interstitial cell-stimulating hormone. The production of testosterone by the interstitial cells of Leydig is regulated by LH. FSH binds selectively to Sertoli's cells surrounding the seminiferous tubules, where it functions in the initiation of spermatogenesis. Under the influence of FSH, Sertoli's cells produce and rogen-binding protein, plasminogen activator, and inhibin. Androgen-binding protein binds testosterone and serves as a carrier of testosterone in Sertoli's cells and as a storage site for testosterone. Although FSH is necessary for the initiation of spermatogenesis, full maturation of the spermatozoa requires testosterone. Androgen-binding protein also serves as a carrier of testosterone from the testes to the epididymis. Plasminogen activator, which converts plasminogen to plasmin, functions in the final detachment of mature spermatozoa from Sertoli's cells.

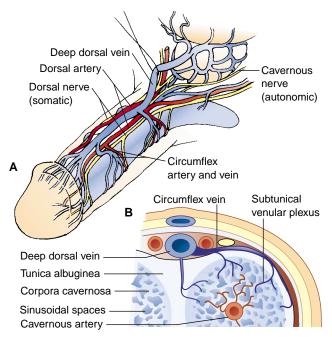
Circulating levels of the gonadotropic hormones are regulated in a negative feedback manner by testosterone. High levels of testosterone suppress LH secretion through a direct action on the pituitary and an inhibitory effect on the hypothalamus. FSH is thought to be inhibited by a substance called *inhibin*, produced by Sertoli's cells. Inhibin suppresses FSH release from the pituitary gland. The pituitary gonadotropic hormones and Sertoli's cells in the testes form a classic negative feedback loop in which FSH stimulates inhibin and inhibin suppresses FSH. Unlike the cyclic hormonal pattern in the female, in the male, FSH, LH, and testosterone secretion and spermatogenesis occur at relatively unchanging rates during adulthood.

Neural Control of Sexual Function

The penis is the male external genital organ through which the urethra passes. It functions both as a sexual organ and as an organ of urine elimination. Anatomically, the external penis consists of a shaft that ends in a tip called the *glans* (Fig. 33-1). The loose skin of the penis shaft folds to cover the glans, forming the prepuce, or foreskin. The glans of the penis contains many sensory nerves, making this the most sensitive portion of the penile shaft. The cylindrical body or shaft of the penis is composed of three masses of erectile tissue held together by fibrous strands and covered with a thin layer of skin (Fig. 33-6). The two lateral masses of tissue are called the *corpora cavernosa*. The third, ventral mass is called the *corpus spongiosum*. The corpora cavernosa and corpus spongiosum are cavernous sinuses that normally are relatively empty but become engorged with blood during penile erection.

The physiology of the male sex act involves a complex interaction between autonomic-mediated spinal cord reflexes, higher neural centers, and the vascular system. It involves erection, emission, ejaculation, and detumescence. Erection involves increased inflow of blood into the corpora cavernosa and penile rigidity. Ejaculation represents the expulsion of the sperm from the urethra. Detumescence, or penile relaxation, results from outflow of blood from the corpora cavernosa.

Erection is a neurovascular process involving the autonomic nervous system, neurotransmitters and endothelial relaxing factors, the vascular smooth muscle of the arteries and veins supplying the penile tissue, and the trabecular smooth muscle of the sinusoids of the corpora cavernosa (Fig. 33-6). The penis is innervated by both the autonomic and somatic nervous systems. In the pelvis, the sympathetic and parasympathetic components of the autonomic nervous system merge to form what are called the *cavernous nerves*.³ Erection is under the control of the parasympathetic nervous system, and ejaculation and detumescence (penile relaxation) are under the control of the sympathetic nervous system. Somatic innervation, which occurs through the pudendal nerve, is responsible for penile sensation and contraction and relaxation of the



■ FIGURE 33-6 ■ Anatomy and mechanism of penile erection. (A) Innervation and arterial and venous blood supply to penis. (B) Cross-section of the sinusoidal system of the corpora cavernosa.

KEY CONCEPTS

DISORDERS OF PENILE ERECTION

- Erection is a neurovascular process involving the autonomic nervous system, the somatic nervous system by way of the pudendal nerve, the vascular system, and the sinusoidal spaces of the corpora cavernosa.
- Parasympathetic innervation through the pelvic nerves initiates relaxation of the trabecular smooth muscle of the corpora cavernosa through the action of nitric oxide, the inflow of arterial blood, and cessation of venous outflow.
- Erectile failure can result from disorders in one or a combination of the neural, vascular, or chemical mediator aspects of the erectile process.

extracorporeal striated muscles (bulbocavernous and ischio-cavernous).³

Penile erection is the first effect of male sexual stimulation, whether psychological or physical. It involves increased inflow of blood into the corpora cavernosa due to relaxation of the trabecular smooth muscle that surrounds the sinusoidal spaces and compression of the veins controlling outflow of blood from the venous plexus. Erection is mediated by parasympathetic impulses that pass from the sacral segments of the spinal cord through the pelvic nerves to the penis. Parasympathetic stimulation results in release of nitric oxide, a nonadrenergic-noncholinergic neurotransmitter, which causes relaxation of trabecular smooth muscle of the corpora cavernosa. This relaxation permits inflow of blood into the sinuses of the cavernosa at pressures approaching those of the arterial system. Because the erectile tissues of the cavernosa are surrounded by a nonelastic fibrous covering, high pressure in the sinusoids causes ballooning of the erectile tissue to such an extent that the penis becomes hard and elongated. At the same time, contraction of the somatic-innervated ischiocavernous muscles forcefully compresses the blood-filled corpora cavernosa, producing a further increase in intercavernous pressures. During this phase of erection, inflow and outflow of blood cease.

Parasympathetic innervation must be intact and nitric oxide synthesis must be active for erection to occur. Nitric oxide activates guanyl cyclase, an enzyme that increases the concentration of cyclic guanosine monophosphate (cGMP), which in turn causes smooth muscle relaxation. Other smooth muscle relaxants (*e.g.*, prostaglandin E_1 analogs and α -adrenergic antagonists), if present in high enough concentrations, can independently cause sufficient cavernosal relaxation to result in erection.³ Many of the drugs that have been developed to treat erectile dysfunction act at the levels of these mediators.

Detumescence or penile relaxation is largely a sympathetic nervous system response. It can result from a cessation of neurotransmitter release, the breakdown of second messengers such as cGMP, or sympathetic discharge during ejaculation. Contraction of the trabecular smooth muscle opens the venous channels so that the trapped blood can be expelled and penile flaccidity return.

Erectile Dysfunction

"Erectile dysfunction is defined as the inability to achieve and maintain an erection sufficient to permit satisfactory sexual intercourse."⁴ It has been estimated that the disorder affects 20 to 30 million men in the United States.⁵ Erectile dysfunction is commonly classified as psychogenic, organic, or mixed psychogenic or organic.^{5,6} The latter is the most common.

Psychogenic causes of erectile dysfunction include performance anxiety, a strained relationship with a sexual partner, depression, and overt psychotic disorders such as schizophrenia. Depression is a common cause of erectile dysfunction.⁵

Organic causes span a wide range of pathologies. They include neurogenic, hormonal, vascular, drug-induced, and cavernous impairment etiologies. Neurogenic disorders such as Parkinson's disease, stroke, and cerebral trauma often contribute to erectile dysfunction by decreasing libido or preventing the initiation of erection. In spinal cord injury, the extent of neural impairment depends on the level, location, and extent of the lesion. Somatosensory involvement of the genitalia is essential to the reflex mechanisms involved in erection; this becomes important with aging and conditions such as diabetes that impair peripheral nerve function.

Hormonal causes of erectile dysfunction include a decrease in androgen levels. Androgen levels may be decreased because of aging. Hypoprolactinemia from any cause interferes with both reproduction and erectile function. This is because prolactin acts centrally to inhibit dopaminergic activity, which is a stimulus for release of the hypothalamic GnRH that controls the release of pituitary gonadotropic hormones.

Common risk factors for generalized penile arterial insufficiency include hypertension, hyperlipidemia, cigarette smoking, diabetes mellitus, and pelvic irradiation.⁵ In hypertension, erectile function is impaired not so much by the increased blood pressure as by the associated stenotic arterial lesions. Focal stenosis of the common penile artery most often occurs in men who sustained blunt pelvic or perineal trauma (*e.g.*, from bicycling accidents). Failure of the veins to close completely during an erection (veno-occlusive dysfunction) may occur in men with large venous channels that drain the corpora cavernosa. Other disorders that impair venous occlusion are degenerative changes involving the tunica albuginea, as in Peyronie's disease. Poor relaxation of the trabecular smooth muscle may accompany anxiety with excessive adrenergic tone.

Many drugs are reported to cause erectile dysfunction, including antidepressant, antipsychotic, and antihypertensive medications. Cigarette smoking can induce vasoconstriction and penile venous leakage because of its effects on cavernous smooth muscle.⁵ Alcohol in small amounts may increase libido and improve erection; however, in large amounts it can cause central sedation, decreased libido, and transient erectile dysfunction.

Aging is known to increase the risk of erectile dysfunction.⁷ Many of the pathologic processes that contribute to erectile dysfunction are more common in older men, including diabetes, hyperlipidemia, vascular disease, and the long-term effects of cigarette smoking. Age-related declines in testosterone also may play a role. Psychosocial problems such as depression, esteem issues, partner relationships, history of substance abuse, and anxiety and fear of performance failure also may contribute to erectile dysfunction in older men.

A diagnosis of erectile dysfunction requires careful history (medical, sexual, and psychosocial), physical examination, and laboratory tests aimed at determining what other tests are needed to rule out organic causes of the disorder.⁶ Because many medications, including prescribed, over-the-counter, and illicit drugs, can cause erectile dysfunction, a careful drug history is indicated.

Treatment methods include psychosexual counseling, androgen replacement therapy, oral and intracavernous drug therapy, vacuum constriction devices, and surgical treatment (prosthesis and vascular surgery).⁶ Among the commonly prescribed drugs are sildenafil, yohimbine, alprostadil, and phentolamine. Sildenafil (Viagra) is a selective inhibitor of phosphodiesterase type 5, the enzyme that inactivates cGMP. Yohimbine, an α_2 -adrenergic receptor antagonist, acts at the adrenergic receptors in brain centers associated with libido and penile erection. Both sildenafil and yohimbine are taken orally. Alprostadil, a prostaglandin E analog, acts by producing relaxation of cavernous smooth muscle. It is either injected directly into the cavernosa or placed in the urethra as a minisuppository. Phentolamine, an α_2 -adrenergic receptor antagonist, also is administered by intracavernous injection.

In summary, spermatogenesis occurs in the Sertoli's cells of the seminiferous tubules of the testes. After formation in seminiferous tubules, the spermatozoa travel through the efferent tubules to the epididymis, then to the ductus deferens, and on to the ampulla, where they are stored until released through the penis during ejaculation. The male accessory organs consist of the seminal vesicles, prostate gland, and bulbourethral glands. Secretions from the genital ducts and accessory organs combine with the spermatozoa to form the semen.

The function of the male reproductive system is under the negative feedback control of the hypothalamus and the anterior pituitary gonadotropic hormones FSH and LH. Spermatogenesis is initiated by FSH, and the production of testosterone is regulated by LH. Testosterone, the major male sex hormone, is produced by the interstitial Leydig's cells in the testes. In addition to its role in the differentiation of the internal and external genitalia in the male embryo, testosterone is essential for the development of secondary male characteristics during puberty, the maintenance of these characteristics during adult life, and spermatozoa maturation.

The male sex act involves erection, emission, ejaculation, and detumescence. The physiology of these functions involves a complex interaction between autonomic-mediated spinal cord reflexes, higher neural centers, and the vascular system. Erection is mediated by the parasympathetic nervous system and emission and ejaculation by the sympathetic nervous system. Erectile dysfunction is defined as the inability to achieve and maintain an erection sufficient to permit satisfactory sexual intercourse. It can be due to psychogenic factors, organic disorders, or mixed psychogenic and organic conditions.

DISORDERS OF THE PENIS, THE SCROTUM AND TESTES, AND THE PROSTATE

Disorders of the Penis

Disorders of the penis include congenital defects (discussed in Disorders in Childhood), acute and chronic inflammatory conditions, Peyronie's disease, priapism, and neoplasms.

Inflammation and Infection

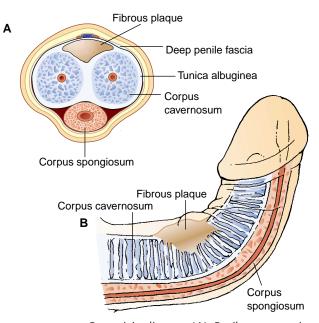
Balantitis and Balanoposthitis. Balanitis is an acute or chronic inflammation of the glans penis. Balanoposthitis refers to inflammation of the glans and prepuce. It usually is encountered in males with phimosis (a tight foreskin) or a large, redundant prepuce that interferes with cleanliness and predisposes to bacterial growth in the accumulated secretions and smegma (*i.e.*, debris from the desquamated epithelia). If left untreated, the condition may cause ulcerations of the mucosal surface of the glans; these ulcerations may lead to inflammatory scarring of the phimosis and further aggravate the condition.

Acute superficial balanoposthitis is characterized by erythema of the glans and prepuce. An exudate in the form of malodorous discharge may be present. Extension of the erythema and edema may result in phimosis. The condition may result from infection, trauma, or irritation. Infective balanoposthitis may be caused by a wide variety of organisms. Chlamydiae and mycoplasmas have been identified as causative organisms in this disease. The inflammatory reaction is nonspecific, and correct identification of the specific agent requires bacterial smears and cultures.

Balanitis xerotica obliterans is a chronic white, patchy lesion that originates on the glans and usually progresses to involve the meatus. It is clinically and histologically similar to the lichen sclerosus that is seen in females. It commonly is observed in middle-aged diabetic men. Treatment measures include topical or intralesional injections of corticosteroids.⁸

Peyronie's Disease

Peyronie's disease involves a localized and progressive fibrosis of unknown origin that affects the tunica albuginea (*i.e.*, the tough, fibrous sheath that surrounds the corpora cavernosa) of the penis. It is named after Francois de la Peyronie, who wrote in 1743, describing a patient who had "rosary beads of scar tissue to cause upward curvature of the penis during erection."9 The disorder is characterized initially by an inflammatory process that results in dense fibrous plaque formation. The plaque usually is on the dorsal midline of the shaft, causing upward bowing of the shaft during erection (Fig. 33-7). Some men may develop scarring on both the dorsal and ventral aspects of the shaft, causing the penis to be straight but shortened or have a lateral bend.9 The fibrous tissue prevents lengthening of the involved area during erection, making intercourse difficult and painful. The disease usually occurs in middle-aged or elderly men. Although the cause of the disorder is unknown, the dense microscopic plaques are consistent with findings of severe vasculitis. As many as 47% of men with Peyronie's disease have another condition associated with fascial tissue fibrosis, such as Dupuytren's contracture (fibrosis of the palmar fascia).9



■ FIGURE 33-7 ■ Peyronie's disease. (A) Penile cross-section showing plaque between the corpora. (B) Penile curvature.

The manifestations of Peyronie's disease include painful erection, bent erection, and the presence of a hard mass at the site of fibrosis. Approximately two thirds of men report pain as a symptom. The pain is thought to be generated by inflammation of the adjacent fascial tissue and usually disappears as the inflammation resolves.¹⁰ During the first year or so after formation of the plaque, while the scar tissue is undergoing the process of remodeling, penile distortion may increase, remain static, or resolve and disappear completely.⁹ In some cases, the scar tissue may progress to calcification and formation of bone-like tissue.

Diagnosis is based on history and physical examination. Doppler ultrasonography may be used to assess causation of the disorder. Although surgical intervention can be used to correct the disorder, it often is delayed, because in many cases the disorder is self-limiting.¹⁰ Less invasive treatments include the administration of oral agents with antioxidant properties (*e.g.*, vitamin D, potassium aminobenzoate, and colchicine).

Priapism

Priapism is an involuntary, prolonged, abnormal and painful erection that is not associated with sexual excitement. Priapism is a true urologic emergency because the prolonged erection can result in ischemia and fibrosis of the erectile tissue with significant risk of subsequent impotence. Priapism can occur at any age, in the newborn as well as other age groups. Sickle cell disease or neoplasms are the most common cause in boys between 5 and 10 years of age.

Priapism is due to impaired blood flow in the corpora cavernosa of the penis. Two mechanisms for priapism have been proposed: low-flow (ischemic) priapism, in which there is stasis of blood flow in the corpora cavernosa with a resultant failure of detumescence, and high-flow (nonischemic) priapism, which involves persistent arterial flow into the corpora cavernosa.¹¹ In high-flow priapism, there is no hypoxia of local tissue, the penis is less rigid, the pain is less than in stasis priapism, and permanent corporal fibrosis and cellular damage are rare.¹¹

Priapism is classified as primary (idiopathic) or secondary to a disease or drug effect. Primary priapism is the result of conditions such as trauma, infection, and neoplasms. Secondary causes include hematologic conditions such as leukemia, sickle cell disease, and thrombocytopenia; neurologic conditions such as stroke, spinal cord injury, and other central nervous system lesions; and renal failure. Between 6% and 12% of males with sickle cell disorders are affected by priapism.¹¹ The relative deoxygenation and stasis of cavernosal blood during erection is thought to increase sickling. Various medications, such as antihypertensive drugs, anticoagulant drugs, antidepressant drugs, alcohol, and marijuana, can contribute to the development of priapism. Androstenedione, sold as an over-the-counter drug to enhance muscle building and athletic performance, has been implicated in the disorder.¹² Currently, intracavernous injection therapy for erectile dysfunction is one of the more common causes of priapism.

The diagnosis of priapism usually is based on clinical findings. Doppler studies of penile blood flow, penile ultrasonography, and computed tomography (CT) scans may be used to determine intrapelvic pathology.

Initial treatment measures include analgesics, sedation, and hydration. Urinary retention may necessitate catheterization. Local measures include application of ice packs and irrigation of the corpus cavernosum with plain or heparinized saline, or instillation of α -adrenergic drugs. If less aggressive treatment does not produce detumescence, a temporary surgical shunt may be established between the corpus cavernosum and the corpus spongiosum.

The prognosis for whether fibrosis or erectile failure will occur is determined by the severity and duration of blood stasis. In high-flow priapism, the damaging effects of decreased oxygen tension and intracavernal blood pressure are less pronounced than in stasis priapism. Normal erectile potency can be restored even after a long duration of high-flow priapism. In contrast, persistent stasis priapism is known to result in impaired erectile function and tissue fibrosis unless resolved within 24 to 48 hours of onset.¹¹

Cancer of the Penis

Squamous cell cancer of the penis is most common in men between 45 and 60 years of age. In the United States, it accounts for less than 1% of male genital tumors; however, in other countries, it accounts for 10% to 20% of male cancers.^{13,14} When it is diagnosed early, penile cancer is highly curable. The greatest hindrance to early diagnosis is a delay in seeking medical attention.

The cause of penile cancer is unknown. Several risk factors have been suggested, including poor hygiene, human papillomavirus (HPV) infections, ultraviolet radiation exposure, and immunodeficiency states. There is an association between penile cancer and poor genital hygiene and phimosis. One theory postulates that smegma accumulation under the phimotic foreskin may produce chronic inflammation, leading to carcinoma. The HPVs have been implicated in the genesis of several genital cancers, including cancer of the penis.¹³ Ultraviolet radiation also is thought to have a carcinogenic effect on the penis. Males who were treated for psoriasis with ultraviolet A or B therapies (*i.e.*, PUVA or PUVB) have had a reported increased incidence of genital squamous cell carcinomas.¹³ Because of this observation, it is suggested that men should shield their genital area when using tanning salons. Immunodeficiency states also may play a role in the pathogenesis of penile cancer. Approximately 18% of men with acquired immunodeficiency syndrome (AIDS)-related Kaposi's sarcoma have lesions of the penis or genitalia.¹³

Dermatologic lesions with precancerous potential include balanitis xerotica obliterans (discussed earlier) and giant condylomata acuminata.¹⁴ Giant condylomata acuminata are cauliflower-like lesions arising from the prepuce or glans that result from HPV infection.

Approximately 95% of penile cancers are squamous cell carcinoma.¹³ It is thought to progress from an in situ lesion to an invasive carcinoma. Bowen's disease and erythroplasia of Queyrat are penile lesions with histologic features of carcinoma in situ.¹⁴ Bowen's disease appears as a solitary, thickened, gray-white, opaque plaque with shallow ulceration and crusting. It commonly involves the skin of the shaft of the penis and the scrotum. Erythroplasia of Queyrat involves the mucosal surface of the glans or prepuce. It is characterized by single or multiple shiny red, sometimes velvety, plaques.¹⁴ These lesions require careful follow-up because of their potential to progress to invasive carcinoma.

Invasive carcinoma of the penis begins as a small lump or ulcer. If phimosis is present, there may be painful swelling, purulent drainage, or difficulty urinating. Palpable lymph nodes may be present in the inguinal region. Diagnosis usually is based on physical examination and biopsy results. CT scans, penile ultrasound studies, and magnetic resonance imaging (MRI) may be used in the diagnostic workup.

Treatment options vary according to stage, size, location, and invasiveness of the tumor. Carcinoma in situ may be treated conservatively with fluorouracil cream application or laser treatment.¹⁴ Surgery remains the mainstay of treatment for invasive carcinoma. Partial or total penectomy is indicated for invasive lesions.

Disorders of the Scrotum and Testes

The testes, or male gonads, are two egg-shaped structures located outside the abdominal cavity in the scrotum. Embryologically, the testes develop in the abdominal cavity and then descend through the inguinal canal into a pouch of peritoneum (which becomes the tunica vaginalis) in the scrotum during the seventh to ninth months of fetal life. As they descend, the testes pull their arteries, veins, lymphatics, nerves, and conducting excretory ducts with them. These structures are encased by the cremaster muscle and layers of fascia that constitute the spermatic cord (Fig. 33-8A). The descent of the testes is thought to be mediated by testosterone, which is active during this stage of development.

After descent of the testes, the inguinal canal closes almost completely. Failure of this canal to close predisposes to the development of an inguinal hernia later in life (Fig. 33-8B). An inguinal hernia or "rupture" is a protrusion of the parietal peritoneum and part of the intestine through an abnormal opening from the abdominal cavity. A loop of small bowel may become incarcerated in an inguinal hernia (strangulated hernia), in which case the lumen may become obstructed, and the vascular supply compromised (see Chapter 27).



The testes and epididymis are completely surrounded by the tunica vaginalis, a serous pouch derived from the peritoneum during fetal descent of the testes into the scrotum. The tunica vaginalis has an outer parietal layer and a deeper visceral layer that adheres to the dense fibrous covering of the testes, the tunica albuginea. The tunica albuginea protects the testes and gives them their ovoid shape. A space exists between these two layers that typically contains a few milliliters of clear fluid. The cremaster muscles, which are bands of skeletal muscle arising from the internal oblique muscles of the trunk, elevate the testes. The testes receive their arterial blood supply from the long testicular arteries, which branch from the aortic artery. The spermatic veins, which drain the testes, arise from a venous network called the pampiniform plexus that surrounds the spermatic artery. The testes are innervated by fibers from both divisions of the autonomic nervous system. Associated sensory nerves transmit pain impulses, resulting in excruciating pain, especially when the testes are hit forcibly.

The scrotum, which houses the testes, is made up of a thin outer layer of skin that forms rugae, or folds, and is continuous with the perineum and outer skin of the groin. Under the outer skin lies a thin layer of fascia and smooth muscle (*i.e.*, dartos muscle). This layer contains a septum that separates the two testes. The dartos muscle responds to changes in temperature. When it is cold, the muscle contracts, bringing the testes closer to the body, and the scrotum becomes shorter and heavily wrinkled. When it is warmer, the muscle relaxes, allowing the scrotum to fall away from the body.

The location of the testes in the scrotum is important for sperm production, which is optimal at 2°C to 3°C below body temperature. Two systems maintain the temperature of the testes at a level consistent with sperm production. One is the pampiniform plexus of spermatic veins that surround the spermatic artery. This plexus absorbs heat from the arterial blood, cooling it as it enters the testes. The other is the cremaster muscles, which respond to decreases in testicular temperature by moving the testes closer to the body. Prolonged exposure to elevated temperatures, as a result of prolonged fever or the dysfunction of thermoregulatory mechanisms, can impair spermatogenesis. Some tight-fitting undergarments hold the testes against the body and are thought to contribute to a decrease in sperm counts and infertility by interfering with the thermoregulatory function of the scrotum. Cryptorchidism, the failure of the testes to descend into the scrotum, also exposes the testes to the higher temperature of the body.

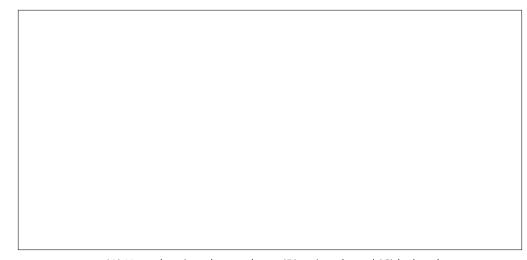
Disorders of the Testicular Tunica

Hydrocele. A hydrocele forms when excess fluid collects between the layers of the tunica vaginalis (Fig. 33-9). It may be unilateral or bilateral and can develop as a primary congenital defect or as a secondary condition. Acute hydrocele may develop after local injury, epididymitis or orchitis, gonorrhea, lymph obstruction, or germ cell testicular tumor, or as a side effect of radiation therapy. Chronic hydrocele is more common. Fluid collects about the testis, and the mass grows gradually. Its cause is unknown, and it usually develops in men older than 40 years.

Most cases of hydrocele in male infants and children are due to a patent processus vaginalis, which is continuous with the peritoneal cavity. In many cases they are associated with an indirect inguinal hernia.¹⁵ Most hydroceles of infancy close spontaneously; therefore, they are not usually repaired before the age of 1 year. Hydroceles that persist beyond 2 years of age may require surgical treatment.

Hydroceles are palpated as cystic masses that may attain massive proportions. If there is enough fluid, the mass may be mistaken for a solid tumor. Transillumination of the scrotum (*i.e.*, shining a light through the scrotum for the purposes of visualizing its internal structures) or ultrasonography can help to determine whether the mass is solid or cystic and whether the testicle is normal.¹⁶ A dense hydrocele that does not illuminate should be differentiated from a testicular tumor. If a hydrocele develops in a young man without apparent cause, careful evaluation is needed to exclude cancer or infection.

In an adult male, a hydrocele is a relatively benign condition. The condition often is asymptomatic, and no treatment is necessary. When symptoms do occur, the feeling may be that of heaviness in the scrotum or pain in the lower back. In cases of secondary hydrocele, the primary condition is treated. If the hydrocele is painful or cosmetically undesirable, surgical correction is indicated. Surgical repair may be done inguinally or transcrotally.



■ FIGURE 33-9 ■ (A) Normal testis and appendages, (B) varicocele, and (C) hydrocele.

Hematocele. A hematocele is an accumulation of blood in the tunica vaginalis, which causes the scrotal skin to become dark red or purple. It may develop as a result of an abdominal surgical procedure, scrotal trauma, a bleeding disorder, or a testicular tumor.

Spermatocele. A spermatocele is a painless, sperm-containing cyst that forms at the end of the epididymis. It is located above and posterior to the testis, is attached to the epididymis, and is separate from the testes. Spermatoceles may be solitary or multiple and usually are less than 1 cm in diameter. They are freely movable and should transilluminate. Spermatoceles rarely cause problems, but a large one may become painful and require excision.

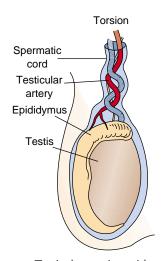
Varicocele. A varicocele is characterized by varicosities of the pampiniform plexus, a network of veins supplying the testes (Fig. 33-9). The left side is more commonly affected because the left internal spermatic vein inserts into the left renal vein at a right angle, whereas the right spermatic vein usually enters the inferior vena cava. Incompetent valves are more common in the left internal spermatic veins, causing a reflux of blood back into the veins of the pampiniform plexus. The force of gravity resulting from the upright position also contributes to venous dilatation. If the condition persists, there may be damage to the elastic fibers and hypertrophy of the vein walls, as occurs in formation of varicose veins in the leg. Sperm concentration and motility are decreased in 65% to 75% of men with varicocele because of changes in testicular temperature resulting from altered blood flow.¹⁶

Varicoceles rarely are found before puberty, and the incidence is highest in men between 15 and 35 years of age. Symptoms of varicocele include an abnormal feeling of heaviness in the left scrotum, although many varicoceles are asymptomatic. Usually, the varicocele is readily diagnosed on physical examination with the patient in the standing and recumbent positions. Typically, the varicocele disappears in the lying position because of venous decompression into the renal vein. Scrotal palpation of a varicocele has been compared to feeling a "bag of worms." Treatment options include surgical ligation or sclerosis using a percutaneous transvenous catheter under fluoroscopic guidance. It has been shown that 40% of men with abnormalities in their semen and a varicocele show some degree of improvement in fertility after obliteration of the dilated veins.¹⁶ Aside from improving fertility, other reasons for surgery include the relief of the sensation of "heaviness" and cosmetic improvement.

Testicular Torsion

Testicular torsion is a twisting of the spermatic cord that suspends the testis (Fig. 33-10). It is the most common acute scrotal disorder in the pediatric and young adult population. Testicular torsion can be divided into two distinct clinical entities, depending on the level of spermatic cord involvement: extravaginal and intravaginal torsion.¹⁷

Extravaginal torsion, which occurs almost exclusively in neonates, is the less common form of testicular torsion. It occurs



■ FIGURE 33-10 ■ Testicular torsion with twisting of the spermatic cord that suspends the testis and the spermatic vessels that supply the testis with blood.

when the testicle and the fascial tunicae that surround it rotate around the spermatic cord at a level well above the tunica vaginalis. Intravaginal torsion is considerably more common than extravaginal torsion. It occurs when the testis rotates on the long axis in the tunica vaginalis. In most cases, congenital abnormalities of the tunica vaginalis or spermatic cord exist.¹⁷ The tunica vaginalis normally surrounds the testes and epididymis, allowing the testicle to rotate freely in the tunica. Although anomalies of suspension vary, the epididymal attachment may be loose enough to permit torsion between the testis and the epididymis. More commonly, the testis rotates about the distal spermatic cord. Because this abnormality is developmental, bilateral anomalies are common.

Intravaginal torsion occurs most frequently between the ages of 8 and 18 years and rarely is seen after 30 years of age. Males usually present in severe distress within hours of onset and often have nausea, vomiting, and tachycardia. The affected testis is large and tender, with pain radiating to the inguinal area. Extensive cremaster muscle contraction causes a thickening of the spermatic cord.

Testicular torsion must be differentiated from epididymitis, orchitis, and trauma to the testis. On physical examination, the testicle often is high in the scrotum and in an abnormal orientation. These changes are due to the twisting and shortening of the spermatic cord. The degree of scrotal swelling and redness depends on the duration of symptoms. The testes are firm and tender. The cremasteric reflex, normally elicited by stroking the medial aspect of the thigh and observing testicular retraction, frequently is absent.¹⁷ Color Doppler ultrasonography is increasingly used in the evaluation of suspected testicular torsion.17

Intravaginal testicular torsion is a true surgical emergency, and early recognition and treatment are necessary if the testicle is to be saved. Treatment includes surgical detorsion (repositioning and fixation) and orchiectomy. Orchiectomy is carried out when the testis is deemed nonviable after surgical detorsion. Testicular salvage rates are directly related to the duration of torsion. Because the opposite testicle usually is affected by the same abnormal attachments, prophylactic fixation of that testis often is performed.

Inflammation and Infection

Epididymitis. Epididymitis is an inflammation of the epididymis. There are two major types of epididymitis: sexually transmitted infections associated with urethritis and primary nonsexually transmitted infections associated with urinary tract infections and prostatitis. Most cases of epididymitis are caused by bacterial pathogens.

In primary nonsexual infections, the pressure associated with voiding or physical strain may force pathogen-containing urine from the urethra or prostate up the ejaculatory duct and through the vas deferens and into the epididymis. Infections also may reach the epididymis through the lymphatics of the spermatic cord. In rare cases, organisms from other foci of infection reach the epididymis through the bloodstream. In children, the disorder usually is associated with congenital urinary tract abnormalities and infection with gram-negative rods.

Sexually transmitted acute epididymitis occurs mainly in young men without underlying genitourinary disease and is most commonly caused by Chlamydia trachomatis and Neisseria gonorrhoeae (singly or in combination). In men older than 35 years, epididymitis often is associated with pathogens such as Escherichia coli, Pseudomonas, and gram-positive cocci.

Epididymitis is characterized by unilateral pain and swelling, accompanied by erythema and edema of the overlying scrotal skin that develops during a period of 24 to 48 hours. Initially, the swelling and induration are limited to the epididymis. However, the distinction between the testis and epididymis becomes less evident as the inflammation progresses, and the testis and epididymis become one mass. There may be tenderness over the groin (spermatic cord) or in the lower abdomen. Fever and reports of dysuria occur in approximately one half of cases. Whether urethral discharge is present depends on the organism causing the infection; it usually accompanies gonorrheal infections, is common in chlamydial infections, and is less common in infections caused by gram-negative organisms.

Laboratory findings usually reveal an elevated white blood cell count. Urinalysis and urine culture are important in the diagnosis of epididymitis, with bacteriuria and pyuria suggestive of the disorder. Treatment during the acute phase (which usually lasts for 3 to 4 days) includes bed rest, scrotal elevation and support, and antibiotics.¹⁸ Bed rest with scrotal support improves lymphatic drainage. The choice of antibiotics is determined by age, physical findings, urinalysis, Gram's stain results, cultures, and sexual history. Oral analgesics and antipyretics usually are indicated. Sexual activity or physical strain may exacerbate the infection and worsen the symptoms and should be avoided.

Orchitis. Orchitis is an infection of the testes. It can be precipitated by a primary infection in the genitourinary tract, or the infection can be spread to testes through the bloodstream or the lymphatics. Epididymitis with subsequent infection of the testis is commonly related to genitourinary tract infections (cystitis, urethritis, genitoprostatitis) that travel to the epididymis and testis through the vas deferens or the lymphatics of the spermatic cord.

Orchitis can develop as a complication of a systemic infection, such as parotitis (i.e., mumps), scarlet fever, or pneumonia. Probably the best known of these complications is orchitis caused by the mumps virus. Mumps orchitis does not occur in prepubertal boys. However, approximately 20% to 35% of adolescent boys and young men with mumps develop this form of orchitis.¹⁸ The onset of mumps orchitis is sudden; it usually occurs approximately 3 to 4 days after the onset of the parotitis and is characterized by fever, painful enlargement of the testes, and small hemorrhages into the tunica albuginea. Unlike epididymitis, the urinary symptoms are absent. The symptoms usually run their course in 7 to 10 days. The residual effects seen after the acute phase include hyalinization of the seminiferous tubules and atrophy of the testes. Spermatogenesis is irreversibly impaired in approximately 30% of testes damaged by mumps orchitis.¹⁸ If both testes are involved, permanent sterility results, but and rogenic hormone function usually is maintained.

Cancer of the Scrotum and Testes

Tumors can develop in the scrotum or the testes. Benign scrotal tumors are common and often do not require treatment. Carcinoma of the scrotum is rare and usually is associated with exposure to carcinogenic agents. Almost all solid tumors of the testes are malignant.

Scrotal Cancer. Cancer of the scrotum was the first cancer directly linked to a specific occupation when, in the 1800s, it was associated with chimney sweeps.¹⁹ Studies have linked this cancer to exposure to tar, soot, and oils. Most squamous cell cancers of the scrotum are linked to poor hygiene and chronic inflammation. Exposure to ultraviolet A radiation (*e.g.*, PUVA) or HPV also has been associated with the disease. The mean age of presentation with the disease is 60 years, often preceded by 20 to 30 years of chronic irritation.

In the early stages, cancer of the scrotum may appear as a small tumor or wartlike growth that eventually ulcerates. The thin scrotal wall lacks the tissue reactivity needed to block the malignant process; more than one half of the cases seen involve metastasis to the lymph nodes. Because this tumor does not respond well to chemotherapy or irradiation, the treatment includes wide local excision of the tumor with inguinal and femoral node dissection.²⁰

Testicular Cancer. Testicular cancer accounts for 1% of all male cancers and 3% of male urogenital cancers. Although relatively rare, it is the most common cause of cancer in the 15- to 34-year-old age group.^{20,21} In the past, testicular cancer was a leading cause of death among males entering their most productive years. However, since the late 1970s, advances in therapy have transformed an almost invariably fatal disease into one that is highly curable. With appropriate treatment, the prognosis for men with testicular cancer is excellent. The 5-year survival rate for patients with early disease exceeds 95%. Even patients with more advanced disease have excellent chances for long-term survival.

Although the cause of testicular cancer is unknown, several predisposing influences may be important: cryptorchidism, genetic factors, and disorders of testicular development.²² The strongest association has been with cryptorchid or undescended testes. The higher the location of the undescended testis, the greater the risk.²² Genetic predisposition also appears to be important. Family clustering of the disorder has been described, although a well-defined pattern of inheritance has not been established. Men with disorders of testicular development, including those with Klinefelter's syndrome and testicular feminization, have a higher risk of germ cell tumors.

Approximately 95% of malignant tumors arising in the testis are germ cell tumors.^{21,23} Germ cell tumors can be classified as seminomas and nonseminomas based on their origin in primordial germ cells and their ability to differentiate in vivo. Because these tumors derive from germ cells in the testis, they are multipotential (able to differentiate into different tissue types) and often secrete polypeptide hormones or enzymes representing earlier stages of development (see Chapter 5). *Seminomas* are the most common type of testicular tumors. They account for approximately 50% of germ cell tumors and occur most frequently during the fourth decade of life.²¹ Seminomas are thought to arise from the seminiferous epithelium of the testes and are the type of germ cell tumor most likely to produce a uniform population of cells.

The nonseminoma tumors include embryonal carcinoma, teratoma, choriocarcinoma, and yolk cell carcinoma derivatives. Nonseminoma tumors usually contain more than one cell type and are less differentiated than seminomas. Embryonal carcinomas are the least differentiated of the tumors, with totipotential capacity to differentiate into other nonseminomatous cell types. They occur most commonly in the 20- to 30-year-old age group. Choriocarcinoma is a rare and highly malignant form of testicular cancer that is identical to tumors that arise in placental tissue. Yolk sac tumors mimic the embryonic yolk sac histologically. They are the most common type of testicular tumors in infants and children to 3 years of age, and in this age group are associated with a very good prognosis.²² Teratomas are composed of somatic cell types from two or more germ-line layers (ectoderm, mesoderm, or endoderm). They constitute less than 2% to 3% of germ cell tumors and can occur at any age from infancy to old age. They usually behave as benign tumors in children; in adults, they often contain minute foci of cancer cells.

Often the first sign of testicular cancer is a slight enlargement of the testicle that may be accompanied by some degree of discomfort. This may be an ache in the abdomen or groin or a sensation of dragging or heaviness in the scrotum. Frank pain may be experienced in the later stages, when the tumor is growing rapidly and hemorrhaging occurs. Testicular cancer can spread when the tumor may be barely palpable. Signs of metastatic spread include swelling of the lower extremities, back pain, cough, hemoptysis, or dizziness. Gynecomastia (breast enlargement) may result from human chorionic gonadotropin (hCG)-producing tumors.

Early diagnosis of testicular cancer is important because a delay in seeking medical attention often results in presentation with a later stage of the disease and decreased treatment effectiveness. Recognition of the importance of prompt diagnosis and treatment has resulted in the development of a procedure for testicular self-examination and an emphasis on public education programs about this type of cancer. The American Cancer Society strongly advocates that every young adult male examine his testes at least once each month as a means of early detection of testicular cancer.

The diagnosis of testicular cancer requires a thorough urologic history and physical examination. A painless testicular mass may be cancer. The examination for masses should include palpation of the testes and surrounding structures, transillumination of the scrotum, and abdominal palpation. Testicular ultrasonography can be used to differentiate testicular masses. CT scans and MRI are used in assessing metastatic spread.

Tumor markers, which measure protein antigens produced by malignant cells, provide information about the existence of a small or undetected tumor and the type of tumor present. Three tumor markers are useful in evaluating the tumor response: α -fetoprotein, a glycoprotein that normally is present in fetal serum in large amounts; hCG, a hormone that normally is produced by the placenta in pregnant women; and lactate dehydrogenase (LDH), a cellular enzyme normally found in muscle, liver, kidney, and brain. The reappearance in the adult of these protein markers suggests activity normally present in the embryo.

The basic treatment of all testicular cancers includes orchiectomy, which is done at the time of diagnostic exploration. Dependent on the histologic characteristics of the tumor and the clinical stage of the disease, radiation or chemotherapy may used after orchiectomy. Rigorous follow-up in all men with testicular cancer is necessary to detect recurrence, which most often occurs within the first year.²³

Disorders of the Prostate

The prostate is a fibromuscular and glandular organ lying just inferior to the bladder. The prostate gland secretes a thin, milky, alkaline fluid containing citric acid, calcium, acid phosphate, a clotting enzyme, and a profibrinolysin. During ejaculation, the capsule of the prostate contracts, and the added fluid increases the bulk of the semen. Both vaginal secretions and the fluid from the vas deferens are strongly acidic. Because sperm mobilization occurs at a pH of 6.0 to 6.5, the alkaline nature of the prostatic secretions is essential for successful fertilization of the ovum. The bulbourethral glands (Fig. 33-1) lie on either side of the membranous urethra and secrete an alkaline mucus, which further aids in neutralizing acids from the urine that remain in the urethra.

The prostate gland, which forms a fibrous capsule that surrounds the urethra where it joins the bladder, also functions in the elimination of urine. The segment of urethra that travels through the prostate gland is called the *prostatic urethra*. The prostatic urethra is lined by a thin layer of smooth muscle that is continuous with the bladder wall. This smooth muscle represents the true involuntary sphincter of the male posterior urethra. Because the prostate surrounds the urethra, enlargement of the gland can produce urinary obstruction.

The prostate gland is made up of many secretory glands arranged in three concentric areas surrounding the prostatic urethra, into which they open. The component glands of the prostate include the (1) small mucosal glands associated with the urethral mucosa, (2) the intermediate submucosal glands that lie peripheral to the mucosal glands, and (3) the large main prostatic glands that are situated toward the outside of the gland. It is the overgrowth of the mucosal glands that causes benign prostatic hyperplasia in older men.

Prostatitis

Prostatitis refers to a variety of inflammatory disorders of the prostate gland, some of which are bacterial and some are not. It may occur spontaneously, as a result of catheterization or instrumentation, or secondary to other diseases of the male genitourinary system. As an outcome of 1995 and 1998 consensus conferences, the National Institutes of Health has established a classification system with four categories of prostatitis syndromes: asymptomatic inflammatory prostatitis, acute bacterial prostatitis, chronic bacterial prostatitis, and chronic prostatitis/ pelvic pain syndrome.²⁴ Men with asymptomatic inflammatory prostatitis have no subjective symptoms and are detected incidentally on biopsy or examination of prostatic fluid.

Acute Bacterial Prostatitis. Acute bacterial prostatitis often is considered a subtype of urinary tract infection. The most likely etiology of acute bacterial prostatitis is an ascending urethral infection or reflux of infected urine into the prostatic ducts. The most common organism is E. coli. Other frequently found species include Pseudomonas, Klebsiella, and Proteus. Less frequently, the infection is caused by Staphylococcus aureus, Streptococcus faecalis, Chlamydia, or anaerobes such as Bacteroides species.^{25,26}

The manifestations of acute bacterial prostatitis include fever and chills, malaise, myalgia, arthralgia, frequent and urgent urination, dysuria, and urethral discharge. Dull, aching pain often is present in the perineum, rectum, or sacrococcygeal region. The urine may be cloudy and malodorous because of urinary tract infection. Rectal examination reveals a swollen, tender, warm prostate with scattered soft areas. Prostatic massage produces a thick discharge with white blood cells that grows large numbers of pathogens on culture.

Acute prostatitis usually responds to appropriate antimicrobial therapy chosen in accordance with the sensitivity of the causative agents in the urethral discharge. Depending on the urine culture results, antibiotic therapy usually is continued for at least 4 weeks. Because acute prostatitis often is associated with anatomic abnormalities, a thorough urologic examination usually is performed after treatment is completed.

A persistent fever indicates the need for further investigation for an additional site of infection or a prostatic abscess. CT scans and transrectal ultrasonography of the prostate are useful in the diagnosis of prostatic abscesses. Prostatic abscesses, which are relatively uncommon since the advent of effective antibiotic therapy, are found more commonly in males with diabetes mellitus. Because prostatic abscesses usually are associated with bacteremia, prompt drainage by transperitoneal or transurethral incision followed by appropriate antimicrobial therapy usually is indicated.²⁵

Chronic Bacterial Prostatitis. In contrast to acute bacterial prostatitis, chronic bacterial prostatitis is a subtle disorder that is difficult to treat. Men with the disorder typically have recurrent urinary tract infections with persistence of the same strain of pathogenic bacteria in prostatic fluid and urine. Organisms responsible for chronic bacterial prostatitis usually are the gram-negative enterobacteria (E. coli, Proteus, or Klebsiella) or *Pseudomonas*. Occasionally, a gram-positive organism such as S. faecalis is the causative organism. Infected prostatic calculi may develop and contribute to the chronic infection.

The symptoms of chronic prostatitis are variable and include frequent and urgent urination, dysuria, perineal discomfort, and low back pain. Occasionally, myalgia and arthralgia accompany the other symptoms. Secondary epididymitis sometimes is associated with the disorder. Many men develop relapsing lower or upper urinary tract infections because of recurrent invasion of the bladder by the prostatic bacteria. Bacteria may exist in the prostate gland even when the prostatic fluid is sterile.

The most accurate method of establishing a diagnosis is by urine cultures. Even after an accurate diagnosis has been established, treatment of chronic prostatitis often is difficult and frustrating. Unlike their action in the acutely inflamed prostate, antibacterial drugs penetrate poorly into the chronically inflamed prostate. Long-term therapy (3 to 4 months) with an appropriate low-dose oral antimicrobial agent often is used to treat the infection. Transurethral prostatectomy may be indicated when the infection is not cured or adequately controlled by medical therapy, particularly when prostate stones are present.

Chronic Prostatitis/Chronic Pelvic Pain Syndrome. Chronic prostatitis/pelvic pain syndrome is both the most common and least understood of the prostatitis syndromes.²⁷ The category is divided into two types, inflammatory and noninflammatory, based on the presence of leukocytes in the prostatic fluid. The inflammatory type was previously referred to as nonbacterial prostatitis, and the noninflammatory type as prostatodynia.

Men with nonbacterial prostatitis have inflammation of the prostate with an elevated leukocyte count, inflammatory cells

594 Unit Nine: Alterations in the Male and Female Reproductive Systems

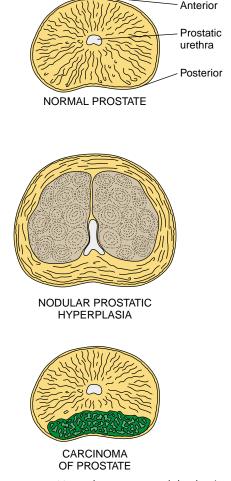
in their prostatic secretions, but no evidence of bacteria. The cause of the disorder is unknown, and efforts to prove the presence of unusual pathogens (*e.g.*, mycoplasmas, chlamydiae, trichomonads, viruses) have been largely unsuccessful. It also is thought that nonbacterial prostatitis may be an autoimmune disorder. Manifestations of *inflammatory prostatitis* include pain along the penis, testicles, and scrotum; painful ejaculation; low back pain; rectal pain along the inner thighs; urinary symptoms; decreased libido; and impotence.

Men with noninflammatory prostatitis have symptoms resembling those of nonbacterial prostatitis but have negative urine culture results and no evidence of prostatic inflammation (*i.e.*, normal leukocyte count). The cause of noninflammatory prostatitis is unknown, but because of the absence of inflammation, the search for the cause of symptoms has been directed toward extraprostatic sources. In some cases, there is an apparent functional obstruction of the bladder neck near the external urethral sphincter; during voiding, this results in higher than normal pressures in the prostatic urethra that cause intraprostatic urine reflux and chemical irritation of the prostate by urine. In other cases, there is an apparent myalgia (*i.e.*, muscle pain) associated with prolonged tension of the pelvic floor muscles. Emotional stress also may play a role.

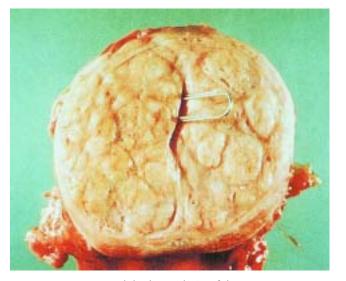
Treatment methods for chronic prostatitis/pelvic pain syndrome are highly variable. Antibiotic therapy is used when an occult infection is suspected. Sitz baths and nonsteroidal antiinflammatory agents may provide some symptom relief. In men with irritative urination symptoms, anticholinergic agents or α -adrenergic-blocking agents may be beneficial.

Benign Prostatic Hyperplasia

Benign prostatic hyperplasia (BPH) is an age-related, nonmalignant enlargement of the prostate gland (Fig. 33-11). It is characterized by the formation of large, nodular lesions in



■ FIGURE 33-12 ■ Normal prostate, nodular benign prostatic hypertrophy, and cancer of the prostate. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 954]. Philadelphia: Lippincott Williams & Wilkins.) (Artist: Dimitri Karetnikov)



■ FIGURE 33-11 ■ Nodular hyperplasia of the prostate. Cut surface of a prostate enlarged by nodular hyperplasia shows numerous, well-circumscribed nodules of prostatic tissue. The prostatic urethra (*paper clip*) has been compressed to a narrow slit. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 955]. Philadelphia: Lippincott Williams & Wilkins)

the periurethral region of the prostate, rather than the peripheral zones, which commonly are affected by prostate cancer (Fig. 33-12). BPH is one of the most common diseases of aging men. It has been reported that 25% of men older than 55 years of age and 50% of men older than 75 years of age experience symptoms of BPH.²⁸

The exact cause of the BPH is unknown. Potential risk factors include age, family history, race, ethnicity, and hormonal factors. The incidence of BPH increases with advancing age, is highest in African Americans, and is lowest in native Japanese. Men with a family history of BPH are reported to have had larger prostates than those of control subjects, and higher rates of BPH were found in monozygotic twins than in dizygotic twins.²⁸

Both androgens (testosterone) and estrogens appear to contribute to the development of BPH. Dihydrotestosterone (DHT), the biologically active metabolite of testosterone, is thought to be the ultimate mediator of prostatic hyperplasia, with estrogen serving to sensitize the prostatic tissue to the growth-producing effects of DHT. Free plasma testosterone enters prostatic cells, where at least 90% is converted into DHT by

KEY CONCEPTS

HYPERPLASIA AND CANCER OF THE PROSTATE

- The prostate gland surrounds the urethra, and periurethral enlargement will cause manifestations of urinary obstruction.
- Benign prostatic hyperplasia is an age-related enlargement of the prostate gland with formation of large, discrete lesions in the periurethral region of the prostate. These lesions compress the urethra and produce symptoms of dysuria or difficulty urinating.
- Prostatic cancer begins in the peripheral zones of the prostate gland and usually is asymptomatic until the disease is far advanced and the tumor has eroded the outer prostatic capsule and spread to adjacent pelvic tissues or metastasized.

the action of 5α -reductase. The discovery that DHT is the active factor in BPH is the rationale for use of 5α -reductase inhibitors in the treatment of the disorder. Although the exact source of estrogen is uncertain, small amounts of estrogen are produced in the male. It has been postulated that a relative increase in estrogen levels that occurs with aging may facilitate the action of androgens in the prostate despite a decline in the testicular output of testosterone.

The anatomic location of the prostate at the bladder neck contributes to the pathophysiology and symptomatology of BPH. There are two prostatic components to the obstructive properties of BPH and the development of lower urinary tract symptoms: dynamic and static.28,29 The static component of BPH is related to an increase in prostatic size and gives rise to symptoms such as a weak urinary stream, postvoid dribbling, frequency of urination, and nocturia. The dynamic component of BPH is related to prostatic smooth muscle tone. The α_1 adrenergic receptors are the main receptors for the smooth muscle component of the prostate. The recognition of the role of α_1 -adrenergic receptor on neuromuscular function in the prostate is the basis for use of α_1 -adrenergic-receptor blockers in treating BPH. A third component, detrusor instability and impaired bladder contractility, may contribute to the symptoms of BPH independent of the outlet obstruction created by an enlarged prostate (see Chapter 25).^{29,30} It has been suggested that some of the symptoms of BPH might be related to a decompensating or aging bladder, rather than being primarily related to outflow obstruction. An example is the involuntary contraction that results in urgency and an attempt to void that occurs because of small bladder volume.29

Clinical Course. The clinical significance of BPH resides in its tendency to compress the urethra and cause partial or complete obstruction of urinary outflow. As the obstruction increases, acute retention may occur with overdistention of the bladder. The residual urine in the bladder causes increased frequency of urination and a constant desire to empty the bladder, which be-

comes worse at night. With marked bladder distention, overflow incontinence may occur with the slightest increase in intra-abdominal pressure. The resulting obstruction to urinary flow can give rise to urinary tract infection, destructive changes of the bladder wall, hydroureter, and hydronephrosis. Hypertrophy and changes in bladder wall structure develop in stages. Initially, the hypertrophied fibers form trabeculations and then herniations, or sacculations; finally, diverticula develop as the herniations extend through the bladder wall (Chapter 25, Fig. 25-4). Because urine seldom is completely emptied from them, these diverticula are readily infected. Back pressure on the ureters and collecting system of the kidneys promotes hydroureter, hydronephrosis, and eventual renal failure.

It is now thought that the single most important factor in the evaluation and treatment of BPH is the man's own experiences related to the disorder. The American Urological Society Symptom Index consists of seven questions about symptoms regarding incomplete emptying, frequency, intermittency, urgency, weak stream, straining, and nocturia.^{31,32} Each question is rated with a score of 0 (mild) to 7 (severe). A maximum score of 35 indicates severe symptoms. Total scores below 7 are considered mild; those between 8 and 20, moderate; and scores over 20, severe. A final question relates to quality of life related to urinary problems.

Diagnosis of BPH is based on history, physical examination, digital rectal examination, urinalysis, blood tests for serum creatinine and prostate-specific antigen (PSA), and urine flow rate. Blood and urine analyses are used as adjuncts to determine BPH complications. Urinalysis is done to detect bacteria, white blood cells, or microscopic hematuria in the presence of infection and inflammation. The serum creatinine test is used as an estimate of the glomerular filtration rate and kidney function. The PSA test is used to screen for prostatic cancer. These evaluation measures, along with the symptom index, are used to describe the extent of obstruction, determine if other diagnostic tests are needed, and establish the need for treatment.

Transabdominal or transrectal diagnostic ultrasonography can be used to evaluate the kidneys, ureters, and bladder. Abdominal radiographs may be used to reveal the size of the gland. Urethrocystoscopy is indicated in men with a history of hematuria, stricture disease, urethral injury, or prior lower urinary tract surgery. It is used to evaluate the length and diameter of the urethra, the size and configuration of the prostate, and bladder capacity. It also detects the presence of trabeculations, bladder stones, and small bladder cancers. CT scans, MRI studies, and radionuclide scans are reserved for rare instances of tumor detection.

Treatment of BPH is determined by the degree of symptoms that the condition produces and complications due to obstruction. When a man develops mild symptoms related to BPH, a "watch and wait" stance often is taken. The condition does not always run a predictable course; it may remain stable or even improve. Until the 1980s, surgery was the mainstay of treatment to alleviate urinary obstruction due to BPH. Currently, there is an emphasis on less invasive methods of treatment, including use of pharmacologic agents. However, when more severe signs of obstruction develop, surgical treatment (*e.g.*, transurethral resection of the prostate [TURP]) usually is indicated to provide comfort and avoid serious renal damage. For men who have heart or lung disease or a condition that precludes major surgery, a stent may be used to widen and maintain the patency of the urethra.

Pharmacologic management includes the use of 5α -reductase inhibitors and α_1 -adrenergic–blocking drugs.^{28,29} The 5α reductase inhibitors such as finasteride reduce prostate size by blocking the effect of androgens on the prostate. The presence of α -adrenergic receptors in prostatic smooth muscle has prompted the use of α_1 -adrenergic–blocking drugs to relieve prostatic obstruction and increase urine flow.

Cancer of the Prostate

Prostatic cancer is the most common male cancer in the United States and is second to lung cancer as a cause of cancer-related death in men. The American Cancer Society estimates that during 2002, approximately 189,000 men in the United States received a diagnosis of prostate cancer, and 32,200 men died of the disorder.³³ The increase in diagnosed cases is thought to reflect earlier diagnosis because of the widespread use of PSA testing since the early 1990s.³⁴ The incidence of prostate cancer varies markedly from country to country and varies among races in the same country.³³ African-American men have the highest reported incidence for prostate cancer at all ages, and Asians and Native American men have the lowest rate. Prostate cancer also is a disease of aging. The incidence increases rapidly after 50 years of age; more than 80% of all prostate cancers are diagnosed in men older than 65 years of age.³³

The precise cause of prostatic cancer is unclear. As with other cancers, it appears that the development of prostate cancer is a multistep process involving genes that control cell differentiation and growth (see Chapter 5). Several risk factors, such as age, race, heredity, and environmental influences, are suspected of playing a role.^{35,36} Male hormone levels also may play a role. There is insufficient evidence linking socioeconomic status, infectious agents, smoking, vasectomy, sexual behavior, or BPH to the pathogenesis of prostate cancer.

The incidence of prostate cancer appears to be higher in relatives of men with prostate cancer. Diet may also play a role. It has been suggested that a diet high in fats may alter the production of sex hormones and increase the risk of prostate cancer. Supporting the role of dietary fats as a risk factor for prostate cancer has been the observation that the diet of Japanese men, who have a low rate of prostate cancer, is much lower in fat content than that of U.S. men, who have a much higher incidence.

In terms of hormonal influence, androgens are believed to play a role in the pathogenesis of prostate cancer.³⁵ Evidence favoring a hormonal influence includes the presence of steroid receptors in the prostate, the requirement of sex hormones for normal growth and development of the prostate, and the fact that prostate cancer almost never develops in men who have been castrated. The response of prostatic cancer to estrogen administration or androgen deprivation further supports a correlation between the disease and testosterone levels.

Prostatic adenocarcinomas, which account for 98% of all primary prostatic cancers, are commonly multicentric and located in the peripheral zones of the prostate³⁶ (see Fig. 33-12). The high frequency of invasion of the prostatic capsule by adenocarcinoma relates to its subcapsular location. Invasion of the urinary bladder is less common and occurs later in the clinical course. Metastasis to the lung reflects lymphatic spread through the thoracic duct and dissemination from the prostatic venous plexus to the inferior vena cava. Bony metastases, particularly to the vertebral column, ribs, and pelvis, produce pain that often presents as a first sign of the disease.

Most men with early-stage prostate cancer are asymptomatic. The presence of symptoms often suggests locally advanced or metastatic disease. Depending on the size and location of prostatic cancer at the time of diagnosis, there may be changes associated with the voiding pattern similar to those found in BPH. These include urgency, frequency, nocturia, hesitancy, dysuria, hematuria, or blood in the ejaculate. On physical examination, the prostate is nodular and fixed. Bone metastasis often is characterized by low back pain. Pathologic fractures can occur at the site of metastasis. Men with metastatic disease may experience weight loss, anemia, or shortness of breath.

Screening. Because early cancers of the prostate usually are asymptomatic, screening tests are important. The screening tests currently available are digital rectal examination, PSA testing, and transrectal ultrasonography. PSA is a glycoprotein secreted into the cytoplasm of benign and malignant prostatic cells that is not found in other normal tissues or tumors. However, a positive PSA test indicates only the possible presence of prostate cancer. It also can be positive in cases of BPH and prostatitis. The American Cancer Society and the American Urological Association recommend that men 50 years of age or older should undergo annual measurement of PSA and rectal examination for early detection of prostate cancer.³³ Men at high risk for prostate cancer, such as blacks and those with a strong family history, should undergo annual screening beginning at 45 years of age.³³

Diagnosis. The diagnosis of prostate cancer is based on history and physical examination and confirmed through biopsy methods. Transrectal ultrasonography, a continuously improving method of imaging, is used to guide a biopsy needle and document the exact location of the biopsied tissue. It also is used for providing staging information.²⁸ Newly developed small probes for transrectal MRI have been shown to be effective in detecting the presence of cancer in the prostate. Radiologic examination of the bones of the skull, ribs, spine, and pelvis can be used to reveal metastases, although radionuclide bone scans are more sensitive. Excretory urograms are used to delineate changes due to urinary tract obstruction and renal involvement.

Cancer of the prostate, like other forms of cancer, is graded and staged (see Chapter 5). Prostatic adenocarcinoma commonly is classified using the Gleason grading system.^{35,36} Welldifferentiated tumors are assigned a grade of 1, and poorly differentiated tumors a grade of 5. Two tumor markers, PSA and serum acid phosphatase, are important in the staging and management of prostatic cancer. In untreated cases, the level of PSA correlates with the volume and stage of disease.³⁷ A rising PSA after treatment is consistent with progressive disease, whether it is locally recurring or metastatic. Measurement of PSA is used to detect recurrence after total prostatectomy. Because the prostate is the source of PSA, levels of the antigen should drop to zero after surgery; a rising PSA indicates recurring disease. Serum acid phosphatase is less sensitive than PSA and is used less frequently. However, it is more predictive of metastatic disease and may be used for that purpose.

Treatment. Cancer of the prostate is treated by surgery, radiation therapy, and hormonal manipulations.^{28,38,39} Chemotherapy has shown limited effectiveness in the treatment of prostate cancer. Treatment decisions are based on tumor grade and stage and on the age and health of the man. Expectant therapy (watching and waiting) may be used if the tumor is not producing symptoms, is expected to grow slowly, and is small and contained in one area of the prostate. This approach is particularly suited for men who are elderly or have other health problems. Most men with an anticipated survival greater than 10 years are considered for surgical or radiation therapy.³⁹ Radical prostatectomy involves complete removal of the seminal vesicles, prostate, and ampullae of the vas deferens. Radiation therapy can be delivered by a variety of techniques, including external beam radiation therapy and transperineal implantation of radioisotopes.

Metastatic disease often is treated with androgen deprivation therapy. Androgen deprivation may be induced at several levels along the pituitary-gonadal axis using a variety of methods or agents. Orchiectomy or estrogen therapy often is effective in reducing symptoms and extending survival. The GnRH analogs (*e.g.*, leuprolide, buserelin, nafarelin) block luteinizing hormone release from the pituitary and reduce testosterone levels without orchiectomy or estrogen therapy. When given continuously and in therapeutic doses, these drugs desensitize GnRH receptors in the pituitary, thereby preventing the release of luteinizing hormone. The antiandrogens block the uptake and actions of androgens in the target tissues. Complete androgen blockade can be achieved by combining an antiandrogen with a GnRH agent or orchiectomy.

In summary, disorders of the penis include balanitis, an acute or chronic inflammation of the glans penis, and balanoposthitis, an inflammation of the glans and prepuce. Peyronie's disease is characterized by the growth of a band of fibrous tissue on top of the penile shaft. Priapism is prolonged, painful, and nonsexual erection that can lead to thrombosis with ischemia and necrosis of penile tissue. Cancer of the penis accounts for less than 1% of male genital cancers in the United States. Although the tumor is slow growing and highly curable when diagnosed early, the greatest hindrance to successful treatment is a delay in seeking medical attention.

Disorders of the scrotum and testes include hydrocele, hematocele, spermatocele, varicocele, and testicular torsion. Inflammatory conditions can involve the scrotal sac, epididymis, or testes. Tumors can arise in the scrotum or the testes. Scrotal cancers usually are associated with exposure to petroleum products such as tar, pitch, and soot. Testicular cancers account for 1% of all male cancers and 3% of cancers of the male genitourinary system. With current treatment methods, a large percentage of men with these tumors can be cured. Testicular self-examination is recommended as a means of early detection of this form of cancer.

The prostate is a firm, glandular structure that surrounds the urethra. Inflammation of the prostate occurs as an acute or a chronic process. Chronic prostatitis probably is the most common cause of relapsing urinary tract infections in men. BPH is a common disorder in men older than 50 years. Because the prostate encircles the urethra, BPH exerts its effect through obstruction of urinary outflow from the bladder. Advances in the treatment of BPH include laser surgery, balloon dilatation, prostatic stents, and pharmacologic treatment.

Prostatic cancer is the most common male cancer in the United States and is second to lung cancer as a cause of cancer-related death in men. A recent increase in diagnosed cases is thought to reflect earlier diagnosis because of widespread use of PSA testing. Most prostate cancers are asymptomatic and are incidentally discovered on rectal examination. Cancer of the prostate, like other forms of cancer, is graded according to the histologic characteristics of the tumor and staged clinically using the TNM system. Treatment, which is based on the extent of the disease, includes surgery, radiation therapy, and hormonal manipulation.

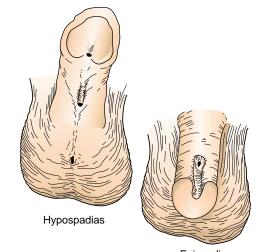
DISORDERS IN CHILDHOOD AND AGING CHANGES

Disorders of Childhood

Disorders of the male reproductive system that present in childhood include hypospadias, epispadias, phimosis and paraphimosis, and cryptorchidism.

Hypospadias and Epispadias

Hypospadias and epispadias are congenital disorders of the penis resulting from embryologic defects in the development of the urethral groove and penile urethra (Fig. 33-13). In hypospadias, which affects approximately 1 in 300 male infants, the termination of the urethra is on the ventral surface of the penis.^{34,40} The testes are undescended in 10% of boys born with hypospadias and chordee (*i.e.*, ventral bowing of the penis), and inguinal hernia also may accompany the disorder. In the newborn with severe hypospadias and cryptorchidism (undescended testes), the differential diagnosis should consider ambiguous genitalia and masculinization



Epispadias

FIGURE 33-13 Hypospadias and epispadias.

that is seen in females with congenital adrenal hyperplasia (see Chapter 31). Because many chromosomal aberrations result in ambiguity of the external genitalia, chromosomal studies often are recommended for male infants with hypospadias and cryptorchidism.³⁴

Surgery is the treatment of choice for hypospadias.³⁴ Circumcision is avoided because the foreskin is used for surgical repair. Factors that influence the timing of surgical repair include anesthetic risk, penile size, and the psychological effects of the surgery on the child. In mild cases, the surgery is done for cosmetic reasons only. In more severe cases, surgical repair becomes essential for normal sexual functioning and to prevent the psychological effects of having malformed genitalia. When indicated, surgical repair is usually done between the ages of 6 to 12 months.

Epispadias, in which the opening of the urethra is on the dorsal surface of the penis, is a less common defect. Although epispadias may occur as a separate entity, it often is associated with exstrophy of the bladder, a condition in which the abdominal wall fails to cover the bladder. The treatment depends on the extent of the developmental defect.

Phimosis and Paraphimosis

Phimosis refers to a tightening of the prepuce or penile foreskin that prevents its retraction over the glans. Embryologically, the foreskin begins to develop during the eighth week of gestation as a fold of skin at the distal edge of the penis that eventually grows forward over the base of the glans.⁴⁰ By the 16th week of gestation, the prepuce and the glans are adherent. Only a small percentage of newborns have a fully retractable foreskin. With growth, a space develops between the glans and foreskin, and by 3 years of age, approximately 90% of male children have retractable foreskins.

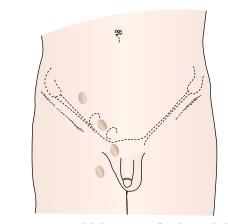
Because the foreskin of many boys cannot be fully retracted in early childhood, it is important that the area be cleaned thoroughly. There is no need to retract the foreskin forcibly because this could lead to infection, scarring, or paraphimosis. As the child grows, the foreskin becomes retractable, and the glans and foreskin should be cleaned routinely. If symptomatic phimosis occurs after childhood, it can cause difficulty with voiding or sexual activity. Circumcision is then the treatment of choice.

In a related condition called *paraphimosis*, the foreskin is so tight and constricted that it cannot cover the glans. A tight foreskin can constrict the blood supply to the glans and lead to ischemia and necrosis. Many cases of paraphimosis result from the foreskin being retracted for an extended period, as in the case of catheterized uncircumcised males.

Cryptorchidism

Cryptorchidism, or undescended testes, occurs when one or both of the testicles fail to move down into the scrotal sac. The condition is bilateral in 10% to 20% of cases. The testes develop intra-abdominally in the fetus and usually descend into the scrotum through the inguinal canal during the seventh to ninth months of gestation.⁴⁰ The undescended testes may remain in the lower abdomen or at a point of descent in the inguinal canal (Fig. 33-14).

The incidence of cryptorchidism is directly related to birth weight and gestational age; infants who are born prematurely



■ FIGURE 33-14 ■ Possible locations of undescended testicles.

or are small for gestational age have the highest incidence of the disorder. Up to one third of premature infants and 3% to 5% of full-term infants are born with undescended testicles.^{35,41} The cause of cryptorchidism in full-term infants is poorly understood. Most cases are idiopathic, but some may result from genetic or hormonal factors.²⁹

The major manifestation of cryptorchidism is the absence of one or more of the testes in the scrotum. The testis either is not palpable or can be felt external to the inguinal ring. Spontaneous descent often occurs during the first 3 months of life, and by 6 months of age the incidence decreases to 0.8%.^{34,41} Spontaneous descent rarely occurs after 6 months of age.

In children with cryptorchidism, histologic abnormalities of the testes reflect intrinsic defects in the testicle or adverse effects of the extrascrotal environment. The undescended testicle is normal at birth, but pathologic changes can be demonstrated at 6 to 12 months.³⁴ There is a delay in germ cell development, changes in the spermatic tubules, and reduced number of Leydig cells. These changes are progressive if the testes remain undescended. When the disorder is unilateral, it also may produce morphologic changes in the contralateral descended testis.

The consequences of cryptorchidism include infertility, malignancy, and the possible psychological effects of an empty scrotum. Indirect inguinal hernias usually accompany the undescended testes but rarely are symptomatic. Recognition of the condition and early treatment are important steps in preventing adverse consequences. The risk of malignancy in the undescended testis is four to six times higher than in the general population.^{34,41} The increased risk of testicular cancer is not significantly affected by orchiopexy, hormonal therapy, or late spontaneous descent after the age of 2 years. However, orchiopexy does allow for earlier detection of a testicular malignancy by positioning the testis in a more easily palpable location.

As a group, males with unilateral or bilateral cryptorchidism usually have decreased sperm counts, poorer-quality sperm, and lower fertility rates than do men whose testicles descend normally. The likelihood of decreased fertility increases when the condition is bilateral. Unlike the risk of testicular cancer, there seems to be some advantage to early orchiopexy for protection of fertility.^{34,41}

Diagnosis is based on careful examination of the genitalia in male infants. Undescended testes due to cryptorchidism should be differentiated from retractable testes that retract into the inguinal canal in response to an exaggerated cremaster muscle reflex. Retractable testes usually are palpable at birth but become nonpalpable later. They can be brought down with careful palpation in a warm room. Retractable testes usually assume a scrotal position during puberty. They have none of the complications associated with undescended testicles due to cryptorchidism.³⁴

Improved techniques for testicular localization include ultrasonography (*i.e.*, visualization of the testes by recording the pulses of ultrasonic waves directed into the tissues), gonadal venography and arteriography (*i.e.*, radiography of the veins and arteries of the testes after the injection of a contrast medium), and laparoscopy (*i.e.*, examination of the interior of the abdomen using a visualization instrument).

The treatment goals for the child with cryptorchidism include measures to enhance future fertility potential, placement of the gonad in a favorable place for cancer detection, and improved cosmetic appearance. Regardless of the type of treatment used, it should be carried out between 6 months and 2 years of age.34,41 Treatment modalities for children with unilateral or bilateral cryptorchidism include initial hormone therapy with hCG or luteinizing hormone-releasing hormone (LHRH), a hypothalamic hormone that stimulates production of the gonadotropic hormones by the anterior pituitary gland. For children who do no respond to hormonal treatment, surgical placement and fixation of the testes in the scrotum (*i.e.*, orchiopexy) have proved effective. Approximately 95% of infants who have orchiopexy for a unilateral undescended testis will be fertile, compared with a 30% to 50% fertility rate in uncorrected males.41

Treatment of males with undescended testis should include lifelong follow-up, considering the sequelae of testicular cancer and infertility. Parents need to be aware of the potential issues of infertility and increased risk of testicular cancer. On reaching puberty, boys should be instructed in the necessity of testicular self-examination.

Aging Changes

Like other body systems, the male reproductive system undergoes degenerative changes as a result of the aging process; it becomes less efficient with age. The declining physiologic efficiency of male reproductive function occurs gradually and involves the endocrine, circulatory, and neuromuscular systems.⁴² Compared with the marked physiologic change in aging females, the changes in the aging male are more gradual and less drastic. Gonadal and reproductive failure usually are not related directly to age because a man remains fertile into advanced age; 80- and 90-year-old men have been known to father children.

As the male ages, his reproductive system becomes measurably different in structure and function from that of the younger male. Male sex hormone levels, particularly of testosterone, decrease with age, with the decline starting later on average than in women. The term *andropause* has been used to describe an illdefined collection of symptoms in aging men, typically those older than 50 years, who may have a low androgen level.⁴³

The sex hormones play a part in the structure and function of the reproductive system and other body systems from conception to old age; they affect protein synthesis, salt and water balance, bone growth, and cardiovascular function. Decreasing levels of testosterone affect sexual energy, muscle strength, and the genital tissues. The testes become smaller and lose their firmness. The seminiferous tubules, which produce spermatozoa, thicken and begin a degenerative process that finally inhibits sperm production, resulting in a decrease of viable spermatozoa. The prostate gland enlarges, and its contractions become weaker. The force of ejaculation decreases because of a reduction in the volume and viscosity of the seminal fluid. The seminal vesicle changes little from childhood to puberty. The pubertal increases in the fluid capacity of the gland remain throughout adulthood and decline after the age of 60 years. After age 60 years, the walls of the seminal vesicles thin, the epithelium decreases, and the muscle layer is replaced by connective tissue. Age-related changes in the penis consist of fibrotic changes in the trabeculae in the corpus spongiosum, with progressive sclerotic changes in arteries and veins. Sclerotic changes also follow in the corpora cavernosa, with the condition becoming generalized in 55- to 60-year-old men.

Erectile dysfunction in the elderly male often is directly related to the general physical condition of the person. Diseases that accompany aging can have direct bearing on male reproductive function. Various cardiovascular, respiratory, hormonal, neurologic, and hematologic disorders can be responsible for secondary impotence. For example, vascular disease affects male potency because it may impair blood flow to the pudendal arteries or their tributaries, resulting in loss of blood volume with subsequent poor distention of the vascular spaces of erectile tissue. Other diseases affecting potency include hypertension, diabetes, cardiac disease, and malignancies of the reproductive organs. In addition, certain medications can have an effect on sexual function.

Testosterone and other synthetic androgens may be used in older males with low androgen levels to improve muscle strength and vigor. Preliminary studies of androgen replacement in aging males with low androgen levels show an increase in lean body mass and a decrease in bone turnover. Before testosterone replacement therapy is initiated, all men should be screened for prostate cancer. Testosterone is available as an injectable form that is administered every 2 to 3 weeks or as a transdermal patch or gel. Side effects of replacement therapy may include acne, gynecomastia, and reduced HDL levels. It also may contribute to a worsening of sleep apnea in men who are troubled by this problem.

In summary, childhood disorders of the male reproductive system include congenital disorders in which the urethral opening is located on the ventral surface of the penis (hypospadias) or on the dorsal surface (epispadias). Phimosis is the condition in which the opening of the foreskin is too tight to permit retraction over the glans. Disorders of the scrotum and testes include cryptorchidism or undescended testicles. Early diagnosis and treatment is important because of the risk of malignancy and infertility. Like other body systems, the male reproductive system undergoes changes as a result of the aging process. The changes occur gradually and involve parallel changes in endocrine, circulatory, and neuromuscular function. Testosterone levels decrease, the size and firmness of the testes decrease, sperm production declines, and the prostate gland enlarges. There usually is a decrease in frequency of intercourse, intensity of sensation, speed of attaining erection, and force of ejaculation.

REVIEW QUESTIONS

Describe the structure and function of the male reproductive system and hormones and relate to the process of spermatogenesis.

Relate the development and descent of testes to the pathogenesis of inguinal hernia and cryptorchidism.

Compare the pathophysiology of cancer of the penis, scrotum, and testes.

Compare the cause, appearance, and significance of hydrocele, hematocele, spermatocele, and varicocele.

State the difference between extravaginal and intravaginal testicular torsion.

Compare the pathology and symptoms of acute bacterial prostatitis, chronic bacterial prostatitis, and chronic prostatitis/pelvic pain syndrome.

Describe the urologic manifestations and treatment of benign prostatic hyperplasia.

List the methods used in the diagnosis and treatment of prostatic cancer.

Describe the autonomic nervous system control of erection, emission, and ejaculation and relate to the pathogenesis of erectile dysfunction and priapism.

State the difference between hypospadias and epispadias.

Describe changes in the male reproductive system that occur with aging.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- Guyton A.C., Hall J.E. (2001). *Textbook of medical physiology* (10th ed., p. 913). Philadelphia: W.B. Saunders.
- Braunstein G.D. (2001). Testes. In Greenspan F.S., Gardner D.G. (Eds.), *Basic and clinical endocrinology* (6th ed., pp. 422–452). New York: Lange Medical Books/McGraw-Hill.
- Anderson K.E., Wagner G. (1995). Physiology of penile erection. Physiology Review 75, 191–236.
- 4. NIH Consensus Development Panel on Impotence. (1993). NIH Consensus Conference: Impotence. *Journal of the American Medical Association* 270, 83–90.

- 5. Lue T.F. (2000). Erectile dysfunction. *New England Journal of Medicine* 342, 1802–1813.
- Levine L.A. (2000). Diagnosis and treatment of erectile dysfunction. *American Journal of Medicine* 109 (Suppl. 9A), 3S–12S.
- 7. Kaiser F.E. (1999). Erectile dysfunction in the aging man. *Medical Clinics of North America* 83, 1267–1278.
- 8. Vohra S., Badlani G. (1992). Balanitis and balanoposthitis. Urologic Clinics of North America 19, 143–147.
- Fitkin J., Ho G.T. (1999). Peyronie's disease: Current management. *American Family Physician* 60, 549–554.
- McAninch J.W. (2000). Disorders of the penis and male urethra. In Tanagho E.A., McAninch J.W. (Eds.), *Smith's general urology* (15th ed., pp. 663–675). New York: Lange Medical Books/McGraw-Hill.
- 11. Harmon W.J., Nehra A. (1997). Priapism: Diagnosis and treatment. *Mayo Clinic Proceedings* 72, 350–355.
- Kachhi P.N., Henderson S.O. (2000). Priapism after androstenedione intake for athletic performance enhancement. *Annals of Emergency Medicine* 35, 391–393.
- 13. Krieg R., Hoffman R. (1999). Current management of unusual genitourinary cancers: Part 1. Penile cancer. *Oncology* 13, 1347–1352.
- Presti J.C., Herr H.W. (2000). Genital tumors. In Tanagho E.A., McAninch J.W. (Eds.), *Smith's general urology* (15th ed., pp. 430–434). New York: Lange Medical Books/McGraw-Hill.
- Kapur P., Caty M.G., Glick P.L. (1998). Pediatric hernias and hydroceles. *Pediatric Clinics of North America* 45, 773–789.
- McAninch J.W. (2000). Disorders of the testis, scrotum, and spermatic cord. In Tanagho E.A., McAninch J.W. (Eds.), *Smith's general urology* (15th ed., pp. 684–693). New York: Lange Medical Books/ McGraw-Hill.
- Galejs L.E., Kass E.J. (1999). Diagnosis and treatment of acute scrotum. American Family Physician 59, 817–824.
- Meares E.M. (2000). Nonspecific infections of the genitourinary tract. In Tanagho E.A., McAninch J.W. (Eds.), *Smith's general urology* (15th ed., pp. 237–238). Norwalk, CT: Appleton & Lange.
- Mebcow M.M. (1975). Percivall Pott (1713–1788): 200th anniversary of first report of occupation-induced cancer of the scrotum in chimney sweepers (1745). Urology 6, 745.
- 20. Lowe F.C. (1992). Squamous cell carcinoma of the scrotum. Urologic Clinics of North America 19, 297–405.
- Bosl G.J. (1997). Testicular germ-cell cancer. New England Journal of Medicine 337, 242–252.
- 22. Pillai S.B., Besner G.E. (1998). Pediatric testicular problems. *Pediatric Clinics of North America* 45, 813–818.
- Kinade S. (1999). Testicular cancer. American Family Physician 59, 2539–2544.
- Krieger J.N., Nyberg L., Nickel J.C. (1999). NIH consensus definition and classification of prostatitis. *Journal of the American Medical Association* 282, 721–725.
- McRae S.N., Shortliffe L.M.D. (2000). Bacterial infections of the genitourinary system. In Tanagho E.A., McAninch J.W. (Eds.), *Smith's* general urology (15th ed., pp. 254–259). New York: Lange Medical Books/McGraw-Hill.
- Stevermer J.J., Easley S.K. (2000). Treatment of prostatitis. American Family Physician 61, 3015–3026.
- Collins M.M., MacDonald R., Wilt T.J. (2000). Diagnosis and treatment of chronic abacterial prostatitis: A systemic review. *Annals of Internal Medicine* 133, 367–381.32.
- Presti J.C. (2000). Neoplasms of the prostate gland. In Tanagho E.A., McAninch J.W. (Eds.), *Smith's general urology* (15th ed., pp. 399–406). New York: Lange Medical Books/McGraw-Hill.
- 29. Zida A., Rosenblum M., Crawford E.D. (1999). Benign prostatic hyperplasia: An overview. *Urology* 53 (Suppl. 3A), 1–6.
- Elbadawi A. (1998). Voiding dysfunction in benign prostatic hyperplasia: Trends, controversies and recent revelations: Pathology and pathophysiology. *Urology* 51 (Suppl. 5A), 73–82.
- 31. Agency of Health Care Policy and Research. (1994). *Clinical practice guidelines for benign prostatic hyperplasia*. AHCPR publication

no. 94-0582. Rockville, MD: U.S. Department of Health and Human Services.

- Barry M.J., et al. (1992). The American Urological Association index of benign prostatic hypertrophy. *Journal of Urology* 148, 1549–1557.
- American Cancer Society. (2002). Prostate cancer resource center. [On-line]. Available: http://www.cancer.org.
- Behrman R.E., Kliegman R.M., Jenson H.B. (2000). Nelson textbook of pediatrics (16th ed., pp. 1546–1549, 1650–1651). Philadelphia: W.B. Saunders.
- 35. Cotran R.S., Kumar V., Collins T. (1999). *Robbins pathologic basis of disease* (6th ed., pp. 1015–1033). Philadelphia: W.B. Saunders.
- Peterson R.O. (1999). The urinary tract and male reproductive system. In Rubin E., Farber J.L. (Eds.), *Pathology* (3rd ed., pp. 956–960). Philadelphia: Lippincott Williams & Wilkins.
- Brawer M.K. (1999). Prostate-specific antigen: Current status. CA— A Cancer Journal for Clinicians 49, 264–265.

- Garnick M.B., Fair W.R. (1998). Combating prostate cancer. Scientific American 279 (6), 74–83.
- Stoller M.L., Presti J.C., Carroll P.R. (2001). Urology. In Tierney L.M., McPhee S.J., Papadakis M.A. (Eds.), *Current medical diagnosis and treatment* (40th ed., pp. 956–961). New York: Lange Medical Books/ McGraw-Hill.
- Moore K.L., Persaud T.V.N. (1998). The developing human: Clinically oriented embryology (6th ed., pp. 338–339). Philadelphia: W.B. Saunders.
- Docimo S.G., Silver R.I., Cromie W. (2000). The undescended testicle: Diagnosis and management. *American Family Physician* 62, 2037–2048.
- Merry B.J., Holehan A.M. (1994). Aging of the male reproductive system. In Timinas P.S. (Ed.), *Physiological basis of aging and geriatrics* (2nd ed., pp. 171–178). Boca Raton, FL: CRC Press.
- Morley J.E., Perry H.M. (1999). Androgen deficiency in aging men. Medical Clinics of North America 83, 1279–1289.

Nine

Alterations in the Male and Female Reproductive Systems

CHAPTER 33

Alterations in the Male Reproductive System

Physiologic Basis of Male Reproductive Function **Spermatogenesis** Hormonal Control of Male Reproductive Function Action of the Hypothalamic and Anterior **Pituitary Hormones Neural Control of Sexual Function Erectile Dysfunction** Disorders of the Penis, the Scrotum and Testes, and the Prostate **Disorders of the Penis** Inflammation and Infection Peyronie's Disease **Priapism** Cancer of the Penis **Disorders of the Scrotum and Testes** Disorders of the Testicular Tunica **Testicular Torsion** Inflammation and Infection Cancer of the Scrotum and Testes **Disorders of the Prostate Prostatitis Benign Prostatic Hyperplasia** Cancer of the Prostate

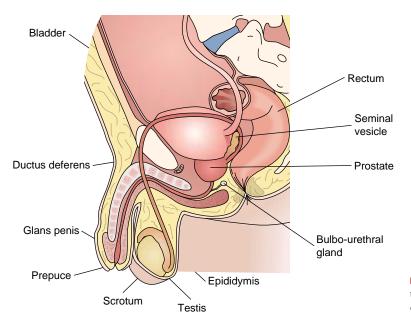
Disorders in Childhood and Aging Changes Disorders of Childhood Hypospadias and Epispadias Phimosis and Paraphimosis Cryptorchidism Aging Changes

he male genitourinary system is subject to structural defects, inflammation, and neoplasms, all of which can affect urine elimination, sexual function, and fertility. This chapter focuses on spermatogenesis and hormonal control of male reproductive function; neural control of sexual function and erectile dysfunction; disorders of the penis, scrotum, testes, and prostate; disorders of the male reproductive system in children; and changes in function as a result of the aging process.

PHYSIOLOGIC BASIS OF MALE REPRODUCTIVE FUNCTION

The male genitourinary system is composed of the paired gonads, or testes, genital ducts, accessory organs, and penis (Fig. 33-1). The dual function of the testes is to produce male sex androgens (*i.e.*, male sex hormones), mainly testosterone, and spermatozoa (*i.e.*, male germ cells). The internal accessory organs produce the fluid constituents of semen, and the ductile system aids in the storage and transport of spermatozoa. The penis functions in urine elimination and sexual function.

582 Unit Nine: Alterations in the Male and Female Reproductive Systems



■ FIGURE 33-1 ■ The structures of the male reproductive system, including the testes, the scrotum, and the excretory ducts.

Spermatogenesis

Spermatogenesis refers to the generation of spermatozoa or sperm. It begins at an average age of 13 years and continues throughout the reproductive years of a man's life. Spermatogenesis occurs in the seminiferous tubules of the testes (Fig. 33-2). Internally, the testes are composed of several hundred compartments or lobules. Each lobule contains one or more coiled seminiferous tubules. The outer layer of the seminiferous tubules is made up of connective tissue and smooth muscle; the inner lining is composed of Sertoli's cells, which are embedded with sperm in

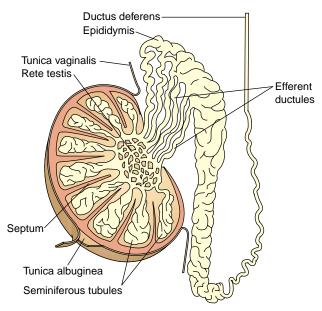
KEY CONCEPTS

MALE REPRODUCTIVE SYSTEM

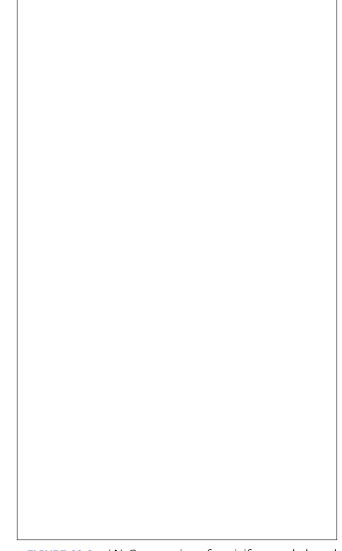
- The male genitourinary system functions in both urine elimination and reproduction.
- The testes function in both production of male germ cells (spermatogenesis) and secretion of the male sex hormone, testosterone.
- The ductile system (epididymides, vas deferens, and ejaculatory ducts) transports and stores sperm and assists in their maturation, and the accessory glands (seminal vesicles, prostate gland, and bulbourethral glands) prepare the sperm for ejaculation.
- Sperm production requires temperatures that are 2° to 3°C below body temperature. The position of the testes in the scrotum and the unique blood flow-cooling mechanisms provide this environment.
- The urethra, which is enclosed in the penis, is the terminal portion of the male genitourinary system. Because it conveys both urine and semen, it serves both urinary and reproductive functions.

various stages of development (Fig. 33-3). Sertoli's cells secrete a special fluid that contains nutrients to bathe and nourish the immature germ cells; they provide digestive enzymes that play a role in spermiation (*i.e.*, converting the spermatocytes to sperm); and they are thought to play a role in shaping the head and tail of the sperm. Sertoli's cells also secrete several hormones, including the principal feminizing sex hormone, estradiol, which seems to be required in the male for spermatogenesis, and inhibin, which controls the function of Sertoli's cells through feedback inhibition of follicle-stimulating hormone (FSH) from the anterior pituitary gland.

After the spermatozoa are formed in the seminiferous tubules, they travel through the efferent tubules to the epididymis, which is the final site for sperm maturation. Because the spermatozoa are not motile at this stage of development,



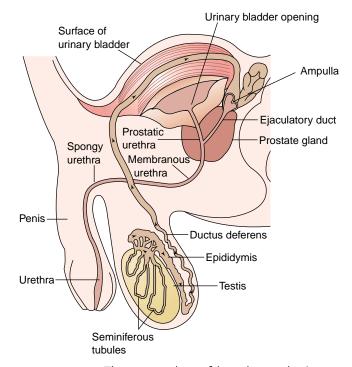
■ FIGURE 33-2 ■ The parts of the testes and epididymis.



■ FIGURE 33-3 ■ (A) Cross-section of seminiferous tubule and (B) stages of development of spermatozoa.

peristaltic movements of the ductal walls of the epididymis aid in their movement.¹ The spermatozoa continue their migration through the ductus deferens, also called the *vas deferens*. The ampulla of the vas deferens serves as a storage reservoir for sperm. Sperm are stored in the ampulla until they are released through the penis during ejaculation (Fig. 33-4). Spermatozoa can be stored in the genital ducts for as long as 42 days and still maintain their fertility. Surgical disconnection of the vas deferens in the scrotal area (*i.e.*, vasectomy) serves as an effective method of male contraception. Because sperm are stored in the ampulla, men can remain fertile for 4 to 5 weeks after performance of a vasectomy.

The seminal vesicles, the prostate gland, and the bulbourethral glands form the accessory reproductive structures of the male genitourinary system. The spermatozoa plus the secretions from the genital ducts and accessory organs make up



■ FIGURE 33-4 ■ The excretory ducts of the male reproductive system and the path that sperm follows as it leaves the testis and travels to the urethra.

the semen (from the Latin word meaning seed). The seminal vesicles consist of two highly tortuous tubes that secrete fluid for the semen. Each of the paired seminal vesicles is lined with secretory epithelium containing an abundance of fructose, prostaglandins, and several other proteins. The fructose secreted by the seminal vesicles provides the energy for sperm motility. The prostaglandins are thought to assist in fertilization by making the cervical mucus more receptive to sperm and by causing reverse peristaltic contractions in the uterus and fallopian tubes to move the sperm toward the ovaries. Each seminal vesicle joins its corresponding vas deferens to form the ejaculatory duct, which enters the posterior part of the prostate and continues through until it ends in the prostatic portion of the urethra. During the emission phase of coitus, each vesicle empties fluid into the ejaculatory duct, adding bulk to the semen. Approximately 70% of the ejaculate originates in the seminal vesicles.

Hormonal Control of Male Reproductive Function

The male sex hormones are called *androgens*. The testes secrete several male sex hormones, including *testosterone*, *dihydrotes*-*tosterone*, and *androstenedione*.² Testosterone, which is the most abundant of these hormones, is considered the main testicular hormone. The adrenal cortex also produces androgens, although in much smaller quantities (<5% of the total male androgens) than those produced in the testes. The testes also secrete small quantities of estradiol and estrone.

Testosterone is produced and secreted by the interstitial Leydig's cells in the testes. It is metabolized in the liver and excreted by the kidneys. In the bloodstream, testosterone exists in a free (unbound) or a bound form. The bound form is attached to plasma proteins, including albumin and the sex hormonebinding protein produced by the liver. Only approximately 2% of circulating testosterone is unbound and therefore able to enter the cell and exert its metabolic effects. Much of the testosterone that becomes fixed to the tissues is converted to dihydrotestosterone, especially in certain target tissues such as the prostate gland. Some of the actions of testosterone depend on this conversion, whereas others do not. Testosterone also can be aromatized or converted to estradiol in the peripheral tissues.

Testosterone exerts a variety of biologic effects in the male (Chart 33-1). In the male embryo, testosterone is essential for the appropriate differentiation of the internal and external genitalia, and it is necessary for descent of the testes in the fetus. Testosterone is essential to the development of primary and secondary male sex characteristics during puberty and for the maintenance of these characteristics during adult life. It causes growth of pubic, chest, and facial hair; it produces changes in the larynx that result in the male bass voice; and it increases the thickness of the skin and the activity of the sebaceous glands, predisposing to acne.

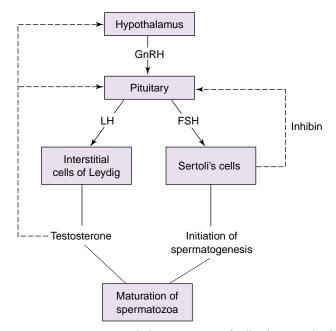
All or almost all of the actions of testosterone and other androgens result from increased protein synthesis in target tissues. Androgens function as anabolic agents in males and females to promote metabolism and musculoskeletal growth. Testosterone and the androgens have a great effect on the development of increasing musculature during puberty, with boys averaging approximately 50% more of an increase in muscle mass than do girls.

Action of the Hypothalamic and Anterior Pituitary Hormones

The hypothalamus and the anterior pituitary gland play an essential role in promoting spermatogenic activity in the testes and maintaining the endocrine function of the testes by means of the gonadotropic hormones. The synthesis and release of the gonadotropic hormones from the pituitary gland are regulated by gonadotropin-releasing hormone (GnRH), which is synthesized by the hypothalamus and secreted into the hypothalamohypophysial portal circulation (Fig. 33-5).

CHART 33-1 Main Actions of Testosterone

- Induces differentiation of the male genital tract during fetal development
- Induces development of primary and secondary sex
 - characteristics
 - Gonadal function
 - External genitalia and accessory organs
- Male voice timbre
- Male skin characteristics
- Male hair distribution
- Anabolic effects
 - Promotes protein metabolism
 - Promotes musculoskeletal growth
- Influences subcutaneous fat distribution
- Promotes spermatogenesis (in FSH-primed tubules) and maturation of sperm



■ **FIGURE 33-5** ■ Hypothalamic-pituitary feedback control of spermatogenesis and testosterone levels in the male. The *dashed line* represents negative feedback.

Two gonadotropic hormones are secreted by the pituitary gland: FSH and luteinizing hormone (LH). In the male, LH also is called interstitial cell-stimulating hormone. The production of testosterone by the interstitial cells of Leydig is regulated by LH. FSH binds selectively to Sertoli's cells surrounding the seminiferous tubules, where it functions in the initiation of spermatogenesis. Under the influence of FSH, Sertoli's cells produce and rogen-binding protein, plasminogen activator, and inhibin. Androgen-binding protein binds testosterone and serves as a carrier of testosterone in Sertoli's cells and as a storage site for testosterone. Although FSH is necessary for the initiation of spermatogenesis, full maturation of the spermatozoa requires testosterone. Androgen-binding protein also serves as a carrier of testosterone from the testes to the epididymis. Plasminogen activator, which converts plasminogen to plasmin, functions in the final detachment of mature spermatozoa from Sertoli's cells.

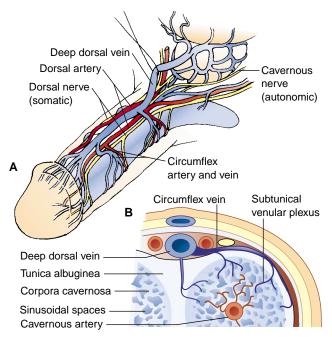
Circulating levels of the gonadotropic hormones are regulated in a negative feedback manner by testosterone. High levels of testosterone suppress LH secretion through a direct action on the pituitary and an inhibitory effect on the hypothalamus. FSH is thought to be inhibited by a substance called *inhibin*, produced by Sertoli's cells. Inhibin suppresses FSH release from the pituitary gland. The pituitary gonadotropic hormones and Sertoli's cells in the testes form a classic negative feedback loop in which FSH stimulates inhibin and inhibin suppresses FSH. Unlike the cyclic hormonal pattern in the female, in the male, FSH, LH, and testosterone secretion and spermatogenesis occur at relatively unchanging rates during adulthood.

Neural Control of Sexual Function

The penis is the male external genital organ through which the urethra passes. It functions both as a sexual organ and as an organ of urine elimination. Anatomically, the external penis consists of a shaft that ends in a tip called the *glans* (Fig. 33-1). The loose skin of the penis shaft folds to cover the glans, forming the prepuce, or foreskin. The glans of the penis contains many sensory nerves, making this the most sensitive portion of the penile shaft. The cylindrical body or shaft of the penis is composed of three masses of erectile tissue held together by fibrous strands and covered with a thin layer of skin (Fig. 33-6). The two lateral masses of tissue are called the *corpora cavernosa*. The third, ventral mass is called the *corpus spongiosum*. The corpora cavernosa and corpus spongiosum are cavernous sinuses that normally are relatively empty but become engorged with blood during penile erection.

The physiology of the male sex act involves a complex interaction between autonomic-mediated spinal cord reflexes, higher neural centers, and the vascular system. It involves erection, emission, ejaculation, and detumescence. Erection involves increased inflow of blood into the corpora cavernosa and penile rigidity. Ejaculation represents the expulsion of the sperm from the urethra. Detumescence, or penile relaxation, results from outflow of blood from the corpora cavernosa.

Erection is a neurovascular process involving the autonomic nervous system, neurotransmitters and endothelial relaxing factors, the vascular smooth muscle of the arteries and veins supplying the penile tissue, and the trabecular smooth muscle of the sinusoids of the corpora cavernosa (Fig. 33-6). The penis is innervated by both the autonomic and somatic nervous systems. In the pelvis, the sympathetic and parasympathetic components of the autonomic nervous system merge to form what are called the *cavernous nerves*.³ Erection is under the control of the parasympathetic nervous system, and ejaculation and detumescence (penile relaxation) are under the control of the sympathetic nervous system. Somatic innervation, which occurs through the pudendal nerve, is responsible for penile sensation and contraction and relaxation of the



■ FIGURE 33-6 ■ Anatomy and mechanism of penile erection. (A) Innervation and arterial and venous blood supply to penis. (B) Cross-section of the sinusoidal system of the corpora cavernosa.

KEY CONCEPTS

DISORDERS OF PENILE ERECTION

- Erection is a neurovascular process involving the autonomic nervous system, the somatic nervous system by way of the pudendal nerve, the vascular system, and the sinusoidal spaces of the corpora cavernosa.
- Parasympathetic innervation through the pelvic nerves initiates relaxation of the trabecular smooth muscle of the corpora cavernosa through the action of nitric oxide, the inflow of arterial blood, and cessation of venous outflow.
- Erectile failure can result from disorders in one or a combination of the neural, vascular, or chemical mediator aspects of the erectile process.

extracorporeal striated muscles (bulbocavernous and ischio-cavernous).³

Penile erection is the first effect of male sexual stimulation, whether psychological or physical. It involves increased inflow of blood into the corpora cavernosa due to relaxation of the trabecular smooth muscle that surrounds the sinusoidal spaces and compression of the veins controlling outflow of blood from the venous plexus. Erection is mediated by parasympathetic impulses that pass from the sacral segments of the spinal cord through the pelvic nerves to the penis. Parasympathetic stimulation results in release of nitric oxide, a nonadrenergic-noncholinergic neurotransmitter, which causes relaxation of trabecular smooth muscle of the corpora cavernosa. This relaxation permits inflow of blood into the sinuses of the cavernosa at pressures approaching those of the arterial system. Because the erectile tissues of the cavernosa are surrounded by a nonelastic fibrous covering, high pressure in the sinusoids causes ballooning of the erectile tissue to such an extent that the penis becomes hard and elongated. At the same time, contraction of the somatic-innervated ischiocavernous muscles forcefully compresses the blood-filled corpora cavernosa, producing a further increase in intercavernous pressures. During this phase of erection, inflow and outflow of blood cease.

Parasympathetic innervation must be intact and nitric oxide synthesis must be active for erection to occur. Nitric oxide activates guanyl cyclase, an enzyme that increases the concentration of cyclic guanosine monophosphate (cGMP), which in turn causes smooth muscle relaxation. Other smooth muscle relaxants (*e.g.*, prostaglandin E_1 analogs and α -adrenergic antagonists), if present in high enough concentrations, can independently cause sufficient cavernosal relaxation to result in erection.³ Many of the drugs that have been developed to treat erectile dysfunction act at the levels of these mediators.

Detumescence or penile relaxation is largely a sympathetic nervous system response. It can result from a cessation of neurotransmitter release, the breakdown of second messengers such as cGMP, or sympathetic discharge during ejaculation. Contraction of the trabecular smooth muscle opens the venous channels so that the trapped blood can be expelled and penile flaccidity return.

Erectile Dysfunction

"Erectile dysfunction is defined as the inability to achieve and maintain an erection sufficient to permit satisfactory sexual intercourse."⁴ It has been estimated that the disorder affects 20 to 30 million men in the United States.⁵ Erectile dysfunction is commonly classified as psychogenic, organic, or mixed psychogenic or organic.^{5,6} The latter is the most common.

Psychogenic causes of erectile dysfunction include performance anxiety, a strained relationship with a sexual partner, depression, and overt psychotic disorders such as schizophrenia. Depression is a common cause of erectile dysfunction.⁵

Organic causes span a wide range of pathologies. They include neurogenic, hormonal, vascular, drug-induced, and cavernous impairment etiologies. Neurogenic disorders such as Parkinson's disease, stroke, and cerebral trauma often contribute to erectile dysfunction by decreasing libido or preventing the initiation of erection. In spinal cord injury, the extent of neural impairment depends on the level, location, and extent of the lesion. Somatosensory involvement of the genitalia is essential to the reflex mechanisms involved in erection; this becomes important with aging and conditions such as diabetes that impair peripheral nerve function.

Hormonal causes of erectile dysfunction include a decrease in androgen levels. Androgen levels may be decreased because of aging. Hypoprolactinemia from any cause interferes with both reproduction and erectile function. This is because prolactin acts centrally to inhibit dopaminergic activity, which is a stimulus for release of the hypothalamic GnRH that controls the release of pituitary gonadotropic hormones.

Common risk factors for generalized penile arterial insufficiency include hypertension, hyperlipidemia, cigarette smoking, diabetes mellitus, and pelvic irradiation.⁵ In hypertension, erectile function is impaired not so much by the increased blood pressure as by the associated stenotic arterial lesions. Focal stenosis of the common penile artery most often occurs in men who sustained blunt pelvic or perineal trauma (*e.g.*, from bicycling accidents). Failure of the veins to close completely during an erection (veno-occlusive dysfunction) may occur in men with large venous channels that drain the corpora cavernosa. Other disorders that impair venous occlusion are degenerative changes involving the tunica albuginea, as in Peyronie's disease. Poor relaxation of the trabecular smooth muscle may accompany anxiety with excessive adrenergic tone.

Many drugs are reported to cause erectile dysfunction, including antidepressant, antipsychotic, and antihypertensive medications. Cigarette smoking can induce vasoconstriction and penile venous leakage because of its effects on cavernous smooth muscle.⁵ Alcohol in small amounts may increase libido and improve erection; however, in large amounts it can cause central sedation, decreased libido, and transient erectile dysfunction.

Aging is known to increase the risk of erectile dysfunction.⁷ Many of the pathologic processes that contribute to erectile dysfunction are more common in older men, including diabetes, hyperlipidemia, vascular disease, and the long-term effects of cigarette smoking. Age-related declines in testosterone also may play a role. Psychosocial problems such as depression, esteem issues, partner relationships, history of substance abuse, and anxiety and fear of performance failure also may contribute to erectile dysfunction in older men.

A diagnosis of erectile dysfunction requires careful history (medical, sexual, and psychosocial), physical examination, and laboratory tests aimed at determining what other tests are needed to rule out organic causes of the disorder.⁶ Because many medications, including prescribed, over-the-counter, and illicit drugs, can cause erectile dysfunction, a careful drug history is indicated.

Treatment methods include psychosexual counseling, androgen replacement therapy, oral and intracavernous drug therapy, vacuum constriction devices, and surgical treatment (prosthesis and vascular surgery).⁶ Among the commonly prescribed drugs are sildenafil, yohimbine, alprostadil, and phentolamine. Sildenafil (Viagra) is a selective inhibitor of phosphodiesterase type 5, the enzyme that inactivates cGMP. Yohimbine, an α_2 -adrenergic receptor antagonist, acts at the adrenergic receptors in brain centers associated with libido and penile erection. Both sildenafil and yohimbine are taken orally. Alprostadil, a prostaglandin E analog, acts by producing relaxation of cavernous smooth muscle. It is either injected directly into the cavernosa or placed in the urethra as a minisuppository. Phentolamine, an α_2 -adrenergic receptor antagonist, also is administered by intracavernous injection.

In summary, spermatogenesis occurs in the Sertoli's cells of the seminiferous tubules of the testes. After formation in seminiferous tubules, the spermatozoa travel through the efferent tubules to the epididymis, then to the ductus deferens, and on to the ampulla, where they are stored until released through the penis during ejaculation. The male accessory organs consist of the seminal vesicles, prostate gland, and bulbourethral glands. Secretions from the genital ducts and accessory organs combine with the spermatozoa to form the semen.

The function of the male reproductive system is under the negative feedback control of the hypothalamus and the anterior pituitary gonadotropic hormones FSH and LH. Spermatogenesis is initiated by FSH, and the production of testosterone is regulated by LH. Testosterone, the major male sex hormone, is produced by the interstitial Leydig's cells in the testes. In addition to its role in the differentiation of the internal and external genitalia in the male embryo, testosterone is essential for the development of secondary male characteristics during puberty, the maintenance of these characteristics during adult life, and spermatozoa maturation.

The male sex act involves erection, emission, ejaculation, and detumescence. The physiology of these functions involves a complex interaction between autonomic-mediated spinal cord reflexes, higher neural centers, and the vascular system. Erection is mediated by the parasympathetic nervous system and emission and ejaculation by the sympathetic nervous system. Erectile dysfunction is defined as the inability to achieve and maintain an erection sufficient to permit satisfactory sexual intercourse. It can be due to psychogenic factors, organic disorders, or mixed psychogenic and organic conditions.

DISORDERS OF THE PENIS, THE SCROTUM AND TESTES, AND THE PROSTATE

Disorders of the Penis

Disorders of the penis include congenital defects (discussed in Disorders in Childhood), acute and chronic inflammatory conditions, Peyronie's disease, priapism, and neoplasms.

Inflammation and Infection

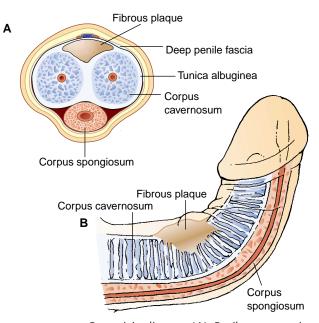
Balantitis and Balanoposthitis. Balanitis is an acute or chronic inflammation of the glans penis. Balanoposthitis refers to inflammation of the glans and prepuce. It usually is encountered in males with phimosis (a tight foreskin) or a large, redundant prepuce that interferes with cleanliness and predisposes to bacterial growth in the accumulated secretions and smegma (*i.e.*, debris from the desquamated epithelia). If left untreated, the condition may cause ulcerations of the mucosal surface of the glans; these ulcerations may lead to inflammatory scarring of the phimosis and further aggravate the condition.

Acute superficial balanoposthitis is characterized by erythema of the glans and prepuce. An exudate in the form of malodorous discharge may be present. Extension of the erythema and edema may result in phimosis. The condition may result from infection, trauma, or irritation. Infective balanoposthitis may be caused by a wide variety of organisms. Chlamydiae and mycoplasmas have been identified as causative organisms in this disease. The inflammatory reaction is nonspecific, and correct identification of the specific agent requires bacterial smears and cultures.

Balanitis xerotica obliterans is a chronic white, patchy lesion that originates on the glans and usually progresses to involve the meatus. It is clinically and histologically similar to the lichen sclerosus that is seen in females. It commonly is observed in middle-aged diabetic men. Treatment measures include topical or intralesional injections of corticosteroids.⁸

Peyronie's Disease

Peyronie's disease involves a localized and progressive fibrosis of unknown origin that affects the tunica albuginea (*i.e.*, the tough, fibrous sheath that surrounds the corpora cavernosa) of the penis. It is named after Francois de la Peyronie, who wrote in 1743, describing a patient who had "rosary beads of scar tissue to cause upward curvature of the penis during erection."9 The disorder is characterized initially by an inflammatory process that results in dense fibrous plaque formation. The plaque usually is on the dorsal midline of the shaft, causing upward bowing of the shaft during erection (Fig. 33-7). Some men may develop scarring on both the dorsal and ventral aspects of the shaft, causing the penis to be straight but shortened or have a lateral bend.9 The fibrous tissue prevents lengthening of the involved area during erection, making intercourse difficult and painful. The disease usually occurs in middle-aged or elderly men. Although the cause of the disorder is unknown, the dense microscopic plaques are consistent with findings of severe vasculitis. As many as 47% of men with Peyronie's disease have another condition associated with fascial tissue fibrosis, such as Dupuytren's contracture (fibrosis of the palmar fascia).9



■ FIGURE 33-7 ■ Peyronie's disease. (A) Penile cross-section showing plaque between the corpora. (B) Penile curvature.

The manifestations of Peyronie's disease include painful erection, bent erection, and the presence of a hard mass at the site of fibrosis. Approximately two thirds of men report pain as a symptom. The pain is thought to be generated by inflammation of the adjacent fascial tissue and usually disappears as the inflammation resolves.¹⁰ During the first year or so after formation of the plaque, while the scar tissue is undergoing the process of remodeling, penile distortion may increase, remain static, or resolve and disappear completely.⁹ In some cases, the scar tissue may progress to calcification and formation of bone-like tissue.

Diagnosis is based on history and physical examination. Doppler ultrasonography may be used to assess causation of the disorder. Although surgical intervention can be used to correct the disorder, it often is delayed, because in many cases the disorder is self-limiting.¹⁰ Less invasive treatments include the administration of oral agents with antioxidant properties (*e.g.*, vitamin D, potassium aminobenzoate, and colchicine).

Priapism

Priapism is an involuntary, prolonged, abnormal and painful erection that is not associated with sexual excitement. Priapism is a true urologic emergency because the prolonged erection can result in ischemia and fibrosis of the erectile tissue with significant risk of subsequent impotence. Priapism can occur at any age, in the newborn as well as other age groups. Sickle cell disease or neoplasms are the most common cause in boys between 5 and 10 years of age.

Priapism is due to impaired blood flow in the corpora cavernosa of the penis. Two mechanisms for priapism have been proposed: low-flow (ischemic) priapism, in which there is stasis of blood flow in the corpora cavernosa with a resultant failure of detumescence, and high-flow (nonischemic) priapism, which involves persistent arterial flow into the corpora cavernosa.¹¹ In high-flow priapism, there is no hypoxia of local tissue, the penis is less rigid, the pain is less than in stasis priapism, and permanent corporal fibrosis and cellular damage are rare.¹¹

Priapism is classified as primary (idiopathic) or secondary to a disease or drug effect. Primary priapism is the result of conditions such as trauma, infection, and neoplasms. Secondary causes include hematologic conditions such as leukemia, sickle cell disease, and thrombocytopenia; neurologic conditions such as stroke, spinal cord injury, and other central nervous system lesions; and renal failure. Between 6% and 12% of males with sickle cell disorders are affected by priapism.¹¹ The relative deoxygenation and stasis of cavernosal blood during erection is thought to increase sickling. Various medications, such as antihypertensive drugs, anticoagulant drugs, antidepressant drugs, alcohol, and marijuana, can contribute to the development of priapism. Androstenedione, sold as an over-the-counter drug to enhance muscle building and athletic performance, has been implicated in the disorder.¹² Currently, intracavernous injection therapy for erectile dysfunction is one of the more common causes of priapism.

The diagnosis of priapism usually is based on clinical findings. Doppler studies of penile blood flow, penile ultrasonography, and computed tomography (CT) scans may be used to determine intrapelvic pathology.

Initial treatment measures include analgesics, sedation, and hydration. Urinary retention may necessitate catheterization. Local measures include application of ice packs and irrigation of the corpus cavernosum with plain or heparinized saline, or instillation of α -adrenergic drugs. If less aggressive treatment does not produce detumescence, a temporary surgical shunt may be established between the corpus cavernosum and the corpus spongiosum.

The prognosis for whether fibrosis or erectile failure will occur is determined by the severity and duration of blood stasis. In high-flow priapism, the damaging effects of decreased oxygen tension and intracavernal blood pressure are less pronounced than in stasis priapism. Normal erectile potency can be restored even after a long duration of high-flow priapism. In contrast, persistent stasis priapism is known to result in impaired erectile function and tissue fibrosis unless resolved within 24 to 48 hours of onset.¹¹

Cancer of the Penis

Squamous cell cancer of the penis is most common in men between 45 and 60 years of age. In the United States, it accounts for less than 1% of male genital tumors; however, in other countries, it accounts for 10% to 20% of male cancers.^{13,14} When it is diagnosed early, penile cancer is highly curable. The greatest hindrance to early diagnosis is a delay in seeking medical attention.

The cause of penile cancer is unknown. Several risk factors have been suggested, including poor hygiene, human papillomavirus (HPV) infections, ultraviolet radiation exposure, and immunodeficiency states. There is an association between penile cancer and poor genital hygiene and phimosis. One theory postulates that smegma accumulation under the phimotic foreskin may produce chronic inflammation, leading to carcinoma. The HPVs have been implicated in the genesis of several genital cancers, including cancer of the penis.¹³ Ultraviolet radiation also is thought to have a carcinogenic effect on the penis. Males who were treated for psoriasis with ultraviolet A or B therapies (*i.e.*, PUVA or PUVB) have had a reported increased incidence of genital squamous cell carcinomas.¹³ Because of this observation, it is suggested that men should shield their genital area when using tanning salons. Immunodeficiency states also may play a role in the pathogenesis of penile cancer. Approximately 18% of men with acquired immunodeficiency syndrome (AIDS)-related Kaposi's sarcoma have lesions of the penis or genitalia.¹³

Dermatologic lesions with precancerous potential include balanitis xerotica obliterans (discussed earlier) and giant condylomata acuminata.¹⁴ Giant condylomata acuminata are cauliflower-like lesions arising from the prepuce or glans that result from HPV infection.

Approximately 95% of penile cancers are squamous cell carcinoma.¹³ It is thought to progress from an in situ lesion to an invasive carcinoma. Bowen's disease and erythroplasia of Queyrat are penile lesions with histologic features of carcinoma in situ.¹⁴ Bowen's disease appears as a solitary, thickened, gray-white, opaque plaque with shallow ulceration and crusting. It commonly involves the skin of the shaft of the penis and the scrotum. Erythroplasia of Queyrat involves the mucosal surface of the glans or prepuce. It is characterized by single or multiple shiny red, sometimes velvety, plaques.¹⁴ These lesions require careful follow-up because of their potential to progress to invasive carcinoma.

Invasive carcinoma of the penis begins as a small lump or ulcer. If phimosis is present, there may be painful swelling, purulent drainage, or difficulty urinating. Palpable lymph nodes may be present in the inguinal region. Diagnosis usually is based on physical examination and biopsy results. CT scans, penile ultrasound studies, and magnetic resonance imaging (MRI) may be used in the diagnostic workup.

Treatment options vary according to stage, size, location, and invasiveness of the tumor. Carcinoma in situ may be treated conservatively with fluorouracil cream application or laser treatment.¹⁴ Surgery remains the mainstay of treatment for invasive carcinoma. Partial or total penectomy is indicated for invasive lesions.

Disorders of the Scrotum and Testes

The testes, or male gonads, are two egg-shaped structures located outside the abdominal cavity in the scrotum. Embryologically, the testes develop in the abdominal cavity and then descend through the inguinal canal into a pouch of peritoneum (which becomes the tunica vaginalis) in the scrotum during the seventh to ninth months of fetal life. As they descend, the testes pull their arteries, veins, lymphatics, nerves, and conducting excretory ducts with them. These structures are encased by the cremaster muscle and layers of fascia that constitute the spermatic cord (Fig. 33-8A). The descent of the testes is thought to be mediated by testosterone, which is active during this stage of development.

After descent of the testes, the inguinal canal closes almost completely. Failure of this canal to close predisposes to the development of an inguinal hernia later in life (Fig. 33-8B). An inguinal hernia or "rupture" is a protrusion of the parietal peritoneum and part of the intestine through an abnormal opening from the abdominal cavity. A loop of small bowel may become incarcerated in an inguinal hernia (strangulated hernia), in which case the lumen may become obstructed, and the vascular supply compromised (see Chapter 27).



The testes and epididymis are completely surrounded by the tunica vaginalis, a serous pouch derived from the peritoneum during fetal descent of the testes into the scrotum. The tunica vaginalis has an outer parietal layer and a deeper visceral layer that adheres to the dense fibrous covering of the testes, the tunica albuginea. The tunica albuginea protects the testes and gives them their ovoid shape. A space exists between these two layers that typically contains a few milliliters of clear fluid. The cremaster muscles, which are bands of skeletal muscle arising from the internal oblique muscles of the trunk, elevate the testes. The testes receive their arterial blood supply from the long testicular arteries, which branch from the aortic artery. The spermatic veins, which drain the testes, arise from a venous network called the pampiniform plexus that surrounds the spermatic artery. The testes are innervated by fibers from both divisions of the autonomic nervous system. Associated sensory nerves transmit pain impulses, resulting in excruciating pain, especially when the testes are hit forcibly.

The scrotum, which houses the testes, is made up of a thin outer layer of skin that forms rugae, or folds, and is continuous with the perineum and outer skin of the groin. Under the outer skin lies a thin layer of fascia and smooth muscle (*i.e.*, dartos muscle). This layer contains a septum that separates the two testes. The dartos muscle responds to changes in temperature. When it is cold, the muscle contracts, bringing the testes closer to the body, and the scrotum becomes shorter and heavily wrinkled. When it is warmer, the muscle relaxes, allowing the scrotum to fall away from the body.

The location of the testes in the scrotum is important for sperm production, which is optimal at 2°C to 3°C below body temperature. Two systems maintain the temperature of the testes at a level consistent with sperm production. One is the pampiniform plexus of spermatic veins that surround the spermatic artery. This plexus absorbs heat from the arterial blood, cooling it as it enters the testes. The other is the cremaster muscles, which respond to decreases in testicular temperature by moving the testes closer to the body. Prolonged exposure to elevated temperatures, as a result of prolonged fever or the dysfunction of thermoregulatory mechanisms, can impair spermatogenesis. Some tight-fitting undergarments hold the testes against the body and are thought to contribute to a decrease in sperm counts and infertility by interfering with the thermoregulatory function of the scrotum. Cryptorchidism, the failure of the testes to descend into the scrotum, also exposes the testes to the higher temperature of the body.

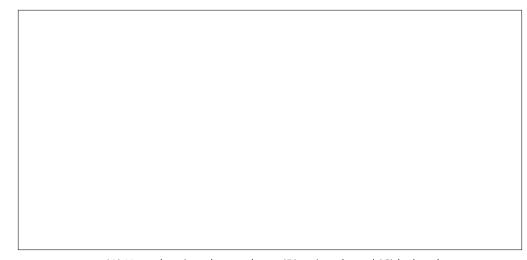
Disorders of the Testicular Tunica

Hydrocele. A hydrocele forms when excess fluid collects between the layers of the tunica vaginalis (Fig. 33-9). It may be unilateral or bilateral and can develop as a primary congenital defect or as a secondary condition. Acute hydrocele may develop after local injury, epididymitis or orchitis, gonorrhea, lymph obstruction, or germ cell testicular tumor, or as a side effect of radiation therapy. Chronic hydrocele is more common. Fluid collects about the testis, and the mass grows gradually. Its cause is unknown, and it usually develops in men older than 40 years.

Most cases of hydrocele in male infants and children are due to a patent processus vaginalis, which is continuous with the peritoneal cavity. In many cases they are associated with an indirect inguinal hernia.¹⁵ Most hydroceles of infancy close spontaneously; therefore, they are not usually repaired before the age of 1 year. Hydroceles that persist beyond 2 years of age may require surgical treatment.

Hydroceles are palpated as cystic masses that may attain massive proportions. If there is enough fluid, the mass may be mistaken for a solid tumor. Transillumination of the scrotum (*i.e.*, shining a light through the scrotum for the purposes of visualizing its internal structures) or ultrasonography can help to determine whether the mass is solid or cystic and whether the testicle is normal.¹⁶ A dense hydrocele that does not illuminate should be differentiated from a testicular tumor. If a hydrocele develops in a young man without apparent cause, careful evaluation is needed to exclude cancer or infection.

In an adult male, a hydrocele is a relatively benign condition. The condition often is asymptomatic, and no treatment is necessary. When symptoms do occur, the feeling may be that of heaviness in the scrotum or pain in the lower back. In cases of secondary hydrocele, the primary condition is treated. If the hydrocele is painful or cosmetically undesirable, surgical correction is indicated. Surgical repair may be done inguinally or transcrotally.



■ FIGURE 33-9 ■ (A) Normal testis and appendages, (B) varicocele, and (C) hydrocele.

Hematocele. A hematocele is an accumulation of blood in the tunica vaginalis, which causes the scrotal skin to become dark red or purple. It may develop as a result of an abdominal surgical procedure, scrotal trauma, a bleeding disorder, or a testicular tumor.

Spermatocele. A spermatocele is a painless, sperm-containing cyst that forms at the end of the epididymis. It is located above and posterior to the testis, is attached to the epididymis, and is separate from the testes. Spermatoceles may be solitary or multiple and usually are less than 1 cm in diameter. They are freely movable and should transilluminate. Spermatoceles rarely cause problems, but a large one may become painful and require excision.

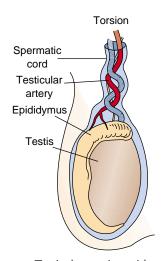
Varicocele. A varicocele is characterized by varicosities of the pampiniform plexus, a network of veins supplying the testes (Fig. 33-9). The left side is more commonly affected because the left internal spermatic vein inserts into the left renal vein at a right angle, whereas the right spermatic vein usually enters the inferior vena cava. Incompetent valves are more common in the left internal spermatic veins, causing a reflux of blood back into the veins of the pampiniform plexus. The force of gravity resulting from the upright position also contributes to venous dilatation. If the condition persists, there may be damage to the elastic fibers and hypertrophy of the vein walls, as occurs in formation of varicose veins in the leg. Sperm concentration and motility are decreased in 65% to 75% of men with varicocele because of changes in testicular temperature resulting from altered blood flow.¹⁶

Varicoceles rarely are found before puberty, and the incidence is highest in men between 15 and 35 years of age. Symptoms of varicocele include an abnormal feeling of heaviness in the left scrotum, although many varicoceles are asymptomatic. Usually, the varicocele is readily diagnosed on physical examination with the patient in the standing and recumbent positions. Typically, the varicocele disappears in the lying position because of venous decompression into the renal vein. Scrotal palpation of a varicocele has been compared to feeling a "bag of worms." Treatment options include surgical ligation or sclerosis using a percutaneous transvenous catheter under fluoroscopic guidance. It has been shown that 40% of men with abnormalities in their semen and a varicocele show some degree of improvement in fertility after obliteration of the dilated veins.¹⁶ Aside from improving fertility, other reasons for surgery include the relief of the sensation of "heaviness" and cosmetic improvement.

Testicular Torsion

Testicular torsion is a twisting of the spermatic cord that suspends the testis (Fig. 33-10). It is the most common acute scrotal disorder in the pediatric and young adult population. Testicular torsion can be divided into two distinct clinical entities, depending on the level of spermatic cord involvement: extravaginal and intravaginal torsion.¹⁷

Extravaginal torsion, which occurs almost exclusively in neonates, is the less common form of testicular torsion. It occurs



■ FIGURE 33-10 ■ Testicular torsion with twisting of the spermatic cord that suspends the testis and the spermatic vessels that supply the testis with blood.

when the testicle and the fascial tunicae that surround it rotate around the spermatic cord at a level well above the tunica vaginalis. Intravaginal torsion is considerably more common than extravaginal torsion. It occurs when the testis rotates on the long axis in the tunica vaginalis. In most cases, congenital abnormalities of the tunica vaginalis or spermatic cord exist.¹⁷ The tunica vaginalis normally surrounds the testes and epididymis, allowing the testicle to rotate freely in the tunica. Although anomalies of suspension vary, the epididymal attachment may be loose enough to permit torsion between the testis and the epididymis. More commonly, the testis rotates about the distal spermatic cord. Because this abnormality is developmental, bilateral anomalies are common.

Intravaginal torsion occurs most frequently between the ages of 8 and 18 years and rarely is seen after 30 years of age. Males usually present in severe distress within hours of onset and often have nausea, vomiting, and tachycardia. The affected testis is large and tender, with pain radiating to the inguinal area. Extensive cremaster muscle contraction causes a thickening of the spermatic cord.

Testicular torsion must be differentiated from epididymitis, orchitis, and trauma to the testis. On physical examination, the testicle often is high in the scrotum and in an abnormal orientation. These changes are due to the twisting and shortening of the spermatic cord. The degree of scrotal swelling and redness depends on the duration of symptoms. The testes are firm and tender. The cremasteric reflex, normally elicited by stroking the medial aspect of the thigh and observing testicular retraction, frequently is absent.¹⁷ Color Doppler ultrasonography is increasingly used in the evaluation of suspected testicular torsion.17

Intravaginal testicular torsion is a true surgical emergency, and early recognition and treatment are necessary if the testicle is to be saved. Treatment includes surgical detorsion (repositioning and fixation) and orchiectomy. Orchiectomy is carried out when the testis is deemed nonviable after surgical detorsion. Testicular salvage rates are directly related to the duration of torsion. Because the opposite testicle usually is affected by the same abnormal attachments, prophylactic fixation of that testis often is performed.

Inflammation and Infection

Epididymitis. Epididymitis is an inflammation of the epididymis. There are two major types of epididymitis: sexually transmitted infections associated with urethritis and primary nonsexually transmitted infections associated with urinary tract infections and prostatitis. Most cases of epididymitis are caused by bacterial pathogens.

In primary nonsexual infections, the pressure associated with voiding or physical strain may force pathogen-containing urine from the urethra or prostate up the ejaculatory duct and through the vas deferens and into the epididymis. Infections also may reach the epididymis through the lymphatics of the spermatic cord. In rare cases, organisms from other foci of infection reach the epididymis through the bloodstream. In children, the disorder usually is associated with congenital urinary tract abnormalities and infection with gram-negative rods.

Sexually transmitted acute epididymitis occurs mainly in young men without underlying genitourinary disease and is most commonly caused by Chlamydia trachomatis and Neisseria gonorrhoeae (singly or in combination). In men older than 35 years, epididymitis often is associated with pathogens such as Escherichia coli, Pseudomonas, and gram-positive cocci.

Epididymitis is characterized by unilateral pain and swelling, accompanied by erythema and edema of the overlying scrotal skin that develops during a period of 24 to 48 hours. Initially, the swelling and induration are limited to the epididymis. However, the distinction between the testis and epididymis becomes less evident as the inflammation progresses, and the testis and epididymis become one mass. There may be tenderness over the groin (spermatic cord) or in the lower abdomen. Fever and reports of dysuria occur in approximately one half of cases. Whether urethral discharge is present depends on the organism causing the infection; it usually accompanies gonorrheal infections, is common in chlamydial infections, and is less common in infections caused by gram-negative organisms.

Laboratory findings usually reveal an elevated white blood cell count. Urinalysis and urine culture are important in the diagnosis of epididymitis, with bacteriuria and pyuria suggestive of the disorder. Treatment during the acute phase (which usually lasts for 3 to 4 days) includes bed rest, scrotal elevation and support, and antibiotics.¹⁸ Bed rest with scrotal support improves lymphatic drainage. The choice of antibiotics is determined by age, physical findings, urinalysis, Gram's stain results, cultures, and sexual history. Oral analgesics and antipyretics usually are indicated. Sexual activity or physical strain may exacerbate the infection and worsen the symptoms and should be avoided.

Orchitis. Orchitis is an infection of the testes. It can be precipitated by a primary infection in the genitourinary tract, or the infection can be spread to testes through the bloodstream or the lymphatics. Epididymitis with subsequent infection of the testis is commonly related to genitourinary tract infections (cystitis, urethritis, genitoprostatitis) that travel to the epididymis and testis through the vas deferens or the lymphatics of the spermatic cord.

Orchitis can develop as a complication of a systemic infection, such as parotitis (i.e., mumps), scarlet fever, or pneumonia. Probably the best known of these complications is orchitis caused by the mumps virus. Mumps orchitis does not occur in prepubertal boys. However, approximately 20% to 35% of adolescent boys and young men with mumps develop this form of orchitis.¹⁸ The onset of mumps orchitis is sudden; it usually occurs approximately 3 to 4 days after the onset of the parotitis and is characterized by fever, painful enlargement of the testes, and small hemorrhages into the tunica albuginea. Unlike epididymitis, the urinary symptoms are absent. The symptoms usually run their course in 7 to 10 days. The residual effects seen after the acute phase include hyalinization of the seminiferous tubules and atrophy of the testes. Spermatogenesis is irreversibly impaired in approximately 30% of testes damaged by mumps orchitis.¹⁸ If both testes are involved, permanent sterility results, but and rogenic hormone function usually is maintained.

Cancer of the Scrotum and Testes

Tumors can develop in the scrotum or the testes. Benign scrotal tumors are common and often do not require treatment. Carcinoma of the scrotum is rare and usually is associated with exposure to carcinogenic agents. Almost all solid tumors of the testes are malignant.

Scrotal Cancer. Cancer of the scrotum was the first cancer directly linked to a specific occupation when, in the 1800s, it was associated with chimney sweeps.¹⁹ Studies have linked this cancer to exposure to tar, soot, and oils. Most squamous cell cancers of the scrotum are linked to poor hygiene and chronic inflammation. Exposure to ultraviolet A radiation (*e.g.*, PUVA) or HPV also has been associated with the disease. The mean age of presentation with the disease is 60 years, often preceded by 20 to 30 years of chronic irritation.

In the early stages, cancer of the scrotum may appear as a small tumor or wartlike growth that eventually ulcerates. The thin scrotal wall lacks the tissue reactivity needed to block the malignant process; more than one half of the cases seen involve metastasis to the lymph nodes. Because this tumor does not respond well to chemotherapy or irradiation, the treatment includes wide local excision of the tumor with inguinal and femoral node dissection.²⁰

Testicular Cancer. Testicular cancer accounts for 1% of all male cancers and 3% of male urogenital cancers. Although relatively rare, it is the most common cause of cancer in the 15- to 34-year-old age group.^{20,21} In the past, testicular cancer was a leading cause of death among males entering their most productive years. However, since the late 1970s, advances in therapy have transformed an almost invariably fatal disease into one that is highly curable. With appropriate treatment, the prognosis for men with testicular cancer is excellent. The 5-year survival rate for patients with early disease exceeds 95%. Even patients with more advanced disease have excellent chances for long-term survival.

Although the cause of testicular cancer is unknown, several predisposing influences may be important: cryptorchidism, genetic factors, and disorders of testicular development.²² The strongest association has been with cryptorchid or undescended testes. The higher the location of the undescended testis, the greater the risk.²² Genetic predisposition also appears to be important. Family clustering of the disorder has been described, although a well-defined pattern of inheritance has not been established. Men with disorders of testicular development, including those with Klinefelter's syndrome and testicular feminization, have a higher risk of germ cell tumors.

Approximately 95% of malignant tumors arising in the testis are germ cell tumors.^{21,23} Germ cell tumors can be classified as seminomas and nonseminomas based on their origin in primordial germ cells and their ability to differentiate in vivo. Because these tumors derive from germ cells in the testis, they are multipotential (able to differentiate into different tissue types) and often secrete polypeptide hormones or enzymes representing earlier stages of development (see Chapter 5). *Seminomas* are the most common type of testicular tumors. They account for approximately 50% of germ cell tumors and occur most frequently during the fourth decade of life.²¹ Seminomas are thought to arise from the seminiferous epithelium of the testes and are the type of germ cell tumor most likely to produce a uniform population of cells.

The nonseminoma tumors include embryonal carcinoma, teratoma, choriocarcinoma, and yolk cell carcinoma derivatives. Nonseminoma tumors usually contain more than one cell type and are less differentiated than seminomas. Embryonal carcinomas are the least differentiated of the tumors, with totipotential capacity to differentiate into other nonseminomatous cell types. They occur most commonly in the 20- to 30-year-old age group. Choriocarcinoma is a rare and highly malignant form of testicular cancer that is identical to tumors that arise in placental tissue. Yolk sac tumors mimic the embryonic yolk sac histologically. They are the most common type of testicular tumors in infants and children to 3 years of age, and in this age group are associated with a very good prognosis.²² Teratomas are composed of somatic cell types from two or more germ-line layers (ectoderm, mesoderm, or endoderm). They constitute less than 2% to 3% of germ cell tumors and can occur at any age from infancy to old age. They usually behave as benign tumors in children; in adults, they often contain minute foci of cancer cells.

Often the first sign of testicular cancer is a slight enlargement of the testicle that may be accompanied by some degree of discomfort. This may be an ache in the abdomen or groin or a sensation of dragging or heaviness in the scrotum. Frank pain may be experienced in the later stages, when the tumor is growing rapidly and hemorrhaging occurs. Testicular cancer can spread when the tumor may be barely palpable. Signs of metastatic spread include swelling of the lower extremities, back pain, cough, hemoptysis, or dizziness. Gynecomastia (breast enlargement) may result from human chorionic gonadotropin (hCG)-producing tumors.

Early diagnosis of testicular cancer is important because a delay in seeking medical attention often results in presentation with a later stage of the disease and decreased treatment effectiveness. Recognition of the importance of prompt diagnosis and treatment has resulted in the development of a procedure for testicular self-examination and an emphasis on public education programs about this type of cancer. The American Cancer Society strongly advocates that every young adult male examine his testes at least once each month as a means of early detection of testicular cancer.

The diagnosis of testicular cancer requires a thorough urologic history and physical examination. A painless testicular mass may be cancer. The examination for masses should include palpation of the testes and surrounding structures, transillumination of the scrotum, and abdominal palpation. Testicular ultrasonography can be used to differentiate testicular masses. CT scans and MRI are used in assessing metastatic spread.

Tumor markers, which measure protein antigens produced by malignant cells, provide information about the existence of a small or undetected tumor and the type of tumor present. Three tumor markers are useful in evaluating the tumor response: α -fetoprotein, a glycoprotein that normally is present in fetal serum in large amounts; hCG, a hormone that normally is produced by the placenta in pregnant women; and lactate dehydrogenase (LDH), a cellular enzyme normally found in muscle, liver, kidney, and brain. The reappearance in the adult of these protein markers suggests activity normally present in the embryo.

The basic treatment of all testicular cancers includes orchiectomy, which is done at the time of diagnostic exploration. Dependent on the histologic characteristics of the tumor and the clinical stage of the disease, radiation or chemotherapy may used after orchiectomy. Rigorous follow-up in all men with testicular cancer is necessary to detect recurrence, which most often occurs within the first year.²³

Disorders of the Prostate

The prostate is a fibromuscular and glandular organ lying just inferior to the bladder. The prostate gland secretes a thin, milky, alkaline fluid containing citric acid, calcium, acid phosphate, a clotting enzyme, and a profibrinolysin. During ejaculation, the capsule of the prostate contracts, and the added fluid increases the bulk of the semen. Both vaginal secretions and the fluid from the vas deferens are strongly acidic. Because sperm mobilization occurs at a pH of 6.0 to 6.5, the alkaline nature of the prostatic secretions is essential for successful fertilization of the ovum. The bulbourethral glands (Fig. 33-1) lie on either side of the membranous urethra and secrete an alkaline mucus, which further aids in neutralizing acids from the urine that remain in the urethra.

The prostate gland, which forms a fibrous capsule that surrounds the urethra where it joins the bladder, also functions in the elimination of urine. The segment of urethra that travels through the prostate gland is called the *prostatic urethra*. The prostatic urethra is lined by a thin layer of smooth muscle that is continuous with the bladder wall. This smooth muscle represents the true involuntary sphincter of the male posterior urethra. Because the prostate surrounds the urethra, enlargement of the gland can produce urinary obstruction.

The prostate gland is made up of many secretory glands arranged in three concentric areas surrounding the prostatic urethra, into which they open. The component glands of the prostate include the (1) small mucosal glands associated with the urethral mucosa, (2) the intermediate submucosal glands that lie peripheral to the mucosal glands, and (3) the large main prostatic glands that are situated toward the outside of the gland. It is the overgrowth of the mucosal glands that causes benign prostatic hyperplasia in older men.

Prostatitis

Prostatitis refers to a variety of inflammatory disorders of the prostate gland, some of which are bacterial and some are not. It may occur spontaneously, as a result of catheterization or instrumentation, or secondary to other diseases of the male genitourinary system. As an outcome of 1995 and 1998 consensus conferences, the National Institutes of Health has established a classification system with four categories of prostatitis syndromes: asymptomatic inflammatory prostatitis, acute bacterial prostatitis, chronic bacterial prostatitis, and chronic prostatitis/ pelvic pain syndrome.²⁴ Men with asymptomatic inflammatory prostatitis have no subjective symptoms and are detected incidentally on biopsy or examination of prostatic fluid.

Acute Bacterial Prostatitis. Acute bacterial prostatitis often is considered a subtype of urinary tract infection. The most likely etiology of acute bacterial prostatitis is an ascending urethral infection or reflux of infected urine into the prostatic ducts. The most common organism is E. coli. Other frequently found species include Pseudomonas, Klebsiella, and Proteus. Less frequently, the infection is caused by Staphylococcus aureus, Streptococcus faecalis, Chlamydia, or anaerobes such as Bacteroides species.^{25,26}

The manifestations of acute bacterial prostatitis include fever and chills, malaise, myalgia, arthralgia, frequent and urgent urination, dysuria, and urethral discharge. Dull, aching pain often is present in the perineum, rectum, or sacrococcygeal region. The urine may be cloudy and malodorous because of urinary tract infection. Rectal examination reveals a swollen, tender, warm prostate with scattered soft areas. Prostatic massage produces a thick discharge with white blood cells that grows large numbers of pathogens on culture.

Acute prostatitis usually responds to appropriate antimicrobial therapy chosen in accordance with the sensitivity of the causative agents in the urethral discharge. Depending on the urine culture results, antibiotic therapy usually is continued for at least 4 weeks. Because acute prostatitis often is associated with anatomic abnormalities, a thorough urologic examination usually is performed after treatment is completed.

A persistent fever indicates the need for further investigation for an additional site of infection or a prostatic abscess. CT scans and transrectal ultrasonography of the prostate are useful in the diagnosis of prostatic abscesses. Prostatic abscesses, which are relatively uncommon since the advent of effective antibiotic therapy, are found more commonly in males with diabetes mellitus. Because prostatic abscesses usually are associated with bacteremia, prompt drainage by transperitoneal or transurethral incision followed by appropriate antimicrobial therapy usually is indicated.²⁵

Chronic Bacterial Prostatitis. In contrast to acute bacterial prostatitis, chronic bacterial prostatitis is a subtle disorder that is difficult to treat. Men with the disorder typically have recurrent urinary tract infections with persistence of the same strain of pathogenic bacteria in prostatic fluid and urine. Organisms responsible for chronic bacterial prostatitis usually are the gram-negative enterobacteria (E. coli, Proteus, or Klebsiella) or *Pseudomonas*. Occasionally, a gram-positive organism such as S. faecalis is the causative organism. Infected prostatic calculi may develop and contribute to the chronic infection.

The symptoms of chronic prostatitis are variable and include frequent and urgent urination, dysuria, perineal discomfort, and low back pain. Occasionally, myalgia and arthralgia accompany the other symptoms. Secondary epididymitis sometimes is associated with the disorder. Many men develop relapsing lower or upper urinary tract infections because of recurrent invasion of the bladder by the prostatic bacteria. Bacteria may exist in the prostate gland even when the prostatic fluid is sterile.

The most accurate method of establishing a diagnosis is by urine cultures. Even after an accurate diagnosis has been established, treatment of chronic prostatitis often is difficult and frustrating. Unlike their action in the acutely inflamed prostate, antibacterial drugs penetrate poorly into the chronically inflamed prostate. Long-term therapy (3 to 4 months) with an appropriate low-dose oral antimicrobial agent often is used to treat the infection. Transurethral prostatectomy may be indicated when the infection is not cured or adequately controlled by medical therapy, particularly when prostate stones are present.

Chronic Prostatitis/Chronic Pelvic Pain Syndrome. Chronic prostatitis/pelvic pain syndrome is both the most common and least understood of the prostatitis syndromes.²⁷ The category is divided into two types, inflammatory and noninflammatory, based on the presence of leukocytes in the prostatic fluid. The inflammatory type was previously referred to as nonbacterial prostatitis, and the noninflammatory type as prostatodynia.

Men with nonbacterial prostatitis have inflammation of the prostate with an elevated leukocyte count, inflammatory cells

594 Unit Nine: Alterations in the Male and Female Reproductive Systems

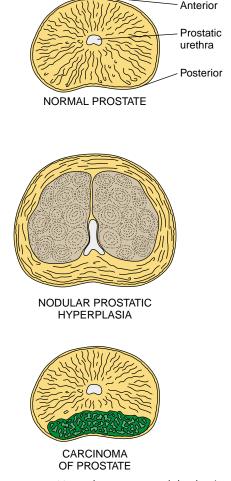
in their prostatic secretions, but no evidence of bacteria. The cause of the disorder is unknown, and efforts to prove the presence of unusual pathogens (*e.g.*, mycoplasmas, chlamydiae, trichomonads, viruses) have been largely unsuccessful. It also is thought that nonbacterial prostatitis may be an autoimmune disorder. Manifestations of *inflammatory prostatitis* include pain along the penis, testicles, and scrotum; painful ejaculation; low back pain; rectal pain along the inner thighs; urinary symptoms; decreased libido; and impotence.

Men with noninflammatory prostatitis have symptoms resembling those of nonbacterial prostatitis but have negative urine culture results and no evidence of prostatic inflammation (*i.e.*, normal leukocyte count). The cause of noninflammatory prostatitis is unknown, but because of the absence of inflammation, the search for the cause of symptoms has been directed toward extraprostatic sources. In some cases, there is an apparent functional obstruction of the bladder neck near the external urethral sphincter; during voiding, this results in higher than normal pressures in the prostatic urethra that cause intraprostatic urine reflux and chemical irritation of the prostate by urine. In other cases, there is an apparent myalgia (*i.e.*, muscle pain) associated with prolonged tension of the pelvic floor muscles. Emotional stress also may play a role.

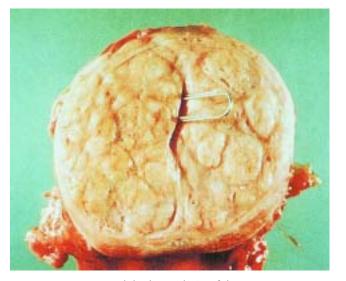
Treatment methods for chronic prostatitis/pelvic pain syndrome are highly variable. Antibiotic therapy is used when an occult infection is suspected. Sitz baths and nonsteroidal antiinflammatory agents may provide some symptom relief. In men with irritative urination symptoms, anticholinergic agents or α -adrenergic-blocking agents may be beneficial.

Benign Prostatic Hyperplasia

Benign prostatic hyperplasia (BPH) is an age-related, nonmalignant enlargement of the prostate gland (Fig. 33-11). It is characterized by the formation of large, nodular lesions in



■ FIGURE 33-12 ■ Normal prostate, nodular benign prostatic hypertrophy, and cancer of the prostate. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 954]. Philadelphia: Lippincott Williams & Wilkins.) (Artist: Dimitri Karetnikov)



■ FIGURE 33-11 ■ Nodular hyperplasia of the prostate. Cut surface of a prostate enlarged by nodular hyperplasia shows numerous, well-circumscribed nodules of prostatic tissue. The prostatic urethra (*paper clip*) has been compressed to a narrow slit. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 955]. Philadelphia: Lippincott Williams & Wilkins)

the periurethral region of the prostate, rather than the peripheral zones, which commonly are affected by prostate cancer (Fig. 33-12). BPH is one of the most common diseases of aging men. It has been reported that 25% of men older than 55 years of age and 50% of men older than 75 years of age experience symptoms of BPH.²⁸

The exact cause of the BPH is unknown. Potential risk factors include age, family history, race, ethnicity, and hormonal factors. The incidence of BPH increases with advancing age, is highest in African Americans, and is lowest in native Japanese. Men with a family history of BPH are reported to have had larger prostates than those of control subjects, and higher rates of BPH were found in monozygotic twins than in dizygotic twins.²⁸

Both androgens (testosterone) and estrogens appear to contribute to the development of BPH. Dihydrotestosterone (DHT), the biologically active metabolite of testosterone, is thought to be the ultimate mediator of prostatic hyperplasia, with estrogen serving to sensitize the prostatic tissue to the growth-producing effects of DHT. Free plasma testosterone enters prostatic cells, where at least 90% is converted into DHT by

KEY CONCEPTS

HYPERPLASIA AND CANCER OF THE PROSTATE

- The prostate gland surrounds the urethra, and periurethral enlargement will cause manifestations of urinary obstruction.
- Benign prostatic hyperplasia is an age-related enlargement of the prostate gland with formation of large, discrete lesions in the periurethral region of the prostate. These lesions compress the urethra and produce symptoms of dysuria or difficulty urinating.
- Prostatic cancer begins in the peripheral zones of the prostate gland and usually is asymptomatic until the disease is far advanced and the tumor has eroded the outer prostatic capsule and spread to adjacent pelvic tissues or metastasized.

the action of 5α -reductase. The discovery that DHT is the active factor in BPH is the rationale for use of 5α -reductase inhibitors in the treatment of the disorder. Although the exact source of estrogen is uncertain, small amounts of estrogen are produced in the male. It has been postulated that a relative increase in estrogen levels that occurs with aging may facilitate the action of androgens in the prostate despite a decline in the testicular output of testosterone.

The anatomic location of the prostate at the bladder neck contributes to the pathophysiology and symptomatology of BPH. There are two prostatic components to the obstructive properties of BPH and the development of lower urinary tract symptoms: dynamic and static.28,29 The static component of BPH is related to an increase in prostatic size and gives rise to symptoms such as a weak urinary stream, postvoid dribbling, frequency of urination, and nocturia. The dynamic component of BPH is related to prostatic smooth muscle tone. The α_1 adrenergic receptors are the main receptors for the smooth muscle component of the prostate. The recognition of the role of α_1 -adrenergic receptor on neuromuscular function in the prostate is the basis for use of α_1 -adrenergic-receptor blockers in treating BPH. A third component, detrusor instability and impaired bladder contractility, may contribute to the symptoms of BPH independent of the outlet obstruction created by an enlarged prostate (see Chapter 25).^{29,30} It has been suggested that some of the symptoms of BPH might be related to a decompensating or aging bladder, rather than being primarily related to outflow obstruction. An example is the involuntary contraction that results in urgency and an attempt to void that occurs because of small bladder volume.29

Clinical Course. The clinical significance of BPH resides in its tendency to compress the urethra and cause partial or complete obstruction of urinary outflow. As the obstruction increases, acute retention may occur with overdistention of the bladder. The residual urine in the bladder causes increased frequency of urination and a constant desire to empty the bladder, which be-

comes worse at night. With marked bladder distention, overflow incontinence may occur with the slightest increase in intra-abdominal pressure. The resulting obstruction to urinary flow can give rise to urinary tract infection, destructive changes of the bladder wall, hydroureter, and hydronephrosis. Hypertrophy and changes in bladder wall structure develop in stages. Initially, the hypertrophied fibers form trabeculations and then herniations, or sacculations; finally, diverticula develop as the herniations extend through the bladder wall (Chapter 25, Fig. 25-4). Because urine seldom is completely emptied from them, these diverticula are readily infected. Back pressure on the ureters and collecting system of the kidneys promotes hydroureter, hydronephrosis, and eventual renal failure.

It is now thought that the single most important factor in the evaluation and treatment of BPH is the man's own experiences related to the disorder. The American Urological Society Symptom Index consists of seven questions about symptoms regarding incomplete emptying, frequency, intermittency, urgency, weak stream, straining, and nocturia.^{31,32} Each question is rated with a score of 0 (mild) to 7 (severe). A maximum score of 35 indicates severe symptoms. Total scores below 7 are considered mild; those between 8 and 20, moderate; and scores over 20, severe. A final question relates to quality of life related to urinary problems.

Diagnosis of BPH is based on history, physical examination, digital rectal examination, urinalysis, blood tests for serum creatinine and prostate-specific antigen (PSA), and urine flow rate. Blood and urine analyses are used as adjuncts to determine BPH complications. Urinalysis is done to detect bacteria, white blood cells, or microscopic hematuria in the presence of infection and inflammation. The serum creatinine test is used as an estimate of the glomerular filtration rate and kidney function. The PSA test is used to screen for prostatic cancer. These evaluation measures, along with the symptom index, are used to describe the extent of obstruction, determine if other diagnostic tests are needed, and establish the need for treatment.

Transabdominal or transrectal diagnostic ultrasonography can be used to evaluate the kidneys, ureters, and bladder. Abdominal radiographs may be used to reveal the size of the gland. Urethrocystoscopy is indicated in men with a history of hematuria, stricture disease, urethral injury, or prior lower urinary tract surgery. It is used to evaluate the length and diameter of the urethra, the size and configuration of the prostate, and bladder capacity. It also detects the presence of trabeculations, bladder stones, and small bladder cancers. CT scans, MRI studies, and radionuclide scans are reserved for rare instances of tumor detection.

Treatment of BPH is determined by the degree of symptoms that the condition produces and complications due to obstruction. When a man develops mild symptoms related to BPH, a "watch and wait" stance often is taken. The condition does not always run a predictable course; it may remain stable or even improve. Until the 1980s, surgery was the mainstay of treatment to alleviate urinary obstruction due to BPH. Currently, there is an emphasis on less invasive methods of treatment, including use of pharmacologic agents. However, when more severe signs of obstruction develop, surgical treatment (*e.g.*, transurethral resection of the prostate [TURP]) usually is indicated to provide comfort and avoid serious renal damage. For men who have heart or lung disease or a condition that precludes major surgery, a stent may be used to widen and maintain the patency of the urethra.

Pharmacologic management includes the use of 5α -reductase inhibitors and α_1 -adrenergic–blocking drugs.^{28,29} The 5α reductase inhibitors such as finasteride reduce prostate size by blocking the effect of androgens on the prostate. The presence of α -adrenergic receptors in prostatic smooth muscle has prompted the use of α_1 -adrenergic–blocking drugs to relieve prostatic obstruction and increase urine flow.

Cancer of the Prostate

Prostatic cancer is the most common male cancer in the United States and is second to lung cancer as a cause of cancer-related death in men. The American Cancer Society estimates that during 2002, approximately 189,000 men in the United States received a diagnosis of prostate cancer, and 32,200 men died of the disorder.³³ The increase in diagnosed cases is thought to reflect earlier diagnosis because of the widespread use of PSA testing since the early 1990s.³⁴ The incidence of prostate cancer varies markedly from country to country and varies among races in the same country.³³ African-American men have the highest reported incidence for prostate cancer at all ages, and Asians and Native American men have the lowest rate. Prostate cancer also is a disease of aging. The incidence increases rapidly after 50 years of age; more than 80% of all prostate cancers are diagnosed in men older than 65 years of age.³³

The precise cause of prostatic cancer is unclear. As with other cancers, it appears that the development of prostate cancer is a multistep process involving genes that control cell differentiation and growth (see Chapter 5). Several risk factors, such as age, race, heredity, and environmental influences, are suspected of playing a role.^{35,36} Male hormone levels also may play a role. There is insufficient evidence linking socioeconomic status, infectious agents, smoking, vasectomy, sexual behavior, or BPH to the pathogenesis of prostate cancer.

The incidence of prostate cancer appears to be higher in relatives of men with prostate cancer. Diet may also play a role. It has been suggested that a diet high in fats may alter the production of sex hormones and increase the risk of prostate cancer. Supporting the role of dietary fats as a risk factor for prostate cancer has been the observation that the diet of Japanese men, who have a low rate of prostate cancer, is much lower in fat content than that of U.S. men, who have a much higher incidence.

In terms of hormonal influence, androgens are believed to play a role in the pathogenesis of prostate cancer.³⁵ Evidence favoring a hormonal influence includes the presence of steroid receptors in the prostate, the requirement of sex hormones for normal growth and development of the prostate, and the fact that prostate cancer almost never develops in men who have been castrated. The response of prostatic cancer to estrogen administration or androgen deprivation further supports a correlation between the disease and testosterone levels.

Prostatic adenocarcinomas, which account for 98% of all primary prostatic cancers, are commonly multicentric and located in the peripheral zones of the prostate³⁶ (see Fig. 33-12). The high frequency of invasion of the prostatic capsule by adenocarcinoma relates to its subcapsular location. Invasion of the urinary bladder is less common and occurs later in the clinical course. Metastasis to the lung reflects lymphatic spread through the thoracic duct and dissemination from the prostatic venous plexus to the inferior vena cava. Bony metastases, particularly to the vertebral column, ribs, and pelvis, produce pain that often presents as a first sign of the disease.

Most men with early-stage prostate cancer are asymptomatic. The presence of symptoms often suggests locally advanced or metastatic disease. Depending on the size and location of prostatic cancer at the time of diagnosis, there may be changes associated with the voiding pattern similar to those found in BPH. These include urgency, frequency, nocturia, hesitancy, dysuria, hematuria, or blood in the ejaculate. On physical examination, the prostate is nodular and fixed. Bone metastasis often is characterized by low back pain. Pathologic fractures can occur at the site of metastasis. Men with metastatic disease may experience weight loss, anemia, or shortness of breath.

Screening. Because early cancers of the prostate usually are asymptomatic, screening tests are important. The screening tests currently available are digital rectal examination, PSA testing, and transrectal ultrasonography. PSA is a glycoprotein secreted into the cytoplasm of benign and malignant prostatic cells that is not found in other normal tissues or tumors. However, a positive PSA test indicates only the possible presence of prostate cancer. It also can be positive in cases of BPH and prostatitis. The American Cancer Society and the American Urological Association recommend that men 50 years of age or older should undergo annual measurement of PSA and rectal examination for early detection of prostate cancer.³³ Men at high risk for prostate cancer, such as blacks and those with a strong family history, should undergo annual screening beginning at 45 years of age.³³

Diagnosis. The diagnosis of prostate cancer is based on history and physical examination and confirmed through biopsy methods. Transrectal ultrasonography, a continuously improving method of imaging, is used to guide a biopsy needle and document the exact location of the biopsied tissue. It also is used for providing staging information.²⁸ Newly developed small probes for transrectal MRI have been shown to be effective in detecting the presence of cancer in the prostate. Radiologic examination of the bones of the skull, ribs, spine, and pelvis can be used to reveal metastases, although radionuclide bone scans are more sensitive. Excretory urograms are used to delineate changes due to urinary tract obstruction and renal involvement.

Cancer of the prostate, like other forms of cancer, is graded and staged (see Chapter 5). Prostatic adenocarcinoma commonly is classified using the Gleason grading system.^{35,36} Welldifferentiated tumors are assigned a grade of 1, and poorly differentiated tumors a grade of 5. Two tumor markers, PSA and serum acid phosphatase, are important in the staging and management of prostatic cancer. In untreated cases, the level of PSA correlates with the volume and stage of disease.³⁷ A rising PSA after treatment is consistent with progressive disease, whether it is locally recurring or metastatic. Measurement of PSA is used to detect recurrence after total prostatectomy. Because the prostate is the source of PSA, levels of the antigen should drop to zero after surgery; a rising PSA indicates recurring disease. Serum acid phosphatase is less sensitive than PSA and is used less frequently. However, it is more predictive of metastatic disease and may be used for that purpose.

Treatment. Cancer of the prostate is treated by surgery, radiation therapy, and hormonal manipulations.^{28,38,39} Chemotherapy has shown limited effectiveness in the treatment of prostate cancer. Treatment decisions are based on tumor grade and stage and on the age and health of the man. Expectant therapy (watching and waiting) may be used if the tumor is not producing symptoms, is expected to grow slowly, and is small and contained in one area of the prostate. This approach is particularly suited for men who are elderly or have other health problems. Most men with an anticipated survival greater than 10 years are considered for surgical or radiation therapy.³⁹ Radical prostatectomy involves complete removal of the seminal vesicles, prostate, and ampullae of the vas deferens. Radiation therapy can be delivered by a variety of techniques, including external beam radiation therapy and transperineal implantation of radioisotopes.

Metastatic disease often is treated with androgen deprivation therapy. Androgen deprivation may be induced at several levels along the pituitary-gonadal axis using a variety of methods or agents. Orchiectomy or estrogen therapy often is effective in reducing symptoms and extending survival. The GnRH analogs (*e.g.*, leuprolide, buserelin, nafarelin) block luteinizing hormone release from the pituitary and reduce testosterone levels without orchiectomy or estrogen therapy. When given continuously and in therapeutic doses, these drugs desensitize GnRH receptors in the pituitary, thereby preventing the release of luteinizing hormone. The antiandrogens block the uptake and actions of androgens in the target tissues. Complete androgen blockade can be achieved by combining an antiandrogen with a GnRH agent or orchiectomy.

In summary, disorders of the penis include balanitis, an acute or chronic inflammation of the glans penis, and balanoposthitis, an inflammation of the glans and prepuce. Peyronie's disease is characterized by the growth of a band of fibrous tissue on top of the penile shaft. Priapism is prolonged, painful, and nonsexual erection that can lead to thrombosis with ischemia and necrosis of penile tissue. Cancer of the penis accounts for less than 1% of male genital cancers in the United States. Although the tumor is slow growing and highly curable when diagnosed early, the greatest hindrance to successful treatment is a delay in seeking medical attention.

Disorders of the scrotum and testes include hydrocele, hematocele, spermatocele, varicocele, and testicular torsion. Inflammatory conditions can involve the scrotal sac, epididymis, or testes. Tumors can arise in the scrotum or the testes. Scrotal cancers usually are associated with exposure to petroleum products such as tar, pitch, and soot. Testicular cancers account for 1% of all male cancers and 3% of cancers of the male genitourinary system. With current treatment methods, a large percentage of men with these tumors can be cured. Testicular self-examination is recommended as a means of early detection of this form of cancer.

The prostate is a firm, glandular structure that surrounds the urethra. Inflammation of the prostate occurs as an acute or a chronic process. Chronic prostatitis probably is the most common cause of relapsing urinary tract infections in men. BPH is a common disorder in men older than 50 years. Because the prostate encircles the urethra, BPH exerts its effect through obstruction of urinary outflow from the bladder. Advances in the treatment of BPH include laser surgery, balloon dilatation, prostatic stents, and pharmacologic treatment.

Prostatic cancer is the most common male cancer in the United States and is second to lung cancer as a cause of cancer-related death in men. A recent increase in diagnosed cases is thought to reflect earlier diagnosis because of widespread use of PSA testing. Most prostate cancers are asymptomatic and are incidentally discovered on rectal examination. Cancer of the prostate, like other forms of cancer, is graded according to the histologic characteristics of the tumor and staged clinically using the TNM system. Treatment, which is based on the extent of the disease, includes surgery, radiation therapy, and hormonal manipulation.

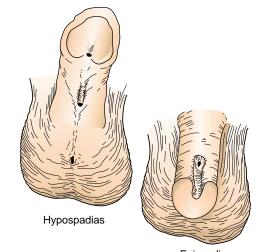
DISORDERS IN CHILDHOOD AND AGING CHANGES

Disorders of Childhood

Disorders of the male reproductive system that present in childhood include hypospadias, epispadias, phimosis and paraphimosis, and cryptorchidism.

Hypospadias and Epispadias

Hypospadias and epispadias are congenital disorders of the penis resulting from embryologic defects in the development of the urethral groove and penile urethra (Fig. 33-13). In hypospadias, which affects approximately 1 in 300 male infants, the termination of the urethra is on the ventral surface of the penis.^{34,40} The testes are undescended in 10% of boys born with hypospadias and chordee (*i.e.*, ventral bowing of the penis), and inguinal hernia also may accompany the disorder. In the newborn with severe hypospadias and cryptorchidism (undescended testes), the differential diagnosis should consider ambiguous genitalia and masculinization



Epispadias

FIGURE 33-13 Hypospadias and epispadias.

that is seen in females with congenital adrenal hyperplasia (see Chapter 31). Because many chromosomal aberrations result in ambiguity of the external genitalia, chromosomal studies often are recommended for male infants with hypospadias and cryptorchidism.³⁴

Surgery is the treatment of choice for hypospadias.³⁴ Circumcision is avoided because the foreskin is used for surgical repair. Factors that influence the timing of surgical repair include anesthetic risk, penile size, and the psychological effects of the surgery on the child. In mild cases, the surgery is done for cosmetic reasons only. In more severe cases, surgical repair becomes essential for normal sexual functioning and to prevent the psychological effects of having malformed genitalia. When indicated, surgical repair is usually done between the ages of 6 to 12 months.

Epispadias, in which the opening of the urethra is on the dorsal surface of the penis, is a less common defect. Although epispadias may occur as a separate entity, it often is associated with exstrophy of the bladder, a condition in which the abdominal wall fails to cover the bladder. The treatment depends on the extent of the developmental defect.

Phimosis and Paraphimosis

Phimosis refers to a tightening of the prepuce or penile foreskin that prevents its retraction over the glans. Embryologically, the foreskin begins to develop during the eighth week of gestation as a fold of skin at the distal edge of the penis that eventually grows forward over the base of the glans.⁴⁰ By the 16th week of gestation, the prepuce and the glans are adherent. Only a small percentage of newborns have a fully retractable foreskin. With growth, a space develops between the glans and foreskin, and by 3 years of age, approximately 90% of male children have retractable foreskins.

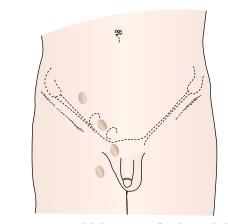
Because the foreskin of many boys cannot be fully retracted in early childhood, it is important that the area be cleaned thoroughly. There is no need to retract the foreskin forcibly because this could lead to infection, scarring, or paraphimosis. As the child grows, the foreskin becomes retractable, and the glans and foreskin should be cleaned routinely. If symptomatic phimosis occurs after childhood, it can cause difficulty with voiding or sexual activity. Circumcision is then the treatment of choice.

In a related condition called *paraphimosis*, the foreskin is so tight and constricted that it cannot cover the glans. A tight foreskin can constrict the blood supply to the glans and lead to ischemia and necrosis. Many cases of paraphimosis result from the foreskin being retracted for an extended period, as in the case of catheterized uncircumcised males.

Cryptorchidism

Cryptorchidism, or undescended testes, occurs when one or both of the testicles fail to move down into the scrotal sac. The condition is bilateral in 10% to 20% of cases. The testes develop intra-abdominally in the fetus and usually descend into the scrotum through the inguinal canal during the seventh to ninth months of gestation.⁴⁰ The undescended testes may remain in the lower abdomen or at a point of descent in the inguinal canal (Fig. 33-14).

The incidence of cryptorchidism is directly related to birth weight and gestational age; infants who are born prematurely



■ FIGURE 33-14 ■ Possible locations of undescended testicles.

or are small for gestational age have the highest incidence of the disorder. Up to one third of premature infants and 3% to 5% of full-term infants are born with undescended testicles.^{35,41} The cause of cryptorchidism in full-term infants is poorly understood. Most cases are idiopathic, but some may result from genetic or hormonal factors.²⁹

The major manifestation of cryptorchidism is the absence of one or more of the testes in the scrotum. The testis either is not palpable or can be felt external to the inguinal ring. Spontaneous descent often occurs during the first 3 months of life, and by 6 months of age the incidence decreases to 0.8%.^{34,41} Spontaneous descent rarely occurs after 6 months of age.

In children with cryptorchidism, histologic abnormalities of the testes reflect intrinsic defects in the testicle or adverse effects of the extrascrotal environment. The undescended testicle is normal at birth, but pathologic changes can be demonstrated at 6 to 12 months.³⁴ There is a delay in germ cell development, changes in the spermatic tubules, and reduced number of Leydig cells. These changes are progressive if the testes remain undescended. When the disorder is unilateral, it also may produce morphologic changes in the contralateral descended testis.

The consequences of cryptorchidism include infertility, malignancy, and the possible psychological effects of an empty scrotum. Indirect inguinal hernias usually accompany the undescended testes but rarely are symptomatic. Recognition of the condition and early treatment are important steps in preventing adverse consequences. The risk of malignancy in the undescended testis is four to six times higher than in the general population.^{34,41} The increased risk of testicular cancer is not significantly affected by orchiopexy, hormonal therapy, or late spontaneous descent after the age of 2 years. However, orchiopexy does allow for earlier detection of a testicular malignancy by positioning the testis in a more easily palpable location.

As a group, males with unilateral or bilateral cryptorchidism usually have decreased sperm counts, poorer-quality sperm, and lower fertility rates than do men whose testicles descend normally. The likelihood of decreased fertility increases when the condition is bilateral. Unlike the risk of testicular cancer, there seems to be some advantage to early orchiopexy for protection of fertility.^{34,41}

Diagnosis is based on careful examination of the genitalia in male infants. Undescended testes due to cryptorchidism should be differentiated from retractable testes that retract into the inguinal canal in response to an exaggerated cremaster muscle reflex. Retractable testes usually are palpable at birth but become nonpalpable later. They can be brought down with careful palpation in a warm room. Retractable testes usually assume a scrotal position during puberty. They have none of the complications associated with undescended testicles due to cryptorchidism.³⁴

Improved techniques for testicular localization include ultrasonography (*i.e.*, visualization of the testes by recording the pulses of ultrasonic waves directed into the tissues), gonadal venography and arteriography (*i.e.*, radiography of the veins and arteries of the testes after the injection of a contrast medium), and laparoscopy (*i.e.*, examination of the interior of the abdomen using a visualization instrument).

The treatment goals for the child with cryptorchidism include measures to enhance future fertility potential, placement of the gonad in a favorable place for cancer detection, and improved cosmetic appearance. Regardless of the type of treatment used, it should be carried out between 6 months and 2 years of age.34,41 Treatment modalities for children with unilateral or bilateral cryptorchidism include initial hormone therapy with hCG or luteinizing hormone-releasing hormone (LHRH), a hypothalamic hormone that stimulates production of the gonadotropic hormones by the anterior pituitary gland. For children who do no respond to hormonal treatment, surgical placement and fixation of the testes in the scrotum (*i.e.*, orchiopexy) have proved effective. Approximately 95% of infants who have orchiopexy for a unilateral undescended testis will be fertile, compared with a 30% to 50% fertility rate in uncorrected males.41

Treatment of males with undescended testis should include lifelong follow-up, considering the sequelae of testicular cancer and infertility. Parents need to be aware of the potential issues of infertility and increased risk of testicular cancer. On reaching puberty, boys should be instructed in the necessity of testicular self-examination.

Aging Changes

Like other body systems, the male reproductive system undergoes degenerative changes as a result of the aging process; it becomes less efficient with age. The declining physiologic efficiency of male reproductive function occurs gradually and involves the endocrine, circulatory, and neuromuscular systems.⁴² Compared with the marked physiologic change in aging females, the changes in the aging male are more gradual and less drastic. Gonadal and reproductive failure usually are not related directly to age because a man remains fertile into advanced age; 80- and 90-year-old men have been known to father children.

As the male ages, his reproductive system becomes measurably different in structure and function from that of the younger male. Male sex hormone levels, particularly of testosterone, decrease with age, with the decline starting later on average than in women. The term *andropause* has been used to describe an illdefined collection of symptoms in aging men, typically those older than 50 years, who may have a low androgen level.⁴³

The sex hormones play a part in the structure and function of the reproductive system and other body systems from conception to old age; they affect protein synthesis, salt and water balance, bone growth, and cardiovascular function. Decreasing levels of testosterone affect sexual energy, muscle strength, and the genital tissues. The testes become smaller and lose their firmness. The seminiferous tubules, which produce spermatozoa, thicken and begin a degenerative process that finally inhibits sperm production, resulting in a decrease of viable spermatozoa. The prostate gland enlarges, and its contractions become weaker. The force of ejaculation decreases because of a reduction in the volume and viscosity of the seminal fluid. The seminal vesicle changes little from childhood to puberty. The pubertal increases in the fluid capacity of the gland remain throughout adulthood and decline after the age of 60 years. After age 60 years, the walls of the seminal vesicles thin, the epithelium decreases, and the muscle layer is replaced by connective tissue. Age-related changes in the penis consist of fibrotic changes in the trabeculae in the corpus spongiosum, with progressive sclerotic changes in arteries and veins. Sclerotic changes also follow in the corpora cavernosa, with the condition becoming generalized in 55- to 60-year-old men.

Erectile dysfunction in the elderly male often is directly related to the general physical condition of the person. Diseases that accompany aging can have direct bearing on male reproductive function. Various cardiovascular, respiratory, hormonal, neurologic, and hematologic disorders can be responsible for secondary impotence. For example, vascular disease affects male potency because it may impair blood flow to the pudendal arteries or their tributaries, resulting in loss of blood volume with subsequent poor distention of the vascular spaces of erectile tissue. Other diseases affecting potency include hypertension, diabetes, cardiac disease, and malignancies of the reproductive organs. In addition, certain medications can have an effect on sexual function.

Testosterone and other synthetic androgens may be used in older males with low androgen levels to improve muscle strength and vigor. Preliminary studies of androgen replacement in aging males with low androgen levels show an increase in lean body mass and a decrease in bone turnover. Before testosterone replacement therapy is initiated, all men should be screened for prostate cancer. Testosterone is available as an injectable form that is administered every 2 to 3 weeks or as a transdermal patch or gel. Side effects of replacement therapy may include acne, gynecomastia, and reduced HDL levels. It also may contribute to a worsening of sleep apnea in men who are troubled by this problem.

In summary, childhood disorders of the male reproductive system include congenital disorders in which the urethral opening is located on the ventral surface of the penis (hypospadias) or on the dorsal surface (epispadias). Phimosis is the condition in which the opening of the foreskin is too tight to permit retraction over the glans. Disorders of the scrotum and testes include cryptorchidism or undescended testicles. Early diagnosis and treatment is important because of the risk of malignancy and infertility. Like other body systems, the male reproductive system undergoes changes as a result of the aging process. The changes occur gradually and involve parallel changes in endocrine, circulatory, and neuromuscular function. Testosterone levels decrease, the size and firmness of the testes decrease, sperm production declines, and the prostate gland enlarges. There usually is a decrease in frequency of intercourse, intensity of sensation, speed of attaining erection, and force of ejaculation.

REVIEW QUESTIONS

Describe the structure and function of the male reproductive system and hormones and relate to the process of spermatogenesis.

Relate the development and descent of testes to the pathogenesis of inguinal hernia and cryptorchidism.

Compare the pathophysiology of cancer of the penis, scrotum, and testes.

Compare the cause, appearance, and significance of hydrocele, hematocele, spermatocele, and varicocele.

State the difference between extravaginal and intravaginal testicular torsion.

Compare the pathology and symptoms of acute bacterial prostatitis, chronic bacterial prostatitis, and chronic prostatitis/pelvic pain syndrome.

Describe the urologic manifestations and treatment of benign prostatic hyperplasia.

List the methods used in the diagnosis and treatment of prostatic cancer.

Describe the autonomic nervous system control of erection, emission, and ejaculation and relate to the pathogenesis of erectile dysfunction and priapism.

State the difference between hypospadias and epispadias.

Describe changes in the male reproductive system that occur with aging.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- Guyton A.C., Hall J.E. (2001). *Textbook of medical physiology* (10th ed., p. 913). Philadelphia: W.B. Saunders.
- Braunstein G.D. (2001). Testes. In Greenspan F.S., Gardner D.G. (Eds.), *Basic and clinical endocrinology* (6th ed., pp. 422–452). New York: Lange Medical Books/McGraw-Hill.
- Anderson K.E., Wagner G. (1995). Physiology of penile erection. Physiology Review 75, 191–236.
- 4. NIH Consensus Development Panel on Impotence. (1993). NIH Consensus Conference: Impotence. *Journal of the American Medical Association* 270, 83–90.

- 5. Lue T.F. (2000). Erectile dysfunction. *New England Journal of Medicine* 342, 1802–1813.
- Levine L.A. (2000). Diagnosis and treatment of erectile dysfunction. *American Journal of Medicine* 109 (Suppl. 9A), 3S–12S.
- 7. Kaiser F.E. (1999). Erectile dysfunction in the aging man. *Medical Clinics of North America* 83, 1267–1278.
- 8. Vohra S., Badlani G. (1992). Balanitis and balanoposthitis. Urologic Clinics of North America 19, 143–147.
- Fitkin J., Ho G.T. (1999). Peyronie's disease: Current management. *American Family Physician* 60, 549–554.
- McAninch J.W. (2000). Disorders of the penis and male urethra. In Tanagho E.A., McAninch J.W. (Eds.), *Smith's general urology* (15th ed., pp. 663–675). New York: Lange Medical Books/McGraw-Hill.
- 11. Harmon W.J., Nehra A. (1997). Priapism: Diagnosis and treatment. *Mayo Clinic Proceedings* 72, 350–355.
- Kachhi P.N., Henderson S.O. (2000). Priapism after androstenedione intake for athletic performance enhancement. *Annals of Emergency Medicine* 35, 391–393.
- 13. Krieg R., Hoffman R. (1999). Current management of unusual genitourinary cancers: Part 1. Penile cancer. *Oncology* 13, 1347–1352.
- Presti J.C., Herr H.W. (2000). Genital tumors. In Tanagho E.A., McAninch J.W. (Eds.), *Smith's general urology* (15th ed., pp. 430–434). New York: Lange Medical Books/McGraw-Hill.
- Kapur P., Caty M.G., Glick P.L. (1998). Pediatric hernias and hydroceles. *Pediatric Clinics of North America* 45, 773–789.
- McAninch J.W. (2000). Disorders of the testis, scrotum, and spermatic cord. In Tanagho E.A., McAninch J.W. (Eds.), *Smith's general urology* (15th ed., pp. 684–693). New York: Lange Medical Books/ McGraw-Hill.
- Galejs L.E., Kass E.J. (1999). Diagnosis and treatment of acute scrotum. American Family Physician 59, 817–824.
- Meares E.M. (2000). Nonspecific infections of the genitourinary tract. In Tanagho E.A., McAninch J.W. (Eds.), *Smith's general urology* (15th ed., pp. 237–238). Norwalk, CT: Appleton & Lange.
- Mebcow M.M. (1975). Percivall Pott (1713–1788): 200th anniversary of first report of occupation-induced cancer of the scrotum in chimney sweepers (1745). Urology 6, 745.
- 20. Lowe F.C. (1992). Squamous cell carcinoma of the scrotum. Urologic Clinics of North America 19, 297–405.
- Bosl G.J. (1997). Testicular germ-cell cancer. New England Journal of Medicine 337, 242–252.
- 22. Pillai S.B., Besner G.E. (1998). Pediatric testicular problems. *Pediatric Clinics of North America* 45, 813–818.
- Kinade S. (1999). Testicular cancer. American Family Physician 59, 2539–2544.
- Krieger J.N., Nyberg L., Nickel J.C. (1999). NIH consensus definition and classification of prostatitis. *Journal of the American Medical Association* 282, 721–725.
- McRae S.N., Shortliffe L.M.D. (2000). Bacterial infections of the genitourinary system. In Tanagho E.A., McAninch J.W. (Eds.), *Smith's* general urology (15th ed., pp. 254–259). New York: Lange Medical Books/McGraw-Hill.
- Stevermer J.J., Easley S.K. (2000). Treatment of prostatitis. American Family Physician 61, 3015–3026.
- Collins M.M., MacDonald R., Wilt T.J. (2000). Diagnosis and treatment of chronic abacterial prostatitis: A systemic review. *Annals of Internal Medicine* 133, 367–381.32.
- Presti J.C. (2000). Neoplasms of the prostate gland. In Tanagho E.A., McAninch J.W. (Eds.), *Smith's general urology* (15th ed., pp. 399–406). New York: Lange Medical Books/McGraw-Hill.
- 29. Zida A., Rosenblum M., Crawford E.D. (1999). Benign prostatic hyperplasia: An overview. *Urology* 53 (Suppl. 3A), 1–6.
- Elbadawi A. (1998). Voiding dysfunction in benign prostatic hyperplasia: Trends, controversies and recent revelations: Pathology and pathophysiology. *Urology* 51 (Suppl. 5A), 73–82.
- 31. Agency of Health Care Policy and Research. (1994). *Clinical practice guidelines for benign prostatic hyperplasia*. AHCPR publication

no. 94-0582. Rockville, MD: U.S. Department of Health and Human Services.

- Barry M.J., et al. (1992). The American Urological Association index of benign prostatic hypertrophy. *Journal of Urology* 148, 1549–1557.
- American Cancer Society. (2002). Prostate cancer resource center. [On-line]. Available: http://www.cancer.org.
- Behrman R.E., Kliegman R.M., Jenson H.B. (2000). Nelson textbook of pediatrics (16th ed., pp. 1546–1549, 1650–1651). Philadelphia: W.B. Saunders.
- 35. Cotran R.S., Kumar V., Collins T. (1999). *Robbins pathologic basis of disease* (6th ed., pp. 1015–1033). Philadelphia: W.B. Saunders.
- Peterson R.O. (1999). The urinary tract and male reproductive system. In Rubin E., Farber J.L. (Eds.), *Pathology* (3rd ed., pp. 956–960). Philadelphia: Lippincott Williams & Wilkins.
- Brawer M.K. (1999). Prostate-specific antigen: Current status. CA— A Cancer Journal for Clinicians 49, 264–265.

- Garnick M.B., Fair W.R. (1998). Combating prostate cancer. Scientific American 279 (6), 74–83.
- Stoller M.L., Presti J.C., Carroll P.R. (2001). Urology. In Tierney L.M., McPhee S.J., Papadakis M.A. (Eds.), *Current medical diagnosis and treatment* (40th ed., pp. 956–961). New York: Lange Medical Books/ McGraw-Hill.
- Moore K.L., Persaud T.V.N. (1998). The developing human: Clinically oriented embryology (6th ed., pp. 338–339). Philadelphia: W.B. Saunders.
- Docimo S.G., Silver R.I., Cromie W. (2000). The undescended testicle: Diagnosis and management. *American Family Physician* 62, 2037–2048.
- Merry B.J., Holehan A.M. (1994). Aging of the male reproductive system. In Timinas P.S. (Ed.), *Physiological basis of aging and geriatrics* (2nd ed., pp. 171–178). Boca Raton, FL: CRC Press.
- Morley J.E., Perry H.M. (1999). Androgen deficiency in aging men. Medical Clinics of North America 83, 1279–1289.

CHAPTER 34

Alterations in the Female Reproductive System

Structure and Function of the Female **Reproductive System External Genitalia Internal Genitalia** Vagina Cervix, Uterus, and Uterine Tubes **Ovaries Ovarian Hormones Ovarian Follicle Development and Ovulation Disorders of the Female Reproductive Organs Disorders of the External Genitalia** Bartholin Gland Cyst and Abscess Non-neoplastic Epithelial Disorders Vulvodynia Premalignant and Malignant Neoplasms **Disorders of the Vagina** Vaginitis Cancer of the Vagina **Disorders of the Uterine Cervix Cervicitis and Cervical Polyps** Cancer of the Cervix **Disorders of the Uterus** Infectious Disorders of the Uterus and **Pelvic Structures Endometriosis** Adenomyosis **Endometrial Cancer** Leiomyomas **Disorders of Pelvic Support and Uterine Position Disorders of Pelvic Support** Variations in Uterine Position **Disorders of the Ovaries Ovarian Cysts** Benign and Functioning Ovarian Tumors **Ovarian Cancer**

Menstrual Disorders Dysfunctional Menstrual Cycles Dysfunctional Bleeding Amenorrhea Dysmenorrhea **Premenstrual Syndrome** Menopause and Aging Changes **Disorders of the Breast Breast Structures Mastitis Ductal Disorders** Fibrocystic Disease **Breast Cancer** Detection **Diagnosis and Classification** Treatment

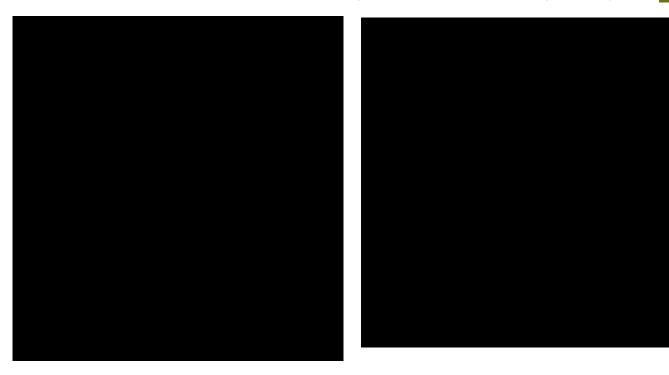
he reproductive function of the female is far more complex than the male. Not only must the female produce germ cells, but she must also nourish the developing embryo and prepare to nurse the infant once childbirth has occurred. This chapter includes a review of the structure and function of the female reproductive system and a discussion of disorders of the internal and external female reproductive organs.

STRUCTURE AND FUNCTION OF THE FEMALE REPRODUCTIVE SYSTEM

The female genitourinary system consists of the external and internal genital organs. The external sex organs of the female are referred to as the genitalia or vulva. The internal genital organs include the vagina, uterus, uterine tubes, and ovaries. These organs are largely located within the pelvic cavity (Fig. 34-1).

External Genitalia

The external genitalia are located at the base of the pelvis in the perineal area. The external genitalia, also called the vulva, include the mons pubis, labia majora, labia minora, clitoris, and



perineal body (Fig. 34-2). Because of their location, the urethra and anus usually are considered in a discussion of the external genitalia.

The *mons pubis* is a rounded skin-covered fat pad located anterior to the symphysis pubis. Running posteriorly from the mons pubis is two elongated hair-covered fatty folds, the *labia majora*. The labia majora are analogous to the male scrotum. The labia majora enclose the *labia minora*, which are smaller than the labia majora and are composed of skin, fat, and some

KEY CONCEPTS

THE FEMALE GENITOURINARY SYSTEM

- The female reproductive system, which consists of the external and internal genitalia, has both sexual and reproductive functions.
- The external genitalia (labia majora, labia minora, clitoris, and vestibular glands) surround the openings of the urethra and vagina. Although the female urinary and genital structures are anatomically separate, their close proximity provides a means for cross-contamination and shared symptomatology.
- The internal genitalia of the female reproductive system are specialized to participate in sexual intercourse (the vagina), to produce and maintain the female egg cells (the ovaries), to transport these cells to the site of fertilization (the fallopian tubes), to provide a favorable environment for development of the offspring (the uterus), and to produce the female sex hormones (the ovaries).

erectile tissue. The *clitoris* is located below the clitoral hood, which is formed by the joining of the two labia minora. The female clitoris is an erectile organ, rich in vascular and nervous supply. Analogous to the male penis, it is a highly sensitive organ that becomes distended during sexual stimulation.

The area between the labia minora is called the *vestibule*. Located in the vestibule are the urethral and vaginal openings and Bartholin's lubricating glands. The urethra is located posterior to the clitoris and usually is closer to the vaginal opening than to the clitoris. The urethral opening is the site of *Skene's glands*, which have a lubricating function. The vaginal orifice, commonly known as the *introitus*, is the opening between the external and internal genitalia.

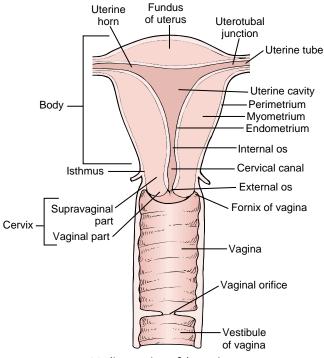
Internal Genitalia

Vagina

The vagina is a fibromuscular tube that connects the external and internal genitalia. The vagina, which is essentially free of sensory nerve fibers, is located behind the urinary bladder and urethra and anterior to the rectum (Fig. 34-3). The uterine cervix projects into the vagina at its upper end, forming recesses called *fornices*. The vagina functions as a route for discharge of menses and other secretions. It also serves as an organ of sexual fulfillment and reproduction.

Cervix, Uterus, and Uterine Tubes

Cervix. The round cervix forms the neck of the uterus. The opening, or os, of the cervix, forms a pathway between the uterus and the vagina. The vaginal opening is called the external os and the uterine opening, the *internal os*. The space between these two openings is the called the *endocervical canal*. Secretions from the columnar epithelium of the endocervix protect the uterus from infection, alter receptivity to sperm, and form a mucoid "plug" during pregnancy. The endocervical canal provides a route for menstrual discharge and entry of sperm.



■ FIGURE 34-3 ■ Median section of the vagina.

604

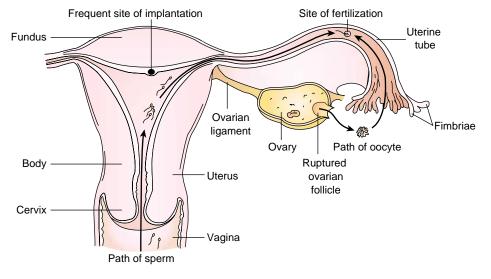
Uterus. The uterus is a thick-walled muscular organ. This pearshaped, hollow structure is located between the bladder and the rectum. The uterus can be divided into three parts: (1) the upper portion above the insertion of the fallopian tubes, called the *fundus*; (2) the central tapering portion, called the *body*; and (3) the inferior constricted part, called the *cervix* (Fig. 34-4).

The wall of the uterus is composed of three layers: the perimetrium, the myometrium, and the endometrium. The *perimetrium* is the outer serous covering that is derived from the abdominal peritoneum. This outer layer merges with the peritoneum that covers the broad ligaments. Anteriorly, the perimetrium is reflected over the bladder wall, forming the vesicouterine pouch; posteriorly, it extends to form the *cul-de-sac*, or *pouch of Douglas* (Fig. 34-3). Because of the proximity of the perimetrium to the urinary bladder, infection of this organ often causes uterine symptoms, particularly during pregnancy.

The middle muscle layer, the *myometrium*, forms the major portion of the uterine wall. It is continuous with the myometrium of the fallopian tubes and the vagina and extends into all the supporting ligaments with the exception of the broad ligaments. The inner fibers of the myometrium run in various directions, giving it an interwoven appearance. Contractions of these muscle fibers help to expel menstrual flow and the products of conception during miscarriage or childbirth.

The endometrium, or inner layer of the uterus, is continuous with the lining of the fallopian tubes and vagina. It consists of two distinct layers, or zones, that are responsive to hormonal stimulation: a basal layer and a functional layer.¹ The basal layer lies adjacent to the myometrium and is not sloughed during menstruation. The functional laver, which can be subdivided into a thin compact superficial layer and a deeper spongiosa layer, arises from the basal layer and undergoes proliferative changes and menstrual sloughing. The endometrial cycle can be divided into three phases: proliferative, secretory, and menstrual. The proliferative, or preovulatory, phase is the period during which the glands and stroma of the superficial layer grow rapidly under the influence of estrogen. The secretory, or postovulatory, phase is the period during which progesterone produces glandular dilation and active mucus secretion and the endometrium becomes highly vascular and edematous. The menstrual phase is the period during which the superficial layer degenerates and sloughs off.

Uterine Tubes. The *fallopian*, or uterine, tubes are slender cylindrical structures attached bilaterally to the uterus and supported by the upper folds of the broad ligament. The end of the fallopian tube nearest the ovary forms a funnel-like opening with fringed fingerlike projections, called *fimbriae*, that pick up the ovum following its release into the peritoneal cavity after ovulation (see Fig. 34-4). The fallopian tubes are formed of smooth muscle and lined with a ciliated, mucus-producing epithelial layer. The beating of the cilia, along with contractile movements of the smooth muscle, propels the nonmobile ovum toward the uterus. Besides providing a passageway for



■ FIGURE 34-4 ■ Schematic drawing of female reproductive organs, showing the path of the oocyte as it moves from the ovary into the fallopian (uterine) tube; the path of sperm is also shown, as is the usual site of fertilization.

ova and sperm, the fallopian tubes provide for drainage of tubal secretions into the uterus.

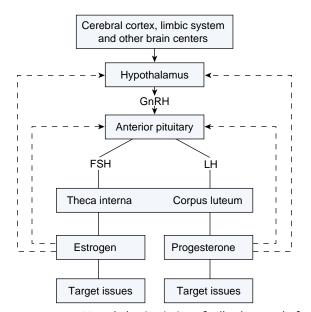
Ovaries

In the adult, the ovaries are flat, almond-shaped structures that are $4 \times 2.5 \times 1.5$ cm in dimension.² They are located on either side of the uterus below the ends of the two fallopian tubes. The ovaries are attached to the posterior surface of the broad ligament and to the uterus by the ovarian ligament (see Fig. 34-4). They are covered with a thin layer of surface epithelium that is continuous with the lining of the peritoneum. The integrity of this covering is periodically broken at the time of ovulation.

The tissues of the adult ovary can be conveniently divided into four compartments, or units: (1) the stroma, or supporting tissue; (2) the interstitial cells; (3) the follicles; and (4) the corpus luteum. The *stroma* is the connective tissue substance of the ovary in which the follicles are distributed. The *interstitial cells* are estrogen-secreting cells that resemble the Leydig's cells, or interstitial cells, of the testes. The *follicles* contain the female germ cells or ovum. The *corpus luteum* (*yellow body*) develops after expulsion of the ovum from the follicle.

Ovarian Hormones

The ovaries produce estrogens, progesterone, and androgens.³ Ovarian hormones are secreted in a cyclic pattern as a result of the interaction between the hypothalamic gonadotrophic releasing hormone (GnRH) and the pituitary gonadotropic hormones, follicle stimulating hormone (FSH), and luteinizing hormone (LH). The secretion of LH and FSH is stimulated by GnRH from the hypothalamus (Fig. 34-5). In addition to LH and FSH, the anterior pituitary secretes a third hormone called *prolactin*. The primary function of prolactin is the stimulation of lactation in the postpartum period. During pregnancy, prolactin and other hormones such as estrogen, progesterone, in-



■ **FIGURE 34-5** ■ Hypothalamic-pituitary feedback control of estrogen and progesterone levels in the female. The *dashed line* represents negative feedback.

sulin, and cortisol contribute to breast development in preparation for lactation.

Estrogens. Estrogens are a family of structurally related female sex hormones synthesized and secreted by cells in the ovaries and, in small amounts, by cells in the adrenal cortex. Androgens can be converted to estrogens peripherally, especially in fat tissue. Three estrogens occur naturally in humans: estrone, estradiol, and estriol. Of these, estradiol is the most biologically potent and the most abundantly secreted product of the ovary. Estrogens are secreted throughout the menstrual cycle. Two peaks occur: one before ovulation and the other in the middle of the luteal phase. Estrogens are transported in the blood bound to specific plasma globulins, inactivated and conjugated in the liver, and then excreted in the bile.

Estrogens are necessary for the normal female physical maturation. In concert with other hormones, estrogens provide for the reproductive processes of ovulation, implantation of the products of conception, pregnancy, parturition, and lactation by stimulating the development and maintaining the growth of the accessory organs. The estrogens stimulate the intrauterine development of the vagina, uterus, and uterine tubes in the embryo. They also stimulate the stromal development and ductal growth of the breasts at puberty, are responsible for the accelerated pubertal skeletal growth phase and for closure of the epiphyses of the long bones, contribute to the growth of axillary and pubic hair, and alter the distribution of body fat to produce the typical female body contours.

Estrogens have a number of important extragenital metabolic effects. They are responsible for maintaining the normal structure of skin and blood vessels in women. Estrogens decrease the rate of bone resorption by antagonizing the effects of calcitonin on bone; for this reason, osteoporosis is a common problem in estrogen-deficient postmenopausal women (see Chapter 43). In the liver, estrogens increase the synthesis of transport proteins for thyroxine, estrogen, testosterone, and other hormones. Estrogens also affect the composition of the plasma lipoproteins. They produce an increase in highdensity lipoproteins (HDLs), a slight reduction in low-density lipoproteins (LDLs), and a reduction in cholesterol levels (see Chapter 15). Estrogens also increase plasma triglyceride levels, and they enhance the coagulability of blood by effecting increased circulating levels of plasminogen and factors II, VII, IX, and X.

The estrogens cause moderate retention of sodium and water. Most women retain sodium and water and gain weight just before menstruation. This occurs because the estrogens facilitate the movement of intravascular fluids into the extracellular spaces, producing edema and increased sodium and water retention by the kidneys because of the decreased plasma volume. The actions of estrogens are summarized in Table 34-1.

Progesterone. Although the word *progesterone* refers to a substance that maintains pregnancy, progesterone is secreted as part of the normal menstrual cycle. The corpus luteum of the ovary secretes large amounts of progesterone after ovulation, and the adrenal cortex secretes small amounts. The hormone circulates in the blood attached to a specific plasma protein. It is metabolized in the liver and conjugated for excretion in the bile.

The local effects of progesterone on reproductive organs include the glandular development of the lobular and alveolar

606 Unit Nine: Alterations in the Male and Female Reproductive Systems

TABLE 34-1 Actions of Estrogens				
General Function	Specific Actions			
Growth and development				
Reproductive organs	Stimulate development of vagina, uterus, and fallopian tubes in utero and of secondary sex characteristics during puberty			
Skeleton	Accelerate growth of long bones and closure of epiphyses at puberty			
Reproductive processes				
Ovulation	Promote growth of ovarian follicles			
Fertilization	Alter the cervical secretions to favor survival and transport of sperm Promote motility of sperm within the fallopian tubes by decreasing mucus viscosity			
Implantation	Promote development of endometrial lining in the event of pregnancy			
Vagina	Proliferate and cornify vaginal mucosa			
Cervix	Increase mucus consistency			
Breasts	Stimulate stromal development and ductal growth			
General metabolic effects				
Bone resorption	Decrease rate of bone resorption			
Plasma proteins	Increase production of thyroid and other binding globulins			
Lipoproteins	Increase high-density and slightly decrease low-density lipoproteins			

tissue of the breasts and the cyclic glandular development of the endometrium. Progesterone also can compete with aldosterone at the level of the renal tubule, causing a decrease in sodium reabsorption, with a resultant increase in secretion of aldosterone by the adrenal cortex (as occurs in pregnancy). Although the mechanism is uncertain, progesterone increases basal body temperature and is responsible for the increase in body temperature that occurs with ovulation. Smooth muscle relaxation under the influence of progesterone plays an important role in maintaining pregnancy by decreasing uterine contractions and is responsible for many of the common discomforts of pregnancy, such as edema, nausea, constipation, flatulence, and headaches. The increased progesterone present during pregnancy and the luteal phase of the menstrual cycle enhances the ventilatory response to carbon dioxide, leading to a measurable change in arterial and alveolar carbon dioxide (PCO₂) levels.

Androgens. The normal female also produces androgens. Approximately 25% of these androgens are secreted from the ovaries, 25% from the adrenal cortex, and 50% from ovarian or adrenal precursors. In the female, androgens contribute to normal hair growth at puberty and may have other important metabolic effects.

Ovarian Follicle Development and Ovulation

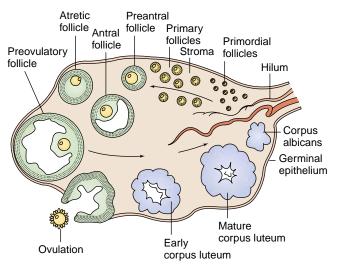
Unlike the male gonads, which produce sperm throughout a man's reproductive life, the female gonads contain a fixed number of ova at birth that diminishes throughout a woman's life. The process of oogenesis begins during the sixth week of fetal life and proceeds to the development of the primary oocytes, which become surrounded by a single layer of granulosa cells. The primary oocytes with their surrounding granulosa cells are referred to as *primordial follicles*. These primitive germ cells provide the 1 to 2 million oocytes that are present in the ovaries at birth. Throughout childhood, the granulosa cells provide nourishment for the ovum and secrete an in-

hibiting factor that keeps the ovum suspended in a primordial state.³ After puberty, when FSH and LH from the anterior pituitary begin to be secreted in sufficient amounts, the entire ovaries, together with some of the follicles within them, begin to grow.

Ovarian Cycle. The monthly series of events associated with the maturation of the ovum is called the ovarian cycle. It consists of two phases: the follicular phase and the luteal phase. The follicular phase, typically days 1 to 14, is the period of follicle growth. The luteal phase, days 14 to 28, is the period of corpus luteum activity. The typical ovarian cycle repeats at intervals of 28 days, with ovulation occurring midcycle. However, cycles as long as 40 days and as short as 21 days are not uncommon.

Follicles at all stages of development can be found in both ovaries, except in menopausal women (Fig. 34-6). Most follicles exist as primary follicles, each of which consists of a round oocyte surrounded by a single layer of flattened, epitheliumderived granulosa cells and a basement membrane. The primary follicles constitute an inactive pool of follicles from which all the ovulating follicles develop.

Under the influence of FSH and LH stimulation, 6 to 12 primary follicles develop into secondary follicles once every ovulatory cycle. During the development of the secondary follicle, the primary oocyte increases in size, and the surrounding granulosa cells proliferate to form a multilayered wall around it. In addition, cells from the surrounding ovarian interstitium align themselves to form a cellular wall called the *theca*. The cells of the theca become differentiated into two layers: an inner theca interna, which lies adjacent to the follicular cells, and an outer theca externa. The cells in the theca interna take on epithelioid characteristics similar to those of the granulosa cells and develop the ability to secrete additional sex hormones (estrogen and progesterone). The theca externa develops into a highly vascular connective tissue capsule that becomes the capsule for the developing follicle.



■ FIGURE 34-6 ■ Schematic diagram of an ovary, showing the sequence of events in the origin, growth, and rupture of an ovarian follicle and the formation and retrogression of a corpus luteum. The atretic follicles are those that show signs of degeneration and death.

As the follicle enlarges, a single large cavity, or *antrum*, is formed, and a portion of the granulosa cells and the oocytes are displaced to one side of the follicle by the fluid that accumulates. The secondary oocyte remains surrounded by a crown of granulosa cells, the *corona radiata*. As the granulosa cells continue to divide and the follicle grows, the thecal and granulosa cells cooperate to secrete a follicular fluid that contains a high concentration of estrogen. Selection of a dominant follicle occurs with the conversion to an estrogen microenvironment. The dominant follicle accumulates a greater mass of granulosa cells, and the theca becomes richly vascular, giving the follicle a hyperemic appearance. The lesser follicles, although continuing to produce some estrogen, atrophy or become atretic.

High levels of estrogen also exert a negative feedback effect on FSH, inhibiting further follicular development and causing an increase in LH levels. This represents the follicular stage of the menstrual cycle. As estrogen suppresses FSH, the actions of LH predominate, and the mature follicle bursts; the oocyte, along with the corona radiata, is ejected from the follicle. The ovum normally is then picked up and transported through the fallopian tube toward the uterus.

After ovulation, the follicle collapses, and the luteal stage of the menstrual cycle begins. The granulosa cells are invaded by blood vessels and yellow lipochrome-bearing cells from the theca layer. A rapid accumulation of blood and fluid forms a mass called the corpus luteum. Leakage of this blood onto the peritoneal surface that surrounds the ovary is thought to contribute to the *mittelschmerz* ("middle [or intermenstrual] pain") of ovulation. During the luteal stage, progesterone is secreted from the corpus luteum. If fertilization does not take place, the corpus luteum atrophies and is replaced by white scar tissue called the *corpus albicans*; the hormonal support of the endometrium is withdrawn, and menstruation occurs. In the event of fertilization, the corpus luteum remains functional for 3 months and provides hormonal support for pregnancy until the placenta is fully functional.

In summary, the genitourinary system as a whole serves sexual and reproductive functions throughout a woman's life. The female reproductive system consists of external and internal genitalia. The internal genitalia consist of the vagina, uterus, uterine tubes, and paired ovaries. The uterus is a thickwalled, muscular organ. The wall of the uterus is composed of three layers: the outer perimetrium; the myometrium or muscle layer, which is continuous with the myometrium of the fallopian tubes and the vagina; and the inner lining or endometrium, which is continuous with the lining of the fallopian tubes and vagina.

The gonads, or ovaries, which are internal in the female (unlike the testes in the male), have the dual function of storing the female germ cells, or ova, and producing the female sex hormones. Through the regulation and release of sex hormones, the ovaries influence the development of secondary sexual characteristics, regulation of menstrual cycles, maintenance of pregnancy, and advent of menopause.

DISORDERS OF THE FEMALE REPRODUCTIVE ORGANS

Disorders of the female genitourinary system have widespread effects on physical and psychological function, affecting sexuality and reproductive function. The reproductive organs are located close to other pelvic structures, particularly those of the urinary system, and disorders of the reproductive system may affect urinary function. This section of the chapter focuses on infections and benign and malignant disorders of the external and internal genitalia.

Disorders of the External Genitalia

Bartholin Gland Cyst and Abscess

A Bartholin gland cyst is a fluid-filled sac that results from the occlusion of the duct system of the gland. When the cyst becomes infected, the contents become purulent; if the infection goes untreated, an abscess can result. Most commonly cyst and abscess formation follows a bacterial, chlamydial, or gonococcal infection. Cysts can attain the size of an orange and frequently recur. Abscesses can be extremely tender and painful. Treatment of symptomatic cysts consists of the administration of appropriate antibiotics, local application of moist heat, and incision and drainage.

Non-neoplastic Epithelial Disorders

The term non-neoplastic epithelial disorders refers to nonmalignant atrophic and hyperplastic changes of the vulvar skin and mucosa. The condition, commonly referred to as leukoplakia, presents as white lesions of the vulva. Itching is the most common symptom, and dyspareunia (painful intercourse) is common.⁴

There are two forms of non-neoplastic epithelial lesions: lichen sclerosus and lichen simplex chronicus. Lichen sclerosus patches are hypopigmented, parchment-thin, and atrophic. Such lesions occur in all age groups but are most common in postmenopausal women. They may also occur on other areas of the skin. Although slowly progressive in its development, lichen sclerosus is not premalignant. Lichen simple chronicus *lesions* are thick, gray-white plaques. The thickened epithelium displays a marked increase in superficial keratin, which imparts a white appearance to the vulva. Because squamous carcinoma may also appear as white plaques, biopsy is required to distinguish benign from malignant lesions.

Vulvodynia

Vulvodynia is a syndrome of unexplained vulvar pain, also referred to as *vulvar pain syndrome* or *burning vulva syndrome*. It is characterized by burning, stinging, irritation, and rawness. Several forms or subsets of vulvodynia have been identified. It may occur as localized tenderness near the vaginal opening at the onset of intercourse or tampon insertion (*i.e.*, insertional vulvodynia), as a sensitivity to tight-fitting pants, bicycling, or prolonged sitting (*vulvar vestibulitis*), or as episodic flares that occur only before menses or after coitus (*cyclic vulvodynia*). Symptoms that are in general noncyclic and pruritic and develop progressively during the perimenopausal or postmenopausal years are characteristic of *vulvar dermatoses*.

Vulvar dysesthesia, also known as idiopathic or essential vulvodynia, involves severe, constant, widespread burning that interferes with daily activities.⁵ Although the cause is unknown, the quality of pain with vulvar dysesthesia resembles reflex sympathetic dystrophy (see Chapter 39). One of the proposed causes is pudendal neuralgia or pain along the pudendal nerve.

Possible causes for other forms of vulvodynia include candidal hypersensitivity related to chronic recurrent yeast infections; chemical irritation or drug effects, especially prolonged use of topical steroid creams; and dermatoses such as lichen sclerosus. Herpes simplex virus may be related to episodic vulvodynia.

Treatment of vulvodynia is aimed at symptom relief and elimination of suspected underlying problems. Careful history taking and physical assessment are essential for differential diagnosis and treatment. Psychosocial support often is needed because this condition can cause strain in sexual, family, and work relationships. Severe forms of the disorder often need to be managed from a multidimensional, chronic pain perspective.⁵

Premalignant and Malignant Neoplasms

Carcinoma of the vulva accounts for approximately 3% of all cancers of the female genitourinary system, occurring most often in women who are 60 years of age or older.⁴ Approximately 90% of vulvar malignancies are squamous cell carcinomas; the remainder are adenocarcinomas, melanomas, and basal cell carcinomas.

Vulvar intraepithelial neoplasia (VIN), which is a precursor lesion of squamous cell carcinoma, represents a spectrum of neoplastic changes that range from minimal cellular atypia to invasive cancer. There has been a significant increase in frequency of one form of VIN among younger women (40 to 60 years of age) during the past several decades. The increase is more common in women who smoke and those with human papillomavirus (HPV) infections, particularly HPV type 16 (see Chapter 35).⁴ A second form of VIN, which is seen more often in older women, often is preceded by chronic vulvar irritation or lichen sclerosus.⁶ This form of VIN does not appear to be associated with HPV.

The initial lesion of squamous cell vulvar carcinoma may appear as an inconspicuous thickening of the skin, a small raised area or lump, or an ulceration that fails to heal. It may be single or multiple and vary in color from white to velvety red or black. The lesions may resemble eczema or dermatitis and may produce few symptoms, other than pruritus, local discomfort, and exudation. A recurrent, persistent, pruritic vulvitis may be the only complaint. The symptoms frequently are treated with various home remedies before medical treatment is sought. The lesion often becomes secondarily infected, and this causes pain and discomfort. The malignant lesion gradually spreads superficially or as a deep furrow involving all of one labial side. Because there are many lymph channels around the vulva, the cancer metastasizes freely to the regional lymph nodes. The most common extension is to the superficial inguinal, deep femoral, and external iliac lymph nodes.

Early diagnosis is important in the treatment of vulvar carcinoma. Outlook for survival diminishes with increasing tumor size and nodal involvement. Treatment is primarily wide surgical excision of the lesion for noninvasive cancer and vulvectomy with node resection for invasive cancer.

Disorders of the Vagina

The normal vaginal ecology depends on the delicate balance of hormones and bacterial flora. The vagina is lined with mucus-secreting stratified squamous epithelial cells. The epithelial cells of the vagina, like other tissues of the reproductive system, respond to changing levels of the ovarian sex hormones. Estrogen stimulates the proliferation and maturation of the vaginal mucosa; this results in a thickening of the vaginal mucosa and an increased glycogen content of the epithelial cells.

Vaginal tissue usually is moist, with a pH maintained within the bacteriostatic range of 3.8 to 4.2. Döderlein's bacilli, part of the normal vaginal flora, metabolize glycogen, and in the process produce the lactic acid that normally maintains the vaginal pH below 4.5. The vaginal ecology can be disrupted at many levels, rendering it susceptible to infection. Pregnancy and the use of oral contraceptive agents increase the amount of estrogen in the system. Diabetes or a prediabetic state may increase the glycogen content of the cells. The use of systemic antibiotics may decrease the number of lactobacilli in the vagina. Decreased estrogen stimulation after menopause causes the vaginal mucosa to become thin and dry, often resulting in dyspareunia (*i.e.*, painful intercourse), atrophic vaginitis, and occasionally in vaginal bleeding.

Vaginitis

Vaginitis is an inflammatory condition of the vagina. It is characterized by vaginal discharge and burning, itching, redness, and swelling of vaginal tissues. Pain often occurs with urination and sexual intercourse. Vaginitis may be caused by chemical irritants, foreign bodies, and infectious agents. The causes of vaginitis differ in various age groups. In premenarchal girls, most vaginal infections have nonspecific causes, such as poor hygiene, intestinal parasites, or the presence of foreign bodies. *C. albicans, Trichomonas vaginalis,* and bacterial vaginosis are the most common causes of vaginitis in the childbearing years, and some of these organisms can be transmitted sexually (see Chapter 35).^{7,8} Atrophic vaginitis, which is caused by a decrease in estrogen levels, is the most common form in postmenopausal women or after removal of the ovaries. Estrogen deficiency results in a lack of regenerative growth of the vaginal epithelium, rendering these tissues more susceptible to infection and irritation.⁷ Döderlein's bacilli disappear, and the vaginal secretions become less acidic.

Every woman has a normal vaginal discharge during the menstrual cycle, but it should not cause burning or itching or have an unpleasant odor. These symptoms suggest inflammation or infection. Because these symptoms are common to the different types of vaginitis, precise identification of the organism is essential for proper treatment. A careful history should include information about systemic disease conditions, the use of drugs such as antibiotics that foster the growth of yeast, dietary habits, stress, and other factors that alter the resistance of vaginal tissue to infections. A physical examination usually is done to evaluate the nature of the discharge and its effects on the genital structures. Treatment is directed at the cause of the disorder.

Cancer of the Vagina

Primary cancers of the vagina are extremely rare. They account for approximately 3% of all cancers of the female reproductive system. Approximately 85% to 90% of primary vaginal cancers are squamous cell carcinomas.⁹ Adenocarcinoma of the vagina is a rare tumor that is seen almost exclusively in women exposed in utero to diethylstilbestrol (DES). DES is a nonsteroidal synthetic estrogen that was commonly prescribed between 1940 and 1971 to prevent miscarriage. Fortunately, less than 0.14% of women exposed to DES actually develop adenocarcinoma.⁴ Vaginal cancers may also result from local extension of cervical cancer, from local irritation such as occurs with prolonged use of a pessary, or from exposure to sexually transmitted herpesvirus or HPV.

Like vulvar carcinoma, carcinoma of the vagina is largely a disease of older women. Approximately half are women 60 years of age or older at the time of diagnosis. The most common symptom of vaginal carcinoma is abnormal bleeding. Twenty percent of women are asymptomatic, with the cancer being discovered during a routine pelvic examination. The anatomic proximity of the vagina to other pelvic structures (*e.g.*, urethra, bladder, rectum) permits early spread to these areas. Pelvic pain, dysuria, constipation, and vaginal discharge can be associated symptoms.

Vaginal cancer is often detected by vaginal cytology (Papanicolaou's test [Pap smear]) or examination of the vagina during a pelvic exam. It is recommended that women who have been exposed to DES have an initial colposcopic examination to identify areas of abnormal vaginal epithelium, followed by yearly Pap smears. It is also important for women who have had a hysterectomy to continue to have vaginal Pap smears every 3 to 5 years for early detection of vaginal cancer. Diagnosis of vaginal cancer requires biopsy of suspect lesions or areas.

Treatment of vaginal cancer must take into consideration the type of cancer; the size, location, and spread of the lesion; and the woman's age. Local excision, laser vaporization, or a loop electrode excision procedure (LEEP) can be considered with stage 0 squamous cell cancer. Radical surgery (a total hysterectomy, pelvic lymph node dissection, partial vaginectomy) and radiation therapy are both curative with more advanced cancers.

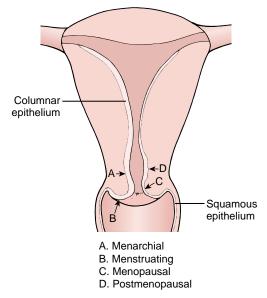
Disorders of the Uterine Cervix

The cervix is composed of two distinct types of tissue. The exocervix, or visible portion, is covered with stratified squamous epithelium, which also lines the vagina. The endocervical canal is lined with columnar epithelium (see Chapter 1). The junction of these two tissue types (*i.e.*, squamocolumnar junction) appears at various locations on the cervix at different points in a woman's life (Fig. 34-7). During periods of high estrogen production, particularly fetal existence, menarche, and the first pregnancy, the cervix everts or turns outward, exposing the columnar epithelium to the vaginal environment. The combination of estrogen and low vaginal pH leads to a gradual transformation from columnar to squamous epithelium-a process called metaplasia (see Chapter 2). The dynamic area of change where metaplasia occurs is called the transformation zone. The transformation zone is a critical area for the development of cervical cancer. During metaplasia, the newly developed squamous epithelial cells are vulnerable to the development of dysplastic changes.

The process of transformation is increased by trauma and infections occurring during the reproductive years.⁴ As the squamous epithelium expands and obliterates the surface columnar papillae, it covers and obstructs crypt openings, with trapping of mucus in the deeper crypts (glands) to form retention cysts, called *nabothian cysts*. These are benign cysts that require no treatment unless they become so numerous that they cause cervical enlargement. The nabothian cyst farthest away from the external cervical os indicates the outer aspect of the transformation zone.

Cervicitis and Cervical Polyps

Cervicitis is an acute or chronic inflammation of the cervix. Acute cervicitis may result from the direct infection of the cervix or may be secondary to a vaginal or uterine infection. It



■ FIGURE 34-7 ■ Location of squamocolumnar junction (transformation zone) in menarchial, menstruating, menopausal, and postmenopausal women.

may be caused by a variety of infective agents, including *C. albicans, T. vaginalis, Neisseria gonorrhoeae, Gardnerella vaginalis, Chlamydia trachomatis, Ureaplasma urealyticum,* and herpes simplex virus. Chronic cervicitis represents a low-grade inflammatory process. It is common in parous women and may be a sequela to minute lacerations that occur during childbirth, instrumentation, or other trauma. The organisms usually are of a nonspecific type, often staphylococcal, streptococcal, or coliform bacteria.

With acute cervicitis, the cervix becomes reddened and edematous. Irritation from the infection results in copious mucopurulent drainage and leukorrhea. The symptoms of chronic cervicitis are less well defined: the cervix may be ulcerated or normal in appearance; it may contain nabothian cysts; the cervical os may be distorted by old lacerations or everted to expose areas of columnar epithelium; and a mucopurulent drainage may be present.

Untreated cervicitis may extend to include the development of pelvic cellulitis, low back pain, painful intercourse, cervical stenosis, dysmenorrhea, and further infection of the uterus or fallopian tubes. Depending on the causative agent, acute cervicitis is treated with appropriate antibiotic therapy. Diagnosis of chronic cervicitis is based on vaginal examination, colposcopy, cytologic (Pap) smears, and occasionally on biopsy to exclude malignant changes. The treatment usually involves cryosurgery or cauterization, which causes the tissues to slough and leads to eradication of the infection.

Polyps are the most common lesions of the cervix. They can be found in women of all ages, but their incidence is higher during the reproductive years. Polyps are soft, velvety red lesions; they usually are pedunculated and often are found protruding through the cervical os. They usually develop as a result of inflammatory hyperplasia of the endocervical mucosa. Polyps typically are asymptomatic but may have associated postcoital bleeding. Most are benign, but they should be removed and examined by a pathologist to exclude malignant change.

Cancer of the Cervix

Cervical cancer is readily detected and, if detected early, is the most easily cured of all the cancers of the female reproductive system. Cervical cancer was the most common cause of cancer deaths in women in the United States in the 1930s.¹¹ However, with the introduction of the Papanicolaou (Pap) smear, early detection and treatment of preinvasive disease became possible. Incidence and mortality rates for cervical cancer in the United States declined dramatically during the remainder of the 20th century. In 2001, there were an estimated 12,900 new cases and 4,400 deaths of cervical cancer.¹¹ By comparison, there were four times as many new cases of cervical carcinoma

in situ (*i.e.*, precancerous lesion) diagnosed, indicating that a large number of potentially invasive cancers are cured by early detection and effective treatment. However, worldwide the incidence of and mortality associated with cervical cancer are second only to breast cancer, and in parts of the developing world, cervical cancer is the major cause of death in women of reproductive age.¹¹

Carcinoma of the cervix is often considered a sexually transmitted disease. A preponderance of evidence suggests a causal link between HPV infection and uterine cancer. Certain strains of HPV have been identified in invasive carcinoma of the cervix, whereas others are associated more often with dysplasia or carcinoma in situ. The strongest link is with HPV types 16, 18, 31, and 33 (see Chapter 35).^{2,7} Because these viruses are spread by sexual contact, their association with cervical cancer provides a tempting hypothesis to explain the relation between sexual practices and cervical cancer. Other factors such as smoking, nutrition, and coexisting sexual partners may play a contributing role in determining whether a woman with HPV infection develops cervical cancer. The recent development of an HPV type 16 vaccine offers hope for the future prevention of cervical cancer.¹²

Precursor Lesions. One of the most important advances in the early diagnosis and treatment of cancer of the cervix was made possible by the observation that this cancer arises from precursor lesions, which begin with the development of atypical (i.e., dysplasia) cervical cells. Atypical cells differ from normal cervical squamous epithelium. There are changes in the nuclear and cytoplasmic parts of the cell and more variation in cell size and shape (*i.e.*, dysplasia). These atypical cells gradually progress to carcinoma in situ, which is localized to the epithelial layer, to invasive cancer of the cervix. Carcinoma in situ is localized to the epithelial layer, whereas invasive cancer of the cervix spreads to deeper layers. A system of grading devised to describe the dysplastic changes of cancer precursors uses the term cervical intraepithelial neoplasia (CIN).^{2,7} The CIN system uses the thickness of the epithelial layer involvement as a means of grading the dysplastic changes (see Table 34-2).

The Pap Smear. There are no specific signs or symptoms of cervical dysplasia or cancer in situ. A presumptive diagnosis is made by Pap smear studies of an asymptomatic population with no visible cervical changes. The purpose of the Pap smear (see Chapter 5) is to detect the presence of abnormal cells on the surface of the cervix or in the endocervix. This test detects precancerous and cancerous lesions. Although the American Cancer Society has suggested that the Pap smear need not be done annually if there have been three negative tests in successions.

TABLE 34-2	Cervical Intraepithelial Neoplasia Grading System		
Grade		Extent of Involvement	Differentiation of Lesion
CIN I (mild dyspla CIN II (moderate o CIN III (severe dys carcinoma in sit	dysplasia) plasia or	Initial one third of epithelial layer Initial two thirds of epithelial layer Full-thickness involvement	Well differentiated Less well differentiated Undifferentiated

(Data from Rubin E., Farber J.L. [1999]. Pathology [3rd ed., pp. 982–983]. Philadelphia: Lippincott Williams & Wilkins)

sion, many clinicians maintain that performing an annual test is the safest course to follow. If the woman has risk factors, such as previous HPV infection, DES exposure in utero, or a strong family history of cervical cancer, more frequent Pap smears may be recommended.

It has been estimated that approximately 15% to 25% of women with intraepithelial lesions have normal Pap smear results.¹¹ New techniques of specimen collection, slide preparation and processing, and computer-assisted evaluation of Pap smears are being evaluated and offer the hope of improved accuracy in the diagnosis of precancerous cervical changes.

Squamous Cell Cervical Carcinoma. The majority of cervical cancers are squamous cell carcinomas. Cancers of the cervix have a long latent period; untreated dysplasia or CIN gradually progresses to carcinoma in situ, which may remain static for 7 to 10 years before it becomes invasive. After the preinvasive period, growth may be rapid, and survival rates decline significantly, depending on the extent of disease at the time of diagnosis.¹³

Diagnosis of cervical cancer requires pathologic confirmation. Although the Pap smear has been shown to be a costeffective cancer prevention method, it remains a screening method, not a diagnostic method. Pap smear results demonstrating squamous intraepithelial lesions often require further evaluation by colposcopy. This is a vaginal examination that is done using a colposcope, an instrument that affords a well-lit and magnified stereoscopic view of the cervix. A biopsy sample may be obtained from suspect areas and examined microscopically. A diagnostic cone biopsy, which involves the removal of a cone-shaped wedge of cervix, may be indicated when a lesion is partly or completely beyond colposcopic view.

KEY CONCEPTS

GYNECOLOGIC CANCERS

- Certain types of sexually transmitted human papillomaviruses are risk factors for cervical intraepithelial neoplasia, which can be a precursor lesion of invasive carcinoma.
- Endometrial cancers, which are seen most frequently in women 55 to 65 years of age, are strongly associated with conditions that produce excessive estrogen stimulation and endometrial hyperplasia.
- Ovarian cancer is the second most common female cancer and the most lethal because there are few early signs and no good screening tests. The most significant risk factors for ovarian cancers are the length of time that a woman's ovarian cycles are not suppressed by pregnancy, lactation, or oral contraceptive use, and family history.
- Breast cancer is the most common cancer in women and the second most common cause of cancer death in women. Risk factors include age, family history, and hormonal influences.

Early treatment of cervical cancer involves removal of the lesion by one of various techniques. Biopsy or local cautery may be therapeutic in and of itself. Electrocautery, cryosurgery, or carbon dioxide laser therapy may be used to treat moderate to severe dysplasia that is limited to the exocervix. Invasive cancer is treated with radiation therapy, surgery, or both. External beam irradiation and intracavitary cesium irradiation (*i.e.*, insertion of a closed metal cylinder containing cesium) can be used in the treatment of cervical cancer. Surgery can include extended hysterectomy (*i.e.*, removal of the uterus, fallopian tubes, ovaries, and upper portion of the vagina) without pelvic lymph node dissection, radical hysterectomy with pelvic lymph node dissection, or pelvic exenteration (*i.e.*, removal of all pelvic organs, including the bladder, rectum, vulva, and vagina).

Disorders of the Uterus

Infectious Disorders of the Uterus and Pelvic Structures

The uterus and pelvic structures are subject to infections by a number of agents, including *N. gonorrhoeae* and *C. trachomatis,* as well as endogenous microorganisms such as anaerobes, *Haemophilus influenzae*, enteric gram-negative rods, and *Streptococci.* Tuberculosis salpingitis is rare in the United States but more common in developing countries.

Endometritis. Inflammation or infection of the endometrium is an ill-defined entity that produces variable symptoms. Endometritis can occur as a postpartum or postabortal infection, with gonococcal or chlamydial salpingitis, or after instrumentation or surgery, or can be associated with an intrauterine device.¹² Abnormal vaginal bleeding, mild to severe uterine tenderness, fever, malaise, and foul-smelling discharge have been associated with endometritis, but the clinical picture is variable. Treatment involves oral or intravenous antibiotic therapy, depending on the severity of the condition.

Pelvic Inflammatory Disease. Pelvic inflammatory disease (PID) is an inflammation of the upper reproductive tract that involves the uterus (endometritis), fallopian tubes (salpingitis), or ovaries (oophoritis). Most women with acute salpingitis have *N. gonorrhoeae* or *C. trachomatis* identified in the reproductive tract. PID is a polymicrobial infection, and the cause varies by geographic location and population.

The organisms ascend through the endocervical canal to the endometrial cavity and then to the tubes and ovaries. The endocervical canal is slightly dilated during menstruation, allowing bacteria to gain entrance to the uterus and other pelvic structures. After entering the upper reproductive tract, the organisms multiply rapidly in the favorable environment of the sloughing endometrium and ascend to the fallopian tube. Factors that predispose women to the development of PID include an age of 16 to 24 years, unmarried status, nulliparity, history of multiple sexual partners, and previous history of PID.

The symptoms of PID include lower abdominal pain, which may start just after a menstrual period; purulent cervical discharge; pelvic tenderness; and an exquisitely painful cervix. Fever (>101°F), increased erythrocyte sedimentation rate, and an elevated white blood cell count (>10,000 cells/mL)

commonly are seen, even though the woman may not appear acutely ill.¹³

Diagnosis is based on the presence of lower abdominal pain and pelvic or cervical tenderness. A newer test involves measurement of C-reactive protein in the blood. Elevated C-reactive protein levels equate with inflammation.¹³ Endocervical cultures may be done to document the presence of *N. gonorrhoeae* or *C. trachomatis.* Transvaginal ultrasonography or imaging techniques may be used. Laparoscopy is often used to confirm a diagnosis of PID.

Treatment is aimed at preventing complications, which can include pelvic adhesions, infertility, ectopic pregnancy, chronic abdominal pain, and tubo-ovarian abscesses. It may involve hospitalization with intravenous administration of antibiotics. If the condition is diagnosed early, outpatient antibiotic therapy may be sufficient.

Endometriosis

Endometriosis is the condition in which functional endometrial tissue is found in ectopic sites outside the uterus. The site may be the ovaries, broad ligaments, pouch of Douglas (cul-de-sac), pelvis, vagina, vulva, perineum, or intestines (Fig. 34-8). Rarely, endometrial implants have been found in the nostrils, umbilicus, lungs, and limbs.

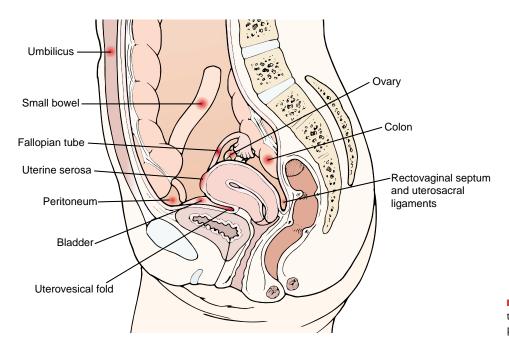
The cause of endometriosis is unknown. There appears to have been an increase in its incidence in the developed Western countries during the past four to five decades. Approximately 10% to 15% of premenopausal women have some degree of endometriosis. The incidence may be higher in women with infertility (25% to 40%)¹⁴ or women younger than 20 years of age with chronic pelvic pain (45% to 70%).¹⁵ It is more common in women who have postponed childbearing. Risk factors for endometriosis may include early menarche; regular periods with shorter cycles (<27 days), longer duration (>7 days), or heavier flow; increased menstrual pain; and other first-degree relatives with the condition.

Several theories attempt to account for endometriosis. One theory suggests that menstrual blood containing fragments of endometrium is forced upward through the fallopian tubes into the peritoneal cavity. Retrograde menstruation is not an uncommon phenomenon, and it is unknown why endometrial cells implant and grow in some women but not in others. Another proposal is that dormant, immature cellular elements from embryonic development persist into adult life and develop into ectopic endometrial tissue. Another theory suggests that the endometrial tissue may metastasize through the lymphatics or vascular system. Altered cellular immunity and genetic components also have been studied as contributing factors to the development of endometriosis.¹⁴

The gross pathologic changes that occur in endometriosis differ with location and duration. In the ovary, the endometrial tissue may form cysts (*i.e.*, endometriomas filled with old blood that resembles chocolate syrup [chocolate cysts]). Rupture of these cysts can cause peritonitis and adhesions. Elsewhere in the pelvis, the tissue may take the form of small hemorrhagic lesions, some of which may be surrounded by scar tissue. These ectopic implants respond to hormonal stimulation in the same way normal endometrium does, becoming proliferative, then secretory, and finally undergoing menstrual breakdown. Bleeding into the surrounding structures can cause pain and the development of significant pelvic adhesions. Ex-

Endometriosis may be difficult to diagnose because its symptoms mimic those of other pelvic disorders, and the severity of the symptoms does not always reflect the extent of the disease. The classic triad of dysmenorrhea, dyspareunia, and infertility strongly suggests endometriosis. Accurate diagnosis can be accomplished only through laparoscopy. This minimally invasive surgery allows direct visualization of pelvic organs to determine the presence and extent of endometrial lesions.

Treatment goals for endometriosis are pain management or restoration of fertility. Treatment modalities fall into three cat-



■ FIGURE 34-8 ■ Common locations of endometriosis within the pelvis and abdomen.

egories: pain relief, endometrial suppression, and surgery. In young women, who have no fertility concerns, simple observation and antiprostaglandin analgesics (*i.e.*, nonsteroidal anti-inflammatory drugs [NSAIDs]) may be sufficient treatment. The use of hormones to induce physiologic amenorrhea is based on the observation that pregnancy affords temporary relief by inducing atrophy of the endometrial tissue.

Surgery is the most definitive therapy for many women with endometriosis. With the advent of lasers, in-depth treatment of endometriosis or pelvic adhesions can be accomplished by means of laparoscopy.¹⁶ Radical treatment involves total hysterectomy and bilateral salpingo-oophorectomy (*i.e.*, removal of the fallopian tubes and ovaries) when the symptoms are unbearable or the woman's childbearing is completed.

Treatment offers relief but not cure. Recurrence of endometriosis is not uncommon, regardless of the treatment (except for radical surgery). Pregnancy may delay, but does not preclude, recurrence.

Adenomyosis

Adenomyosis is the condition in which endometrial glands and stroma are found within the myometrium, interspersed between the smooth muscle fibers.⁴ It is thought that events associated with repeated pregnancies, deliveries, and uterine involution may cause the endometrium to be displaced throughout the myometrium. In contrast to endometriosis, which usually is a problem of young, infertile women, adenomyosis typically is found in multiparous women in their late fourth or fifth decade. Up to 70% of these women have a retrospective history of painful, heavy periods.¹⁷

The diagnosis of adenomyosis often occurs as an incidental finding in a uterus removed for symptoms indicative of myoma or hyperplasia. Adenomyosis resolves with menopause. Conservative therapy using oral contraceptives or gonadotropinreleasing hormone (GnRH) agonists is the first choice for treatment. Hysterectomy (with preservation of the ovaries in premenopausal women) is considered when this approach fails.¹⁷

Endometrial Cancer

Endometrial cancer is the most common invasive cancer of the female reproductive tract. It accounts for 7% of all invasive cancers in women, excluding skin cancers.² There is now an estimated 34,000 new diagnoses of endometrial cancer a year, compared with 13,000 new invasive cervical cancers.²

Endometrial cancer occurs more frequently in older women (peak ages of 55 to 65 years) and is uncommon in women younger than 40 years of age. A major risk factor for endometrial cancer is prolonged estrogenic stimulation with excessive growth (*i.e.*, hyperplasia) of the endometrium. Obesity, anovulatory cycles, conditions that alter estrogen metabolism, estrogen-secreting neoplasms, and unopposed estrogen therapy all increase the risk of endometrial cancer.^{20,21} Estrogens are synthesized in body fats from adrenal and ovarian androgen precursors; endometrial hyperplasia and endometrial cancer appear to be related to obesity. The degree of risk correlates with body weight, with the risk increasing 10-fold for women who are more than 50 pounds overweight.7 Ovulatory dysfunction that causes infertility at any age or occurs with declining ovarian function in perimenopausal women also can result in unopposed estrogen and increase the risk of endometrial cancer. Diabetes mellitus, hypertension, and polycystic ovary syndrome are conditions that alter estrogen metabolism and elevate estrogen levels.

Endometrial cancer risk also is increased in women with estrogen-secreting granulosa cell tumors and in those receiving unopposed estrogen therapy. It is the presence of progesterone in the second half of the normal menstrual cycle that matures the endometrium and the withdrawal of progesterone that ultimately results in endometrial sloughing. Long-term unopposed estrogen exposure without periodic addition of progesterone allows for continued endometrial growth and the development of hyperplasia, with or without the presence of atypical cells. A sharp increase in endometrial cancer was seen during the 1970s among middle-aged women who had received unopposed estrogen therapy (i.e., estrogen therapy without progesterone) for menopausal symptoms. It was later determined that it was not the estrogen exposure that increased the risk of cancer, but administration of estrogen without administration of progesterone. Tamoxifen, a drug that blocks estrogen receptor sites and is used in the treatment of breast cancer, exerts a weak estrogenic effect on the endometrium and represents another exogenous risk factor for endometrial cancer.

A small subset of women in whom endometrial cancer develops do not exhibit increased estrogen levels or pre-existing hyperplasia. These women usually acquire the disease at an older age. These tumors arise from clones of cancer-initiated mutant cells and are more poorly differentiated. This type of endometrial cancer usually is associated with a poorer prognosis than is the endometrial cancer that is associated with prolonged estrogen stimulation and endometrial hyperplasia.²¹

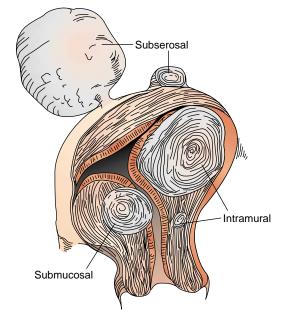
The major symptom of endometrial hyperplasia or overt endometrial cancer is abnormal, painless bleeding. In menstruating women, this takes the form of bleeding between periods or excessive, prolonged menstrual flow. In postmenopausal women, any bleeding is abnormal and warrants investigation. Abnormal bleeding is an early warning sign of the disease, and because endometrial cancer tends to be slow growing in its early stages, the chances of cure are good if prompt medical care is sought. Later signs of uterine cancer may include cramping, pelvic discomfort, postcoital bleeding, lower abdominal pressure, and enlarged lymph nodes.

Although the Pap smear can identify a small percentage of endometrial cancers, it is not a good screening test for the tumor. Endometrial biopsy is far more accurate. Dilatation and curettage (D & C), which consists of dilating the cervix and scraping the uterine cavity, is the definitive procedure for diagnosis because it provides a more thorough evaluation. Transvaginal ultrasonography may be used to determine the endometrial thickness as an indicator of hypertrophy and possible neoplastic change.

The prognosis for endometrial cancer depends on the clinical stage of the disease when it is discovered and its histologic grade and type. Surgery and radiation therapy are the most successful methods of treatment for endometrial cancer. With early diagnosis and treatment, the 5-year survival rate is approximately 90%. This decreases to 20% for more advanced stages of the disease.²

Leiomyomas

Leiomyomas are benign neoplasms of smooth muscle origin. They also are known as *myomas* and sometimes are called *fibroids*. These are the most common form of pelvic tumor and are believed to occur in one of every four or five women older than 35 years of age. They are seen more often and their rate of growth is more rapid in black women than in white women. Leiomyomas usually develop in the corpus of the uterus; they may be submucosal, subserosal, or intramural (Fig. 34-9). Intramural fibroids are embedded in the myometrium. They are the most common type of fibroid and present as a symmetric enlargement of the nonpregnant uterus. Subserosal tumors are located beneath the perimetrium of the uterus. These tumors are recognized as irregular projections on the uterine surface; they may become pedunculated, displacing or impinging on other genitourinary structures and causing hydroureter or bladder problems. Submucosal fibroids displace endometrial tissue and are more likely to cause bleeding, necrosis, and infection than either of the other types.







В

■ FIGURE 34-9 (A) Submucosal, intramural, and subserosal leiomyomas. (A redrawn from Green T.H. [1977]. *Gynecology: Essentials of clinical practice* [3rd ed.]. Boston: Little, Brown.) (B) A bisected uterus displays a prominent, sharply circumscribed, fleshy tumor. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 999]. Philadelphia: Lippincott Williams & Wilkins)

Leiomyomas are asymptomatic approximately half of the time and may be discovered during a routine pelvic examination, or they may cause menorrhagia (excessive menstrual bleeding), anemia, urinary frequency, rectal pressure/constipation, abdominal distention, and infrequently pain. Their rate of growth is variable, but they may increase in size during pregnancy or with exogenous estrogen stimulation (*i.e.*, oral contraceptives or menopausal estrogen replacement therapy). Interference with pregnancy is rare unless the tumor is submucosal and interferes with implantation or obstructs the cervical outlet. These tumors may outgrow their blood supply, become infarcted, and undergo degenerative changes.

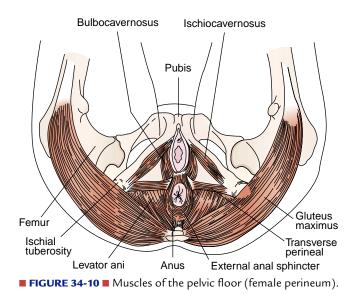
Most leiomyomas regress with menopause, but if bleeding, pressure on the bladder, pain, or other problems persist, hysterectomy may be required. Myomectomy (removal of just the tumors) can be done to preserve the uterus for future childbearing. Cesarean section may be recommended if the uterine cavity is entered during myomectomy. Hypothalamic GnRH may be used to suppress leiomyoma growth before surgery. Uterine artery embolization is a nonsurgical therapy for management of heavy bleeding.²²

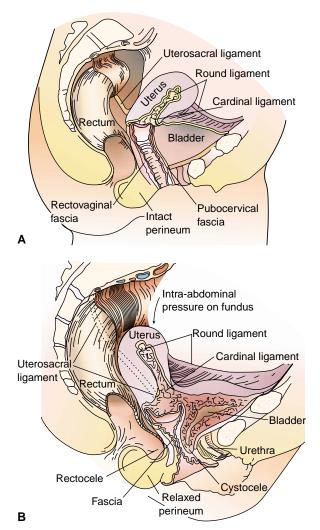
Disorders of Pelvic Support and Uterine Position

The muscular floor of the pelvis is a strong, slinglike structure that supports the uterus, vagina, urinary bladder, and rectum (Fig. 34-10). In the female anatomy, nature is faced with the problems of supporting the pelvic viscera against the force of gravity and increases in intra-abdominal pressure associated with coughing, sneezing, defecation, and laughing while at the same time allowing for urination, defecation, and normal reproductive tract function, especially the delivery of an infant.

Disorders of Pelvic Support

The uterus and the pelvic structures are maintained in proper position by the uterosacral ligaments, round ligaments, broad ligament, and cardinal ligaments. The two cardinal ligaments maintain the cervix in its normal position. The uterosacral ligaments normally hold the uterus in a forward position





■ FIGURE 34-11 ■ (A) Normal support of the uterus and vagina. (B) Relaxation of pelvic support structures with descent of the uterus as well as formation of cystocele and rectocele. (Rock J.A., Thompson J.D. [1992]. *Te Linde's operative gynecology* [7th ed.]. Philadelphia: J.B. Lippincott)

(Fig. 34-11 A). The broad ligament suspends the uterus, fallopian tubes, and ovaries in the pelvis.

Three supporting structures are provided for the abdominal pelvic diaphragm. The bony pelvis provides support and protection for parts of the digestive tract and genitourinary structures, and the peritoneum holds the pelvic viscera in place. However, the main support for the viscera is the pelvic diaphragm, made up of muscles and connective tissue that stretch across the bones of the pelvic outlet. The openings that must exist for the urethra, rectum, and vagina cause an inherent weakness in the pelvic diaphragm. Congenital or acquired weakness of the pelvic diaphragm results in widening of these openings, particularly the vagina, with the possible herniation of pelvic viscera through the pelvic floor (*i.e.*, prolapse).

Relaxation of the pelvic outlet usually comes about because of overstretching of the perineal supporting tissues during pregnancy and childbirth. Although the tissues are stretched only during these times, there may be no difficulty until later in life, such as the fifth or sixth decade, when further loss of elasticity and muscle tone occurs. Even in a woman who has not borne children, the combination of aging and postmenopausal changes may give rise to problems related to relaxation of the pelvic support structures. It also may result from pelvic tumors and neurologic conditions, such as spina bifida and diabetic neuropathy, that interrupt the innervation of pelvic muscles. The three most common conditions associated with this relaxation are cystocele, rectocele, and uterine prolapse. These may occur separately or together.

Cystocele is a herniation of the bladder into the vagina (Fig. 34-11 B). It occurs when the normal muscle support for the bladder is weakened, and the bladder sags below the uterus. The vaginal wall stretches and bulges downward because of the force of gravity and the pressure from coughing, lifting, or straining at stool. The bladder herniates through the anterior vaginal wall, and a cystocele forms. The symptoms include an annoying bearing-down sensation, difficulty in emptying the bladder, frequency and urgency of urination, and cystitis. Stress incontinence may occur at times of increased abdominal pressure, such as during squatting, straining, coughing, sneezing, laughing, or lifting.

Rectocele is the herniation of the rectum into the vagina (Fig 34-11 A). It occurs when the posterior vaginal wall and underlying rectum bulge forward, ultimately protruding through the introitus as the pelvic floor and perineal muscles are weakened. The symptoms include discomfort because of the protrusion of the rectum and difficulty in defecation. Digital pressure (*i.e.*, splinting) on the bulging posterior wall of the vagina may become necessary for defecation.

Uterine prolapse is the bulging of the uterus into the vagina that occurs when the primary supportive ligaments (*i.e.*, cardinal ligaments) are stretched. Prolapse is ranked as first, second, or third degree, depending on how far the uterus protrudes through the introitus. First-degree prolapse shows some descent, but the cervix has not reached the introitus. In seconddegree prolapse, the cervix or part of the uterus has passed through the introitus. The entire uterus protrudes through the vaginal opening in third-degree prolapse (*i.e.*, procidentia). Prolapse often is accompanied by perineal relaxation, cystocele, or rectocele. The symptoms associated with uterine prolapse result from irritation of the exposed mucous membranes of the cervix and vagina and the discomfort of the protruding mass.

Most of the disorders of pelvic relaxation require surgical correction. These are elective surgeries and usually are deferred until after the childbearing years. The symptoms associated with the disorders often are not severe enough to warrant surgical correction. In other cases, the stress of surgery is contraindicated because of other physical disorders; this is particularly true of older women, in whom many of these disorders occur. Kegel exercises, which strengthen the pubococcygeus muscle, may be helpful in cases of mild cystocele or rectocele or after surgical repair to help maintain the improved function (see Chapter 25). In women with uterine prolapse, a pessary may be inserted to hold the uterus in place and may stave off surgical intervention in women who want to have children or in older women for whom the surgery may pose a significant health risk.

Variations in Uterine Position

Variations in the position of the uterus are common. Some variations are innocuous; others, which may be the result of weakness and relaxation of the perineum, give rise to various problems that compromise the structural integrity of the pelvic floor, particularly after childbirth.

The uterus usually is flexed approximately 45 degrees anteriorly, with the cervix positioned posteriorly and downward in the anteverted position. When the woman is standing, the angle of the uterus is such that it lies practically horizontal, resting lightly on the bladder. Asymptomatic, normal variations in the axis of the uterus in relation to the cervix (i.e., flexion) and physiologic displacements that arise after pregnancy or with pathology of the cul-de-sac include anteflexion, retroflexion, and retroversion (Fig. 34-12). An anteflexed uterus is flexed forward on itself. Retroflexion is flexion backward at the isthmus. Retroversion describes the condition in which the uterus inclines posteriorly while the cervix remains tilted forward. Simple retroversion is the most common variation. It usually is a congenital condition caused by a short anterior vaginal wall and relaxed uterosacral ligaments. Retroversion also can follow certain diseases, such as endometriosis and PID, which produce fibrous tissue adherence with retraction of the fundus posteriorly. Large leiomyomas can also cause the uterus to move into a posterior position. Symptoms of retroversion include dyspareunia with deep penetration and low back pain with menses.

Disorders of the Ovaries

Disorders of the ovaries frequently cause menstrual and fertility problems. Benign conditions of the ovaries can present as primary lesions of the ovarian structures or as secondary disorders related to hypothalamic, pituitary, or adrenal dysfunction.

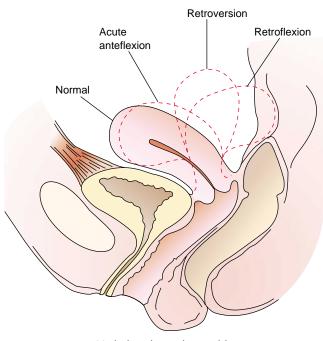
Ovarian Cysts

Cysts are the most common form of ovarian tumor.⁷ Many are benign. A follicular cyst is one that results from occlusion of the duct of the follicle. Each month, several follicles begin to de-

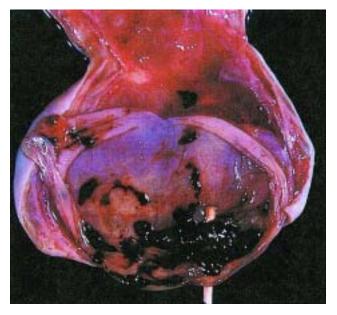
velop and are blighted at various stages of development. These follicles form cavities that fill with fluid, producing a cyst. The dominant follicle normally ruptures to release the egg (*i.e.*, ovulation) but occasionally persists and continues growing. Likewise, a luteal cyst is a persistent cystic enlargement of the corpus luteum that is formed after ovulation and does not regress in the absence of pregnancy. Functional cysts are asymptomatic unless there is substantial enlargement or bleeding into the cyst. This can cause considerable discomfort or a dull, aching sensation on the affected side. The cyst may become twisted or may rupture into the intra-abdominal cavity (Fig. 34-13). However, usually these regress spontaneously.

Polycystic Ovary Syndrome. Ovarian dysfunction associated with infrequent or absent menses in obese, infertile women was first reported in the 1930s by Stein and Leventhal, for whom the syndrome was originally named. Polycystic ovary syndrome (PCOS) is characterized by numerous cystic follicles or follicular cysts. Once thought to be relatively rare, it appears that this clinical entity is one of the most common endocrine disorders among women during the reproductive years. PCOS is characterized by varying degrees of hirsutism, acne, obesity, and anovulation and infertility, and often is associated with hyperinsulinemia or insulin resistance (see Chapter 31).^{7,23} Chronic anovulation can increase a woman's risk of endometrial cancer, cardiovascular disease, and hyperinsulinemia leading to diabetes mellitus.⁷

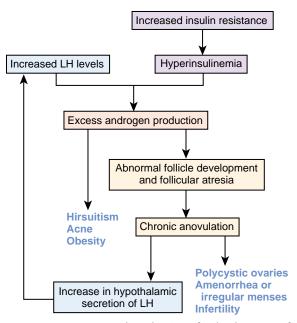
This syndrome has been the subject of considerable research. Chronic anovulation, causing amenorrhea or irregular menses, is now thought to be the underlying cause of the bilaterally enlarged "polycystic" ovaries (Fig. 34-14). Thus, the appearance of the ovary is a sign of the disorder, not the disease itself.²⁴ The precise etiology of this condition is still being de-



■ FIGURE 34-12 ■ Variations in uterine position.



■ FIGURE 34-13 ■ Follicle cyst of the ovary. The rupture of this thin-walled follicular cyst led to intra-abdominal hemorrhage. The cyst has been opened and rupture site indicated by the dowel stick. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 1002]. Philadelphia: Lippincott Williams & Wilkins)



■ FIGURE 34-14 ■ Proposed mechanisms for development of the polycystic ovary syndrome.

bated. A possible genetic basis has been suggested with an autosomal dominant mode of inheritance and premature balding as the phenotype in males.²⁴

Most women with PCOS have elevated luteinizing hormone (LH) levels with normal estrogen and follicle-stimulating hormone (FSH) production. Elevated levels of testosterone, dehydroepiandrosterone sulfate (DHAS), or androstenedione are not uncommon, and occasionally hyperprolactinemia or hypothyroidism is present. Persistent anovulation results in an estrogen environment that alters the hypothalamic release of GnRH and with a resultant increase in LH secretion and suppression of FSH secretion by the pituitary. This altered LH: FSH ratio often is used as a diagnostic criterion for this condition, but it is not universally present. Although the presence of some FSH allows for new follicular development, full maturation is not attained, and ovulation does not occur. The elevated LH level also results in increased androgen production, which in turn prevents normal follicular development and contributes to the vicious cycle of anovulation.24 The association between insulin resistance and increased androgen levels is complex. Evidence suggests that the hyperinsulinemia may lead to the excess androgen production. Lowering insulin levels by weight loss or treatment with several of the oral antidiabetic drugs (thiazolidinediones or metformin) may reduce testosterone levels and restore ovulatory cycles in many women with the disorder.25,26

The diagnosis can be suspected from the clinical picture. Confirmation with ultrasonography or laparoscopic visualization of the ovaries is not required. When fertility is desired, the condition usually is treated by the administration of the hypothalamic-pituitary–stimulating drug clomiphene citrate or injectable gonadotropins to induce ovulation. These drugs must be used carefully because they can induce extreme enlargement of the ovaries. An insulin-sensitizing drug may be used before or concurrent with ovulation-inducing medications.²⁶ Weight loss also may be beneficial in restoring normal ovulation when obesity is present. When medication is ineffective, laser surgery to puncture the multiple follicles may restore normal ovulatory function, although adhesion formation is a potential problem. If fertility is not desired, oral contraceptives or cyclic progesterone can induce regular menses and prevent the development of endometrial hyperplasia caused by unopposed estrogen.

Benign and Functioning Ovarian Tumors

Benign ovarian tumors can be composed of epithelial tissue, endometriosis tissue, fibrocytes and collagen fibers, or primordial germ cells. Benign epithelial tumors are almost always serous or mucinous.^{2,7} They generally occur in women between the ages of 20 and 60 years and are often large, growing 15 to 30 cm in diameter. They are often cystic, thus the term cystadenomas. However, some of the cystadenomas are considered to have low malignant potential. Endometriomas are the "chocolate cysts" that develop secondary to ovarian endometriosis (see the endometriosis section earlier in this chapter). Ovarian fibromas are connective tissue tumors composed of fibrocytes and collagen. They range in size from 6 to 20 cm. Cystic teratomas, or dermoid cysts, are derived from primordial germ cells and are composed of various combinations of well-differentiated ectodermal, mesodermal, and endodermal elements. Not uncommonly, they contain sebaceous material, hair, or teeth.

Benign ovarian tumors are usually asymptomatic unless their size is sufficient to cause abdominal enlargement. Treatment for all ovarian tumors is surgical excision. Ovarian tissue that is not affected by the tumor can be left intact if frozensection analysis does not reveal malignancy. When ovarian tumors are very large, as is frequently the case with serous or mucinous cystadenomas, the entire ovary must be removed.

The three types of functioning ovarian tumors are estrogen secreting, androgen secreting, and mixed estrogen-androgen secreting. These tumors may be benign or cancerous. One such tumor, the granulosa cell tumor, is associated with excess estrogen production. When it develops during the reproductive period, the persistent and uncontrolled production of estrogen interferes with the normal menstrual cycle, causing irregular and excessive bleeding, endometrial hyperplasia, or amenorrhea and fertility problems. When it develops after menopause, it causes postmenopausal bleeding, stimulation of the glandular tissues of the breast, and other signs of renewed estrogen production. Androgen-secreting tumors inhibit ovulation and estrogen production. They tend to cause hirsutism and development of masculine characteristics, such as baldness, acne, oily skin, breast atrophy, and deepening of the voice. The treatment of functioning ovarian tumors is surgical removal.

Ovarian Cancer

Ovarian cancer is the second most common female genitourinary cancer and the most lethal. In 2000, 23,100 new cases of ovarian cancer were reported in the United States, two thirds of which were in advanced stages of the disease. Of these women, it is estimated that 14,000 will die of the disease, accounting for 53% of all deaths caused by gynecological cancers.²⁷ The incidence of ovarian cancer increases with age, being greatest between 65 and 84 years of age. Ovarian cancer is difficult to diagnose, and as many as 75% of women have metastatic disease before the time of discovery. The most significant risk factor for ovarian cancer appears to be ovulatory age—the length of time during a woman's life when her ovarian cycle is not suppressed by pregnancy, lactation, or oral contraceptive use. The incidence of ovarian cancer is much lower in countries where women bear numerous children than in the United States. Family history also is a significant risk factor for ovarian cancer. Women with two or more first- or second-degree relatives who have had *site-specific ovarian cancer* have as much as a 50% risk for development of the disease. A high-fat Western diet and use of powders containing talc in the genital area are other factors that have been linked to the development of ovarian cancer.

Cancer of the ovary is complex because of the diversity of tissue types that originate in the ovary. As a result of this diversity, there are several types of ovarian cancers. Malignant neoplasms of the ovary can be divided into three categories: epithelial tumors, germ cell tumors, and gonadal stromal tumors. Epithelial tumors account for approximately 90% of cases. These different cancers display various degrees of virulence, depending on the type of tumor and degree of differentiation involved. A well-differentiated cancer of the ovary may have produced symptoms for many months and still be found operable at the time of surgery. A poorly differentiated tumor may have been clinically evident for only a few days but found to be widespread and inoperable. Often no correlation exists between the duration of symptoms and the extent of the disease.

Most cancers of the ovary produce no symptoms or the symptoms are so vague that the woman seldom seeks medical care until the disease is far advanced. These vague discomforts include abdominal distress, flatulence, and bloating, especially after ingesting food. These gastrointestinal manifestations may precede other symptoms by months. It is not fully understood why the initial symptoms of ovarian cancer are manifested as gastrointestinal disturbances. It is thought that biochemical changes in the peritoneal fluids may irritate the bowel or that pain originating in the ovary may be referred to the abdomen and be interpreted as a gastrointestinal disturbance. Clinically evident ascites (*i.e.*, fluid in the peritoneal cavity) is seen in approximately one fourth of women with malignant ovarian tumors and is associated with a worse prognosis.

At present, there are no good screening tests or other early methods of detection for ovarian cancer. The serum tumor marker CA-125 is a cell surface antigen that can be used in monitoring therapy and recurrences when preoperative levels have been elevated. Because it lacks sensitivity and specificity, CA-125 has limited value as a single screening test. Transvaginal ultrasonography (TVS) can be used to evaluate ovarian masses for malignant potential. However, cost has prohibited its use as a universal screening tool. The National Institutes of Health Consensus Panel convened in 1995 recommended against widespread screening of women for ovarian cancer.²⁸ CA-125 with TVS is suggested only for women who are part of a family with hereditary ovarian cancer syndrome (*i.e.*, two or more affected first-degree relatives).

When ovarian cancer is suspected, surgical evaluation is required for diagnosis, complete and accurate staging, and cytoreduction and debulking procedures to reduce the size of the tumor. Recommendations regarding treatment beyond surgery and prognosis depend on the stage of the disease. In summary, the external genitalia are affected by disorders that affect skin on other parts of the body. Bartholin's cysts are the result of occluded ducts in Bartholin's glands. They often are painful and can become infected. Vulvar dystrophies are characterized by thinning and hyperplastic thickening of vulvar tissues. Vulvodynia is a chronic vulvar pain syndrome with several classifications and variable treatment results. Cancer of the vulva, which accounts for 4% of all female genitourinary cancers, is associated with HPV infections.

The normal vaginal ecology depends on the delicate balance of hormones and bacterial flora. Vaginitis or inflammation of the vagina is characterized by vaginal discharge and burning, itching, redness, and swelling of vaginal tissues. It may be caused by chemical irritants, foreign bodies, and infectious agents. Primary cancers of the vagina are uncommon, accounting for 3% of all cancers of the female reproductive system. Daughters of women treated with DES to prevent miscarriage are at increased risk for the development of adenocarcinoma of the vagina.

Disorders of the cervix and uterus include inflammatory conditions (*i.e.*, cervicitis and endometritis), cancer (*i.e.*, cervical and endometrial cancer), endometriosis, and leiomyomas. Acute cervicitis, which may be caused by a number of agents, may result from the direct infection of the cervix or may be secondary to a vaginal or uterine infection. Chronic cervicitis represents a low-grade inflammatory process resulting from trauma or nonspecific infectious agents. Cervical cancer arises from precursor lesions that can be detected on a Pap smear, and if detected early, is the most easily cured of all the cancers of the female reproductive system.

PID is an inflammation of the upper reproductive tract that involves the uterus (endometritis), fallopian tubes (salpingitis), or ovaries (oophoritis). It is most commonly caused by *N. gonorrhoeae* or *C. trachomatis*. Endometriosis is the condition in which functional endometrial tissue is found in ectopic sites outside the uterus. It causes dysmenorrhea, dyspareunia, and infertility.

Adenomyosis is the condition in which endometrial glands and stroma are found in the myometrium interspersed between the smooth muscle fibers. Leiomyomas are benign uterine wall neoplasms of smooth muscle origin. They can develop in the corpus of the uterus and can be submucosal, subserosal, or intramural. Submucosal fibroids displace endometrial tissue and are more likely to cause bleeding, necrosis, and infection than either of the other types. Endometrial cancer is the most common cancer found in the female pelvis; it occurs more than twice as often as cervical cancer. Prolonged estrogen stimulation with hyperplasia of the endometrium has been identified as a major risk factor for endometrial cancer.

Alterations in pelvic support frequently occur because of weaknesses and relaxation of the pelvic floor and perineum. Cystocele and rectocele involve herniation of the bladder or rectum into the vagina. Uterine prolapse occurs when the uterus bulges into the vagina. Pelvic relaxation disorders typically result from overstretching of the perineal supporting muscles during pregnancy and childbirth. The loss of elasticity in these structures that is a normal accompaniment of aging contributes to these problems. Variations in uterine position are common; they include anteflexion, in which the uterus is flexed forward on itself; retroflexion, in which the uterus is flexed backward at the isthmus; and retroversion, in which the uterus inclines posteriorly while the cervix remains tilted forward.

Disorders of the ovaries include benign cysts, functioning ovarian tumors, and cancer of the ovary; they usually are asymptomatic unless there is substantial enlargement or bleeding into the cyst, or the cyst becomes twisted or ruptures. Polycystic ovary syndrome is characterized by numerous cystic follicles or follicular cysts; it causes various degrees of hirsutism, obesity, and infertility. Benign ovarian tumors consist of endometriomas, which are chocolate cysts that develop secondarily to ovarian endometriosis; ovarian fibromas, which are connective tissue tumors composed of fibrocytes and collagen; and cystic teratomas or dermoid cysts, which are derived from primordial germ cells and are composed of various combinations of well-differentiated ectodermal, mesodermal, and endodermal elements. Functioning ovarian tumors are of three types: estrogen secreting, androgen secreting, and mixed estrogen-androgen secreting, and may be benign or cancerous. Cancer of the ovary is the second most common female genitourinary cancer and the most lethal. It can be divided into three categories: epithelial tumors, germ cell tumors, and gonadal stromal tumors. There are no effective screening methods for ovarian cancer, and often the disease is well advanced at the time of diagnosis.

MENSTRUAL DISORDERS

Between menarche (i.e., first menstrual bleeding) and menopause (*i.e.*, last menstrual bleeding), the female reproductive system undergoes cyclic changes called the *menstrual cycle*. This includes the maturation and release of oocytes from the ovary during ovulation and periodic vaginal bleeding resulting from the shedding of the endometrial lining. It is not necessary for a woman to ovulate to menstruate; anovulatory cycles do occur. The menstrual cycle produces changes in the breasts, uterus, skin, ovaries, and perhaps other unidentified tissues. The maintenance of the cycle affects biologic and sociologic aspects of a woman's life, including fertility, reproduction, sexuality, and femaleness.

The hormonal control of the menstrual cycle is complex. For example, the biosynthesis of estrogens that occurs in adipose tissue may be a significant source of the hormone. There is evidence that a certain minimum body weight (48 kg) and fat content (16% to 24%) are necessary for menarche to occur and for the menstrual cycle to be maintained. This is supported by the observation of amenorrhea in women with anorexia nervosa, chronic disease, and malnutrition and in those who are longdistance runners. In women with anorexia nervosa, gonadotropin and estradiol secretion, including LH release and responsiveness to the GnRH, can revert to prepubertal levels. With resumption of weight gain and attainment of sufficient body mass, the normal hormonal pattern usually is reinstated. Obesity or significant weight gain also is associated with oligomenorrhea or amenorrhea and infertility, although the mechanism is not well understood.

In addition to their effects on the growth of uterine muscle, estrogens play an important role in the development of the endometrial lining. During anovulatory cycles, continued exposure to estrogens for prolonged periods leads to abnormal hyperplasia of the endometrium and abnormal bleeding patterns. When estrogen production is poorly coordinated during the normal menstrual period, inappropriate bleeding and shedding of the endometrium also can occur.

Dysfunctional Menstrual Cycles

Normal menstrual function results from interactions among the central nervous system, hypothalamus, anterior pituitary, ovaries, and associated target tissues. Although each part of the system is essential to normal function, the ovaries are primarily responsible for controlling the cyclic changes and the length of the menstrual cycle. In most women in the middle reproductive years, menstrual bleeding occurs every 25 to 35 days, with a median length of 28 days.

Dysfunctional Bleeding

Although unexplained uterine bleeding can occur for many reasons, such as pregnancy, abortion, bleeding dyscrasias, and neoplasms, the most frequent cause in the nonpregnant female is what is commonly called *dysfunctional menstrual cycles* or *bleeding*. Dysfunctional cycles may take the form of amenorrhea (absence of menstruation), hypomenorrhea (scanty menstruation), oligomenorrhea (infrequent menstruation, periods more than 35 days apart), menorrhagia (excessive menstruation), or metrorrhagia (bleeding between periods). Menometrorrhagia is heavy bleeding during and between menstrual periods.

Dysfunctional menstrual cycles are related to alterations in the hormones that support normal cyclic endometrial changes. Estrogen deprivation causes retrogression of a previously builtup endometrium and bleeding. Such bleeding often is irregular

KEY CONCEPTS

DYSFUNCTIONAL MENSTRUAL CYCLES

- The pattern of menstrual bleeding tends to be fairly consistent in most healthy women with regard to frequency, duration, and amount of flow.
- Dysfunctional bleeding in postpubertal women can take the form of absent or scanty periods, infrequent periods, excessive and irregular periods, excessive bleeding during periods, and bleeding between periods.
- When the basic pattern of bleeding is changed, it is most often due to a lack of ovulation and disturbances in the pattern of hormone secretion.
- When the basic pattern is undisturbed and there are superimposed episodes of bleeding or spotting, the etiology is more likely to be related to organic lesions or hematologic disorders.

in amount and duration, with the flow varying with the time and degree of estrogen stimulation and with the degree of estrogen withdrawal. A lack of progesterone can cause abnormal menstrual bleeding; in its absence, estrogen induces development of a much thicker endometrial layer with a richer blood supply. The absence of progesterone results from the failure of any of the developing ovarian follicles to mature to the point of ovulation, with the subsequent formation of the corpus luteum and production and secretion of progesterone.

Periodic bleeding episodes alternating with amenorrhea are caused by variations in the number of functioning ovarian follicles present. If sufficient follicles are present and active and if new follicles assume functional capacity, high levels of estrogen develop, causing the endometrium to proliferate for weeks or even months. In time, estrogen withdrawal and bleeding develop. This can occur for two reasons: an absolute estrogen deficiency may develop when several follicles simultaneously degenerate, or a relative deficiency may develop as the needs of the enlarged endometrial tissue mass exceed the capabilities of the existing follicles, even though estrogen levels remain constant. Estrogen and progesterone deficiency are associated with the absence of ovulation, thus the term *anovulatory bleeding*. Because the vasoconstriction and myometrial contractions that normally accompany menstruation are caused by progesterone, anovulatory bleeding seldom is accompanied by cramps, and the flow frequently is heavy. Anovulatory cycles are common among adolescents during the first several years after menarche, when ovarian function is becoming established, and among perimenopausal women, whose ovarian function is beginning to decline.

Dysfunctional menstrual cycles can originate as a primary disorder of the ovaries or as a secondary defect in ovarian function related to hypothalamic-pituitary stimulation. The latter can be initiated by emotional stress, marked variation in weight (*i.e.*, sudden gain or loss), or nonspecific endocrine or metabolic disturbances. Organic causes of irregular menstrual bleeding include endometrial polyps, submucosal myoma (*i.e.*, fibroid), blood dyscrasia, infection, endometrial cancer, polycystic ovarian disease, and pregnancy.

The treatment of dysfunctional bleeding depends on what is identified as the probable cause. The minimum evaluation should include a detailed history with emphasis on bleeding pattern and a physical examination. Endocrine studies (FSH:LH ratio, prolactin, testosterone, DHAS), β -hCG pregnancy test, endometrial biopsy, D & C with or without hysteroscopy, and progesterone withdrawal tests may be needed for diagnosis. If organic problems are excluded and alterations in hormone levels are the primary cause, treatment may include the use of oral contraceptives, cyclic progesterone therapy, or long-acting progesterone injections.

Amenorrhea

There are two types of amenorrhea: primary and secondary. Primary amenorrhea is the failure to menstruate by 16 years of age, or by 14 years of age if failure to menstruate is accompanied by absence of secondary sex characteristics. Secondary amenorrhea is the cessation of menses for at least 6 months in a woman who has established normal menstrual cycles. Primary amenorrhea usually is caused by gonadal dysgenesis, congenital müllerian agenesis, testicular feminization, or a hypothalamic-pituitary-ovarian axis disorder. Causes of secondary amenorrhea include ovarian, pituitary, or hypothalamic dysfunction and destruction of the endometrial cavity by chronic infections such as tuberculosis or destruction of the endometrium by curettage (surgical scraping). Another cause is anorexia nervosa or participation in athletic activities to the extent that there is an alteration in the critical body fatmuscle ratio needed for menses to occur.²⁹

Diagnostic evaluation of amenorrhea resembles that for dysfunctional uterine bleeding, with the possible addition of a computed tomographic scan to exclude a pituitary tumor. Treatment is based on correcting the underlying cause and inducing menstruation with cyclic progesterone or combined estrogen-progesterone regimens.

Dysmenorrhea

Dysmenorrhea is pain or discomfort with menstruation. Although not usually a serious medical problem, it causes some degree of monthly disability for a significant number of women. There are two forms of dysmenorrhea: primary and secondary. Primary dysmenorrhea is menstrual pain that is not associated with a physical abnormality or pathology.³⁰ It usually occurs with ovulatory menstruation beginning 6 months to 2 years after menarche. Symptoms may begin 1 to 2 days before menses, peak on the first day of flow, and subside within several hours to several days. Severe dysmenorrhea may be associated with systemic symptoms such as headache, nausea, vomiting, diarrhea, fatigue, irritability, dizziness, and syncope. The pain typically is described as dull, lower abdominal aching or cramping, spasmodic or colicky in nature, often radiating to the lower back, labia majora, or upper thighs.

Secondary dysmenorrhea is menstrual pain caused by specific organic conditions, such as endometriosis, uterine fibroids, adenomyosis, pelvic adhesions, or PID. Laparoscopy often is required for diagnosis of secondary dysmenorrhea if medication for primary dysmenorrhea is ineffective.

Treatment for primary dysmenorrhea is directed at symptom control. Although analgesic agents such as aspirin and acetaminophen may relieve minor uterine cramping or low back pain, prostaglandin synthetase inhibitors, such as ibuprofen, naproxen, mefenamic acid, and indomethacin, are more specific for dysmenorrhea. Ovulation suppression and symptomatic relief of dysmenorrhea can be instituted simultaneously with the use of oral contraceptives. Relief of secondary dysmenorrhea depends on identifying the cause of the problem. Medical or surgical intervention may be needed to eliminate the problem.

Premenstrual Syndrome

The *premenstrual syndrome* (PMS) is a distinct clinical entity characterized by a cluster of physical and psychological symptoms limited to 3 to 14 days preceding menstruation and relieved by onset of the menses. The incidence of PMS seems to increase with age. It is less common in women in their teens and 20s, and most of the women seeking help for the problem are in their mid-30s.³¹

Although the causes of PMS are poorly documented, they probably are multifactorial. Like dysmenorrhea, it is only recently that PMS has been recognized as a bona fide disorder, rather than merely a psychosomatic illness.

The physical symptoms of PMS include painful and swollen breasts, bloating, abdominal pain, headache, and backache. Psychologically, there may be depression, anxiety, irritability, and behavioral changes. In some cases, there are puzzling alterations in motor function, such as clumsiness and altered handwriting. Women with PMS may report one or several symptoms, with symptoms varying from woman to woman and from month to month in the same patient. Signs and symptoms associated with this disorder are summarized in Table 34-3. PMS can significantly affect a woman's ability to perform at normal levels. She may lose time from or function ineffectively at work. Family responsibilities and relationships may suffer. More crimes are committed by females during the premenstrual phase of the cycle, and more lives are lost to suicide during this period. The term premenstrual dysphoric disorder is a psychiatric diagnosis that has been developed to distinguish women whose symptoms are severe enough to interfere significantly with activities of daily living or in whom the symptoms are not relieved with the onset of menstruation, as is usually the case with PMS.31,32

Diagnosis focuses on identification of the symptom clusters by means of prospective charting for at least 3 months. A complete history and physical examination are necessary to exclude other physical causes of the symptoms. Depending on the symptom pattern, blood studies, including thyroid hormones and glucose tests, may be done. Psychosocial evaluation is helpful to exclude emotional illness that is merely exacerbated premenstrually.

Treatment of PMS is directed toward an integrated program of regular exercise, avoidance of caffeine, and a diet emphasizing complex carbohydrates. Foods high in simple sugars and alcohol should be avoided. Drug therapy should be used cautiously until well-controlled studies establish criteria for use and effective treatment results.

Menopause and Aging Changes

Menopause is the cessation of menstrual cycles. Like menarche, it is more of a process than a single event. Most women stop menstruating between 48 and 55 years of age. Perimenopause (the years immediately surrounding menopause) precedes menopause by approximately 4 years and is characterized by menstrual irregularity and other menopausal symptoms. Climacteric is a more encompassing term that refers to the entire transition to the nonreproductive period of life. Premature

ovarian failure describes the approximately 1% of women who experience menopause before the age of 40 years. A woman who has not menstruated for a full year or has an FSH level greater than 30 mIU/mL is considered menopausal.33

Menopause results from the gradual cessation of ovarian function and the resultant diminished levels of estrogen. Although estrogens derived from the adrenal cortex continue to circulate in a woman's body, they are insufficient to maintain the secondary sexual characteristics in the same manner as ovarian estrogens. As a result, breast tissue, body hair, skin elasticity, and subcutaneous fat decrease; the ovaries and uterus diminish in size; and the cervix and vagina become pale and friable. Problems that can arise as a result of this urogenital atrophy include vaginal dryness, urinary stress incontinence, urgency, nocturia, vaginitis, and urinary tract infection. The woman may find intercourse painful and traumatic, although some type of vaginal lubrication may be helpful.

Systemically a woman may experience significant vasomotor instability secondary to the decrease in estrogens and the relative increase in other hormones, including FSH, LH, GnRH, dehydroepiandrosterone, and androstenedione, epinephrine, corticotropin, β -endorphin, growth hormone, and calcitonin gene-related peptide. This instability may give rise to "hot flashes," palpitations, dizziness, and headaches as the blood vessels dilate. Despite the association with these biochemical changes, the underlying cause of hot flashes is unknown.³⁴ Tremendous variation exists in the onset, frequency, severity, and length of time that women experience hot flashes. When they occur at night and are accompanied by significant perspiration, they are referred to as *night sweats*. Insomnia as well as frequent awakening because of vasomotor symptoms can lead to sleep deprivation. A woman may experience irritability, anxiety, and depression as a result of these uncontrollable and unpredictable events.

Consequences of long-term estrogen deprivation include osteoporosis, due to an imbalance in bone remodeling (*i.e.*, bone resorption occurs at a faster rate than bone formation), and an increased risk for cardiovascular disease (atherosclerosis is accelerated), which is the leading cause of death for women after menopause.

Hormone replacement therapy (HRT) has recently come under scrutiny. Topical hormone preparations are available to treat symptoms related to vaginal atrophy. Selective estrogen receptor modulators (SERMs) may be used in place of estrogen to prevent osteoporosis.

TABLE 34-3	Symptoms of Premenstrual Syndrome (PMS) by System
Body System	Symptoms
Cerebral	Irritability, anxiety, nervousness, fatigue, and exhaustion; increased physical and mental activity; lability; crying spells; depressions; inability to concentrate
Gastrointestinal	Craving for sweets or salts, lower abdominal pain, bloating, nausea, vomiting, diarrhea, constipation
Vascular	Headache, edema, weakness, or fainting
Reproductive	Swelling and tenderness of the breasts, pelvic congestion, ovarian pain, altered libido
Neuromuscular	Trembling of the extremities, changes in coordination, clumsiness, backache, leg aches
General	Weight gain, insomnia, dizziness, acne

In summary, between the menarche and menopause, the female reproductive system undergoes cyclic changes called the *menstrual cycle*. The normal menstrual function results from complex interactions among the hypothalamus, which produces GnRH; the anterior pituitary gland, which synthesizes and releases FSH, LH, and prolactin; the ovaries, which synthesize and release estrogens, progesterone, and androgens; and associated target tissues, such as the endometrium and the vaginal mucosa. Although each component of the system is essential for normal functioning, the ovarian hormones are largely responsible for controlling the cyclic changes and length of the menstrual cycle.

Menstrual disorders include dysfunctional menstrual cycles, dysmenorrhea, and PMS. Dysfunctional menstrual cycles produce amenorrhea, oligomenorrhea, metrorrhagia, or menorrhagia. Dysmenorrhea is characterized by pain or discomfort during menses. It can occur as a primary or secondary disorder. Primary dysmenorrhea is not associated with other disorders and begins soon after menarche. Secondary dysmenorrhea is caused by a specific organic condition, such as endometriosis or pelvic adhesions. It occurs in women with previously painless menses. PMS represents a cluster of physical and psychological symptoms that precede menstruation by 1 to 2 weeks.

Menopause is the cessation of ovarian function and menstrual cycles. It is accompanied by a decline in secondary sexual characteristics, vasomotor instability, and long-term consequences, including osteoporosis and increased rate of heart disease.

DISORDERS OF THE BREAST

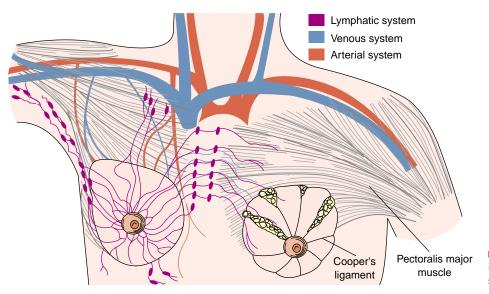
Although anatomically separate, the breasts are functionally related to the female genitourinary system in that they respond to the cyclic changes in sex hormones and produce milk for infant nourishment. Most breast diseases may be described as benign or cancerous. Benign breast conditions are nonprogressive; however, some forms of benign disease increase the risk of malignant disease.

Breast Structures

The breasts, or mammary tissues, are located between the third and seventh ribs of the anterior chest wall and are supported by the pectoral muscles and superficial fascia. They are specialized glandular structures that have an abundant shared nervous, vascular, and lymphatic supply (Fig. 34-15). Structurally the breast consists of fat, fibrous connective tissue, and glandular tissue. The superficial fibrous connective tissue is attached to the skin, a fact that is important in the visual observation of skin movement over the breast during breast self-examination. The breast mass is supported by the fascia of the pectoralis major and minor muscles and by the fibrous connective tissue of the breast. Fibrous tissue ligaments, called Cooper's ligaments, extend from the outer boundaries of the breast to the nipple area in a radial manner, like the spokes on a wheel. These ligaments further support the breast and form septa that divide the breast into 15 to 25 lobes. Each lobe consists of grapelike clusters, alveoli or glands, which are interconnected by ducts. Estrogen stimulates the growth of the ductile system, while progesterone stimulates the growth and development of the ductile and alveolar secretory epithelium.

The alveoli are lined with secretory cells capable of producing milk or fluid under the proper hormonal conditions (Fig. 34-16). The route of descent of milk and other breast secretions is from alveoli to duct, to intralobar duct, to lactiferous duct and reservoir, to nipple. Breast milk is produced secondary to complex hormonal changes associated with pregnancy. The breasts respond to the cyclic changes in the menstrual cycle with fullness and discomfort.

The nipple is made up of epithelial, glandular, erectile, and nervous tissue. Areolar tissue surrounds the nipple and is recognized as the darker, smooth skin between the nipple and the breast. The small bumps or projections on the areolar surface known as *Montgomery's tubercles* are sebaceous glands that keep the nipple area soft and elastic. At the time of puberty and during pregnancy, increased levels of estrogen and progesterone



■ FIGURE 34-15 ■ The breasts, showing the shared vascular and lymphatic supply as well as the pectoral muscles.

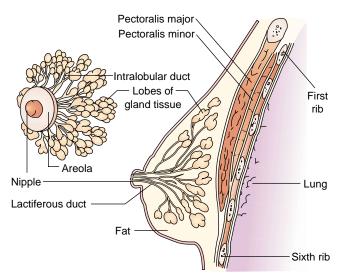


FIGURE 34-16 The breast, showing the glandular tissue and ducts of the mammary glands.

cause the areola and nipple to become darker and more prominent and Montgomery's glands to become more active. The erectile tissue of the nipple is responsive to psychological and tactile stimuli, which contributes to the sexual function of the breasts.

Mastitis

Mastitis is inflammation of the breast. It most frequently occurs during lactation but may also result from other conditions. In the lactating woman, inflammation results from an ascending infection that travels from the nipple to the ductile structures. The most common organisms isolated are *Staphylococcus* and *Streptococcus*.⁷ The offending organisms originate from the suckling infant's nasopharynx or the mother's hands. During the early weeks of nursing, the breast is particularly vulnerable to bacterial invasion because of minor cracks and fissures that occur with vigorous suckling. Infection and inflammation cause obstruction of the ductile system. The breast area becomes hard, inflamed, and tender if not treated early. Without treatment, the area becomes walled off and may abscess, requiring incision and drainage. It is advisable for the mother to continue breastfeeding during antibiotic therapy to prevent this.

Mastitis is not confined to the postpartum period; it can occur as a result of hormonal fluctuations, tumors, trauma, or skin infection. Cyclic inflammation of the breast occurs most frequently in adolescents, who commonly have fluctuating hormone levels. Tumors may cause mastitis secondary to skin involvement or lymphatic obstruction. Local trauma or infection may develop into mastitis because of ductal blockage of trapped blood, cellular debris, or the extension of superficial inflammation. The treatment for mastitis symptoms may include application of heat or cold, excision, aspiration, mild analgesics, antibiotics, and a supportive brassiere or breast binder.

Ductile Disorders

Ductal ectasia manifests in older women as a spontaneous, intermittent, usually unilateral, grayish-green nipple discharge. Palpation of the breast increases the discharge. Ectasia occurs during or after menopause and is symptomatically associated with burning, itching, pain, and a pulling sensation of the nipple and areola. The disease results in inflammation of the ducts and subsequent thickening. The treatment requires removal of the involved ductal mass.

Intraductal papillomas are benign epithelial tissue tumors that range in size from 2 mm to 5 cm. Papillomas usually manifest with a bloody nipple discharge. The tumor may be palpated in the areolar area. The papilloma is probed through the nipple, and the involved duct is removed.

Fibrocystic Disease

The term *fibrocystic breast disease* is the most common lesion of the breast. It is most common in women 30 to 50 years of age and is rare in postmenopausal women not receiving hormone replacement.³⁵ Fibrocystic disease usually presents as nodular (*i.e.*, "shotty"), granular breast masses that are more prominent and painful during the luteal or progesterone-dominant portion of the menstrual cycle. Discomfort ranges from heaviness to exquisite tenderness, depending on the degree of vascular engorgement and cystic distention.

Fibrocystic disease encompasses a wide variety of lesions and breast changes. Microscopically, fibrocystic disease refers to a constellation of morphologic changes manifested by (1) cystic dilation of terminal ducts, (2) relative increase in fibrous tissue, and (3) variable proliferation of terminal duct epithelial elements.⁷ Autopsy studies have demonstrated some degree of fibrocystic change in 60% to 80% of adult women in the United States.⁷ Symptomatic fibrocystic disease, in which large, clinically detectable cysts are present, is much less common, occurring in approximately 10% of the adult women between 35 and 50 years of age.⁷ Although fibrocystic disease often has been thought to increase the risk of breast cancer, only certain variants in which proliferation of the epithelial components is demonstrated represent a true risk.

Diagnosis of fibrocystic disease is made by physical examination, mammography, ultrasound, and biopsy (*i.e.*, aspiration or tissue sample). Ultrasound is useful in differentiating a cystic from a solid mass. Because a mass caused by fibrocystic disease may be indistinguishable from carcinoma on the basis of clinical findings, suspicious lesions should undergo biopsy. Fineneedle aspiration may be used, but if a suspect mass that was nonmalignant on cytologic examination does not resolve during the course of several months, it should be removed surgically.

Treatment for fibrocystic breast disease usually is symptomatic. Aspirin, mild analgesics, and local application of heat or cold may be recommended. Women should be encouraged to wear a good supporting brassiere and advised to avoid foods that contain xanthines (*e.g.*, coffee, cola, chocolate, and tea) in their daily diets, particularly premenstrually. Aspiration of a discrete mass indicative of a cyst may be used to relieve pain and confirm the cystic nature of the lesion. Danazol, a synthetic androgen, may be used for treatment of severe pain.

Breast Cancer

Cancer of the breast is the most common female cancer. One in eight women in the United States will have breast cancer in her lifetime. In 2000, breast cancer affected 182,800 American women and killed almost 40,800 women.³⁶ Although the breast cancer mortality rate has shown a slight decline, it is second only to lung cancer as a cause of cancer-related deaths in women. Incidence rates for carcinoma in situ have increased dramatically since the mid-1970s because of recommendations regarding mammography screening. The decline in the breast cancer mortality rate since 1989 is attributable to this earlier diagnosis as well as improvements in cancer treatments.³⁷

Risk factors for breast cancer include sex, increasing age, personal or family history of breast cancer (i.e., at highest risk are those with multiple affected first-order relatives), history of benign breast disease (*i.e.*, primary "atypical" hyperplasia), and hormonal influences that promote breast maturation and may increase the chance of cell mutation (i.e., early menarche, late menopause, and no term pregnancies or first child after 30 years of age).³⁸ Most women with breast cancer have no identifiable risk factors.

Approximately 8% of all breast cancers are hereditary.³⁸ Two breast cancer susceptibility genes-BRCA1 on chromosome 17 and BRCA2 on chromosome 13-may account for most inherited forms of breast cancer (see Chapter 5). Known carriers of these genes should begin monthly breast self-examination (BSE) at 18 years of age and begin having annual mammograms at 25 years of age.39

Breast cancer may involve the lobular or ductile structures. Invasive, or infiltrating, ductile carcinoma is the most common form of breast cancer. Invasive lobular carcinoma is the second most common form of breast cancer. It may occur alone or mixed with ductile carcinoma. Paget's disease accounts for 1% of all breast cancers. The disease presents as an eczematoid lesion of the nipple and areola. Paget's disease usually is associated with an infiltrating, intraductal carcinoma. Ductile carcinoma in situ (including Paget's disease) has increased in the past several decades from less than 5% of all carcinomas before mammography screening to 15% to 20% in well screened populations. The lesion consists of a malignant population of cells that lack the capacity to invade the basement membrane and metastasize. However, they can spread throughout the ductile system involving an entire sector of the breast.²

Detection

Cancer of the breast may manifest clinically as a mass, a puckering, nipple retraction, or unusual discharge. Many cancers are found by women themselves through BSE-sometimes when only a thickening or subtle change in breast contour is noticed. The variety of symptoms and potential for self-discovery underscore the need for regular, systematic self-examination. BSE should be done routinely by women older than 20 years of age. As an adjunct to BSE, women should have a clinical examination by a trained health professional at least every 3 years between 20 and 40 years of age, and annually after 40 years of age.

Mammography is the only effective screening technique for the early detection of clinically inapparent lesions. A generally slow-growing form of cancer, breast cancer may have been present for 2 to 9 years before it reaches 1 cm, the smallest mass normally detected by palpation. Mammography can disclose lesions as small as 1 mm and the clustering of calcifications that may warrant biopsy to exclude cancer. The American Cancer Society recommends annual evaluation for women after 40 years of age. The most comprehensive approach to screening is a combination of BSE, clinical evaluation by a health professional, and mammography.

Diagnosis and Classification

Procedures used in the diagnosis of breast cancer include physical examination, mammography, ultrasonography, percutaneous needle aspiration, stereotactic needle biopsy (i.e., core biopsy), and excisional biopsy. Figure 34-17 illustrates the appearance of breast cancer on mammography. Breast cancer often manifests as a solitary, painless, firm, fixed lesion with poorly defined borders. It can be found anywhere in the breast but is most common in the upper outer quadrant. Because of the variability in presentation, any suspect change in breast tissue warrants further investigation. The diagnostic use of mammography enables additional definition of the clinically suspect area (e.g., appearance, character, calcification). Placement of a wire marker under radiographic guidance can ensure accurate surgical biopsy of nonpalpable suspect areas. Ultrasonography is useful as a diagnostic adjunct to differentiate cystic from solid tissue in women with nonspecific thickening.

Fine-needle aspiration is a simple in-office procedure that can be performed repeatedly in multiple sites and with minimal discomfort. It can be used to identify the presence of malignant cells, but it cannot differentiate in situ from infiltrating cancers. Stereotactic needle biopsy is an outpatient procedure done with the guidance of a mammography machine. After the lesion is localized radiologically, a large-bore needle is mechanically thrust quickly into the area, removing a core of tissue. Excisional biopsy to remove the entire lump provides the only definitive diagnosis of breast cancer and often is therapeutic without additional surgery.

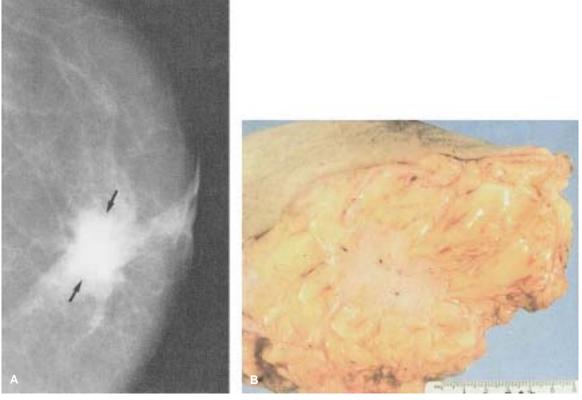
Tumors are classified histologically according to tissue characteristics and staged clinically according to tumor size, nodal involvement, and presence of metastasis. It is recommended that estrogen and progesterone receptor analysis be performed on surgical specimens. Information about the presence or absence of estrogen and progesterone receptors can be used in predicting tumor responsiveness to hormonal manipulation. High levels of both receptors improve the prognosis and increase the likelihood of remission.

Treatment

The treatment methods for breast cancer include surgery, chemotherapy, radiation therapy, and hormonal manipulation.⁴⁰ Radical mastectomy (*i.e.*, removal of the entire breast, underlying muscles, and all axillary nodes) rarely is used today as a primary surgical therapy unless breast cancer is advanced at the time of diagnosis. Modified surgical techniques (*i.e.*, mastectomy plus axillary dissection or lumpectomy for breast conservation) accompanied by chemotherapy or radiation therapy have achieved outcomes comparable with those obtained with radical surgical methods and constitute the preferred treatment methods.

The prognosis and need for adjuvant systemic therapy is related more to the extent of nodal involvement than to the extent of breast involvement. A newer technique for evaluating lymph node involvement is sentinel lymph node biopsy. A radioactive substance or dye is injected into the region of the tumor. In theory, the dye is carried to the first (sentinel) node to receive lymph from the tumor. This would therefore be the node most likely to contain cancer cells if the cancer has spread. If the sentinel node biopsy is positive, more nodes are removed. If it is negative, further lymph node evaluation may not be needed. It is not always possible to identify the sentinel node.41

624



■ FIGURE 34-17 ■ Carcinoma of the breast. (A) Mammogram. An irregularly shaped, dense mass (*arrows*) is seen in this otherwise fatty breast. (B) Mastectomy specimen. The irregular white, firm mass in the center is surrounded by fatty tissue. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 1042]. Philadelphia: Lippincott Williams & Wilkins)

Adjuvant systemic therapy refers to the administration of chemotherapy or hormonal therapy to women without detectable metastatic disease. The goal of this therapy depends on nodal involvement, menopausal status, and hormone receptor status. Tamoxifen is a nonsteroidal antiestrogen that binds to estrogen receptors and blocks the effects of estrogens on the growth of malignant cells in the breast. Studies have shown decreased cancer recurrence, decreased mortality rates, and increased 5-year survival rates in women with estrogen receptorpositive tissue samples who have been treated with the drug. Autologous bone marrow transplantation and peripheral stem cell transplantation are experimental therapies that may be used for the treatment of advanced disease or in women at increased risk for recurrence.

In summary, the breasts are subject to benign and malignant disease. Mastitis is inflammation of the breast, occurring most frequently during lactation. Ductal ectasia and intraductal papilloma cause abnormal drainage from the nipple. Fibroadenoma and fibrocystic disease are characterized by abnormal masses in the breast that are benign. By far the most important disease of the breast is breast cancer, which is a significant cause of death of women. BSE and mammography afford a woman the best protection against breast cancer. They provide the means for early detection of breast cancer and, in many cases, allow early treatment and cure.

REVIEW QUESTIONS

List the actions of estrogen and progesterone and explain the interactions among the gonadotropin-releasing hormones, LH, and FSH; ovarian follicle development; and estrogen and progesterone levels.

- Compare the extragenital abnormalities associated with Bartholin cyst and abscess, vulvar dystrophy, vulvodynia, and cancer of the vulva.
- Characterize the development of cervical cancer, from the appearance of atypical cells to the development of invasive cervical cancer, and the role of the Pap smear in early detection.
- State the proposed relationship between unopposed estrogen stimulation of the endometrium and endometrial cancer.
- List the common causes and symptoms of pelvic inflammatory disease.
- State the underlying cause of ovarian cysts.
- Differentiate benign ovarian cyst from polycystic ovary syndrome.
- State the reason that ovarian cancer may be difficult to detect at an early stage.
- Describe the manifestations of cystocele, rectocele, and enterocele.
- Compare the symptoms of primary dysmenorrhea with those of secondary dysmenorrhea.

Characterize the manifestations of the premenstrual syndrome, its possible causes, and the methods of treatment.

Describe the physiology of normal menopause.

Describe changes in breast function that occur with mastitis and ductal ectasia.

Describe the manifestations of fibrocystic disease and state why it is often referred to as a "catchall" for breast irregularities.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- Goldfein A. (2001). The ovary. In Greenspan F.S., Gardner D.G. (Eds.), *Basic and clinical endocrinology* (6th ed., pp. 453–495). New York: Lange Medical Books/McGraw-Hill.
- Crum C. (1999). The female genital tract. In Cotran R.S., Kumar V., Collins T. (Eds.), *Robbins pathologic basis of disease* (6th ed., pp. 1034– 1035, 1042–1044, 1061–1062, 1067–1069). Philadelphia: W.B. Saunders.
- Guyton A.C., Hall J.E. (2001). *Textbook of medical physiology* (10th ed., pp. 929–940). Philadelphia: W.B. Saunders.
- Wilkinson E.J., Stone I.K. (1995). Atlas of vulvar disease (pp. 16, 33–35, 101). Baltimore: Williams & Wilkins.
- Glazer H.I. (2000). Dysesthetic vulvodynia: Long term follow-up after treatment with surface electromyography-assisted pelvic floor muscle rehabilitation. *Journal of Reproductive Medicine* 45, 798–802.
- Masheb R.M., Nash J.M., Brondolo E., et al. (2000). Vulvodynia: An introduction and critical review of a chronic pain condition. *Pain* 86, 3–10.
- Robboy S., Duggan M.A., Kurman R.J. (1999). The female reproductive system. In Rubin E., Farber J.L. (Eds.), *Pathology* (3rd ed., pp. 963– 1026): Philadelphia: Lippincott Williams & Wilkins.
- Sobel J.D. (1997). Vaginitis. New England Journal of Medicine 337, 1896–1903.
- Egan M.E., Lipsky M. (2000). Diagnosis of vaginitis. American Family Physician 62, 1095–1104.
- National Cancer Institute, National Institute of Child Health and Human Development, National Institutes of Health. (1995). Clear cell carcinoma: Resource guide for DES-exposed daughters and their families (pp. 3, 5, 24). Rockville, MD: National Institutes of Health.
- Janicek M.F., Averette H.E. (2001). Cervical cancer: Prevention, diagnosis, and therapeutics. CA—A Cancer Journal for Clinicians 51 (2), 92–114.
- Crum C.P. (2002). The beginning of the end of cervical cancer. New England Journal of Medicine 347 (21), 1703–1705.
- Scott J.R., DiSaia P.J., Hammond C.B., et al. (1999). Danforth's obstetrics and gynecology (8th ed., pp. 589, 815, 818). Philadelphia: Lippincott Williams & Wilkins.
- 14. Mott A.M. (2000). Prevention and management of pelvic inflammatory disease by primary care providers. *American Journal of Nurse Practitioners* 8, 7–13.
- Lessey B.A. (2000). Medical management of endometriosis and infertility. *Fertility and Sterility* 73, 1089–1096.
- 16. Propst A.M., Laufer M.R. (2000). Diagnosing and treating adolescent endometriosis. *Contemporary Nurse Practitioner* Spring, 11–18.
- Feste J.R., Schattman G.L. (2000). Laparoscopy and endometriosis: Preventing complications and improving outcomes. In *Changing perspectives: A new outlook on gynecologic disorders. Symposia proceedings* 2000 (pp. 40–42). Medical Education Collaborative.

- American College of Obstetricians and Gynecologists. (1999). Medical management of endometriosis. ACOG Practice Bulletin no. 11. In 2001 Compendium of selected publications (pp. 982, 986). Washington, DC: Author.
- Seltzer V.L., Pearse W.H. (1999). Women's primary health care: Office practice and procedures (2nd ed., pp. 230–231, 320, 325). New York: McGraw-Hill.
- Finan M.A., Kline R.C. (1998). Current management of endometrial cancer. Contemporary Obstetrics and Gynecology 12, 86–97.
- 21. Canavan T.P., Doshi N.R. (1999). Endometrial cancer. American Family Practitioner 59, 3069–3077.
- 22. Hutchins F.J. (1998). Fibroids in primary care. The Clinical Advisor 9, 29.
- Franks S. (1995). Polycystic ovary syndrome. New England Journal of Medicine 333, 853–861.
- Speroff L., Glass R.H., Kase N.G. (1999). Clinical gynecologic endocrinology and infertility (6th ed., pp. 497–503), Philadelphia: Lippincott Williams & Wilkins.
- Velzquez E., Acosta A., Mendoza S.G. (1997). Menstrual cyclicity after metformin in polycystic ovary syndrome. *Obstetrics and Gynecology* 90, 392–395.
- 26. American Society for Reproductive Medicine. (2000). Use of insulin sensitizing agents in the treatment of polycystic ovary syndrome. Practice Committee Report. Birmingham, AL: Author.
- Fields A.L., Jones J.G., Thomas G.M., et. al. (2001). Gynecological cancer. In Lenhard R.E. Jr., Osteen R.T., Gansler T. (Eds.), *The American Cancer Society's clinical oncology* (pp. 455–496). Atlanta: American Cancer Society.
- National Institutes of Health Consensus Development Panel on Ovarian Cancer. (1995). Ovarian cancer: Screening and follow-up. *Journal of the American Medical Association* 273, 491–497.
- American College of Obstetricians and Gynecologists. (2000). Management of anovulatory bleeding. ACOG Practice Bulletin no. 14. In 2001 Compendium of selected publications (pp. 961–968). Washington, DC: Author.
- Coco A.S. (1999). Primary dysmenorrhea. American Family Practitioner 60, 489–496.
- Star W.L., Lommel L.L., Shannon M.T. (1995). Women's primary health care: Protocols for practice (pp. 12–153). Washington, DC: American Nurses Publishing.
- Frye G.M., Silverman S.D. (2000). Is it premenstrual syndrome? Keys to focused diagnosis, therapies for multiple symptoms. *Postgraduate Medicine* 107 (5), 151–159.
- Rebar R., Gass M. (2000). Hormonal changes and symptomatic management. Clinical Bulletins in Menopause 1, 2.
- Speroff L. (2000). Postmenopausal HRT and CHD: Clinical implications of recent randomized trial results. *Contemporary Obstetrics/ Gynecology* 8, 89–104.
- Giuliano A.E. (2001). Benign breast disorders. In Tierney L.M., McPhee S.J., Papadakis M.A. (Eds.), *Current medical diagnosis and treatment* (40th ed., pp. 706–707). New York: Lange Medical Books/ McGraw-Hill.
- Osteen R.T. (2001). Breast cancer. In Lenhard R.E. Jr., Osteen R.T., Gansler T. (Eds.), *The American Cancer Society's clinical oncology* (pp. 251–268). Atlanta: American Cancer Society.
- McCance K.L., Jorde L.B. (1998). Evaluating the genetic risk of breast cancer. Nurse Practitioner 23 (8), 14–16, 19–20, 23–27.
- Apantaku L.M. (2000). Breast cancer diagnosis and screening. American Family Physician 62, 596–606.
- Burke W., Daly M., Garber J., et al. (1997). Recommendations for follow-up care of individuals with an inherited predisposition to cancer: II. BRCA1 and BRCA2. Cancer Genetics Studies Consortium. *Journal of the American Medical Association* 227, 997–1003.
- 40. Hortobagyi G.N. (1998). Treatment of breast cancer. New England Journal of Medicine 339, 974–984.
- McMasters K.M., Guilliano A.E., Ross M.I., et al. (1998). Sentinellymph-node biopsy for breast cancer: Not yet the standard of care. *New England Journal of Medicine* 339, 990–996.

35

CHAPTER

Sexually Transmitted Diseases

Infections of the External Genitalia Human Papillomavirus (Condylomata Acuminata) Genital Herpes Chancroid Lymphogranuloma Venereum Vaginal Infections Candidiasis Trichomoniasis Bacterial Vaginosis (Nonspecific Vaginitis) Vaginal-Urogenital-Systemic Infections Chlamydial Infections Gonorrhea Syphilis

he incidence and types of sexually transmitted diseases (STDs), as reported in the professional literature and public health statistics, are increasing. However, the incidence of disease is based on clinical reports, and many STDs are not reportable or not reported. The agents of transmission include bacteria, chlamydiae, viruses, fungi, protozoa, parasites, and unidentified microorganisms. Portals of entry include the mouth, genitalia, urinary meatus, rectum, and skin. All STDs are more common in persons who have more than one sexual partner, and it is not uncommon for a person to be concurrently infected with more than one type of STD.

This chapter discusses the manifestations of STDs in men and women in terms of infections of the external genitalia, vaginal infections, and infections that have systemic effects and genitourinary manifestations. Human immunodeficiency virus (HIV) infection is presented in Chapter 10.

INFECTIONS OF THE EXTERNAL GENITALIA

Some STDs primarily affect the mucocutaneous tissues of the external genitalia. These include human papillomavirus infection, genital herpes, chancroid, granuloma inguinale, and lymphogranuloma venereum (LGV).

Human Papillomavirus (Condylomata Acuminata)

Condylomata acuminata, or *genital warts*, are caused by the human papillomavirus (HPV). Although recognized for centuries, HPV-induced genital warts have become one of the fastest-growing STDs of the past decade. The Centers for Disease Control and Prevention (CDC) estimates that 20 million Americans carry the virus and that as many as 5.5 million new cases are diagnosed each year.¹ The current prevalence of HPV is difficult to determine because it is not a reportable disease in all states.

A 1998 American Medical Association consensus conference on external genital warts identified four specific types of warts: *condyloma acuminata* (cauliflower-shaped lesions that tend to appear on moist skin surfaces such as the vaginal introitus or anus); *keratotic warts* (display a thick, horny layer; develop on dry, fully keratinized skin such as the penis, scrotum, or labia majora); *papular warts* (smooth surface; typically develop on fully keratinized skin); and *flat warts* (macular, sometimes faintly raised, usually invisible to the naked eye; occur on either fully or partially keratinized skin). Biopsy may be required to differentiate warts from other hyperkeratotic or precancerous lesions.²

A relation between HPV and genital (*i.e.*, cervix, vulva, and penis) neoplasms has become increasingly apparent since the early 1980s. One hundred types of HPV have been identified, more than 30 of which affect the anogenital area. Types 6 and 11 are found in most external genital warts but usually are benign, with only a low potential for dysplasia. Persons with visible genital warts may be infected simultaneously with multiple HPV types. Other HPV types of the anogenital region (*e.g.*, types 16, 18, 31, 33, and 35) have been strongly associated with cervical neoplasia.³ HPV type 16 is present in 50%

KEY CONCEPTS

SEXUALLY TRANSMITTED DISEASE

- Sexually transmitted diseases (STDs) are spread by sexual contact and involve both male and female partners. Portals of entry include the mouth, genitalia, urinary meatus, rectum, and skin. All STDs are more common in persons who have more than one sexual partner, and it is not uncommon for a person to be concurrently infected with more than one type of STD.
- In general, STDs due to bacterial pathogens can be successfully treated and the pathogen eliminated by antimicrobial therapy. However, many of these pathogens are developing antibiotic resistance.
- STDs due to viral pathogens, such as the human papillomavirus (HPV) and genital herpes simplex virus infections (HSV-1 and HSV-2), are not eliminated by current treatment modalities and persist with risk of recurrence (HSV infections) or increased cancer risk (HPV).
- Untreated, STDs such as chlamydial infection and gonorrhea can spread to involve the internal genital organs with risk of complications and infertility.
- Intrauterine or perinatally transmitted STDs can have potentially fatal or severely debilitating effects on a fetus or an infant.

of cervical cancers and in 25% of low-grade cervical intraepithelial neoplasias⁴ (see Chapter 34). However, only a subset of women with HPV go on to develop cancer, suggesting that there may be variants of even the most virulent HPV, type 16, with differing oncogenic potential. Cofactors that may increase the risk for cancer include smoking, immunosuppression, and exposure to hormonal alteration (*e.g.*, pregnancy, oral contraceptives).⁵ The association with premalignant and malignant changes has increased the concern about diagnosis and treatment of this viral infection.

HPV infection begins with viral inoculation into a stratified squamous epithelium, where infection stimulates the replication of the squamous epithelium, producing the various HPV-proliferative lesions. The incubation period for HPVinduced genital warts ranges from 6 weeks to 8 months. Subclinical infection occurs more frequently than visible genital warts among men and women. Infection often is indirectly diagnosed on the cervix by Papanicolaou testing (Pap smear), colposcopy, or biopsy. Both spontaneous resolution and infection with new HPV types are common. Although reinfection from sexual partners has been considered as a reason for the high prevalence of this disease, it is now thought that reinfection with the same HPV type is infrequent. Instead, it is thought that HPV may be a lifelong infection.

Genital condylomas should be considered in any woman who presents with the primary complaint of vulvar pruritus or who has had an abnormal Pap smear. Microscopic examination of a wet-mount slide preparation and cultures are used to exclude associated vaginitis. Acetic acid soaks may be used before inspecting the vulva under magnification, and specimens for biopsy can be taken from questionable areas. Colposcopic examination of the cervix and vagina may be advised as a followup measure when there is an abnormal Pap smear or when HPV lesions are identified on the vulva. Evaluation and treatment of sexual partners may be suggested, although this may be difficult considering that warts often do not become clinically apparent for several years after exposure.

The recent development and controlled trial of a vaccine to protect against HPV type 16 may eventually reduce the risk of cervical cancer associated with this strain of HPV.⁴ However, currently there is no treatment to eradicate the virus once a person has become infected. Thus, treatment goals are aimed at elimination of symptomatic warts, surveillance for malignancy and premalignant changes, and education and counseling to decrease psychosocial distress.⁶ Prevention of HPV transmission through condom use has not been adequately demonstrated.

The CDC recommends several pharmacologic agents for symptomatic removal of visible genital warts, including patientapplied therapies (podofilox and imiquimod) and provideradministered therapies (podophyllin and trichloroacetic acid).³ Podophyllin, a topical cytotoxic agent, has long been used for the treatment of visible external growths. Trichloroacetic acid is a weak destructive agent that produces an initial burning in the affected area, followed in several days by a sloughing of the superficial tissue. Imiquimod cream is a new type of therapeutic agent that stimulates the body's immune system (*i.e.*, production of interferon- α and other cytokines).

Genital warts also may removed using cryotherapy, laser surgery, or electrocautery. Because it can penetrate deeper than other forms of therapy, cryotherapy (*i.e.*, freezing therapy) often is the treatment of choice for cervical HPV lesions. Laser surgery can be used to remove large or widespread lesions of the cervix, vagina, or vulva, or lesions that have failed to respond to other first-line methods of treatment. Electrosurgical treatment has become more widespread for these types of lesions because it is more readily available in outpatient settings and is much less expensive than laser.

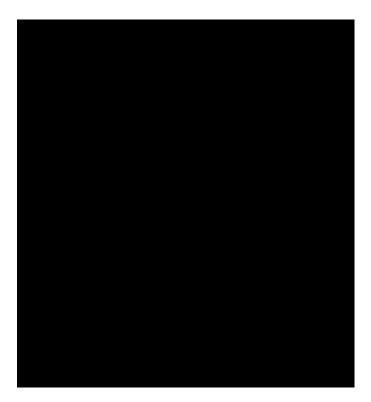
Genital Herpes

Herpesviruses are large, encapsulated viruses that have a doublestranded genome. There are nine types of herpesviruses, belonging to three groups, that cause infections in humans: (1) neurotropic α -group viruses, including herpes simplex virus type 1 (HSV-1; usually associated with cold sores) and HSV-2 (usually associated with genital herpes); (2) varicella-zoster virus (causes chickenpox and shingles); and (3) lymphotropic β -group viruses, including cytomegalovirus (causes cytomegalic inclusion disease), Epstein-Barr virus (causes infectious mononucleosis and Burkitt's lymphoma), and human herpesvirus type 8 (the apparent cause of Kaposi's sarcoma).⁷

Genital herpes is caused by the herpes simplex virus. Because herpesvirus infection is not reportable in all states, reliable data on its true incidence (estimated number of new cases every year) and prevalence (estimated number of people currently infected) are lacking. From the late 1970s to early 1990s, genital herpes prevalence increased 30%. Incidence rates have been relatively stable since 1990, with an estimated 1 million new cases occurring each year. Recent estimates in the United States indicate 50 million people (one in five adolescents or adults) are infected with genital herpes.³ Women have a greater mucosal surface area exposed in the genital area and therefore are at greater risk of acquiring the infection.

HSV-1 and HSV-2 are genetically similar, both cause a similar set of primary and recurrent infections, and both can cause genital lesions. Both viruses replicate in the skin and mucous membranes at the site of infection (oropharynx or genitalia), where they cause vesicular lesions of the epidermis and infect the neurons that innervate the area. HSV-1 and HSV-2 are neurotropic viruses, meaning that they grow in neurons and share the biologic property of latency. Latency refers to the ability to maintain disease potential in the absence of clinical signs and symptoms. In genital herpes, the virus ascends through the peripheral nerves to the sacral dorsal root ganglia (Fig. 35-1). The virus can remain dormant in the dorsal root ganglia, or it can reactivate, in which case the viral particles are transported back down the nerve root to the skin, where they multiply and cause a lesion to develop. During the dormant or latent period, the virus replicates in a different manner so that the immune system or available treatments have no effect on it. It is not known what reactivates the virus. It may be that the body's defense mechanisms are altered. Numerous studies have shown that host responses to infection influence initial development of the disease, severity of infection, development and maintenance of latency, and the frequency of HSV recurrences.

HSV is transmitted by contact with infectious lesions or secretions. HSV-1 is transmitted by oral secretions, and infections frequently occur in childhood, with most persons (50% to 90%) being infected by adulthood.⁸ HSV-1 may be spread to the genital area by autoinoculation after poor hand washing or through oral intercourse. HSV-2 usually is transmitted by sex-



ual contact but can be passed to an infant during childbirth if the virus is actively being shed from the genital tract. Most cases of HSV-2 infection are subclinical, manifesting as truly asymptomatic or symptomatic but unrecognized infections. These subclinical infections can occur in people who have never had a symptomatic outbreak or between recognized clinical recurrences. Up to 70% of genital herpes cases are spread through asymptomatic shedding by people who do not realize they have the infection.³ This "unknown" transmission of the virus to sex partners explains why this infection has reached epidemic proportions throughout the world.

The incubation period for HSV is 2 to 10 days. Genital HSV infection may manifest as a primary, nonprimary, or recurrent infection. *Primary infections* are infections that occur in a person who is seronegative for antibody to HSV-1 or HSV-2. *Initial nonprimary infections* refer to the first clinical episode in a person who is seropositive for antibodies to the opposite HSV type (usually genital herpes in someone seropositive to HSV-1). *Recurrent infections* refer to the second or subsequent outbreak caused by the same virus type. HSV-2 is responsible for more than 90% of recurrent genital herpes infections.⁶

The initial symptoms of primary genital herpes infections include tingling, itching, and pain in the genital area, followed by eruption of small pustules and vesicles. These lesions rupture on approximately the fifth day to form wet ulcers that are excruciatingly painful to touch and can be associated with dysuria, dyspareunia, and urine retention. Involvement of the cervix and urethra is seen in more than 80% of women with primary infections.⁹ In men, the infection can cause urethritis and lesions of the penis and scrotum. Rectal and perianal infections are possible with anal contact. Systemic symptoms associated with primary infections include fever, headache, malaise, muscle ache, and lymphadenopathy. Primary infections may be debilitating enough to require hospitalization, particularly in women.

Untreated primary infections typically are self-limited and last for approximately 2 to 4 weeks. The symptoms usually worsen for the first 10 to 12 days. This period is followed by a 10- to 12-day interval during which the lesions crust over and gradually heal. Nonprimary episodes of genital herpes manifest with less severe symptoms that usually are of shorter duration and have fewer systemic manifestations. Except for the greater tendency of HSV-2 to recur, the clinical manifestations of HSV-2 and genital HSV-1 are similar. Recurrent HSV infection results from reactivation of the virus stored in the dorsal root ganglia of the infected dermatomes. An outbreak may be preceded by a prodrome of itching, burning, or tingling at the site of future lesions. Because immune lymphocytes have already developed from the primary infection, recurrent episodes have fewer lesions, fewer systemic symptoms, less pain, and a shorter duration (7 to 10 days). The frequency and severity of recurrences vary from person to person. Numerous factors, including emotional stress, lack of sleep, overexertion, other infections, vigorous or prolonged coitus, and premenstrual or menstrual distress have been identified as triggering mechanisms.

Diagnosis of genital herpes is based on the symptoms, appearance of the lesions, and identification of the virus from cultures taken from the lesions. The likelihood of obtaining a positive culture decreases with each day that has elapsed after a lesion develops. The chance of obtaining a positive culture from a crusted lesion is slight, and patients suspected of having genital herpes should be instructed to have a culture within 48 hours of development of new lesions. Type specific (HSV-1 and HSV-2) serologic tests are available for determining past infection. Because almost all HSV-2 infections are sexually acquired, the presence of type-specific HSV-2 antibodies usually indicates anogenital infection; whereas the presence of HSV-1 antibodies does not distinguish between anogenital and orolabial infections. The CDC recommends that serologic assays for HSV-2 be available for persons who request them but does not recommend they be used for screening of the general population.³

There is no known cure for genital herpes, and the methods of treatment are largely symptomatic. The antiviral drugs acyclovir, valacyclovir, and famciclovir have become the cornerstone for management of genital herpes. By interfering with viral DNA replication, these drugs decrease the frequency of recurrences, shorten the duration of active lesions, reduce the number of new lesions formed, and decrease viral shedding with primary infections. Good hygiene is essential to prevent secondary HSV infection. Fastidious hand washing is recommended to avoid hand-to-eye spread of the infection. HSV infection of the eye is the most common cause of corneal blindness in the United States. To prevent spread of the disease, intimate contact should be avoided until lesions are completely healed.

Approximately 30% to 50% of infants born vaginally to mothers experiencing a primary HSV infection at the time of delivery will be infected, compared with only 1% of those born to women with recurrent infection.³ The risk of mortality in HSV-infected neonates ranges from 15% to 57%, and a significant number of survivors have significant sequelae.¹⁰ Active infection during labor may necessitate cesarean delivery.

Chancroid

Chancroid (*i.e.*, soft chancre) is a disease of the external genitalia and lymph nodes. The causative organism is the gramnegative bacterium *Haemophilus ducreyi*, which causes acute ulcerative lesions with profuse discharge. This disease has become uncommon in the United States, with only 143 reported cases in 1999.¹ It typically occurs in discrete outbreaks, rather than as an endemic disease in this country. It is more prevalent in Southeast Asia, the West Indies, and North Africa. A highly infectious disease, chancroid usually is transmitted by sexual intercourse or through skin and mucous membrane abrasions. Autoinoculation may lead to multiple chancres.

Lesions begin as macules, progress to pustules, and then rupture. This painful ulcer has a necrotic base and jagged edges. In contrast, the syphilitic chancre is nontender and indurated. Subsequent discharge can lead to further infection of self or others. On physical examination, lesions and regional lymphadenopathy (*i.e.*, buboes) may be found. Secondary infection may cause significant tissue destruction. Diagnosis usually is made clinically but may be confirmed through culture. Gram's stain rarely is used today because it is insensitive and nonspecific. Polymerase chain reaction (PCR) methods may soon be available commercially for definitive identification of *H. ducreyi*. The organism has shown resistance to treatment with sulfamethoxazole alone and to tetracycline. The CDC recommends treatment with azithromycin, erythromycin, ciprofloxacin, or ceftriaxone.³

Lymphogranuloma Venereum

Lymphogranuloma venereum (LGV) is an acute and chronic venereal disease caused by *Chlamydia trachomatis* types L1, L2, and P3. The disease, although found worldwide, has a low incidence outside the tropics. Most cases reported in the United States are in men.

The lesions of LGV can incubate for a few days to several weeks and thereafter cause small, painless papules or vesicles that may go undetected. An important characteristic of the disease is the early (1 to 4 weeks later) development of large, tender, and sometimes fluctuant inguinal lymph nodes called buboes. There may be flulike symptoms with joint pain, rash, weight loss, pneumonitis, tachycardia, splenomegaly, and proctitis. In later stages of the disease, a small percentage of affected persons develop elephantiasis of the external genitalia, caused by lymphatic obstruction or fibrous strictures of the rectum or urethra from inflammation and scarring. Urethral involvement may cause pyuria and dysuria. Cervicitis is a common manifestation of primary LGV and could extend to perimetritis or salpingitis, which are known to occur in other chlamydial infections.⁵ Anorectal structures may be compromised to the point of incontinence. Complications of LGV may be minor or extensive, involving compromise of whole systems or progression to a cancerous state.

Diagnosis usually is accomplished by means of a complement fixation test for LGV-specific *Chlamydia* antibodies. High titers for this antibody differentiate this group from other chlamydial subgroups. Treatment involves 3 weeks of doxycycline or erythromycin.³ Surgery may be required to correct sequelae such as strictures or fistulas or to drain fluctuant lymph nodes.

In summary, STDs that primarily affect the external genitalia include HPV (condyloma acuminata), genital herpes (HSV-2), chancroid, and lymphogranuloma venereum. The lesions of these infections occur on the external genitalia of male and female sexual partners. Of concern is the relation between HPV and genital neoplasms. Genital herpes is caused by a neurotropic virus (HSV-2) that ascends through the peripheral nerves to reside in the sacral dorsal root ganglia. The herpesvirus can be reactivated, producing recurrent lesions in genital structures that are supplied by the peripheral nerves of the affected ganglia. There is no permanent cure for herpes infections. Chancroid and lymphogranuloma venereum produce external genital lesions with various degrees of inguinal lymph node involvement.

VAGINAL INFECTIONS

Candidiasis, trichomoniasis, and bacterial vaginosis are vaginal infections that can be sexually transmitted. Although these infections can be transmitted sexually, the male partner usually is asymptomatic.

Candidiasis

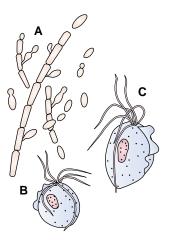
Also called *yeast infection, thrush,* and *moniliasis,* candidiasis is the second leading cause of vulvovaginitis in the United States. Approximately 75% of reproductive-age women in the United

States experience one episode in their lifetime; 40% to 45% experience two or more infections.5

The causative organism is Candida, a genus of yeastlike fungi. The species most commonly identified is Candida albicans, but other candidal species, such as Candida glabrata and Candida tropicalis, have caused symptoms. Although vulvovaginal candidiasis usually is not transmitted sexually, it is included in the CDC STD treatment guidelines because it often is diagnosed in women being evaluated for STDs.3 The possibility of sexual transmission has been recognized for many years; however, candidiasis requires a favorable environment for growth. The gastrointestinal tract also serves as a reservoir for this organism, and candidiasis can develop through autoinoculation in women who are not sexually active. Although studies have documented the presence of Candida on the penis of male partners of women with vulvovaginal candidiasis, few men develop balanoposthitis that requires treatment.

Causes for the overgrowth of C. albicans include antibiotic therapy, which suppresses the normal protective bacterial flora; high hormone levels associated with pregnancy or the use of oral contraceptives, which cause an increase in vaginal glycogen stores; and diabetes mellitus or HIV infection because they compromise the immune system. In obese persons, Candida may grow in skin folds underneath the breast tissue, the abdominal flap, and the inguinal folds. Vulvar pruritus accompanied by irritation, dysuria, dyspareunia, erythema, and an odorless, thick, cheesy vaginal discharge are the predominant symptoms of the infection. Accurate diagnosis is made by identification of budding yeast filaments (*i.e.*, hyphae) or spores on a wet-mount slide using 20% potassium hydroxide (Fig. 35-2). The pH of the discharge, which is checked with litmus paper, typically is less than 4.5. When the wet-mount technique is negative but the clinical manifestations are indicative of candidiasis, a culture may be necessary.

Antifungal agents such as clotrimazole, miconazole, butoconazole, and terconazole, in various forms, are effective in treating candidiasis. These drugs, with the exception of terconazole, are available without prescription for use by women who have had a previously confirmed diagnosis of candidiasis.



■ FIGURE 35-2 ■ Organisms that cause vaginal infections. (A) Candida albicans (blastospores and pseudohyphae). (B, C) Trichomonas vaginalis.

Oral fluconazole has been shown to be as safe and effective as the standard intravaginal regimens.³ Tepid sodium bicarbonate baths, clothing that allows adequate ventilation, and the application of cornstarch to dry the area may increase comfort during treatment. Chronic vulvovaginal candidiasis, defined as four or more mycologically confirmed episodes within 1 year, affects approximately 5% of women and is difficult to manage. Subsequent prophylaxis (maintenance therapy) often is required for the long-term management of this problem.¹¹

Trichomoniasis

An anaerobic protozoan that can be transmitted sexually, Trichomonas vaginalis is shaped like a turnip and has three or four anterior flagella (see Fig. 35-2). Trichomonads can reside in the paraurethral glands of both sexes. Males harbor the organism in the urethra and prostate and are asymptomatic. Although 10% to 25% of women are asymptomatic, trichomoniasis is a common cause of vaginitis when some imbalance allows the protozoan to proliferate. Five million cases of trichomoniasis were diagnosed in 1999.¹ This extracellular parasite feeds on the vaginal mucosa and ingests bacteria and leukocytes. The infection causes a copious, frothy, malodorous, green or yellow discharge. There commonly is erythema and edema of the affected mucosa, with occasional itching and irritation. Sometimes, small hemorrhagic areas, called strawberry spots, appear on the cervix.

Diagnosis is made microscopically by identification of the protozoan on a wet-mount slide preparation. The pH of the discharge usually is greater than 6.0. Special culture media are available for diagnosis but are costly and not needed for diagnosis.

Because the organism resides in urogenital structures other than the vagina, systemic treatment is recommended. The treatment of choice is oral metronidazole (Flagyl), a medication that is effective against anaerobic protozoans.³ Metronidazole is chemically similar to disulfiram (Antabuse), a drug used in the treatment of alcohol addiction that causes nausea, vomiting, flushing of the skin, headache, palpitations, and lowering of the blood pressure when alcohol is ingested. Alcohol should be avoided during and for 24 to 48 hours after treatment. Gastrointestinal disturbances and a metallic taste in the mouth are potential adverse effects of the drug. Metronidazole has not been proven safe for use during pregnancy and is used only after the first trimester for fear of potential teratogenic effects. Sexual partners should be treated to avoid reinfection, and abstinence is recommended until the full course of therapy is completed.

Bacterial Vaginosis (Nonspecific Vaginitis)

Bacterial vaginosis is a vaginal infection that produces a characteristic fishy- or ammonia-smelling discharge yet fails to produce an inflammatory response that is characteristic of most infections.

Bacterial vaginosis represents an upheaval in the complex vaginal bacterial flora with disappearance of the normal lactobacillus species in the vagina and an overgrowth of other organisms, including Gardnerella vaginalis and resident anaerobic vaginal bacterial.¹² It has been suggested that the presence of anaerobes, which produce ammonia or amines from amino

acids, favors the growth of *G. vaginalis* by raising vaginal pH. Because of the presence of anaerobic bacteria and the lack of an inflammatory response, the disorder has come to be called *bacterial vaginosis*.⁵

Bacterial vaginosis is the most prevalent form of vaginal infection seen by health care professionals. Its relation to sexual activity is not clear. Sexual activity is believed to be a catalyst, rather than a primary mode of transmission, and endogenous factors play a role in the development of symptoms. The predominant symptom of bacterial vaginosis is a thin, grayishwhite discharge that has a foul, fishy odor. Burning, itching, and erythema usually are absent because the bacteria has only minimal inflammatory potential. Bacterial vaginosis may be carried asymptomatically by men and women.

The diagnosis is made when at least three of the following characteristics are present: homogeneous, white, noninflammatory discharge that smoothly coats the vaginal walls; production of a fishy, amine odor when a 10% potassium hydroxide solution is dropped onto the secretions; vaginal pH greater than 4.5 (usually 5.0 to 6.0); and the appearance of characteristic "clue cells" on wet-mount microscopic studies.³ *Clue cells* are squamous epithelial cells covered with masses of coccobacilli, often with large clumps of organisms floating free from the cell. Because *G. vaginalis* can be a normal vaginal flora, cultures should not be done routinely.

The mere presence of *G. vaginalis* in an asymptomatic woman is not an indication for treatment. When indicated, treatment is aimed at eradicating the anaerobic component of bacterial vaginosis to re-establish the normal balance of the vaginal flora. The CDC recommends oral metronidazole. Alternative therapies include metronidazole vaginal gel, clindamycin vaginal cream, or oral clindamycin. Treatment of sexual partners is not recommended.³

Bacterial vaginosis is associated with adverse pregnancy outcomes across all gestational ages. It has been linked to first and second trimester fetal losses, preterm delivery, low-birth-weight infants, and maternal/neonatal infections. Oral or cream clindamycin formulations can be used for treatment during the first trimester of pregnancy; oral or vaginal metronidazole can be used after the first trimester for treatment failures.

In summary, candidiasis, trichomoniasis, and bacterial vaginosis are common vaginal infections that become symptomatic because of changes in the vaginal ecosystem. Only trichomoniasis is spread through sexual contact. Trichomoniasis is caused by an anaerobic protozoan. The infection incites the production of a copious, frothy, yellow or green, malodorous discharge. Candidiasis, also called a yeast infection, is the form of vulvovaginitis with which women are most familiar. Candida can be present without producing symptoms; usually some host factor, such as altered immune status, contributes to the development of vulvovaginitis. It often can be treated with over-the-counter medications. Bacterial vaginosis is the most common cause of vaginal discharge. It is a nonspecific type of infection that produces a characteristic fishy-smelling discharge. The infection is thought to be caused by the combined presence of G. vaginalis and anaerobic bacteria. The anaerobe raises the vaginal pH, thereby favoring the growth of G. vaginalis.

VAGINAL-UROGENITAL-SYSTEMIC INFECTIONS

Some sexually transmitted diseases (STDs) infect male and female genital and extragenital structures. Among the infections of this type are chlamydial infections, gonorrhea, and syphilis. Many of these infections also pose a risk to infants born to infected mothers. Syphilis may be spread to the infant while in utero, whereas chlamydial and gonorrheal infections can be spread to the infant during the birth process.

Chlamydial Infections

Chlamydia trachomatis is an obligate intracellular bacterial pathogen that is closely related to gram-negative bacteria. It resembles a virus in that it requires tissue culture for isolation, but like a bacteria, it has RNA and DNA and is susceptible to some antibiotics. C. trachomatis causes a wide variety of genitourinary infections, including nongonococcal urethritis in men and pelvic inflammatory disease (PID) in women. The closely related organisms Chlamydia pneumoniae and Chlamydia psittaci cause mild and severe pneumonia, respectively. C. trachomatis can be serologically subdivided into types A, B, and C, which are associated with trachoma and chronic keratoconjunctivitis; types D through K, which are associated with genital infections and their complications; and types L1, L2, and L3, which are associated with LGV. C. trachomatis can cause significant ocular disease in neonates; it is a leading cause of blindness in underdeveloped countries. In these countries, the organism is spread primarily by flies, fomites, and nonsexual personal contact. In industrial countries, the organism is spread almost exclusively by sexual contact and therefore affects primarily the genitourinary structures.

Chlamydial infection is the most prevalent STD in the United States. Although chlamydial infections are not reportable in all states, their incidence is estimated to be more than twice that of gonorrhea. According to CDC estimates, chlamydial infections occur at a rate of 3 million new cases each year, predominantly among individuals younger than 25 years.¹ Reported rates for chlamydial infections are higher in women, largely because of the increased use of screening tests, although actual occurrence rates are thought to be the same for men and women.¹³ In the United States, costs associated with managing chlamydial infections and their complications exceed \$2 billion annually.¹³ Rates are believed to be declining because of increased efforts to screen and treat this infection.

Chlamydiae exist in two forms: (1) the elementary bodies, which are the infectious particles capable of entering uninfected cells, and (2) the initiator or reticulate bodies, which multiply by binary fission to produce the inclusions identified in stained cells. The 48-hour growth cycle starts with attachment of the elementary body to the susceptible host cell, after which it is ingested by a process that resembles phagocytosis (Fig. 35-3). Once inside the cell, the elementary body is organized into the reticulate body, the metabolically active form of the organism that is capable of reproduction. The reticulate body is not infectious and cannot survive outside the body. The reticulate bodies divide in the cell for as long as 36 hours and then condense to form new elementary bodies, which are released when the infected cell bursts.



In women, chlamydial infections may cause urinary frequency, dysuria, and vaginal discharge. The most common symptom is a mucopurulent cervical discharge. The cervix itself frequently hypertrophies and becomes erythematous, edematous, and extremely friable. Seventy-five percent of women with chlamydial infection have no symptoms; therefore, most cases are undiagnosed, unreported, and untreated.¹³ This can lead to greater fallopian tube damage and increase the reservoir for further chlamydial infections. Approximately 40% of women with an untreated chlamydial infection develop PID, and 1 in 5 of these women becomes infertile. Research has identified a possible link between three specific serotypes of chlamydiae and an increased risk for cervical cancer. The mechanism by which this occurs is unclear.¹⁴

In men, chlamydial infections cause urethritis, including meatal erythema and tenderness, urethral discharge, dysuria, and urethral itching. Prostatitis and epididymitis with subsequent infertility may develop. However, approximately 50% of men are asymptomatic. The most serious complication that can develop with nongonococcal urethritis is Reiter's syndrome, a systemic condition characterized by urethritis, conjunctivitis, arthritis, and mucocutaneous lesions (see Chapter 43).

Routine screening for adolescents and young adults has been suggested by the CDC in an effort to minimize these serious sequelae of asymptomatic infection.³ Between 25% and 50% of infants born to mothers with cervical chlamydial infections develop ocular disease (*i.e.*, inclusion conjunctivitis), and 10% to 20% having chlamydial pneumonitis.

Diagnosis of chlamydial infections takes several forms. The identification of polymorphonuclear leukocytes on Gram's stain of male discharge or cervical discharge is presumptive evidence. The direct fluorescent antibody test and the enzymelinked immunosorbent assay that use antibodies against an antigen in the *Chlamydia* cell wall are rapid tests that are highly sensitive and specific. The positive predictive value of these tests is excellent among high-risk groups, but false-positive results occur more often in populations with lower risks. Amplified DNA probe assays, such as the PCR, ligase chain reaction (LCR), or transcription-mediated amplification (TMA), have demonstrated specificities of near 100%.

The CDC recommends the use of azithromycin or doxycycline as the treatment of choice for chlamydial infection; penicillin is ineffective. Erythromycin or amoxicillin is the preferred choice in pregnancy.³ Antibiotic treatment of both sexual partners simultaneously is recommended. Abstinence from sexual activity is encouraged to facilitate cure.

Gonorrhea

Gonorrhea is a reportable disease caused by the bacterium Neisseria gonorrhoeae. In 1999, there were 360,076 reported cases of gonorrhea in the United States.¹ Of these reported cases, more than 90% involved persons between 15 and 44 years of age, with the heaviest concentration among young adults (15 to 24 years of age).¹³ There are an estimated 600,000 new cases every year.³ Although the incidence of gonorrhea has declined steadily from its peak in 1975, there was an increase in occurrence between 1997 and 1998. Improved screening efforts as well as greater use of more sensitive nonculture methods of testing may have contributed to this increase. Higher rates of occurrence among homosexual men were documented in several states, leading to a concern that an increase in unsafe sexual behavior may be occurring because of the availability of highly active antiretroviral agents for treatment of HIV infection.15

The gonococcus is a pyogenic (*i.e.*, pus-forming), gramnegative diplococcus that evokes inflammatory reactions characterized by purulent exudates. Humans are the only natural host for *N. gonorrhoeae*. The organism grows best in warm, mucus-secreting epithelia. The portal of entry can be the genitourinary tract, eyes, oropharynx, anorectum, or skin.

Transmission usually is by heterosexual or homosexual intercourse. Autoinoculation of the organism to the conjunctiva is possible. Neonates born to infected mothers can acquire the infection during passage through the birth canal and are in danger of experiencing gonorrheal conjunctivitis, with resultant blindness, unless treated promptly. An amniotic infection syndrome characterized by premature rupture of the membranes, premature delivery, and increased risk of infant morbidity and mortality has been identified as an additional complication of gonococcal infections in pregnancy. Genital gonorrhea in young children should raise the possibility of sexual abuse.

The infection commonly manifests 2 to 7 days after exposure. It typically begins in the anterior urethra, accessory urethral glands, Bartholin's or Skene's glands, and the cervix. If untreated, gonorrhea spreads from its initial sites upward into the genital tract. In males, it spreads to the prostate and epididymis; in females, it commonly moves to the fallopian tubes. Pharyngitis may follow oral-genital contact. The organism also can invade the bloodstream (*i.e.*, disseminated gonococcal infection), causing serious sequelae such as bacteremic involvement of joint spaces, heart valves, meninges, and other body organs and tissues.

Persons with gonorrhea may be asymptomatic and may unwittingly spread the disease to their sexual partners. Men are more likely to be symptomatic than women. In men, the initial symptoms include urethral pain and a creamy, yellow, sometimes bloody discharge. The disorder may become chronic and affect the prostate, epididymis, and periurethral glands. Rectal infections are common in homosexual men. In women, recognizable symptoms include unusual genital or urinary discharge, dysuria, dyspareunia, pelvic pain or tenderness, unusual vaginal bleeding (including bleeding after intercourse), fever, and proctitis. Symptoms may occur or increase during or immediately after menses because the bacterium is an intracellular diplococcus that thrives in menstrual blood but cannot survive long outside the human body. There may be infections of the uterus and development of acute or chronic infection of the fallopian tubes (*i.e.*, salpingitis), with ultimate scarring and sterility.

Diagnosis is based on the history of sexual exposure and symptoms. It is confirmed by identification of the organism on Gram's stain or culture. An enzyme immunoassay for detecting gonococcal antigens (Gonozyme) is available but has several requirements that limit its usefulness. Detection by means of amplified DNA probes (PCR, LCR, TMA) is possible using urine and urethral swab specimens. The sensitivity of these probes is similar to that of culture, and they may be cost effective in high-risk populations.

Testing for other STDs, particularly syphilis and chlamydial infections, is suggested at the time of examination. Pregnant women are routinely screened at the time of their first prenatal visit; high-risk populations should have repeat cultures during the third trimester. Neonates are routinely treated with various antibacterial agents applied to the conjunctiva within 1 hour of birth to protect against undiagnosed gonorrhea and other diseases.

Penicillin-resistant strains of *N. gonorrhoeae* are prevalent worldwide, and strains with other kinds of antibiotic resistance continue to evolve and spread. The current treatment recommendation to combat tetracycline- and penicillin-resistant strains of *N. gonorrhoeae* is ceftriaxone in a single injection or cefixime, ciprofloxacin, levofloxacin, or ofloxacin in a single oral dose. All are equally effective and should be followed with azithromycin or doxycycline for chlamydiae. All sex partners 60 days prior to discovery of the infection should be con-

tacted, tested, and treated. Test of cure is not required with observed single-dose therapy. Patients are instructed to refrain from intercourse until therapy is completed and symptoms are no longer present.³

Syphilis

Syphilis is a reportable disease caused by a spirochete, Treponema pallidum. During 1999, 6657 new cases of primary and secondary syphilis were reported in the United States (2.5 per 100,000 population). This is the lowest rate ever reported, and it now appears that syphilis transmission is primarily concentrated in a few geographic areas.¹⁶ This represents a steady decline from the 50,000 cases reached in 1990 after an epidemic resurgence of this problem between 1985 and 1990. Syphilis continues to disproportionately affect minority populations. Although the 1999 rate for blacks declined by 10% from that of the previous year, it was still 30 times the rate reported in whites. The rate for Hispanics increased 20%, primarily among men. The National Plan to Eliminate Syphilis was launched in 1999 with a goal of reducing primary and secondary syphilis to fewer than 1000 cases and increasing the number of syphilisfree counties to 90% by 2005. Federal funding is available to support this effort.16

T. pallidum is spread by direct contact with an infectious, moist lesion, usually through sexual intercourse. Bacteria-laden secretions may transfer the organism during kissing or intimate contact. Skin abrasions provide another possible portal of entry. There is rapid transplacental transmission of the organism from the mother to the fetus after 16 weeks' gestation, so that active disease in the mother during pregnancy can produce congenital syphilis in the fetus. Untreated syphilis can cause prematurity, stillbirth, and congenital defects and active infection in the infant. Once treated for syphilis, a pregnant woman usually is followed up throughout pregnancy by repeat testing of serum titers.

The clinical disease is divided into three stages: primary, secondary, and tertiary. Primary syphilis is characterized by the appearance of a chancre at the site of exposure. Chancres typically appear within 3 weeks of exposure but may incubate for 1 week to 3 months. The primary chancre begins as a single, indurated, button-like papule up to several centimeters in diameter that erodes to create a clean-based ulcerated lesion on an elevated base. These lesions usually are painless and located at the site of sexual contact. Primary syphilis is readily apparent in the male, where the lesion is on the penis or scrotum. Although chancres can develop on the external genitalia in females, they are more common on the vagina or cervix, so primary syphilis may go untreated. There usually is an accompanying regional lymphadenopathy. The disease is highly contagious at this stage, but because the symptoms are mild, it frequently goes unnoticed. The chancre usually heals within 3 to 12 weeks, with or without treatment.

The timing of the second stage of syphilis varies even more than that of the first, lasting from 1 week to 6 months. The symptoms of a rash (especially on the palms and soles), fever, sore throat, stomatitis, nausea, loss of appetite, and inflamed eyes may come and go for a year but usually last for 3 to 6 months. Secondary manifestations may include alopecia and genital condylomata lata. Condylomata lata are elevated, redbrown lesions that may ulcerate and produce a foul discharge.

After the second stage, syphilis frequently enters a latent phase that may last the lifetime of the person or progress to tertiary syphilis at some point. Persons can be infective during the first 1 to 2 years of latency.

Tertiary syphilis is a delayed response of the untreated disease. It can occur as long as 20 years after the initial infection. Only approximately one third of those with untreated syphilis progress to the tertiary stage of the disease, and symptoms develop in approximately one half of these. Approximately one third undergo spontaneous cure, and the remaining one third continue to have positive serologic tests but do not have structural lesions.¹⁷ When syphilis does progress to the symptomatic tertiary stage, it commonly takes one of three forms: (1) development of localized destructive lesions called gummas, (2) development of cardiovascular lesions, or (3) development of central nervous system lesions. The syphilitic gumma is a peculiar, rubbery, necrotic lesion that is caused by noninflammatory tissue necrosis. Gummas can occur singly or multiply and vary in size from microscopic lesions to large, tumorous masses. They most commonly are found in the liver, testes, and bone. Central nervous system lesions can produce dementia, blindness, or injury to the spinal cord, with ataxia and sensory loss (*i.e.*, tabes dorsalis). Cardiovascular manifestations usually result from scarring of the medial layer of the thoracic aorta with aneurysm formation. These aneurysms produce enlargement of the aortic valve ring with aortic valve insufficiency.

T. pallidum does not produce endotoxins or exotoxins but evokes a humoral immune response that provides the basis for serologic tests. Two types of antibodies-nonspecific and specific—are produced. The nonspecific antibodies can be detected by flocculation tests such as the Venereal Disease Research Laboratory (VDRL) test or the rapid plasma reagin (RPR) test. Because these tests are nonspecific, positive results can occur with diseases other than syphilis. The tests are easy to perform, rapid, and inexpensive and frequently are used as screening tests for syphilis. Results become positive 4 to 6 weeks after infection or 1 to 3 weeks after the appearance of the primary lesion. Because these tests are quantitative, they can be used to measure the degree of disease activity or treatment effectiveness. The VDRL titer usually is high during the secondary stage of the disease and becomes less so during the tertiary stage. A falling titer during treatment suggests a favorable response. The fluorescent treponemal antibody absorption test or microhemagglutinin test is used to detect specific antibodies to T. pallidum. These qualitative tests are used to determine whether a positive result on a nonspecific test such as the VDRL is attributable to syphilis. The test results remain positive for life.

T. pallidum cannot be cultured. The diagnosis of syphilis is based on serologic tests or dark-field microscopic examination with identification of the spirochete in specimens collected from lesions. Because the disease's incubation period may delay test sensitivity, serologic tests usually are repeated after 6 weeks if the initial test results were negative.

The treatment of choice for syphilis is penicillin.³ Because of the spirochetes' long generation time, effective tissue levels of penicillin must be maintained for several weeks. Long-acting injectable forms of penicillin are used. Tetracycline or doxycycline is used for treatment in persons who are sensitive to penicillin. Pregnant patients should be desensitized and treated

with penicillin because erythromycin does not treat fetal infection. Sexual partners should be evaluated and treated prophylactically even though they may show no sign of infection. All treated individuals should be re-examined clinically and serologically at 6 and 12 months after completing therapy; more frequent monitoring (3-month intervals) is suggested for individuals with HIV infection.3

In summary, the vaginal-urogenital-systemic STDschlamydial infections, gonorrhea, and syphilis-can severely involve the genital structures and manifest as systemic infections. Gonorrheal and chlamydial infections can cause a wide variety of genitourinary complications in men and women, and both can cause ocular disease and blindness in neonates born to infected mothers. Syphilis is caused by a spirochete, T. pallidum. It can produce widespread systemic effects and is transferred to the fetus of infected mothers through the placenta.

REVIEW QUESTIONS

- Define what is meant by a sexually transmitted disease (STD) and cite the common portals of entry for STDs.
- State the significance of infection with the human papillomavirus.
- Explain the recurrent infections in genital herpes.
- Compare the signs and symptoms of infections caused by Candida albicans, Trichomonas vaginalis, and bacterial vaginosis.
- Compare the signs and symptoms of gonorrhea in the male and female.
- State the genital and nongenital complications that can occur with chlamydial infections, gonorrhea, and syphilis.

connection-

Visit the Connection site at connection.lww.com/go/porth for links to chapter-related resources on the Internet.

REFERENCES

- 1. Cates W. Jr. (1999). Estimates of the incidence and prevalence of STDs in the US. Sexually Transmitted Diseases 4 (Suppl.), S2-S7.
- 2. Beutner K.R, Reitano M.V., Richwald G.A, et al. (1998). External genital warts: Report of the American Medical Association Consensus Conference. Clinical Infectious Diseases 27, 796-806.
- 3. Centers for Disease Control and Prevention. (2002). Sexually transmitted diseases: Treatment guidelines 2002. Morbidity and Mortality Weekly Report 51 (RR-6), 1-118.
- 4. Koutsky L.A., Ault K.A., Wheeler C.M. (2002). A controlled trial of human papillomavirus type 16 vaccine. New England Journal of Medicine 347, 1645-1651.
- 5. Holmes K.K, Per-Anders M., Sparling P.F., et al. (1999). Sexually transmitted disease (3rd ed., pp. 287, 290, 347, 424, 563-564, 629, 820-821). New York: McGraw-Hill.
- 6. Handsfield H.H. (2001). Color atlas and synopsis of sexually transmitted diseases (pp. 13, 23, 71, 87, 163). New York: McGraw-Hill.

636 Unit Nine: Alterations in the Male and Female Reproductive Systems

- Cotran R.S., Kumar V., Collins T. (1999). Robbins pathologic basis of disease (6th ed., pp. 359–361). Philadelphia: W.B. Saunders.
- Rubin E., Farber J.L. (1999). Pathology (3rd ed.). Philadelphia: Lippincott Williams & Wilkins.
- 9. Scott J.R., DiSaia P.J., Hammond C.B., et al. (1999). *Danforth's obstetrics and gynecology* (8th ed., pp. 402–403). Philadelphia: Lippincott Williams & Wilkins.
- Kohl S. (2000). Herpes simplex virus. In Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.), *Nelson textbook of pediatrics* (16th ed., pp. 969–972). Philadelphia: W.B. Saunders.
- 11. Ringdahl E.N. (2000). Treatment of recurrent vulvovaginal candidiasis. American Family Physician 61, 3306–3317.
- 12. Sobel J.D. (2000). Bacterial vaginosis. Annual Review of Medicine 51, 349–356.

- Centers for Disease Control and Prevention. (2000). Tracking the hidden epidemics: Trends in STDs in the United States. [On-line]. Available: http://www.cdc.gov/nchstp/dstd/dstdp.html.
- Antilla T., Saiku P., Koskela P., Bloigu A., et al. (2001). Serotypes of *Chlamydia trachomatis* and risk for development of cervical squamous cell carcinoma. *Journal of the American Medical Association* 285, 47–51.
- Centers for Disease Control and Prevention. (2000). Gonorrhea— United States, 1998. Morbidity and Mortality Weekly Report 49, 538.
- Centers for Disease Control and Prevention. (2001). Primary and secondary syphilis—United States, 1999. Morbidity and Mortality Weekly Report 50, 113–117.
- 17. Chapin K. (1999). Probing the STDs. American Journal of Nursing 99 (7), 24AAA-24DDD.

Ten

CHAPTER

Alterations in the Nervous System

Organization and Control of the Nervous System

Nervous Tissue Cells **Neurons** Supporting Cells Supporting Cells of the Peripheral Nervous System Supporting Cells of the Central Nervous System Metabolic Requirements of Nervous Tissue Nerve Cell Communication **Action Potentials** Synaptic Transmission Messenger Molecules Development and Organization of the **Nervous System Embryonic Development** Segmental Organization **Cell Columns Longitudinal Tracts** The Spinal Cord and Brain The Spinal Cord Spinal Nerves The Brain Hindbrain Midbrain Forebrain

Meninges Ventricular System and Cerebrospinal Fluid Blood-Brain and Cerebrospinal Fluid–Brain Barriers **The Autonomic Nervous System** *Autonomic Efferent Pathways* Sympathetic Nervous System Parasympathetic Nervous System *Central Integrative Pathways Autonomic Neurotransmission* Acetylcholine and Cholinergic Receptors Catecholamines and Adrenergic Receptors

he nervous system, in coordination with the endocrine system, provides the means by which cell and tissue functions are integrated into a solitary, surviving organism. It controls skeletal muscle movement and helps to regulate cardiac and visceral smooth muscle activity. The nervous system enables the reception, integration, and perception of sensory information; it provides the substratum necessary for intelligence, anticipation, and judgment; and it facilitates adjustment to an ever-changing external environment.

NERVOUS TISSUE CELLS

The nervous system can be divided into two parts: the central nervous system (CNS) and the peripheral nervous system (PNS). The CNS consists of the brain and spinal cord, which is

protected by the skull and vertebral column. The PNS is found outside these structures. Inherent in the basic design of the nervous system is the provision for the concentration of computational and control functions in the CNS. In this design, the PNS functions as an input-output system for relaying input to the CNS and for transmitting output messages that control effector organs, such as muscles and glands.

Nervous tissue contains two types of cells: neurons and supporting cells. The neurons are the functional cells of the nervous system. They exhibit membrane excitability and conductivity and secrete neurotransmitters and hormones, such as epinephrine and antidiuretic hormone. The supporting cells, such as Schwann cells in the PNS and the glial cells in the CNS, protect the nervous system and provide metabolic support for the neurons.

Neurons

The functioning cells of the nervous system are called *neurons*. Neurons have three distinct parts: the soma or cell body and its cytoplasm-filled processes, the dendrites and axons (Fig. 36-1). These processes form the functional connections, or synapses, with other nerve cells, with receptor cells, or with effector cells. Axonal processes are particularly designed for rapid communication with other neurons and the many body structures innervated by the nervous system. Afferent, or sensory, neurons transmit information from the PNS to the CNS. Efferent neurons, or motoneurons, carry information away from the CNS (Fig. 36-1). Interspersed between the afferent and efferent neurons is a network of interconnecting neurons (interneurons or internuncial neurons) that modulate and control the body's response to changes in the internal and external environments.

The cell body of a neuron contains a large, vesicular nucleus with one or more distinct nucleoli and a well-developed rough endoplasmic reticulum. A neuron's nucleus has the same DNA

KEY CONCEPTS

THE STRUCTURAL ORGANIZATION OF THE NERVOUS SYSTEM

- The nervous system is divided into two parts: the CNS, which consists of the brain and spinal cord, and the PNS, which contains the input and output neurons that lie outside the CNS.
- The neurons, which are functioning cells of the nervous system, consist of a cell body with cytoplasmfilled processes, the dendrites, and the axons, which form the functional connections with other nerve or effector cells.
- There are two types of neurons: afferent neurons or sensory neurons, which carry information to the CNS, and efferent neurons or motoneurons, which carry information from the CNS to the effector organs.

and genetic code content that is present in other cells of the body, and its nucleolus, which is composed of portions of several chromosomes, produces RNA associated with protein synthesis. The cytoplasm contains large masses of ribosomes that are prominent in most neurons. These acidic RNA masses, which are involved in protein synthesis, stain as dark Nissl bodies with basic histologic stains (see Fig. 36-1).

The dendrites (*i.e.*, "treelike") are multiple, branched extensions of the nerve cell body; they conduct information toward the cell body and are the main source of information for the neuron. The dendrites and cell body are studded with synaptic terminals that communicate with axons and dendrites of other neurons.

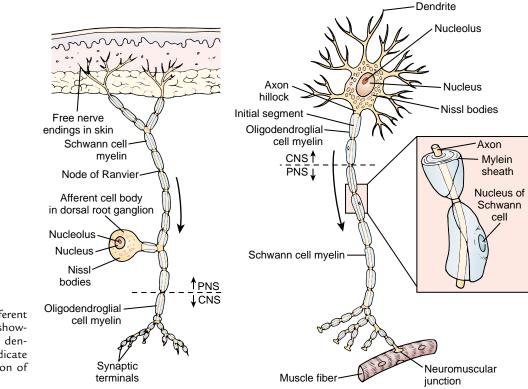
Axons are long efferent processes that project from the cell body and carry impulses away from the cell. Most neurons have only one axon; however, axons may exhibit multiple branching that results in many axonal terminals. The cytoplasm of the cell body extends to fill the dendrites and the axon. The proteins and other materials used by the axon are synthesized in the cell body and then flow down the axon through its cytoplasm.

The cell body of the neuron is equipped for a high level of metabolic activity. This is necessary because the cell body must synthesize the cytoplasmic and membrane constituents required to maintain the function of the axon and its terminals. Some of these axons extend for a distance of 1 to 1.5 m and have a volume that is 200 to 500 times greater than the cell body itself. Two axonal transport systems, one slow and one rapid, move molecules from the cell body through the cytoplasm of the axon to its terminals. Replacement proteins and nutrients slowly diffuse from the cell body, where they are synthesized, down the axon, moving at the rate of approximately 1 mm/day. Other molecules, such as some neurosecretory granules (e.g., neurotransmitters, neuromodulators, and neurohormones) or their precursors, are conveyed by a rapid, energy-dependent active transport system, moving at the rate of approximately 400 mm/day. For example, antidiuretic hormone and oxytocin, which is synthesized by neurons in the hypothalamus, is carried by rapid axonal transport to the posterior pituitary, where the hormones are released into the blood. A reverse rapid (*i.e.*, retrograde) axonal transport system moves materials, including target cell messenger molecules, from axonal terminals back to the cell body.

Supporting Cells

Supporting cells of the nervous system, the Schwann and satellite cells of the PNS and the several types of glial cells of the CNS, give the neurons protection and metabolic support. The supporting cells segregate the neurons into isolated metabolic compartments, which are required for normal neural function. Astrocytes, together with the tightly joined endothelial cells of the capillaries in the CNS, contribute to what is called the *blood-brain barrier*. This term is used to emphasize the impermeability of the nervous system to large or potentially harmful molecules.

The many-layered myelin wrappings of Schwann cells of the PNS and the oligodendroglia of the CNS produce the myelin sheaths that serve to increase the velocity of nerve impulse conduction in axons. Myelin has a high lipid content, which gives



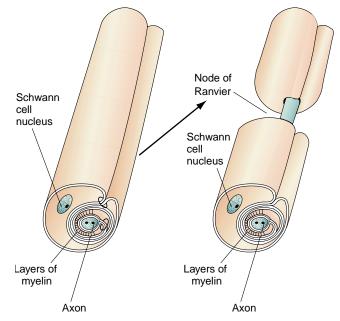
■ FIGURE 36-1 ■ (A) Afferent and (B) efferent neurons, showing the soma or cell body, dendrites, and axon. Arrows indicate the direction for conduction of action potential.

it a whitish color, and the name white matter is given to the masses of myelinated fibers of the spinal cord and brain. Besides its role in increasing conduction velocity, the myelin sheath is essential for the survival of larger neuronal processes, perhaps by the secretion of neurotrophic compounds. In some pathologic conditions, such as multiple sclerosis in the CNS and Guillain-Barré syndrome in the PNS, the myelin may degenerate or be destroyed, leaving a section of the axonal process without myelin while leaving the nearby Schwann or oligodendroglial cells intact. Unless remyelination takes place, the axon eventually dies.

Supporting Cells of the Peripheral Nervous System

Schwann cells and satellite cells are the two types of supporting cells in the PNS. Normally, the nerve cell bodies in the PNS are collected into ganglia, such as the dorsal root and autonomic ganglia. The cell bodies and processes of the peripheral nerves are separated from the connective tissue framework of the ganglion by a single layer of flattened capsular cells called *satellite cells*. Satellite cells secrete a basement membrane that protects the cell body from the diffusion of large molecules.

The processes of larger afferent and efferent neurons are surrounded by the cell membrane and cytoplasm of Schwann cells, which are close relatives of the satellite cells. During myelination, the Schwann cell wraps around the nerve process many times in a "jelly roll" fashion (Fig. 36-2). Schwann cells line up along the neuronal process, and each of these cells forms its own discrete myelin segment. The end of each myelin segment attaches to the cell membrane of the axon by means



■ FIGURE 36-2 ■ The Schwann cell forms a myelin sheath around the larger nerve fibers in the peripheral nervous system. During myelination, the Schwann cell wraps around the nerve process many times in a "jelly roll" fashion. The end of each myelin segment attaches to the cell membrane of the axon by means of intercellular junctions. Successive Schwann cells are separated by short extracellular fluid gaps called the *nodes of Ranvier*, where the myelin is missing and the voltage-gated sodium channels are concentrated.

640 Unit Ten: Alterations in the Nervous System

of intercellular junctions. Successive Schwann cells are separated by short extracellular fluid gaps called the *nodes of Ranvier*, where the myelin is missing and voltage-gated sodium channels are concentrated. The nodes of Ranvier increase nerve conduction by allowing the impulse to jump from node to node through the extracellular fluid in a process called *saltatory conduction* (from the Latin *saltare*, "to jump"). In this way, the impulse can travel more rapidly than it could if it were required to move systematically along the entire nerve process. This increased conduction velocity greatly reduces reaction time, or time between the application of a stimulus and the subsequent motor response. The short reaction time is of particular importance in peripheral nerves with long distances (sometimes 1 to 1.5 m) for conduction between the CNS and distal effector organs.

Each of the Schwann cells along a peripheral nerve is encased in a continuous tube of basement membrane, which in turn is surrounded by a multilayered sheath of loose connective tissue known as the *endoneurium* (Fig. 36-3). The endoneurial sheath, which is essential to the regeneration of injured peripheral nerves, provides a collagenous tube through which a regenerating axon can again reach its former target. The endoneurial sheath does not penetrate the CNS. The absence of the endoneurial sheath is thought to be a major factor in the limited axonal regeneration of CNS nerves compared with those of the PNS.

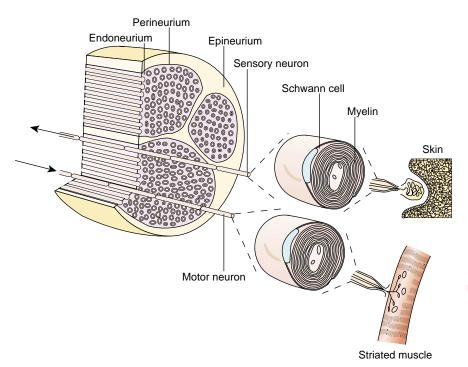
The endoneurial sheaths are bundled with blood vessels into small bundles or clusters of nerves called *fascicles*. In the nerve, the fascicles consisting of bundles of nerve fibers are surrounded by another protective covering called the *perineurium*. Usually, several fascicles are further surrounded by the heavy, protective *epineurial sheath* of the peripheral nerve. The protective layers that surround the peripheral nerve processes are continuous with the connective tissue capsule of the sensory nerve endings and the connective tissue that surrounds the effector structures, such as the skeletal muscle cell. Centrally, the connective tissue layers continue along the dorsal and ventral roots of the nerve and fuse with the meninges that surround the spinal cord and brain.

Supporting Cells of the Central Nervous System

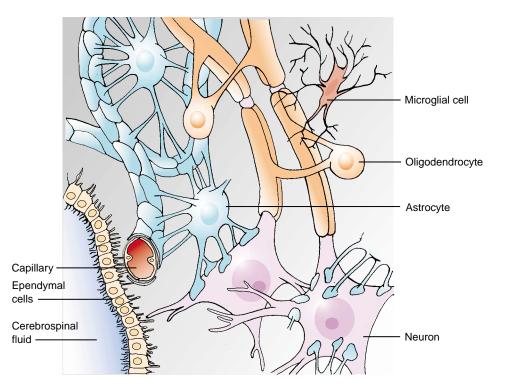
Supporting cells of the CNS consist of the oligodendroglia, astroglia, microglia, and ependymal cells (Fig. 36-4). The *oligodendroglial* cells form the myelin in the CNS. Instead of forming a myelin covering for a single axon, these cells reach out with several processes, each wrapping around and forming a multilayered myelin segment around several different axons. The coverings of axons in the CNS function in increasing the velocity of nerve conduction, similar to the peripheral myelinated fibers.

A second type of glial cell, the *astroglia*, is particularly prominent in the gray matter of the CNS. These large cells have many processes, some reaching to the surface of the capillaries, others reaching to the surface of the nerve cells, and still others filling most of the intercellular space of the CNS. The astrocytic linkage between the blood vessels and the neurons may provide a transport mechanism for the exchange of oxygen, carbon dioxide, and metabolites. The astrocytes also have an important role in sequestering cations such as calcium and potassium from the intercellular fluid. Astrocytes can fill their cytoplasm with microfibrils (*i.e.*, fibrous astrocytes), and masses of these cells form the special type of scar tissue called *gliosis* that develops in the CNS when tissue is destroyed.

A third type of glial cell, the *microglia*, is a small phagocytic cell that is available for cleaning up debris after cellular damage, infection, or cell death. The fourth type of cell, the *ependymal* cell, forms the lining of the neural tube cavity, the ventricular system. In some areas, these cells combine with a rich vascular network to form the *choroid plexus*, where production of the cerebrospinal fluid (CSF) takes place.



■ FIGURE 36-3 ■ Section of a peripheral nerve containing axons of both afferent (sensory) and efferent (motor) neurons. (Modified from Cormack D.H. [1987]. *Ham's histology* [9th ed., p. 374]. Philadelphia: J.B. Lippincott)



■ FIGURE 36-4 ■ The cells of the central nervous system (CNS). Diagrammatic view of relationships between the glial elements (astrocyte, oligodendrocyte, microglial cell, and the ependymal cells), the capillaries, the cerebral spinal fluid, and the cell bodies of CNS neurons.

Metabolic Requirements of Nervous Tissue

Nervous tissue has a high rate of metabolism. Although the brain comprises only 2% of the body's weight, it receives approximately 15% of the resting cardiac output and consumes 20% of its oxygen. Despite its substantial energy requirements, the brain can neither store oxygen nor engage in anaerobic metabolism. An interruption in the blood or oxygen supply to the brain rapidly leads to clinically observable signs and symptoms. Without oxygen, brain cells continue to function for approximately 10 seconds. Unconsciousness occurs almost simultaneously with cardiac arrest, and the death of brain cells begins within 4 to 6 minutes. Interruption of blood flow also leads to the accumulation of metabolic by-products that are toxic to neural tissue.

Glucose is the major fuel source for the nervous system, but neurons have no provision for storing glucose. Ketones can provide for limited temporary energy requirements; however, these sources are rapidly depleted. Unlike muscle cells, neurons have no glycogen stores and must rely on glucose from the blood or the glycogen stores of supporting glial cells. Persons receiving insulin for diabetes may experience signs of neural dysfunction and unconsciousness (*i.e.*, insulin reaction or shock) when blood glucose drops because of insulin excess (see Chapter 32).

In summary, nervous tissue is composed of two types of cells: neurons and supporting cells. Neurons are composed of three parts: a cell body, which controls cell activity; the dendrites, which conduct information toward the cell body; and the axon, which carries impulses from the cell body. The supporting cells consist of Schwann and satellite cells of the PNS and the glial cells of the CNS. Supporting cells protect and

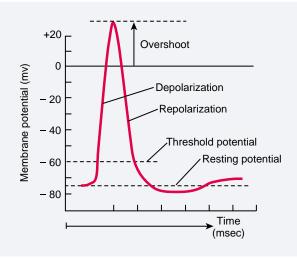
provide metabolic support for the neurons and aid in segregating them into isolated compartments, which is necessary for normal neuronal function. The Schwann cells of the PNS and the oligodendroglial cells of the CNS form the myelin sheath that allows for rapid conduction of impulses. The nervous system has a high level of metabolic activity, requiring a continuous supply of oxygen and glucose. Although the brain comprises only 2% of the body's weight, it receives approximately 15% of the resting cardiac output and consumes 20% of its oxygen.

NERVE CELL COMMUNICATION

Neurons are characterized by the ability to communicate with other neurons and body cells through pulsed electrical signals called *impulses*. An impulse, or action potential (discussed in Chapter 1), represents the movement of electrical charge along the axon membrane. This phenomenon, sometimes called *conductance*, is based on the rapid flow of charged ions through the plasma membrane. In excitable tissue, ions such as sodium, potassium, and calcium move through the membrane channels and carry the electrical charges involved in the initiation and transmission of such impulses.

Action Potentials

Nerve signals are transmitted by action potentials, which are abrupt, pulsatile changes in the membrane potential that last a few ten thousandths to a few thousandths of a second. Action potentials can be divided into three phases: the resting or polarized state, depolarization, and repolarization (Fig. 36-5).



■ **FIGURE 36-5** ■ Time course of the action potential recorded at one point of an axon with one electrode inside and one on the outside of the plasma membrane.

The *resting membrane potential* represents the undisturbed period of the action potential during which the nerve is not transmitting impulses. During this time the inside of the membrane is negatively charged with respect to the outside, and the membrane is said to be *polarized*. The resting phase of the membrane potential continues until some event causes the membrane to increase its permeability to sodium. The *threshold potential* represents the membrane potential at which neurons or other excitable tissues are stimulated to fire. When the threshold potential is reached, the gatelike structures in the ion channels open. The gates are either fully open or fully closed (all-or-none). Under ordinary circumstances, the threshold stimulus is sufficient to open large numbers of ion channels, triggering massive depolarization of the membrane (the action potential).

Depolarization is characterized by the flow of electrically charged ions and the reversal of the membrane potential. During the depolarization phase, the membrane suddenly becomes permeable to sodium ions. The rapid inflow of sodium ions produces local currents that travel through the adjacent cell membrane, causing the sodium channels in this part of the membrane to open. Thus, the impulse moves longitudinally along the nerve, moving from one part of the axon to another. Repolarization is the phase during which the polarity of the resting membrane potential is re-established. This is accomplished with closure of the sodium channels and opening of the potassium channels. The outflow of positively charged potassium ions across the cell membrane returns the membrane potential to negativity. The sodium-potassium pump gradually re-establishes the resting ionic concentrations on each side of the membrane.

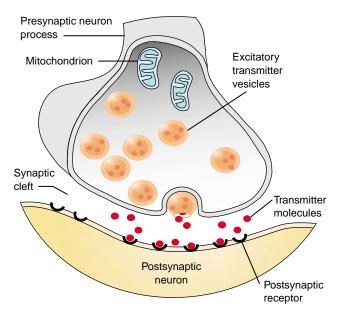
The excitability of neurons can be affected by conditions that alter the resting membrane potential, moving it either closer to or further from the threshold potential. *Hypopolarization* increases the excitability of the postsynaptic neuron by bringing the membrane potential closer to the threshold potential so that a smaller subsequent stimulus is needed to cause the neuron to fire. *Hyperpolarization* brings the membrane potential further from threshold and has the opposite effect. It has an inhibitory effect and decreases the likelihood that an action potential will be generated.

Synaptic Transmission

Neurons communicate with each other through structures known as synapses. Two types of synapses are found in the nervous system: electrical and chemical. *Electrical synapses* permit the passage of current-carrying ions through small openings called *gap junctions* that penetrate the cell junction of adjoining cells and allow current to travel in either direction. The gap junctions allow an action potential to pass directly and quickly from one neuron to another. They may link neurons having close functional relationships into circuits.

The most common type of synapse is the *chemical synapse*. Chemical synapses involve special presynaptic and postsynaptic membrane structures, separated by a synaptic cleft (Fig. 36-6). The presynaptic terminal secretes one and often several chemical transmitter molecules (i.e., neurotransmitters or neuromodulators) into the synaptic cleft. The neurotransmitters diffuse into the synaptic cleft and unite with receptors on the postsynaptic membrane. In contrast to an electrical synapse, a chemical synapse serves as a rectifier, permitting only one-way communication. One-way conduction is a particularly important characteristic of chemical synapses. Chemical synapses are divided into two types: excitatory and inhibitory. In excitatory synapses, binding of the neurotransmitter to the receptor produces depolarization of the postsynaptic membrane. Binding of the neurotransmitter to the receptor in an inhibitory synapse reduces the postsynaptic neuron's ability to generate an action potential. Most inhibitory neurotransmitters induce hyperpolarization of the postsynaptic membrane by making the membrane more permeable to potassium or chloride, or both (see Chapter 6).

Chemical synapses are the slowest component in progressive communication through a sequence of neurons, such as in



■ FIGURE 36-6 ■ A chemical synapse showing the synaptic vesicles in the presynaptic neuron, release of the transmitter, and binding of the transmitter to the receptors on the membrane surface of the postsynaptic neuron.

a spinal reflex. In contrast to the conduction of electrical action potentials, each successive event at the chemical synapse transmitter secretion, diffusion across the synaptic cleft, interaction with postsynaptic receptors, and generation of a subsequent action potential in the postsynaptic neuron—consumes time. On average, conduction across a chemical synapse requires approximately 0.3 milliseconds.

A neuron's cell body and dendrites are covered by thousands of synapses, any or many of which can be active at any moment. Because of the interaction of this rich synaptic input, each neuron resembles a little integrator, in which circuits of many neurons interact with one another. It is the complexity of these interactions and the subtle integrations involved in producing behavorial responses that gives the system its intelligence.

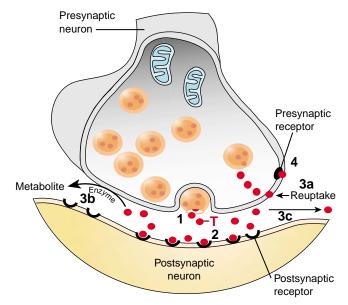
Messenger Molecules

Neurotransmitters are the chemical messenger molecules of the nervous system. The messenger molecules of the nervous system include the neurotransmitters, neurohumoral mediators such as epinephrine, neuromodulators, and neurotrophic or nerve growth factors.

Neurotransmitters are small molecules that incorporate a positively charged nitrogen atom; they include several amino acids, peptides, and monoamines. Amino acids are the building blocks of proteins and are present in body fluids. Peptides are low-molecular-weight molecules that are made up of two or more amino acids. They include substance P and the endorphins and enkephalins, which are involved in pain sensation and perception (see Chapter 39). A monoamine is an amine molecule containing one amino group (NH₂). Serotonin, dopamine, norepinephrine, and epinephrine are monoamines synthesized from amino acids. Fortunately, the blood-brain barrier protects the nervous system from circulating amino acids and other molecules with potential neurotransmitter activity.

The process of neurotransmission involves the synthesis, storage, and release of a neurotransmitter; the reaction of the neurotransmitter with a receptor; and termination of the receptor action (Fig. 36-7). Neurotransmitters are synthesized in the cytoplasm of the axon terminal. The synthesis of transmitters may require one or more enzyme-catalyzed steps (e.g., one for acetylcholine and three for norepinephrine). Neurons are limited as to the type of transmitter they can synthesize by their enzyme systems. After synthesis, the neurotransmitter molecules are stored in the axon terminal in tiny, membrane-bound sacs called synaptic vesicles. There may be thousands of vesicles in a single terminal, each containing 10,000 to 100,000 transmitter molecules. The vesicle protects the neurotransmitters from enzyme destruction in the nerve terminal. The arrival of an impulse at a nerve terminal causes the vesicles to move to the cell membrane and release their transmitter molecules into the synaptic space.

Neurotransmitters exert their actions through specific proteins, called *receptors*, embedded in the postsynaptic membrane. These receptors are tailored precisely to match the size and shape of the transmitter. In each case, the interaction between a transmitter and receptor results in a specific physiologic response. The action of a transmitter is determined by the type of receptor (excitatory or inhibitory) to which it binds. For example, acetylcholine is excitatory when it is released at a myoneural junction, and it is inhibitory when it is released at the sinoatrial node in the heart. Receptors are named according to the type of



■ FIGURE 36-7 ■ Schematic illustration of (1) neurotransmitter (T) release; (2) binding of transmitter to postsynaptic receptor; termination of transmitter action by (3a) reuptake of transmitter into the presynaptic terminal, (3b) enzymatic degradation, or (3c) diffusion away from the synapse; and (4) binding of transmitter to presynaptic receptors for feedback regulation of transmitter release.

neurotransmitter with which they interact. For example, a *cholinergic receptor* is a receptor that binds acetylcholine.

Rapid removal of a transmitter, once it has exerted its effects on the postsynaptic membrane, is necessary to maintain precise control of neural transmission. A released transmitter can undergo one of three fates: it can be broken down into inactive substances by enzymes; it can be taken back up into the presynaptic neuron in a process called *reuptake*; or it can diffuse away into the intercellular fluid until its concentration is too low to influence postsynaptic excitability. For example, acetylcholine is rapidly broken down by acetylcholinesterase into acetic acid and choline, with the choline being taken back into the presynaptic neuron for reuse in acetylcholine synthesis. The catecholamines are largely taken back into the neuron in an unchanged form for reuse. Catecholamines also can be degraded by enzymes in the synaptic space or in the nerve terminals.

Neurohumoral mediators reach their target cells through the bloodstream and produce an even slower action than do the neuromodulators. Both the nervous system and the endocrine system use chemical molecules as messengers. As more information is obtained about the chemical messengers of these systems, the distinction between them becomes less evident. Many neurons, such as those in the adrenal medulla, secrete transmitters into the bloodstream, and it has been found that other neurons possess receptor sites for hormones. Many hormones have turned out to be neurotransmitters. Vasopressin (also known as *antidiuretic hormone*), a peptide hormone released from the posterior pituitary gland, acts as a hormone in the kidney and as a neurotransmitter for nerve cells in the hypothalamus. More than a dozen of these cell-to-cell and blood-borne messengers can relay signals in the nervous system or the endocrine system.

Other classes of messenger molecules, known as *neuro-modulators*, also may be released from axon terminals. Neuro-modulator molecules react with presynaptic or postsynaptic

644 Unit Ten: Alterations in the Nervous System

receptors to alter the release of or response to neurotransmitters. Neuromodulators may act on postsynaptic receptors to produce slower and longer-lasting changes in membrane excitability. This alters the action of the faster-acting neurotransmitter molecules by enhancing or decreasing their effectiveness. By combining with autoreceptors on its own presynaptic membrane, a transmitter can act as a neuromodulator to augment or inhibit further nerve activity. In some nerves, such as the peripheral sympathetic nerves, a messenger molecule can have both transmitter and modulator functions. For example, norepinephrine can activate an α_1 -adrenergic postsynaptic receptor to produce vasoconstriction or stimulate an α_2 -adrenergic presynaptic receptor to inhibit further norepinephrine release.

Neurotrophic or *nerve growth factors* are required to maintain the long-term survival of the postsynaptic cell and are secreted by axon terminals independent of action potentials. Examples include neuron-to-neuron trophic factors in the sequential synapses of CNS sensory neurons. Trophic factors from target cells that enter the axon and are necessary for the long-term survival of presynaptic neurons also have been demonstrated. Target cell-to-neuron trophic factors probably have great significance in establishing specific neural connections during normal embryonic development.

In summary, neurons are characterized by the ability to communicate with other neurons and body cells through pulsed electrical signals called *action potentials*. The cell membranes of neurons contain ion channels that are responsible for generating action potentials. Action potentials are divided into three parts: (1) the resting membrane potential, during which the membrane is polarized but no electrical activity occurs; the (2) depolarization phase, during which sodium channels open, allowing rapid inflow of the sodium ions that generate the electrical impulse; and (3) the repolarization phase, during which the membrane is permeable to the potassium ion, allowing for the efflux of potassium ions and return to the resting membrane potential.

Synapses are structures that permit communication between neurons. Two types of synapses have been identified: electrical and chemical. Electrical synapses consist of gap junctions between adjacent cells that allow action potentials to move rapidly from one cell to another. Chemical synapses involve special presynaptic and postsynaptic structures, separated by a synaptic cleft. They rely on chemical messengers, released from the presynaptic neuron, that cross the synaptic cleft and then interact with receptors on the postsynaptic neuron.

Neurotransmitters are chemical messengers that control neural function; they selectively cause excitation or inhibition of action potentials. Three major types of neurotransmitters are known: amino acids such as glutamic acid and GABA, peptides such as the endorphins and enkephalins, and monoamines such as epinephrine and norepinephrine. Neurotransmitters interact with cell membrane receptors to produce either excitatory or inhibitory actions. Neuromodulators are chemical messengers that react with membrane receptors to produce slower and longer-acting changes in membrane permeability. Neurotrophic or growth factors, also released from presynaptic terminals, are required to maintain the long-term survival of postsynaptic neurons.

DEVELOPMENT AND ORGANIZATION OF THE NERVOUS SYSTEM

The development of the nervous system can be traced far back into evolutionary history. During its development, newer functional features and greater complexity resulted from the modification and enlargement of more primitive structures. Survival of the species depended on the rapid reaction to environmental danger, to potential food sources, or to a sexual partner.

The front, or *rostral*, end of the CNS became specialized for sensing the external environment and controlling reactions to it. In time, the ancient organization, which is largely retained in the spinal cord segments, was expanded in the forward segments of the nervous system. Of these, the most forward segments have undergone the most radical modification and have developed into the forebrain: the diencephalon and the cerebral hemispheres. The dominance of the front end of the CNS is reflected in a hierarchy of control levels: brain stem over spinal cord, and forebrain over brain stem. Throughout evolution, newer functions were added to the surface of functionally more ancient systems. As newer functions became concentrated at the rostral end of the nervous system, they also became more vulnerable to injury.

Embryonic Development

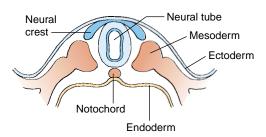
The nervous system appears very early in embryonic development (week 3). This early development is essential because it influences the development and organization of many other body systems, including the axial skeleton, skeletal muscles, and sensory organs such as the eyes and ears. During later fetal life and thereafter, the nervous system provides communication, signal processing, integrative, and memory functions. The early induction and later lifelong communication functions of the nervous system are at the center of the integrity, survival, and individuality of each person.

All body tissues and organs have developed from the three embryonic layers (*i.e.*, endoderm, ectoderm, and mesoderm) that were present during the third week of embryonic life (Fig. 36-8). The body is organized into the soma and viscera.

KEY CONCEPTS

HIERARCHY OF NERVOUS SYSTEM CONTROL

- No part of the nervous system functions independently of other parts.
- Progressively greater complexity in the responses and greater precision in their control occur at each higher level of the nervous system.
- The more rostral, recently developed parts of the neural tube gain dominance or control over lower levels.
- As newer functions became concentrated at the rostral end of the nervous system, they also became more vulnerable to injury.



■ FIGURE 36-8 ■ Transverse section of a 22-23-day-old embryo showing the ectoderm, mesoderm, endoderm, neural crest cells, notocord, and neural tube. During development, the neural tube develops into the CNS and the notocord becomes the foundation upon which the vertebral column develops. The neural crest cells become the progenitors of the neurons and supporting cells of the peripheral nervous system.

The soma, or body wall, includes all of the structures derived from the embryonic ectoderm, such as the epidermis of the skin and the CNS. The mesodermal connective tissues of the soma include the dermis of the skin, skeletal muscle, bone, and the outer lining of the body cavity (*i.e.*, parietal pleura and peritoneum). The nervous system innervates all somatic structures as well as the internal structures making up the viscera. The viscera includes the great vessels derived from the intermediate mesoderm, the urinary system, and the gonadal structures; it also includes the inner lining of the body cavities, such as the visceral pleura and peritoneum, and the mesodermal tissues that surround the endoderm-lined gut and its derivative organs (e.g., lungs, liver, pancreas).

There are both somatic and visceral nerves. The somatic nerves innervate the skeletal muscles and the smooth muscle and glands of skin and body wall. The visceral nerves supply the visceral organs of the body, transmitting information through the autonomic nerves in the PNS to control the smooth muscle and cardiac muscle as well as the glands of the visceral organs.

Segmental Organization

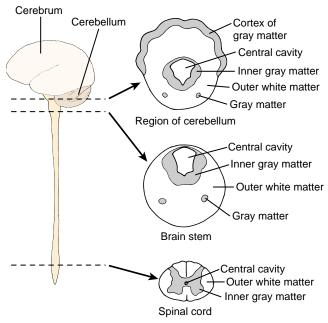
The early pattern of segmental development is presented as a framework for understanding the nervous system. Although the early muscular, skeletal, vascular, and excretory systems and the nerves that supply the somatic and visceral structures have the same segmental pattern, it is the nervous system that most clearly retains this organization in postnatal life. Developmentally, the basic organizational pattern of the body is that of a longitudinal series of segments, each repeating the same fundamental pattern. The CNS and its associated peripheral nerves consist of approximately 43 segments, 33 of which form the spinal cord and spinal nerves, and 10 of which form the brain and its cranial nerves.

The basic pattern of the CNS is that seen in the spinal cord a central cavity surrounded by inner core of gray matter and a superficial layer of white matter (Fig. 36-9). The brain retains this organization, but it also contains additional regions of gray matter that are not evident in the spinal cord. The gray matter is functionally divided into longitudinal columns of nerve cell bodies called the *cell columns*. The superficial white matter region contains the longitudinal tract systems of the CNS. The dorsal half of the gray matter is called the dorsal horn.

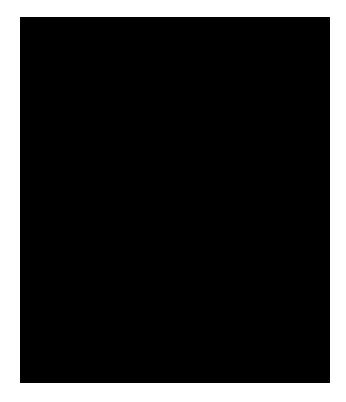
KEY CONCEPTS

THE DEVELOPMENTAL ORGANIZATION OF THE NERVOUS SYSTEM

- On cross section, the embryonic neural tube develops into a central canal surrounded by gray matter, or cellular portion (cell columns), and white matter, or tract system of the central nervous system (CNS).
- As the nervous system develops, it becomes seqmented, with a repeating pattern of afferent neuron axons forming the dorsal roots of each succeeding segmental nerve and the exiting efferent neurons forming the ventral roots of each succeeding segmental nerve.
- The nerve cells in the gray matter are arranged longitudinally in cell columns, with afferent sensory neurons located in the dorsal columns and efferent motor neurons located in the ventral columns.
- The axons of the cell column neurons project out into the white matter of the CNS, forming the longitudinal tract systems.



■ FIGURE 36-9 ■ Segmental organization of gray and white matter in the CNS (highly simplified). From top to bottom, the diagrams represent cross-sections at the levels of cerebellum, brain stem, and spinal cord. In each section, the dorsal aspect is on top. In general, white matter lies external to gray matter; however, collections of gray matter migrate externally into the white matter in the developing brain (see arrows). The cerebrum resembles the cerebellum in its external cortex of gray matter. (Marieb E.N. [1995]. Human anatomy and physiology. [3rd edition, p. 383]. Redwood City, CA. The Benjamin/Cummings Publishing Company)



The ventral portion, or *ventral horn*, contains efferent neurons that communicate by way of the ventral roots with effector cells of the body segment. Many CNS neurons develop axons that grow longitudinally as tract systems that communicate between neighboring and distal segments of the neural tube.

Each segment of the CNS is accompanied by bilateral pairs of bundled nerve fibers, or roots, a ventral pair and a dorsal pair (Fig. 36-10). The paired dorsal roots connect a pair of dorsal root ganglia and their corresponding CNS segment. The dorsal root ganglia contain many afferent nerve cell bodies, each having two axon-like processes—one that ends in a peripheral receptor and the other that enters the central neural segment. The axon-like process that enters the central neural segment communicates with a neuron called an *input association (IA) neuron*. The paired ventral roots of each segment are bundles of axons that provide efferent output to effector sites such as the muscles and glandular cells of the body segment.

Cell Columns

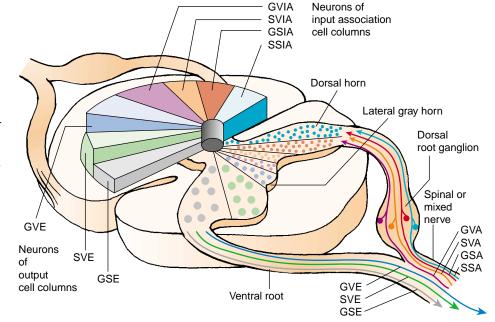
The organizational structure of the nervous system can be best explained and simplified as a pattern in which functionally specific PNS and CNS neurons are repeated as parallel cell columns running lengthwise along the nervous system. In this organizational pattern, afferent neurons, dorsal horn cells, and ventral horn cells are organized as a bilateral series of 11 cell columns.

The cell columns on each side can be further grouped according to their location in the PNS: four in the dorsal ganglia that contain sensory neurons; four in the dorsal horn containing sensory IA neurons; and three in the ventral horn that contain motoneurons (Fig. 36-11). Each column of dorsal root ganglia projects to its particular column of IA neurons in the dorsal horn. The IA neurons distribute afferent information to local reflex circuitry and to more rostral and elaborate segments of the CNS. The ventral horns contain output association (OA) neurons and lower motoneurons, which project to the effector muscles. The afferent and efferent cell columns of the PNS and CNS, their projections, and the type of information they transmit are summarized in Table 36-1.

Between the IA neurons and the OA neurons are networks of small internuncial neurons arranged in complex circuits. Internuncial neurons provide the discreteness, appropriateness, and intelligence of responses to stimuli. Most of the billions of CNS cells in the spinal cord and brain gray matter are internuncial neurons.

Dorsal Horn Cell Columns. Four columns of afferent (sensory) neurons in the dorsal root ganglia directly innervate four

■ FIGURE 36-11 ■ Cell columns of the central nervous system. The cell columns in the dorsal horn contain input association (IA) neurons for the general visceral afferent (GVA), special visceral afferent (SVA), special sensory afferent (SSA), and general somatic afferent (GSA) neurons with cell bodies in the dorsal root ganglion. The cell columns in the ventral horn contain the general visceral efferent (GVE), pharyngeal efferent (PE), and general somite efferent (GSE) neurons and their output association (OA) neurons.



Segment and Nerve	Component	Innervation	Function
1. Forebrain			
I. Olfactory	SVA	Receptors in olfactory mucosa	Reflexes, olfaction (smell)
2. II. Optic nerve		Optic nerve and retina (part of brain system, not a peripheral nerve)	
3. Midbrain			
V. Trigeminal (V ₁) ophthalmic division	SSA GSA	Muscles: upper face: forehead, upper lid Skin, subcutaneous tissue; conjunctiva; frontal/ethmoid sinuses	Facial expression, proprioceptio Somesthesia Reflexes (blink)
III. Oculomotor	GVE	Iris sphincter Ciliary muscle	Pupillary constriction Accommodation
	GSE	Extrinsic eye muscles	Eye movement, lid movement
4. Pons			
V. Trigeminal (V ₂) maxillary division	SSA	Muscles: facial expression	Proprioception Reflexes (sneeze), somesthesia
	GSA	Skin, oral mucosa, upper teeth, hard palate, maxillary sinus	
V. Trigeminal (V ₃) mandibular division	SSA GSA	Lower jaw, muscles: mastication Skin, mucosa, teeth, anterior ½ of tongue	Proprioception, jaw jerk Reflexes, somesthesia
	PE	Muscles: mastication tensor tympani tensor veli palantini	Mastication: speech Protects ear from loud sound Tenses soft palate
IV. Trochlear	GSE	Extrinsic eye muscle	Moves eye down and in
5. Caudal Pons			
VIII. Vestibular, cochlear (vestibulocochlear)	SSA	Vestibular end organs Organ of Corti	Reflexes, sense of head position Reflexes, hearing
VII. Facial nerve, intermedius	GSA	External auditory meatus	Somesthesia
portion	GVA	Nasopharynx	Gag reflex: sensation
	SVA GVE	Taste buds of anterior ⅔ of tongue Nasopharynx	Reflexes: gustation (taste) Mucous secretion, reflexes
Facial nerve	PE	Lacrimal, sublingual, submandibular glands Muscles: facial expression, stapedius	Lacrimation, salivation Facial expression
VI. Abducens	GSE	Extrinsic eye muscle	Protects ear from loud sounds Lateral eye deviation
6. Middle Medulla			
IX. Glossopharyngeal	SSA	Stylopharyngeus muscle	Proprioception
1 7 5	GSA	Posterior external ear	Somesthesia
	SVA	Taste buds of posterior ½ of tongue	Gustation (taste)
	GVA	Oral pharynx	Gag reflex: sensation
	GVE PE	Parotid gland; pharyngeal mucosa Stylopharyngeus muscle	Salivary reflex: mucous secretion Assists swallowing
	12	stylopharyngeus musee	Assists swallowing
7,8,9,10. Caudal Medulla	66A		D
X. Vagus	SSA	Muscles: pharynx, larynx	Proprioception
	GSA SVA	Posterior external ear	Somesthesia Reflexes gustation
	GVA	Taste buds, pharynx, larynx Visceral organs (esophagus to midtransverse colon, liver, pancreas, heart, lungs)	Reflexes, gustation Reflexes, sensation
	GVE	Visceral organs as above	Parasympathetic efferent
	PE	Muscles: pharynx, larynx	Swallowing, phonation, emesis
XIII. Hypoglossal	GSE	Muscles of tongue	Tongue movement, reflexes
Spinal Segments			
C1–C4 Upper Cervical XI. Spinal accessory nerve	PE	Muscles: sternocleidomastoid, trapezius	Head, shoulder movement
Spinal nerves	SSA	Muscles of neck	Proprioception, DTRs
	GSA	Neck, back of head	Somesthesia
	GSE	Neck muscles	Head, shoulder movement

TABLE 36-1 The Segmental Nerves and Their Components (Continued)				
Segment and Ner	ve	Component	Innervation	Function
C5–C8 Lower Cervica	I	SSA GSA GSE	Upper limb muscles Upper limbs Upper limb muscles	Proprioception, DTRs Reflexes, somesthesia Movement, posture
T1–L2 Thoracic, Upp	er Lumbar	SSA GSA GVA GVE GSE	Muscles: trunk, abdominal wall Trunk, abdominal wall All of viscera All of viscera Muscles: trunk, abdominal wall, back	Proprioception Reflexes, somesthesia Reflexes and sensation Sympathetic reflexes, vasomotor control, sweating, piloerection Movement, posture, respiration
L2–S1 Lower Lumbar	. Upper Sacral	SSA GSA GSE	Lower limb muscles Lower trunk, limbs, back Muscles: trunk, lower limbs, back	Proprioception, DTRs Reflexes, somesthesia Movement, posture
S2–S4 Lower Sacral		SSA GSA GVA GVE	Muscles: pelvis, perineum Pelvis, genitalia Hindgut, bladder, uterus Hindgut, visceral organs	Proprioception Reflexes, somesthesia Reflexes, sensation Visceral reflexes, defecation, urination, erection
S5–Co2 Lower Sacral	, Coccygeal	SSA GSA GSE	Perineal muscles Lower sacrum, anus Perineal muscles	Proprioception Reflexes, somesthesia Reflexes, posture

Afferent (sensory) components: SSA, special somatic afferent; GSA, general somatic afferent; SVA, special visceral afferent; GVA, general visceral afferent.

Efferent (motor) components: GVE, general visceral efferent (autonomic nervous system); PE, pharyngeal efferent; GSE, general somatic efferent; DTRs, deep tendon reflexes.

corresponding columns of SIA neurons in the dorsal horn. These columns are categorized as special and general afferents: special somatic afferent (SSA), general somatic afferent (GSA), special visceral afferent (SVA), and general visceral afferent (GVA).

The SSA fibers are concerned with internal sensory information such as joint and tendon sensation (*i.e.*, proprioception). Neurons in the special sensory IA cell columns relay their information to local reflexes concerned with posture and movement. These neurons also relay information to the cerebellum, contributing to coordination of movement, and to the forebrain, contributing to experience. Afferents innervating the vestibular system of the inner ear also belong to the special somatic afferent category.

The GSA neurons innervate the skin and other somatic structures, responding to stimuli such as those that produce pressure or pain. General sensory IA column neurons relay information to protective and other local reflex circuits and project the information to the forebrain, where it is perceived as painful, warm, cold, and the like.

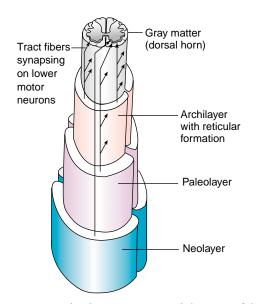
The SVAs innervate specialized gut-related receptors, such as the taste buds and receptors of the olfactory mucosa. Their central processes communicate with special visceral IA column neurons that project to reflex circuits producing salivation, chewing, swallowing, and other responses. Forebrain projection fibers from these association cells provide sensations of taste (*i.e.*, gustation) and smell (*i.e.*, olfaction).

GVA neurons innervate visceral structures such as the gastrointestinal tract, urinary bladder, and heart and great vessels; they project to the general visceral IA column, which relays information to vital reflex circuits and sends information to the forebrain regarding visceral sensations such as stomach fullness and bladder pressure.

Ventral Horn Cell Columns. The ventral horn contains three longitudinal cell columns: general visceral efferent (GVE), pharyngeal efferent (PE), and general somatic efferent (GSE) (Fig. 36-11). The efferent neurons for the OA in the ventral horn originate in brain centers (motor cortex and autonomic nervous system centers) that control skeletal muscle and visceral function. Each of these cell columns contains OA and efferent neurons. The OA neurons coordinate and integrate the function of the efferent motoneurons of its column.

GVE neurons transmit the efferent output of the autonomic nervous system and are called preganglionic neurons. These neurons are structurally and functionally divided between either the sympathetic or the parasympathetic nervous systems (discussed later in this chapter). Their axons project through the segmental ventral roots to innervate smooth and cardiac muscle and glandular cells of the body, most of which are in the viscera. In the viscera, three additional cell columns are present on each side of the body. These become the postganglionic neurons of the autonomic nervous system. In the sympathetic nervous system, the columns are represented by the paravertebral or sympathetic chain ganglia and the prevertebral series of ganglia (e.g., celiac ganglia) associated with the dorsal aorta. For the parasympathetic nervous system, these become the enteric plexus in the wall of the gut-derived organs and a series of ganglia in the head.

The PE neurons innervate the muscles of mastication, facial expression, and muscles of the pharynx and larynx. PE neurons also innervate the muscles responsible for moving the head.



■ **FIGURE 36-12** The three concentric subdivisions of the tract systems of the white matter. Migration of neurons into the archilayer converts it into the reticular formation of the white matter.

The GSE neurons supply skeletal muscles of the body and head, including those of the body, limbs, tongue, and extrinsic eye muscles. These efferent neurons transmit the commands of the CNS to peripheral effectors, the skeletal muscles. They are the "final common pathway neurons" in the sequence leading to motor activity. They often are called *lower motoneurons* (*LMNs*) because they are under the control of higher levels of the CNS, including precise control by *upper motoneurons* (*UMNs*).

Longitudinal Tracts

The gray matter of the cell columns in the CNS is surrounded by bundles of myelinated axons (*i.e.*, white matter) and unmyelinated axons that travel longitudinally along the length of the neural axis. This white matter can be divided into three layers: an inner, a middle, and an outer layer (Fig. 36-12). The inner layer contains short fibers that project for a maximum of approximately five segments before re-entering the gray matter. The middle layer projects to six or more segments. The inner and middle layer fibers have many branches, or collaterals, that enter the gray matter of intervening segments. The outer layer contains large-diameter axons that that can travel the entire length of the nervous system (Table 36-2). *Suprasegmental* is a term that refers to higher levels of the CNS, such as the brain stem and cerebrum and structures above a given CNS segment. The middle and outer layer fibers have suprasegmental projections.

The longitudinal layers are arranged in bundles, or fiber tracts, that contain axons that have the same destination, origin, and function (Fig. 36-13). These longitudinal tracts are named systematically to reflect their origin and destination; the origin is named first, and the destination is named second. For example, the spinothalamic tract originates in the spinal cord and terminates in the thalamus. The corticospinal tract originates in the cerebral cortex and ends in the spinal cord.

The Inner Layer. The inner layer of white matter contains the axons of neurons that connect neighboring segments of the nervous system. Axons of this layer permit the pool of motoneurons of several segments to work together as a functional unit. They also allow the afferent neurons of one segment to trigger reflexes that activate motor units in neighboring and in the same segments. In terms of evolution, this is the oldest of the three layers, and it is sometimes called the archilayer. It is the first of the longitudinal layers to become functional, and its circuitry may be limited to reflex types of movements, including reflex movements of the fetus (*i.e.*, quickening) that begin during the fifth month of intrauterine life.

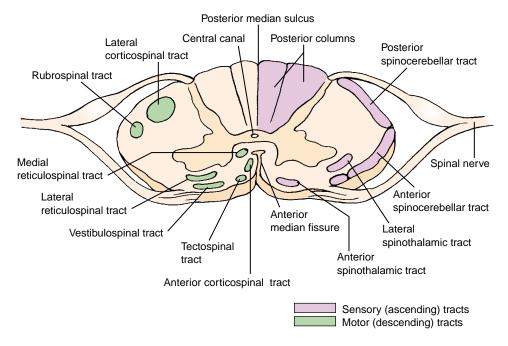
The archilayer of the white matter differs from the other two layers in one important aspect. Many neurons in the embryonic gray matter migrate out into this layer, resulting in a rich mixture of neurons and local fibers called the *reticular formation*. The circuitry of most reflexes is contained in the reticular formation. In the brain stem, the reticular formation becomes quite large and contains major portions of vital reflexes, such as those controlling respiration, cardiovascular function, swallowing, and vomiting. A functional system called the *reticular activating system* operates in the lateral portions of the reticular formation of the medulla, pons, and especially the midbrain. Information converging from all sensory modalities, including those of the somesthetic, auditory, visual, and visceral afferent nerves, bombards the neurons of this system.

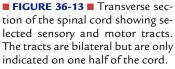
The reticular activating system has descending and ascending portions. The descending portion communicates with all spinal segmental levels through middle layer reticulospinal tracts and serves to facilitate many cord-level reflexes. For example, it speeds reaction time and stabilizes postural reflexes. The ascending portion accelerates brain activity, particularly thalamic and cortical activity. This is reflected by the appearance of awake brain-wave patterns. Sudden stimuli result in protective and attentive postures and cause increased awareness.

TABLE 36-2	Characteristics of the Concentric Subdivisions of the Longitudinal
	White Matter of the Central Nervous System

aleolayer Tracts Neolayer Tracts
uprasegmental (≥5 segments) Suprasegmental
Aultisynaptic but fewer than Monosynaptic with target archilayer tracts structures
ast Fastest
pinothalamic tracts Corticospinal tracts







The Middle Layer. The middle layer of the white matter contains most of the major fiber tract systems required for sensation and movement. It contains the ascending spinoreticular and spinothalamic tracts. This layer consists of larger-diameter and longer suprasegmental fibers, which ascend to the brain stem and are largely functional at birth. In terms of evolutionary development, these tracts are quite old, and this layer is sometimes called the *paleolayer*. It facilitates many primitive functions, such as the auditory startle reflex, which occurs in response to loud noises. This reflex consists of turning the head and body toward the sound, dilating the pupils of the eyes, catching of the breath, and quickening of the pulse.

The Outer Layer. The outer layer of the tract systems is the newest of the three layers with respect to evolutionary development, and it is sometimes called the *neolayer*. It becomes functional approximately the second year of life, and it includes the pathways needed for bladder training. Myelination of these suprasegmental tracts, which include many pathways required for delicate and highly coordinated skills, is not complete until approximately the fifth year of life. This includes the development of tracts needed for fine manipulative skills, such as the finger-thumb coordination required for using tools and the toe movements needed for acrobatics. Neolayer tracts are the most recently evolved systems and, being more superficial on the brain and spinal cord, are the most vulnerable to injury.

Collateral Communication Pathways. Axons in the archilayer and paleolayer characteristically possess many collateral branches that move into the gray cell columns or synapse with fibers of the reticular formation as the axon passes each succeeding CNS segment. Should a major axon be destroyed at some point along its course, these collaterals provide multisynaptic alternative pathways that bypass the local damage. Damage usually is followed by slow return of function, presumably through the collateral connections.

Neolayer tracts do not possess these collaterals but instead project mainly to the target neurons with which they communicate. When neolayer tracts are damaged, the paleolayer and archilayer tracts often remain functional, and rehabilitation methods can result in effective use of the older systems. Delicacy and refinement may be lost, but basic function remains. For example, when the corticospinal system, an important neolayer system that permits the fine manipulative control required for writing, is damaged, the remaining paleolayer systems, if intact, permit the grasping and holding of objects. The hand can still be used to perform its basic function, but the individual manipulation of the fingers is permanently lost.

In summary, the nervous system can be divided into two parts: the CNS, which consists of the brain and spinal cord, and the PNS, which contains the neurons and neuronal processes that leave the CNS and innervate skeletal muscles and visceral structures of the body. The brain stem and spinal cord are subdivided into the dorsal horn, which contains neurons that receive and process incoming or afferent information, and the ventral horn, which contains efferent motoneurons that handle the final stages of output processing.

Throughout life, the organization of the nervous system retains many patterns established during early embryonic life. Each of the 43 or more body segments is connected to corresponding CNS or neural tube segments by segmental afferent and efferent neurons. Afferent or sensory neuronal processes that carry afferent or sensory information enter the CNS through the dorsal root ganglia and the dorsal roots. Afferent neurons of the dorsal root ganglia are of four types: GSA, SSA, GVA, and SVA. Each of these afferent neurons synapses with its appropriate IA neurons in the cell columns of the dorsal horn. There are three cells columns in the ventral horn. The OA neurons in these columns synapse with ventral horn motoneurons that exit the CNS in the ventral roots. GSE neurons are LMNs that innervate skeletal muscles; PE neurons innervate pharyngeal muscles; and GVE neurons innervate visceral structures. This pattern of afferent and efferent neurons, which in general is repeated in each segment of the body, forms parallel cell columns running lengthwise through the CNS and PNS.

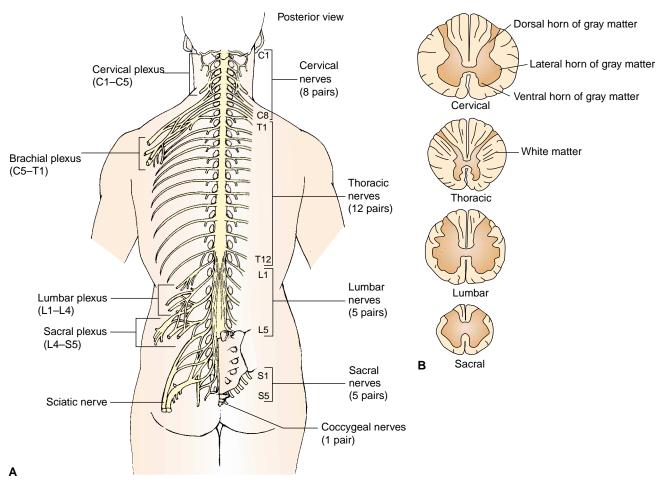
Longitudinal communication between CNS segments is provided by tracts that are arranged in three layers: an inner, middle, and outer layer. The inner layer of white matter contains the axons of neurons that connect neighboring seqments of the nervous system. It contains a mix of nerve cells and axons called the reticular formation and is the site of many important spinal cord and brain stem reflex circuits. The middle layer tracts provide the longitudinal communication between more distant segments of the nervous system; it contains most of the major fiber tract systems required for sensation and movement. The recently evolved outer or neolayer systems, which become functional during infancy and childhood, provide the means for very delicate and discriminative function. The outer position of the neolayer tracts and their lack of collateral and redundant pathways make them the most vulnerable to injury.

THE SPINAL CORD AND BRAIN

The Spinal Cord

In the adult, the spinal cord is found in the upper two thirds of the spinal canal of the vertebral column (Fig. 36-14A). It extends from the foramen magnum at the base of the skull to a cone-shaped termination, the conus medullaris, usually at the level of the first or second lumbar vertebra (L1 or L2) in the adult. The dorsal and ventral roots of the more caudal portions of the cord elongate during development and angle downward from the cord, forming what is called the *cauda equina* (from the Latin for "horse's tail"). The filum terminale, which is composed of non-neural tissues and the pia mater, continues caudally and attaches to the second sacral vertebra (S2).

The spinal cord is somewhat oval on transverse section. Internally, the gray matter has the appearance of a butterfly or the letter "H" on cross section (Fig. 36-14B). The extensions of the gray matter that form the letter "H" are called the *horns*. Those that extend posteriorly are called the *dorsal horns*, and those that extend anteriorly are called the *ventral horns*. The central portion of the cord, which connects the dorsal and ventral horns, is called the *intermediate gray matter*. The inter-



■ **FIGURE 36-14** ■ (**A**) Dorsal view of the spinal cord including portions of the major spinal nerves and some of the components of the major nerve plexuses. (**B**) Cross-sectional views of the spinal cord, showing regional variations in gray matter and increasing white matter as the cord ascends.

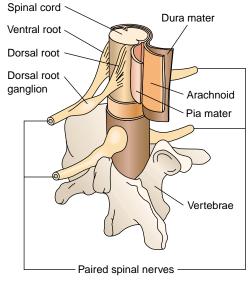
mediate gray matter surrounds the central canal. In the thoracic area, the small, slender projections that emerge from the intermediate gray matter are called the *intermediolateral columns* of the horns. These columns contain the visceral OA neurons and the efferent neurons of the sympathetic nervous system.

The gray matter is proportional to the amount of tissue innervated by a given segment of the cord (Fig. 36-14B). Larger amounts of gray matter are present in the lower lumbar and upper sacral segments, which supply the lower extremities, and in the fifth cervical segment to the first thoracic segment, which supply the upper limbs. The white matter in the spinal cord also increases progressively toward the brain because ever more ascending fibers are added and the number of descending axons is greater.

The spinal cord and the dorsal and ventral roots are covered by a connective tissue sheath, the pia mater, which also contains the blood vessels that supply the white and gray matter of the cord (Fig. 36-15). On the lateral sides of the spinal cord, extensions of the pia mater, the denticulate ligaments, attach the sides of the spinal cord to the bony walls of the spinal canal. Thus, the cord is suspended by both the denticulate ligaments and the segmental nerves. A fat- and vessel-filled epidural space intervenes between the spinal dura mater and the inner wall of the spinal canal. Each vertebral body has two pedicles that extend posteriorly and support the laterally oriented transverse processes of the neural laminae, which arch medially and fuse to continue as the spinal processes.

The spaces between the vertebral bodies are filled with fibrocartilaginous discs and stabilized with tough ligaments. A gap, the intervertebral foramen, occurs between each two succeeding pedicles, allowing for the exit of the segmental nerves and passage of blood vessels. The spinal cord lives in the protective confines of this series of concentric flexible tissue and body sheaths. Supporting structures of the spinal cord are discussed further in Chapter 37.

Early in fetal life, the spinal cord extends the entire length of the vertebral column and the spinal nerves exit through the intervertebral foramina (openings) near their level of origin. Because the vertebral column and spinal dura grow faster than



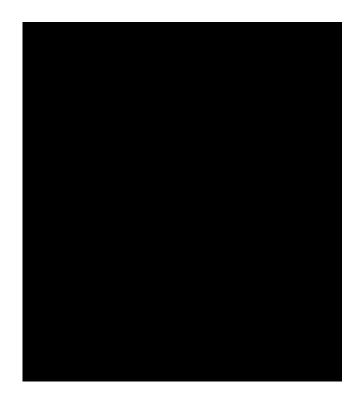
■ FIGURE 36-15 ■ Spinal cord and meninges.

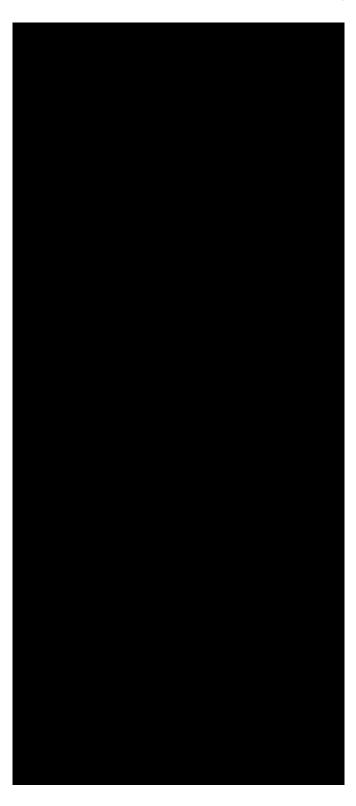
the spinal cord, a disparity develops between each succeeding cord segment and the exit of its dorsal and ventral nerve roots through the corresponding intervertebral foramina. In the newborn, the cord terminates at the level of L2 or L3. In the adult, the cord usually terminates in the inferior border of L1, and the arachnoid and its enclosed subarachnoid space, which is filled with CSF, do not close down on the filum terminale until they reach the level of S2 (Fig. 36-16). This results in the formation of a pocket of CSF, the dural cisterna spinalis, which extends from approximately L2 to S2. Because this area contains an abundant supply of CSF and the spinal cord does not extend this far, the area often is used for sampling the CSF. A procedure called a *spinal tap*, or puncture, can be done by inserting a special needle into the dural sac at L3 or L4. The spinal roots, which are covered with pia mater, are in little danger of trauma from the needle used for this purpose.

Spinal Nerves

The peripheral nerves that carry information to and from the spinal cord are called *spinal nerves*. There are 32 or more pairs of spinal nerves (*i.e.*, 8 cervical, 12 thoracic, 5 lumbar, 5 sacral, and 2 or more coccygeal); each pair is named for the segment of the spinal cord from which it exits. Because the first cervical spinal nerve exits the spinal cord just above the first cervical vertebra (C1), the nerve is given the number of the bony vertebra just below it. However, the numbering is changed for all lower levels. An extra cervical nerve, the C8 nerve, exits above the T1 vertebra, and each subsequent nerve is numbered for the vertebra just above its point of exit (Fig. 36-17).

Each spinal cord segment communicates with its corresponding body segment through the paired segmental spinal nerves. Each spinal nerve, accompanied by the blood vessels supplying the spinal cord, enters the spinal canal through an intervertebral foramen, where it divides into two branches, or





roots. One branch enters the dorsolateral surface of the cord (*i.e.*, dorsal root), carrying the axons of afferent neurons into the CNS. The other branch leaves the ventrolateral surface of the cord (*i.e.*, ventral root), carrying the axons of efferent neurons into the periphery. These two branches or roots fuse at the intervertebral foramen, forming the mixed spinal nerve—"mixed" because it has both afferent and efferent axons.

After emerging from the vertebral column, the spinal nerve divides into two branches or *rami* (singular, *ramus*): a small dorsal primary ramus and a larger ventral primary ramus (Fig. 36-18). The thoracic and upper lumbar spinal nerves also lead to a third branch, the ramus communicans, which contains sympathetic axons supplying the blood vessels, the genitourinary system, and the gastrointestinal system. The dorsal ramus contains sensory fibers from the skin and motor fibers to muscles of the back. The anterior primary ramus contains motor fibers that innervate the skeletal muscles of the anterior body wall and the legs and arms.

Spinal nerves do not go directly to skin and muscle fibers; instead, they form complicated nerve networks called *plexuses* (Fig. 36-14A). A plexus is a site of intermixing nerve branches. Many spinal nerves enter a plexus and connect with other spinal nerves before exiting from the plexus. Nerves emerging from a plexus form progressively smaller branches that supply the skin and muscles of the various parts of the body. The PNS contains four major plexuses: the cervical plexus, the brachial plexus, the lumbar plexus, and the sacral plexus.

The Brain

The brain is divided into three regions, the hindbrain, the midbrain, and the forebrain (Fig. 36-19A). The hindbrain includes the medulla oblongata, the pons, and its dorsal outgrowth, the cerebellum (see Chapter 38). Midbrain structures include two pairs of dorsal enlargements: the superior and inferior colliculi. The forebrain, which consists of two hemispheres and is covered by the cerebral cortex, contains central masses of gray matter, the basal ganglia (discussed in Chapter 38), and the rostral end of the neural tube, the diencephalon with its adult derivatives—the thalamus and hypothalamus.

An important concept is that the more rostral, recently developed parts of neural tube gain dominance or control over regions and functions at lower levels. They do not replace the more ancient circuitry but merely dominate it. After damage to

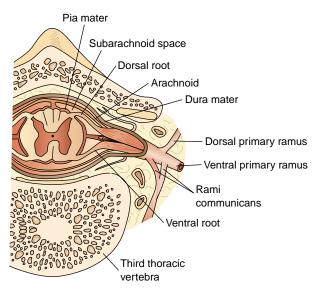
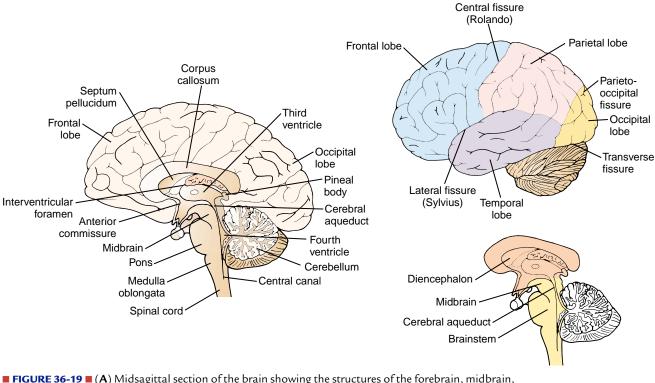


FIGURE 36-18 Cross-section of vertebral column at the level of the third thoracic vertebra, showing the meninges, the spinal cord, and the origin of a spinal nerve and its branches or rami.



■ FIGURE 36-19 ■ (A) Midsagittal section of the brain showing the structures of the forebrain, midbrain, and hindbrain. (B) Lateral view of the cerebral hemispheres. (C) Diencephalon, brain stem, and the cerebral aqueduct connecting the 3rd and 4th ventricles.

the more vulnerable parts of the forebrain, as occurs with brain death, a brain stem-controlled organism remains that is capable of breathing and may survive if the environmental temperature is regulated and nutrition and other aspects of care are provided. However, all aspects of intellectual function, experience, perception, and memory usually are permanently lost. The organization of content in this section moves from the more ancient circuitry of the hindbrain to the more dominant and recently developed structures of the forebrain.

Hindbrain

The term *brain stem* often is used to include the hindbrain, pons, and the midbrain. These regions of the neural tube have the organization of spinal cord segments, except that more of the longitudinal cell columns are present, reflecting the increased complexity of the cranial segmental nerves. In the brain stem, the structure and function of the reticular formation have been greatly expanded. In the pons and medulla, the reticular formation contains networks controlling basic breathing, eating, and locomotion functions. Higher-level integration of these functions occurs in the midbrain. The reticular formation is surrounded on the outside by the long tract systems that connect the forebrain with lower parts of the CNS.

Medulla. The *medulla oblongata* represents the caudal five segments of the brain part of the neural tube; the cranial nerve branches entering and leaving it have functions similar to the spinal segmental nerves. Although the ventral horn areas in the medulla are quite small, the dorsal horn areas are enlarged,

processing a large amount of the information pouring through the cranial nerves. Cranial nerves XII (hypoglossal), X (vagus), and (IX) glossopharyngeal have their origin in the medulla (see Table 36-1).

Pons. The pons, which develops from the fifth neural tube segment, is located between the medulla oblongata and the midbrain. Dorsally, it forms part of the anterior wall of the fourth ventricle (Fig. 36-19B).

As the name implies (pons = bridge), the pons is composed chiefly of conduction fibers. The enlarged area on the ventral surface of the pons contains the pontine nuclei, which receive information from all parts of the cerebral cortex. The axons of these neurons form a massive bundle that swings around the lateral side of the fourth ventricle to enter the cerebellum. In the pons, the reticular formation is large and contains the circuitry for masticating food and manipulating the jaws during speech. Cranial nerves VIII, VII, and VI have their origin in the pons (see Table 36-1).

Midbrain

The midbrain develops from the fourth segment of the neural tube, and its organization is similar to that of a spinal segment. The central canal is re-established as the cerebral aqueduct, connecting the fourth ventricle with the third ventricle (see Fig. 36-19B).

Two prominent bundles of nerve fibers, the *cerebral peduncles*, pass along the ventral surface of the midbrain. These fibers include the corticospinal tracts and are the main motor pathways between the forebrain and the pons. On the dorsal surface, four "little hills," the *superior* and *inferior colliculi*, are areas of cortical formation. The inferior colliculus is involved in directional turning and, to some extent, in experiencing the direction of sound sources. The superior colliculi are essential to the reflex mechanisms that control conjugate eye movements when the visual environment is surveyed. Two general somatic efferent cranial nerves, the oculomotor nerve, or cranial nerve III, and the trochlear nerve, or cranial nerve IV, exit the midbrain (see Table 36-1).

Forebrain

The most rostral part of the brain, the forebrain consists of the telencephalon, or "end brain," and the diencephalon, or "between brain." The diencephalon forms the core of the forebrain, and the telencephalon forms the cerebral hemispheres.

Diencephalon. Three of the most forward brain segments form an enlarged dorsal horn and ventral horn with a narrow, deep, enlarged central canal—the third ventricle—separating the two sides. This region is called the *diencephalon*. The dorsal horn part of the diencephalon is the thalamus and subthalamus, and the ventral horn part is the hypothalamus (Fig. 36-20). The optic nerve, or cranial nerve II, and retina are outgrowths of the diencephalon.

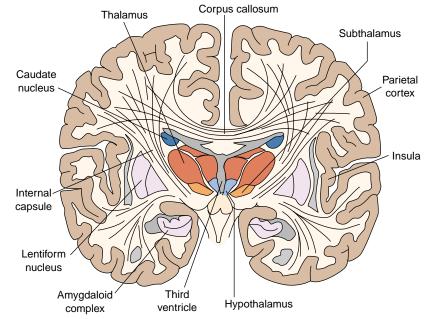
The thalamus consists of two large, egg-shaped masses, one on either side of the third ventricle. The thalamus is divided into several major parts, and each part is divided into distinct nuclei, which are the major relay stations for information going to and from the cerebral cortex. All sensory pathways have direct projections to thalamic nuclei, which convey the information to restricted areas of the sensory cortex. Coordination and integration of peripheral sensory stimuli occur in the thalamus, along with some crude interpretation of highly emotion-laden auditory experiences that not only occur but can be remembered. For example, a person can recover from a deep coma in which cerebral cortex activity is minimal and remember some of what was said at the bedside. The thalamus also plays a role in relaying critical information regarding motor activities to and from selected areas of the motor cortex. Two neuronal circuits are significant in this regard. One is the pathway from the cerebral cortex to the pons and cerebellum and then, by way of the thalamus, back to the motor cortex. The second is the feedback circuit that travels from the cortex to the basal ganglia, then to the thalamus, and from the thalamus back to the cortex. The subthalamus also contains movement control systems related to the basal ganglia.

Through its connections with the ascending reticular activating system, the thalamus processes neural influences that are basic to cortical excitatory rhythms (*i.e.*, those recorded on the electroencephalogram), to essential sleep-wakefulness cycles, and to the process of attending to stimuli. Besides their cortical connections, the thalamic nuclei have connections with each other and with neighboring nonthalamic brain structures such as the limbic system. Through their connections with the limbic system, some thalamic nuclei are involved in the relation between stimuli and the emotional responses they evoke.

The ventral horn portion of the diencephalon is the hypothalamus, which borders the third ventricle and includes a ventral extension, the neurohypophysis (*i.e.*, posterior pituitary). The hypothalamus is the area of master-level integration of homeostatic control of the body's internal environment. Maintenance of blood gas concentration, water balance, food consumption, and major aspects of endocrine and autonomic nervous system control require hypothalamic function.

The internal capsule is a broad band of projection fibers that lies between the thalamus medially and the basal ganglia laterally (see Fig. 36-20). It contains all of the fibers that connect the cerebral cortex with deeper structures, including the basal ganglia, thalamus, midbrain, pons, medulla, and spinal cord.

Cerebral Hemispheres. The two cerebral hemispheres are lateral outgrowths of the diencephalon. The cerebral hemispheres contain the lateral ventricles (*i.e.*, ventricles I and II), which are



■ FIGURE 36-20 ■ Frontal section of the brain passing through the third ventricle, showing the thalamus, subthalamus, hypothalamus, internal capsule, corpus callosum, basal ganglia (caudate nucleus, lenticular nucleus), amygdaloid complex, insula, and parietal cortex.

656 Unit Ten: Alterations in the Nervous System

connected with the third ventricle of the diencephalon by a small opening called the *interventricular foramen* (*i.e.*, foramen of Monro). Axons of the olfactory nerve, or cranial nerve I, terminate in the most ancient portion of the cerebrum—the olfactory bulb, where initial processing of olfactory information occurs. Projection axons from the olfactory bulb relay information through the olfactory tracts to the thalamus and to other parts of the cerebral cortex (*i.e.*, orbital cortex), where olfactory-related reflexes and olfactory experience occur.

The *corpus callosum* is a massive commissure, or bridge, of myelinated axons that connects the cerebral cortex of the two sides of the brain. Two smaller commissures, the anterior and posterior commissures, connect the two sides of the more specialized regions of the cerebrum and diencephalon.

The surfaces of the hemispheres are lateral (side), medial (area between the two sides of the brain), and basal (ventral). The cerebral cortex observed laterally is the recently evolved sixlayered neocortex. The surface of the hemispheres contains many ridges and grooves. A *gyrus* is the ridge between two grooves, and the groove is called a *sulcus* or *fissure*. The cerebral cortex is arbitrarily divided into lobes named after the bones that cover them: the frontal, parietal, temporal, and occipital lobes (Fig. 36-19C).

Frontal Lobe. The frontal lobe extends from the frontal pole to the central sulcus (*i.e.*, fissure) and is separated from the temporal lobe by the lateral sulcus. The frontal lobe can be subdivided rostrally into the frontal pole and laterally into the superior, middle, and inferior gyri, which continue on the undersurface over the eyes as the orbital cortex. These areas are associated with the medial thalamic nuclei, which also are related to the limbic system. In terms of function, the prefrontal cortex is thought to be involved in anticipation and prediction of consequences of behavior.

The precentral gyrus (area 4), next to the central sulcus, is the *primary motor cortex* (Fig. 36-21). This area of the cortex provides precise movement control for distal flexor muscles of the hands and feet and of the phonation apparatus required for speech. Just rostral to the precentral gyrus is a region of the frontal cortex called the *premotor* or *motor association cortex*. This region (area 8 and rostral area 6) is involved in the planning of complex learned movement patterns. The primary motor cortex and the association motor cortex are connected with lateral thalamic nuclei, through which they receive feedback information from the basal ganglia and cerebellum. On the medial surface of the hemisphere, the premotor area includes a *supplementary motor cortex* involved in the control of bilateral movement patterns requiring great dexterity.

Parietal Lobe. The parietal lobe of the cerebrum lies behind the central sulcus (*i.e.*, postcentral gyrus) and above the lateral sulcus. The strip of cortex bordering the central sulcus is called the primary somatosensory cortex (areas 3, 1, and 2) because it receives very discrete sensory information from the lateral nuclei of the thalamus. Just behind the primary sensory cortex is the somatosensory association cortex (areas 5 and 7), which is connected with the thalamic nuclei and with the primary sensory cortex. This region is necessary for perceiving the meaningfulness of integrated sensory information from various sensory systems, especially the perception of "where" the stimulus is in space and in relation to body parts. Localized lesions of this region can result in the inability to recognize the meaningfulness of an object (i.e., agnosia). With the person's eyes closed, a screwdriver can be felt and described as to shape and texture. Nevertheless, the person cannot integrate the sensory information required to identify it as a screwdriver (discussed further in Chapter 39).

Temporal Lobe. The temporal lobe lies below the lateral sulcus and merges with the parietal and occipital lobes. It includes the temporal pole and three primary gyri: the superior, middle, and inferior gyri. The primary auditory cortex (area 41) involves the part of the superior temporal gyrus that extends into the lateral sulcus (see Fig. 36-21). This area is particularly important in discrimination of sounds entering opposite ears. It receives auditory input projections by way of the inferior colliculus of the midbrain and a ventrolateral thalamic nucleus. The more exposed part of the superior temporal gyrus involves the auditory association area (area 22). The recognition of certain sound patterns and their meaning requires the function



of this area. The remaining portion of the temporal cortex is less defined functionally but apparently is important in longterm memory recall. This is particularly true with respect to perception and memory of complex sensory patterns, such as geometric figures and faces (*i.e.*, recognition of "what" or "who" the stimulus is). Irritation or stimulation can result in vivid hallucinations of long-past events.

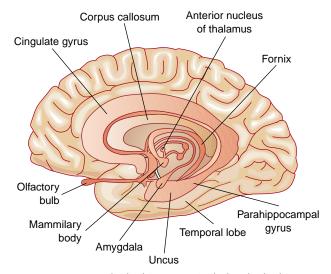
Occipital Lobe. The occipital lobe lies posterior to the temporal and parietal lobes and is only arbitrarily separated from them. The medial surface of the occipital lobe contains a deep sulcus extending from the limbic lobe to the occipital pole, the *calcarine sulcus*, which is surrounded by the primary visual cortex (area 17). Stimulation of this cortex causes the experience of bright lights (phosphenes) in the visual field. Just superior and inferior and extending onto the lateral side of the occipital pole is the *visual association cortex* (areas 18 and 19). This area is closely connected with the primary visual cortex and with complex nuclei of the thalamus. Integrity of the association cortex is required for gnostic visual function, by which the meaningfulness of visual experience, including experiences of color, motion, depth perception, pattern, form, and location in space, occurs.

The neocortical areas of the parietal lobe, between the somatosensory and the visual cortices, have a function in relating the texture, or "feel," and location of an object with its visual image. Between the auditory and visual association areas, the *parieto-occipital region* is necessary for relating the meaningfulness of a sound and image to an object or person.

Limbic System. The medial aspect of the cerebrum is organized into concentric bands of cortex, the limbic system (lim*bic* = borders), which surrounds the connection between the lateral and third ventricles. The innermost band just above and below the cut surface of the corpus callosum is folded out of sight but is an ancient, three-layered cortex ending as the hippocampus in the temporal lobe. Just outside the folded area is a band of transitional cortex, which includes the cingulate and the parahippocampal gyri (Fig. 36-22). This limbic lobe has reciprocal connections with the medial and the intralaminar nuclei of the thalamus, with the deep nuclei of the cerebrum (e.g., amygdaloid nuclei, septal nuclei), and with the hypothalamus. Overall, this region of the brain is involved in emotional experience and in the control of emotion-related behavior. Stimulation of specific areas in this system can lead to feelings of dread, high anxiety, or exquisite pleasure. It also can result in violent behaviors, including attack, defense, or explosive and emotional speech.

Meninges

Inside the skull and vertebral column, the brain and spinal cord are loosely suspended and protected by several connective tissue sheaths called the *meninges* (Fig. 36-23). The surfaces of the spinal cord, brain, and segmental nerves are covered with a delicate connective tissue layer called the *pia mater* (Latin for "delicate mother"). The surface blood vessels and those that penetrate the brain and spinal cord are encased in this protective tissue layer. A second, very delicate, nonvascular, and waterproof layer, called the *arachnoid*, encloses the entire CNS. The arachnoid layer is named for its spider web appearance. The CSF is contained in the subarachnoid space. Immediately



■ FIGURE 36-22 ■ The limbic system includes the limbic cortex (cingulated gyrus, parahippocampal gyrus, uncus) and associated subcortical structures (thalamus, hypothalamus, amygdala).

outside the arachnoid is a continuous sheath of strong connective tissue, the *dura mater* (*i.e.*, "tough mother"), which provides the major protection for the brain and spinal cord. The cranial dura often splits into two layers, with the outer layer serving as the periosteum of the inner surface of the skull.

The inner layer of the dura forms two major folds. The first, a longitudinal fold called the *falx cerebri*, separates the cerebral hemispheres and fuses with a second transverse fold, called the *tentorium cerebelli* (Fig. 36-24). The tentorium cerebelli separates the anterior and middle depression in the skull (cranial fossae), which contains the cerebral hemispheres, from the posterior fossa, found interiorly and containing the brain stem and cerebellum. A semicircular gap, or incisura, is formed at the midline to permit the midbrain to pass forward from the posterior fossa.

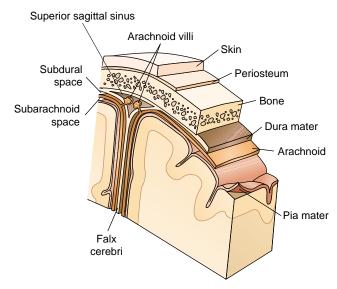


FIGURE 36-23 The cranial meninges. Arachnoid villi, shown within the superior sagittal sinus, are one site of cerebrospinal fluid absorption into the blood.

■ FIGURE 36-24 ■ Cranial dura mater. The skull is open to show the falx cerebri and the tentorium cerebelli, as well as some of the cranial venous sinuses.

Ventricular System and Cerebrospinal Fluid

The ventricular system is a series of CSF-filled cavities in the brain (Fig. 36-25). The CSF provides a supporting and protective fluid in which the brain and spinal cord float. CSF helps maintain a constant ionic environment that serves as a medium for diffusion of nutrients, electrolytes, and metabolic end-products into the extracellular fluid surrounding CNS neurons and glia. Filling the ventricles, the CSF supports the mass of the brain. Because it fills the subarachnoid space surrounding the CNS, a physical force delivered to either the skull or spine is to some extent diffused and cushioned.

The lining of the ventricles and central canal of the spinal cord is called the *ependyma*. There is a tremendous expansion of the ependyma in the roof of the lateral, third and fourth ventricles. The CSF is produced by tiny reddish masses of specialized capillaries from the pia mater, called the *choroid plexus*, that project into the ventricles. CSF is an ultrafiltrate of blood plasma, composed of 99% water with other constituents, mak-

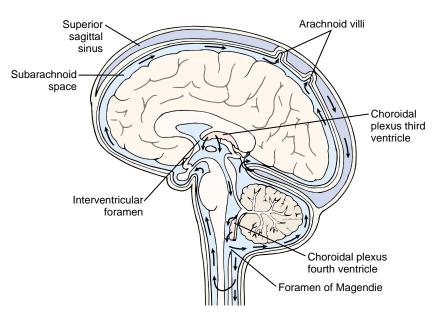
ing it close to the composition of the brain extracellular fluid. Humans secrete approximately 500 mL of CSF each day. However, only approximately 150 mL is in the ventricular system at any one time, meaning that the CSF is continuously being absorbed.

Once produced, the CSF flows freely through the ventricles. Three openings, or foramina, allow the CSF to pass into the subarachnoid space. Two of these, the foramina of Luschka, are located at the lateral corners of the fourth ventricle. The third, the medial foramen of Magendie, is in the midline at the caudal end of the fourth ventricle (see Fig. 36-25). Approximately 30% of the CSF passes down into the subarachnoid space that surrounds the spinal cord, mainly on its dorsal surface, and moves back up to the cranial cavity along its ventral surface.

Reabsorption of CSF into the vascular system occurs along the sides of the superior sagittal sinus in the anterior and middle fossa. Here, the waterproof arachnoid has protuberances, the *arachnoid villi*, that penetrate the inner dura and venous walls of the superior sagittal sinus. The arachnoid villi function as one-way valves, permitting CSF outflow into the blood but not allowing blood to pass into the arachnoid spaces.

Blood-Brain and Cerebrospinal Fluid–Brain Barriers

Maintenance of a chemically stable environment is essential to the function of the brain. In most regions of the body, extracellular fluid undergoes small fluctuations in pH and concentrations of hormones, amino acids, and potassium ions during routine daily activities such as eating and exercising. If the brain were to undergo such fluctuations, the result would be uncontrolled neural activity because some substances such as amino acids act as neurotransmitters, and ions such as potassium influence the threshold for neural firing. Two barriers, the blood-brain barrier and the CSF-brain barrier, provide the means for maintaining the stable chemical environment of the brain. Only water, carbon dioxide, and oxygen enter the brain with relative ease; the transport of other substances between the brain and the blood is slow.

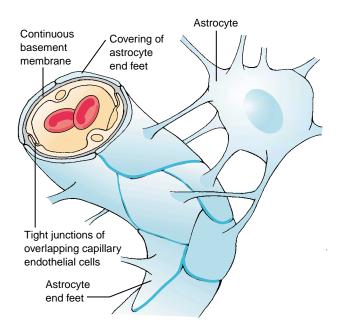


■ FIGURE 36-25 ■ The flow of cerebrospinal fluid from the time of its formation from blood in the choroid plexuses until its return to the blood in the superior sagittal sinus. Plexuses in the lateral ventricles are not illustrated.

Blood-Brain Barrier. The blood-brain barrier depends on the unique characteristics of the brain capillaries. The endothelial cells of brain capillaries are joined by continuous tight junctions. In addition, most brain capillaries are completely surrounded by a basement membrane and by the processes of supporting cells of the brain, called *astrocytes* (Fig. 36-26). The blood-brain barrier permits passage of essential substances while excluding unwanted materials. Reverse transport systems remove materials from the brain. Large molecules such as proteins and peptides are largely excluded from crossing the blood-brain barrier. Acute cerebral lesions, such as trauma and infection, increase the permeability of the blood-brain barrier and alter brain concentrations of proteins, water, and electrolytes.

The blood-brain barrier prevents many drugs from entering the brain. Most highly water-soluble compounds are excluded from the brain, especially molecules with high ionic charge, such as many of the catecholamines. In contrast, many lipidsoluble molecules cross the lipid layers of the blood-brain barrier with ease. Some drugs, such as the antibiotic chloramphenicol, are highly lipid soluble and therefore enter the brain readily. Other medications have a low solubility in lipids and enter the brain slowly or not at all. Alcohol, nicotine, and heroin are very lipid soluble and rapidly enter the brain. Some substances that enter the capillary endothelium are converted by metabolic processes to a chemical form incapable of moving into the brain.

The cerebral capillaries are much more permeable at birth than in adulthood, and the blood-brain barrier develops during the early years of life. In severely jaundiced infants, bilirubin can cross the immature blood-brain barrier, producing kernicterus and brain damage (see Chapter 15). In adults, the mature blood-brain barrier prevents bilirubin from entering the brain, and the nervous system is not affected.



■ FIGURE 36-26 ■ The three components of the blood-brain barrier: the astrocyte and astrocytic feet that encircle the capillary, the capillary basement membrane, and the tight junctions that join the overlapping capillary endothelial cells.

Cerebrospinal Fluid–Brain Barrier. The ependymal cells covering the choroid plexus are linked together by tight junctions, forming a blood-CSF barrier to diffusion of many molecules from the blood plasma of choroid plexus capillaries to the CSF. Water is transported through the choroid epithelial cells by osmosis. Oxygen and carbon dioxide move into the CSF by diffusion, resulting in partial pressures roughly equal to those of plasma. The high sodium and low potassium contents of the CSF are actively regulated and kept relatively constant. Lipids and nonpeptide hormones diffuse through the barrier rather easily, but most large molecules, such as proteins, peptides, many antibiotics, and other medications, do not normally get through. Many substances such as proteins; sodium ions; a number of micronutrients such as vitamins C, B_6 (pyridoxine), and folate are actively secreted into the CSF by the choroid epithelium. Because the resultant CSF has a relatively high sodium content, the negatively charged chloride and bicarbonate diffuse into the CSF along an ionic gradient. The choroid cells also generate bicarbonate from carbon dioxide in the blood. The generation of bicarbonate is important to the regulation of the pH of the CSF.

Mechanisms exist that facilitate the transport of other molecules such as glucose without energy expenditure. Ammonia, a toxic metabolite of neuronal activity, is converted to glutamine by astrocytes. Glutamine moves by facilitated diffusion through the choroid epithelium into the plasma. This exemplifies a major function of the CSF, that of providing a means of removal of toxic waste products from the CNS. Because the brain and spinal cord have no lymphatic channels, the CSF serves this function.

There are several specific areas of the brain where the blood-CSF barrier does not exist. One area is at the caudal end of the fourth ventricle, where specialized receptors for the carbon dioxide level of the CSF influence respiratory function. Another area consists of the walls of the third ventricle, which permit hypothalamic neurons to monitor blood glucose levels. This mechanism permits hypothalamic centers to respond to these blood glucose levels, contributing to hunger and eating behaviors.

In summary, in the adult, the spinal cord is in the upper two thirds of the spinal canal of the vertebral column. On transverse section, the spinal cord has an oval shape, and the internal gray matter has the appearance of a butterfly or letter "H." The dorsal horns contain the IA neurons and receive afferent information from dorsal root and other connecting neurons. The ventral horns contain the OA neurons and efferent LMNs that leave the cord by the ventral roots. Thirty-two pairs of spinal nerves (i.e., 8 cervical, 12 thoracic, 5 lumbar, 5 sacral, and 2 or more coccygeal) are present. Each pair communicates with its corresponding body segments. The spinal nerves and the blood vessels that supply the spinal cord enter the spinal canal through an intervertebral foramen. After entering the foramen, they divide into two branches, or roots, one of which enters the dorsolateral surface of the cord (i.e., dorsal root), carrying the axons of afferent neurons into the CNS. The other root leaves the ventrolateral surface of the cord (i.e., ventral root), carrying the axons of efferent neurons into the periphery. These two roots fuse at the intervertebral foramen, forming the mixed spinal nerve.

The brain can be divided into three regions: the hindbrain, the midbrain, and the forebrain. The hindbrain, consisting of the medulla oblongata, pons, and cerebellum, contains the neuronal circuits for the eating, breathing, and locomotive functions required for survival. Cranial nerves XII, XI, X, IX, VIII, VII, VII, and V are located in the hindbrain. The midbrain contains cranial nerves III and IV.

The forebrain is the most rostral part of the brain; it consists of the diencephalon and the telencephalon. The diencephalon forms the core of the forebrain, and telencephalon forms the cerebral hemispheres. The dorsal horn part of the diencephalon contains the thalamus and subthalamus, and the ventral horn contains the hypothalamus. All sensory pathways have direct projections to the thalamic nuclei, which convey the information to restricted parts of the sensory cortex. The hypothalamus functions in the homeostatic control of the internal environment. The cerebral hemispheres are the lateral outgrowths of the diencephalon and are arbitrarily divided into lobes-the frontal, parietal, temporal, and occipital lobes. The prefrontal premotor area and primary motor cortex are located in the frontal lobe; the primary sensory cortex and somatosensory association area are in the parietal cortex; the primary auditory cortex and the auditory association area are in the temporal lobe; and the primary visual cortex and association visual cortex are in the occipital lobe. The limbic system, which is involved in emotional experience and release of emotional behaviors, is located in the medial aspect of the cerebrum.

The brain is enclosed and protected by the pia mater, arachnoid, and dura mater. The protective CSF in which the brain and spinal cord float isolates them from minor and moderate trauma. The CSF is secreted into the ventricles, circulates through the ventricular system, passes outside to surround the brain, and is reabsorbed into the venous system through the arachnoid villi. The CSF-brain barrier and the blood-brain barrier protect the brain from substances in the blood that would disrupt brain function.

THE AUTONOMIC NERVOUS SYSTEM

The ability to maintain homeostasis and perform the activities of daily living in an ever-changing physical environment is largely vested in the autonomic nervous system (ANS). The ANS functions at the subconscious level and is involved in regulating, adjusting, and coordinating vital visceral functions such as blood pressure and blood flow, body temperature, respiration, digestion, metabolism, and elimination. The ANS is strongly affected by emotional influences and is involved in many of the expressive aspects of behavior. Blushing, pallor, palpitations of the heart, clammy hands, and dry mouth are several emotional expressions that are mediated through the ANS.

As with the somatic nervous system, the ANS is represented in both the CNS and the PNS. Traditionally, the ANS has been defined as a general efferent system innervating visceral organs. The efferent outflow from the ANS has two divisions: the sympathetic nervous system and the parasympathetic nervous system. The afferent input to the ANS is pro-

KEY CONCEPTS

THE AUTONOMIC NERVOUS SYSTEM (ANS)

- The ANS functions at the subconscious level and is responsible for maintaining homeostatic functions of the body.
- The ANS has two divisions: the sympathetic and parasympathetic systems. Although the two divisions function in concert, they are generally viewed as having opposite and antagonistic actions.
- The sympathetic nervous system functions in maintaining vital functions and responding when there is a critical threat to the integrity of the individual the "fight-or-flight" response.
- The parasympathetic nervous system functions in conserving energy, resource replenishment, and maintenance of organ function during periods of minimal activity.

vided by visceral afferent neurons, usually not considered to be part of the ANS.

The functions of the sympathetic nervous system include maintaining body temperature and adjusting blood flow and blood pressure to meet the changing needs of the body that occur with activities of daily living, such as moving from the supine to the standing position. The sympathoadrenal system also can discharge as a unit when there is a critical threat to the integrity of the individual-the so-called fight-or-flight response. During a stress situation, the heart rate accelerates; the blood pressure rises; blood flow shifts from the skin and gastrointestinal tract to the skeletal muscles and brain; blood sugar increases; the bronchioles and pupils dilate; the sphincters of the stomach and intestine and the internal sphincter of the urethra constrict; and the rate of secretion of exocrine glands that are involved in digestion diminishes. Emergency situations often require vasoconstriction and shunting of blood away from the skin and into the muscles and brain, a mechanism that, should a wound occur, provides for a reduction in blood flow and preservation of vital functions needed for survival. Sympathetic function often is summarized as catabolic in that its actions predominate during periods of pronounced energy expenditure, such as when survival is threatened.

In contrast to the sympathetic nervous system, the functions of the parasympathetic nervous system are concerned with conservation of energy, resource replenishment and storage (*i.e.*, anabolism), and maintenance of organ function during periods of minimal activity. The parasympathetic nervous system slows heart rate, stimulates gastrointestinal function and related glandular secretion, promotes bowel and bladder elimination, and contracts the pupil, protecting the retina from excessive light during periods when visual function is not vital to survival. The two divisions of the ANS usually are viewed as having opposite and antagonistic actions (*i.e.*, if one activates, the other inhibits a function). Exceptions are functions, such as sweating and regulation of arteriolar blood vessel diameter, that are controlled by a single division of the ANS, in this case the sympathetic nervous system.

The sympathetic and parasympathetic nervous systems are continually active. The effect of this continual or basal (baseline) activity is referred to as tone. The tone of an effector organ or system can be increased or decreased and usually is regulated by a single division of the ANS. For example, vascular smooth muscle tone is controlled by the sympathetic nervous system. Increased sympathetic activity produces local vasoconstriction from increased vascular smooth muscle tone, and decreased activity results in vasodilatation caused by decreased tone. In structures such as the sinoatrial node and atrioventricular node of the heart, which are innervated by both divisions of the ANS, one division predominates in controlling tone. In this case, the tonically active parasympathetic nervous system exerts a constraining or braking effect on heart rate, and when parasympathetic outflow is withdrawn, similar to releasing a brake, the heart rate increases. The increase in heart rate that occurs with vagal withdrawal can be further augmented by sympathetic stimulation.

Autonomic Efferent Pathways

The outflow of both divisions of the ANS follows a two-neuron pathway. The first motoneuron, called the *preganglionic neuron*, lies in the intermediolateral cell column in the ventral horn of the spinal cord or its equivalent location in the brain stem. The second motoneuron, called the *postganglionic neuron*, synapses with a preganglionic neuron in an autonomic ganglion located in the PNS. The two divisions of the ANS differ in terms of location of preganglionic cell bodies, relative length of preganglionic fibers, general function, nature of peripheral responses, and preganglionic and postganglionic neuromediators (Table 36-3). This two-neuron outflow pathway and the interneurons in the autonomic ganglia that add further modulation to ANS function are features distinctly different from the arrangement in somatic motor innervation.

Most visceral organs are innervated by both sympathetic and parasympathetic fibers. Exceptions include structures such

as blood vessels and sweat glands that have input from only one division of the ANS. The fibers of the sympathetic nervous system are distributed to effectors throughout the body, and as a result, sympathetic actions tend to be more diffuse than those of the parasympathetic nervous system, in which there is a more localized distribution of fibers. The preganglionic fibers of the sympathetic nervous system may traverse a considerable distance and pass through several ganglia before synapsing with postganglionic neurons, and their terminals make contact with a large number of postganglionic fibers. In some ganglia, the ratio of preganglionic to postganglionic cells may be 1:20; because of this, the effects of sympathetic stimulation are diffuse. There is considerable overlap, and one ganglion cell may be supplied by several preganglionic fibers. In contrast to the sympathetic nervous system, the parasympathetic nervous system has its postganglionic neurons located very near or in the organ of innervation. Because the ratio of preganglionic to postganglionic communication often is 1:1, the effects of the parasympathetic nervous system are much more circumscribed.

Sympathetic Nervous System

The preganglionic neurons of the sympathetic nervous system are located primarily in the thoracic and upper lumbar segments (T1 to L2) of the spinal cord; thus, the sympathetic nervous system often is referred to as the thoracolumbar division of the ANS. These preganglionic neurons, which are located primarily in the ventral horn intermediolateral cell column, have axons that are largely myelinated and relatively short. The postganglionic neurons of the sympathetic nervous system are located in the paravertebral ganglia of the sympathetic chain that lie on either side of the vertebral column, or in prevertebral sympathetic ganglia such as the celiac ganglia (Fig. 36-27). In addition to postganglionic efferent neurons, the sympathetic ganglia contain neurons of the internuncial, short-axon type, similar to those associated with complex circuitry in the brain and spinal cord. Many of these inhibit and others modulate preganglionic-to-postganglionic transmission.

The axons of the preganglionic neurons leave the spinal cord through the ventral root of the spinal nerves (T1 to L2),

TABLE 36-3 Charao

Characteristics of the Sympathetic and Parasympathetic

Nervous Systems		
Characteristic	Sympathetic Outflow	Parasympathetic Outflow
Location of preganglionic cell bodies	T1–T12, L1 and L2	Cranial nerves: III, VII (intermedius), IX, X; sacral segments 2, 3, and 4
Relative length of preganglionic fibers	Short—to paravertebral chain of ganglia or to aortic prevertebral of ganglia	Long—to ganglion cells near or in the innervated organ
General function	Catabolic—mobilizes resources in anticipation of chal- lenge for survival (preparation for "fight-or-flight" response)	Anabolic—concerned with conser- vation, renewal, and storage of resources
Nature of peripheral response	Generalized	Localized
Transmitter between preganglionic ter- minals and postganglionic neurons	ACh	ACh
Transmitter of postganglionic neuron	ACh (sweat glands and skeletal muscle vasodilator fibers); norepinephrine (most synapses); norepineph- rine and epinephrine (secreted by adrenal gland)	ACh

ACh, acetylcholine.



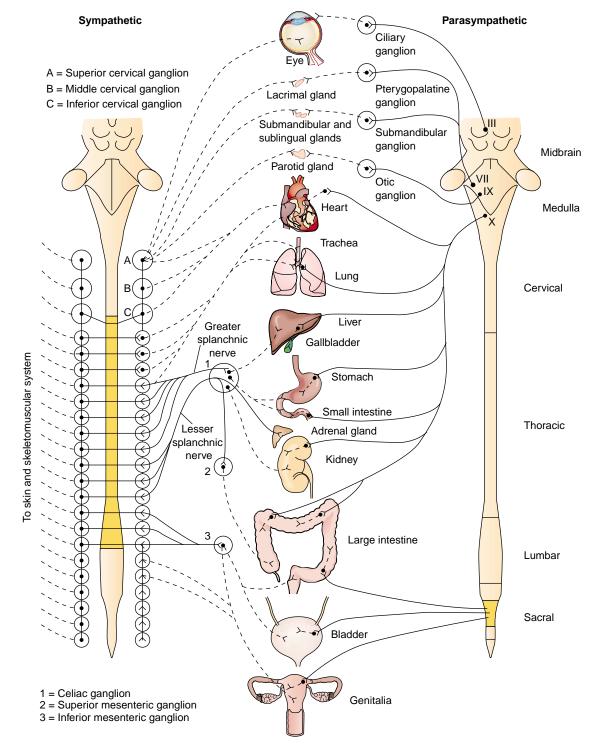
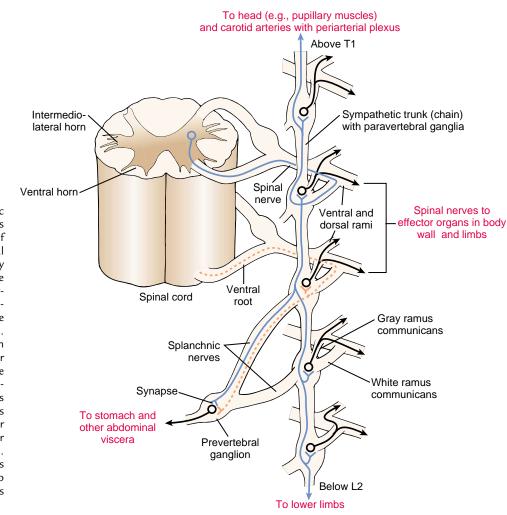


FIGURE 36-27 The autonomic nervous system. The involuntary organs are depicted with their parasympathetic innervation (craniosacral) indicated on the right and sympathetic innervation (thoracolumbar) on the left. Preganglionic fibers are *solid lines*; postganglionic fibers are *dashed lines*. For purposes of illustration, the sympathetic outflow to the skin and skeletomuscular system is shown separately (to the far left); effectors include sweat glands, pilomotor muscles and blood vessels of the skin, and blood vessels of the skeletal muscles and bones. (Modified from Hemer L. [1983]. *The human brain and spinal cord: Functional neuroanatomy and dissection guide.* New York: Springer-Verlag)



■ FIGURE 36-28 ■ Sympathetic pathways. Sympathetic fibers leave the spinal cord by way of the ventral root of the spinal nerves, enter the ventral primary rami, and pass through the white rami to the prevertebral or paravertebral ganglia of the sympathetic chain, where they synapse with postganglionic neurons. Some postganglionic fibers from the paravertebral ganglia reenter the segmental nerves through the gray rami and are then distributed in the spinal nerve branches that innervate the effector organs (eg., sweat glands and arrector pili muscles of skin and vascular smooth muscle of blood vessels). Other postganglionic neurons (dotted lines) travel directly to their destination in the various effector organs.

enter the ventral primary rami, and leave the spinal nerve through white rami of the rami communicantes to reach the paravertebral ganglionic chain (Fig. 36-28). In the sympathetic chain of ganglia, preganglionic fibers may synapse with neurons of the ganglion it enters, pass up or down the chain and synapse with one or more ganglia, or pass through the chain and move outward through a splanchnic nerve to terminate in one of the prevertebral ganglia (*i.e.*, celiac, superior mesenteric, or inferior mesenteric) that are scattered along the dorsal aorta and its branches. The adrenal medulla, which is part of the sympathetic nervous system, contains postganglionic sympathetic neurons that secrete sympathetic neurotransmitters directly into the bloodstream.

Parasympathetic Nervous System

The preganglionic fibers of the parasympathetic nervous system, also referred to as the *craniosacral division* of the ANS, originate in some segments of the brain stem and sacral segments of the spinal cord (see Fig. 36-27). The central regions of origin are the midbrain, pons, medulla oblongata, and the sacral part of the spinal cord. Outflow from the midbrain passes through the oculomotor nerve (cranial nerve III) to supply the pupillary sphincter muscle of each eye and the ciliary muscles that control lens thickness for accommodation. Caudal pontine outflow comes from branches of the facial nerve (cranial nerve VII) that supply the lacrimal and nasal glands. The medullary outflow develops from cranial nerves VII, IX, and X. Fibers in the glossopharyngeal nerve (cranial nerve IX) supply the parotid salivary glands. Approximately 75% of parasympathetic efferent fibers are carried in the vagus nerve (cranial nerve X). The vagus nerve provides parasympathetic innervation for the heart, trachea, lungs, esophagus, stomach, small intestine, proximal half of the colon, liver, gallbladder, pancreas, kidneys, and upper portions of the ureters.

Sacral preganglionic axons leave the S2 to S4 segmental nerves by gathering into the pelvic nerves. The pelvic nerves leave the sacral plexus on each side of the cord and distribute their peripheral fibers to the bladder, uterus, urethra, prostate, distal portion of the transverse colon, descending colon, and rectum. The sacral parasympathetic fibers also supply the venous outflow from the external genitalia to facilitate erectile function.

With the exception of cranial nerves III, VII, and IX, which synapse in discrete ganglia, the long parasympathetic preganglionic fibers pass uninterrupted to short postganglionic fibers located in the organ wall. In the walls of these organs, postganglionic neurons send axons to smooth muscle and glandular cells that modulate their functions.

664 Unit Ten: Alterations in the Nervous System

The gastrointestinal tract has its own intrinsic network of ganglionic cells located between the smooth muscle layers, called the *enteric* (or *intramural*) *plexus*, which controls local peristaltic movements and secretory functions. This network of parasympathetic postganglionic neurons and interneurons runs from the upper portion of the esophagus to the internal anal sphincter. Local afferent sensory neurons respond to mechanical and chemical stimuli and communicate these influences to motor fibers in the enteric plexus. The number of neurons in the enteric neural network (10^8) is so large that it approximates that of the spinal cord. It is thought that this enteric nervous system is capable of independent function without control from CNS fibers. The CNS has a modulating role, by way of preganglionic innervation of the plexus, converting local peristalsis to longer-distance movements, thereby speeding the transit of intestinal contents.

Central Integrative Pathways

General visceral afferent fibers accompany the sympathetic and parasympathetic outflow into the spinal and cranial nerves, bringing chemoreceptor, pressure, and nociceptive information from organs of the viscera to the brain stem, thoracolumbar cord, and sacral cord. Local reflex circuits relating visceral afferent and autonomic efferent activity are integrated into a hierarchic control system in the spinal cord and brain stem. Progressively greater complexity in the responses and greater precision in their control occur at each higher level of the nervous system. Most visceral reflexes contain contributions from the LMNs that innervate skeletal muscles as part of their response patterns.

For most autonomic-mediated functions, the hypothalamus serves as the major control center. The hypothalamus, which has connections with the cerebral cortex, the limbic system, and the pituitary gland, is in a prime position to receive, integrate, and transmit information to other areas of the nervous system. The neurons concerned with thermoregulation, thirst, and feeding behaviors are found in the hypothalamus. The hypothalamus also is the site for integrating neuroendocrine function. Hypothalamic releasing and inhibiting hormones control the secretion of anterior pituitary hormones (see Chapter 30).

The organization of many life-support reflexes occurs in the reticular formation of the medulla and pons. These areas of reflex circuitry, often called *centers*, produce complex combinations of autonomic and somatic efferent functions required for the cough, sneeze, swallow, and vomit reflexes, as well as for the more purely autonomic control of the cardiovascular system. At the hypothalamic level, these reflexes are integrated into more general response patterns, such as rage, defensive behavior, eating, drinking, voiding, and sexual function. Forebrain and especially limbic system control of these behaviors involves inhibiting or facilitating release of the response patterns according to social pressures during learned emotion-provoking situations.

Reflex adjustments of cardiovascular and respiratory function occur at the level of the brain stem. A prominent example is the carotid sinus baroreflex (see Chapter 14). One of the striking features of ANS function is the rapidity and intensity with which it can change visceral function. Within 3 to 5 seconds, it can increase heart rate to approximately twice its resting level. Bronchial smooth muscle tone is largely controlled by parasympathetic fibers carried in the vagus nerve. These nerves produce mild to moderate constriction of the bronchioles.

Other important ANS reflexes are located at the level of the spinal cord. As with other spinal reflexes, these reflexes are modulated by input from higher centers. When there is loss of communication between the higher centers and the spinal reflexes, as occurs in spinal cord injury, these reflexes function in an unregulated manner (see Chapter 38).

Autonomic Neurotransmission

The generation and transmission of impulses in the ANS occur in the same manner as in the CNS. There are self-propagating action potentials with transmission of impulses across synapses and other tissue junctions by way of neurohumoral transmitters. The postganglionic fibers of the ANS form a diffuse neural plexus at the site of innervation. The membranes of the cells of many smooth muscle fibers are connected by conductive protoplasmic bridges, called *gap junctions*, that permit rapid conduction of impulses through whole sheets of smooth muscle, often in repeating waves of contraction. Autonomic neurotransmitters released near a limited portion of these fibers provide a modulating function extending to a large number of effector cells. The muscle layers of the gut and of the bladder wall are examples.

The main neurotransmitters of the autonomic nervous system are acetylcholine and the catecholamines, epinephrine and norepinephrine. Acetylcholine is released at all of the sites of preganglionic transmission in the autonomic ganglia of sympathetic and parasympathetic nerve fibers and at the sites of postganglionic transmission in parasympathetic nerve endings. It also is released at sympathetic nerve endings that innervate the sweat glands and cholinergic vasodilator fibers found in skeletal muscle. Norepinephrine is released at most sympathetic nerve endings. The adrenal medulla, which is a modified prevertebral sympathetic ganglion, produces epinephrine along with small amounts of norepinephrine. Dopamine, which is an intermediate compound in the synthesis of norepinephrine, also acts as a neurotransmitter. It is the principal inhibitory transmitter of internuncial neurons in the sympathetic ganglia. It also has vasodilator effects on renal, splanchnic, and coronary blood vessels when given intravenously and is sometimes used in the treatment of shock (see Chapter 18).

Acetylcholine and Cholinergic Receptors

Acetylcholine is synthesized in the cholinergic neurons from choline and acetyl coenzyme A (acetyl CoA). After acetylcholine is secreted by the cholinergic nerve endings, it is rapidly broken down by the enzyme acetylcholinesterase. The choline molecule is transported back into the nerve ending, where it is used again in the synthesis of acetylcholine.

Receptors that respond to acetylcholine are called *cholinergic receptors*. There are two types of cholinergic receptors: muscarinic and nicotinic. Muscarinic receptors are present on the innervational targets of postganglionic fibers of the parasympathetic nervous system and the sweat glands, which are innervated by the sympathetic nervous system. Nicotinic receptors are found in autonomic ganglia and the end plates of skeletal muscle. Acetylcholine has an excitatory effect on muscarinic and nicotinic receptors, except for those in the heart and lower esophagus, where it has an inhibitory effect. The drug atropine is an antimuscarinic or muscarinic cholinergic-blocking drug that prevents the action of acetylcholine at excitatory and inhibitory muscarinic receptor sites. Because it is a muscarinicblocking drug, it exerts little effect at nicotinic receptor sites.

Catecholamines and Adrenergic Receptors

The catecholamines, which include norepinephrine, epinephrine, and dopamine, are synthesized in the sympathetic nervous system. Synthesis of dopamine and norepinephrine begins in the axoplasm of sympathetic nerve terminals with the conversion of the amino acid tyrosine to dopa; dopa to dopamine; and dopamine to norepinephrine. In the adrenal gland an additional step takes place during which approximately 80% the norepinephrine is transformed into epinephrine.

Each of the steps in sympathetic neurotransmitter synthesis requires a different enzyme, and the type of neurotransmitter that is produced depends on the types of enzymes that are available in a nerve terminal. For example, the postganglionic sympathetic neurons that supply blood vessels have the needed enzymes for the synthesis of norepinephrine, whereas those in the adrenal medulla have the enzymes needed to convert norepinephrine into epinephrine. As the catecholamines are synthesized, they are stored in vesicles. The final step of norepinephrine synthesis occurs in these vesicles. When an action potential reaches an axon terminal, the neurotransmitter molecules are released from the storage vesicles. The storage vesicles provide a means for concentrated storage of the catecholamines and protect them from the cytoplasmic enzymes that degrade the neurotransmitters.

In addition to neuronal synthesis, there is a second major mechanism for replenishment of norepinephrine in sympathetic nerve terminals. This mechanism consists of the active reuptake of the released neurotransmitter into the nerve terminal. Between 50% and 80% of the norepinephrine that is released during an action potential is removed from the synaptic area by an active reuptake process. This process terminates the action of the neurotransmitter and allows it to be reused by the neuron. The remainder of the released catecholamines diffuses into the surrounding tissue fluids or is degraded by two special enzymes: catechol-O-methyltransferase, which is diffusely present in all tissues, and monoamine oxidase (MAO), which is found in the nerve endings themselves.

Catecholamines can cause excitation or inhibition of smooth muscle contraction, depending on the site, dose, and type of receptor present. The excitatory or inhibitory responses of organs to sympathetic neurotransmitters are mediated by interaction with special structures in the cell membrane called *receptors*. There are two types of sympathetic receptors: α and β receptors. In vascular smooth muscle, excitation of α receptors causes vasoconstriction, and excitation of β -adrenergic receptors causes vasodilatation. Constriction of blood vessels in the skin, kidneys, and splanchnic circulation is mediated by α -adrenergic receptors. The β -adrenergic receptors are most prevalent in the heart, the blood vessels of skeletal muscle, and the bronchioles.

 α -Adrenergic receptors have been further subdivided into α_1 and α_2 receptors, and β -adrenergic receptors into β_1 and β_2 receptors. β_1 -Adrenergic receptors are found primarily in the heart and can be selectively blocked by β_1 -receptor–blocking drugs. β_2 -Adrenergic receptors are found in the bronchioles and in other sites that have β -mediated functions. The α_1 receptors are found primarily in postsynaptic effector sites; they mediate responses in vascular smooth muscle. The α_2 receptors are mainly located presynaptically and can inhibit the release of norepinephrine from sympathetic nerve terminals. The α_2 receptors are abundant in the CNS and are thought to influence the central control of blood pressure.

In summary, the ANS regulates, adjusts, and coordinates the visceral functions of the body. The ANS, which is divided into the sympathetic and parasympathetic systems, is an efferent system. It receives its afferent input from visceral afferent neurons. The ANS has CNS and PNS components. The outflow of the sympathetic and parasympathetic nervous system follows a two-neuron pathway, which consists of a preganglionic neuron located in the CNS and a postganglionic neuron located outside the CNS. Sympathetic fibers leave the CNS at the thoracolumbar level, and the parasympathetic fibers leave at the craniosacral level.

In general, the sympathetic and parasympathetic nervous systems have opposing effects on visceral function—if one excites, the other inhibits. The hypothalamus serves as the major control center for most ANS functions; local reflex circuits relating visceral afferent and autonomic efferent activity are integrated in a hierarchic control system in the spinal cord and brain stem.

The main neurotransmitters for the ANS are acetylcholine, the catecholamines, epinephrine, and norepinephrine. Acetylcholine is the transmitter for all preganglionic neurons, for postganglionic parasympathetic neurons, and for selected postganglionic sympathetic neurons. The catecholamines are the neurotransmitters for most postganglionic sympathetic neurons. The ANS neurotransmitters exert their action through specialized cell surface receptors—cholinergic receptors that bind acetylcholine and adrenergic receptors are divided into nicotinic and muscarinic receptors, and adrenergic receptors are divided into α and β receptors.

REVIEW QUESTIONS

The nervous system functions in much the same way as a computer system. Using this analogy, describe the relationship between the PNS and CNS in terms of input, output, and computational functions.

The transmission of impulses in the nervous system occurs by way of chemical or electrical synapses. Compare the two types of synapses in terms of mechanism of impulse generation, speed of conduction, one-way conduction, and in the case of chemical synapses, the mechanism whereby the neurotransmitters are synthesized, stored, released, and inactivated.

The poliovirus produces paralysis of skeletal muscles but does not affect somatosensory or visceral function. Based on your knowledge of cell columns, what neurons are susceptible to attack by the virus.

The inner tract layer of the nervous system contains many of the neuronal processes for reflexes such as those involved in

666 Unit Ten: Alterations in the Nervous System

swallowing and vomiting. Explain the advantages of this arrangement.

State the structures innervated by general somatic afferent, special visceral afferent, general visceral afferent, special somatic afferent, general visceral efferent, pharyngeal efferent, and general somatic efferent neurons.

List the structures of the hindbrain, midbrain, and forebrain and describe their functions.

Name the cranial nerves and cite their location and function.

Describe the characteristics of the CSF and trace its passage through the ventricular system.

State the function of the autonomic nervous system.

Compare the anatomic location and functions of the sympathetic and parasympathetic nervous systems.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

BIBLIOGRAPHY

- Araque A., Parpura V., Sanzgiri R.P., et al. (1999). Tri-partite synapses: Glia, the unacknowledged partner. *Trends in Neuroscience* 22, 208–215.
- Atkins D.L. (1998). Synapses. Step 4: Exploration of the neuron. [On-line]. Available: http://gwis2.circ.gwu.edu/~atkins/ Neuroweb/synapse.html. Accessed August 10, 2001.
- Brodal P. (1998). The central nervous system: Structure and function (2nd ed.). New York: Oxford University Press.
- Dambska M., Wisniewski K.E. (1999). Normal and pathologic development of the human brain and spinal cord. London: John Libbey & Company.
- Gartner L.P., Hiatt J.L. (1997). *Color textbook of histology* (pp. 155–185). Philadelphia: W.B. Saunders.
- Guyton A.C., Hall J.E. (2001). Textbook of medical physiology (10th ed.). Philadelphia: W.B. Saunders.
- Haines D.E. (Ed.). (1997). Fundamental neuroscience (pp. 115–121, 126– 127, 146–148, 443–454). New York: Churchill Livingstone.
- Matthews G.G. (1998). Neurobiology: Molecules, cells, and systems. Malden, MA: Blackwell Science.

Moore K.L, Persaud T.V.N. (1998). The developing human: Clinically oriented embryology (6th ed., pp. 63–82, 451–490). Philadelphia: W.B. Saunders.

- Parent A. (1996). Carpenter's human neuroanatomy (9th ed., pp. 186–192, 268–292, 748–756). Baltimore: Williams & Wilkins.
- Zigmond M.J., Bloom E.F., Landis S.C., et al. (1999). Fundamental neuroscience. San Diego: Academic Press.

Alterations in Brain Function

Brain Injury

CHAPTER

37

Mechanisms of Brain Injury Hypoxia and Ischemia **Excitatory Amino Acid Injury Cerebral Edema Increased Intracranial Volume and Pressure Brain Herniation** Hydrocephalus Traumatic Head and Brain Injury **Skull Fractures Traumatic Brain Injury Manifestations of Global Brain Injury** Anatomic and Physiologic Basis of Consciousness **Reticular Activating System** Levels of Consciousness Other Manifestation of Deteriorating Brain Function **Brain Death Persistent Vegetative State** Cerebrovascular Disease **The Cerebral Circulation Regulation of Cerebral Blood Flow** Stroke (Brain Attack) **Ischemic Stroke** Hemorrhagic Stroke **Transient Ischemic Attacks** Acute Manifestations of Stroke **Diagnosis and Treatment** Aneurysmal Subarachnoid Hemorrhage Arteriovenous Malformations

Infections and Neoplasms Infections Meningitis **Encephalitis Neoplasms Types of Tumors Clinical Course Seizure Disorders** Provoked and Unprovoked Seizures Classification **Partial Seizures Generalized Seizures Unclassified Seizures Diagnosis and Treatment** Anticonvulsant Medications Surgical Therapy **Generalized Convulsive Status Epilepticus Dementias** Alzheimer's Disease Pathophysiology **Clinical Course Diagnosis and Treatment Other Types of Dementia** Vascular Dementia **Pick's Disease** Creutzfeldt-Jakob Disease Wernicke-Korsakoff Syndrome Huntington's Disease

A natomically and functionally the brain is the most complex structure in the body. It controls our ability to think, our awareness of things around us, and our interactions with the outside world. Signals to and from various parts of the body are controlled by very specific areas within the brain. This renders the brain much more vulnerable to focal lesions than other organs in the body. For example, an isolated renal infarct would not be expected to have a significant effect on kidney function; whereas, an infarct of comparable size in specific areas of the brain could produce complete paralysis on one side of the body.¹

BRAIN INJURY

The brain is protected from external forces by the rigid confines of the skull and the cushioning afforded by the cerebrospinal fluid (CSF). The metabolic stability required by its electrically active cells is maintained by a number of regulatory mechanisms, including the blood-brain barrier and autoregulatory mechanisms that ensure its blood supply. Nonetheless, the brain remains remarkably vulnerable to injury.

Mechanisms of Injury

Injury to brain tissue can result from a number of conditions, including trauma, tumors, stroke, and metabolic derangements. Brain damage resulting from these disorders involves several common pathways, including the effects of hypoxia and ischemia, cerebral edema, and injury caused by increased intracranial pressure. In many cases, the mechanisms of injury are interrelated.

Hypoxia and Ischemia

The brain relies on the ability of the cerebral circulation to deliver sufficient oxygen for its energy needs. Although the brain makes up only 2% of the body weight, it receives one sixth of the resting cardiac output and accounts for 20% of the oxygen consumption.^{1,2} By definition, *hypoxia* denotes a deprivation of oxygen with maintained blood flow and ischemia, a situation of greatly reduced or interrupted blood flow. The cellular effects of hypoxia and ischemia are quite different, and the brain tends to have different sensitivities to the two conditions. Hypoxia interferes with the delivery of oxygen, and ischemia interferes with the delivery of oxygen and glucose as well as the removal of metabolic wastes.

Hypoxia usually is seen in conditions such as exposure to reduced atmospheric pressure, carbon monoxide poisoning, severe anemia, and failure to oxygenate the blood. Contrary to popular belief, hypoxia is fairly well tolerated, particularly in situations of chronic hypoxia. Neurons are capable of substantial anaerobic metabolism and are fairly tolerant of pure hypoxia; it commonly produces euphoria, listlessness, drowsiness, and impaired problem solving. Unconsciousness and convulsions may occur when hypoxia is sudden and severe. However, the effects of severe hypoxia (*i.e.*, anoxia) on brain function seldom are seen because the condition rapidly leads to cardiac arrest and ischemia.

Ischemia can be global or focal. Focal ischemia involves a single area of the brain, as in stroke. Collateral circulation may provide low levels of blood flow during focal ischemia. The residual perfusion may provide sufficient substrates to maintain a low level of metabolic activity, preserving neuronal integrity.

Global ischemia occurs when blood flow is inadequate to meet the metabolic needs of the entire brain. In contrast to persons with focal ischemia, those with global ischemia have no collateral circulation during the ischemic event.^{1,3} The result is a spectrum of neurologic disorders. Unconsciousness occurs within seconds of severe global ischemia, such as that resulting from complete cessation of blood flow, as in cardiac arrest. If circulation is restored immediately, consciousness is regained quickly. However, if blood flow is not promptly restored, severe pathologic changes take place. Energy sources (*i.e.*, glucose and glycogen) are exhausted in 2 to 4 minutes, and cellular ATP stores are depleted in 4 to 5 minutes. Approximately 50% to 75% of the total energy requirement of neuronal tissue is spent on mechanisms for maintenance of ionic gradients across the cell membrane (*e.g.*, sodium-potassium pump), resulting in fluxes of sodium, potassium, and calcium ions⁴ (Table 37-1). Excessive influx of sodium results in neuronal and interstitial edema. The influx of calcium initiates a cascade of events, including release of intracellular and nuclear enzymes that cause cell destruction.

Within the brain, certain regions and cell populations are more susceptible than others to hypoxic-ischemic injury.¹ For example, neurons are more susceptible to injury than are the glial cells. Among the neurons, the pyramidal cells of the hippocampus, the Purkinje cells of the cerebellum, and the neurons of the globus pallidus of the basal ganglia are particularly sensitive to generalized ischemic-hypoxic injury. The reason for this selectivity is uncertain but appears to be related at least to some extent on local levels and metabolism of certain excitatory neurotransmitters such as glutamate.

Excitatory Amino Acid Injury

Glutamate is the principal excitatory neurotransmitter in the brain, and its interaction with specific receptors is responsible for many higher-order functions, including memory, cognition, movement, and sensation.⁵ Normally, extracellular concentrations of glutamate are tightly regulated, with excess amounts removed and actively transported into astrocytes and neurons. During prolonged ischemia, these transport mechanisms become immobilized, causing extracellular glutamate to accumulate. In the case of cell injury and death, intracellular glutamate is released from the damaged cells.

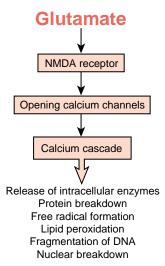
Many of the actions of glutamate are coupled with receptoroperated ion channels. One type of glutamate receptor, called the *glutamate N-methyl-D*-aspartate (NMDA) receptor, has been implicated in causing central nervous system (CNS) injury. The uncontrolled opening of NMDA receptor-operated channels produces an increase in intracellular calcium and leads to a series of calcium-mediated processes called the *calcium cascade* (Fig. 37-1). Activation of the calcium cascade leads to the release of intracellular enzymes that cause protein breakdown,

TABLE 37-1 Pathophysiologic

c of Imp

Cerebral Perfusion	
Consequences	Timing
Depletion of oxygen Depletion of glucose Conversion to anaerobic metabolism Exhaustion of cellular ATP Consequences Efflux of potassium Influx of sodium	10 sec 2–4 min 2–4 min 4–5 min
Influx of calcium	

(Adapted from Richmond T.S. [1997]. Cerebral resuscitation after global brain ischemia: Linking research to practice. AACN Clinical Issues 8 [2], 173)



■ FIGURE 37-1 ■ The role of the glutamate-NMDA receptor in brain cell injury.

free radical formation, lipid peroxidation, fragmentation of DNA, and nuclear breakdown. Drugs called neuroprotectants are being developed to interfere with the glutamate-NMDA pathway and thus reduce brain cell injury.

Cerebral Edema

Cerebral edema, or brain swelling, is an increase in tissue volume secondary to abnormal fluid accumulation. There are basically two types of brain edema: vasogenic or cytotoxic.1

Vasogenic Edema. Vasogenic edema results from an increase in the extracellular fluid that surrounds brain cells. It occurs with conditions such as tumors, prolonged ischemia, hemorrhage, brain injury, and infectious processes (e.g., meningitis) that impair the function of the blood-brain barrier and allow water and plasma proteins to leave the capillary and move into the interstitium. Vasogenic edema occurs primarily in the white matter of the brain, possibly because the white matter is more compliant than the gray matter and offers less resistance to fluid accumulation. Vasogenic edema can be localized, as in the case of abscesses or neoplasms, or it may be more generalized. The functional manifestations of vasogenic edema include focal neurologic deficits, disturbances in consciousness, and severe intracranial hypertension.

Cytotoxic Edema. Cytotoxic edema involves the swelling of brain cells. It involves an increase in fluid in the intracellular space, chiefly the gray matter, although the white matter may be involved. Cytotoxic edema can result from hypo-osmotic states, such as water intoxication or severe ischemia, that impair the function of the sodium-potassium membrane pump. This causes rapid accumulation of sodium in the cell, followed by movement of water along the osmotic gradient. Depending on the nature of the insult, cellular edema can occur in the vascular endothelium or smooth muscle cells, astrocytes, the myelin-forming processes of oligodendrocytes, or neurons. Major changes in cerebral function, such as stupor and coma, occur with cytotoxic edema. The edema associated with ischemia may be severe enough to produce cerebral infarction with necrosis of brain tissue.

Increased Intracranial Volume and Pressure

Increased intracranial pressure (ICP) is a common pathway for brain injury from different types of insults and agents. Excessive ICP can obstruct cerebral blood flow, destroy brain cells, displace brain tissue as in herniation, and otherwise damage delicate brain structures.

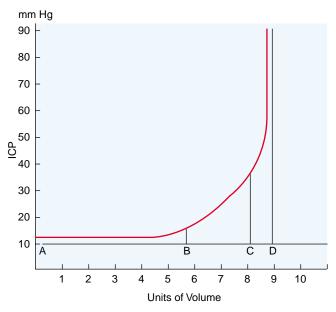
The cranial cavity contains blood (approximately 10%), brain tissue (approximately 80%), and CSF (approximately 10%) within the rigid confines of a nonexpandable skull.⁶ Each of these three volumes contributes to the ICP, which normally is maintained within a range of 0 to 15 mm Hg when measured in the lateral ventricles. The volumes of each of these components can vary slightly without causing marked changes in ICP. This is because small increases in the volume of one component can be compensated for by a decrease in the volume of one or both of the other two components.7 This association is called the Monro-Kellie hypothesis.

Abnormal variation in intracranial volume with subsequent changes in ICP can be caused by a volume change in any of the three intracranial components. For example, an increase in tissue volume can result from a brain tumor, brain edema, or bleeding into brain tissue. An increase in blood volume develops when there is vasodilatation of cerebral vessels or obstruction of venous outflow. Excess production, decreased absorption, or obstructed circulation of CSF affords the potential for an increase in the CSF component.

According to the modified Monro-Kellie hypothesis, reciprocal compensation occurs among the three intracranial compartments.6 Of the three intracranial volumes, the tissue volume is relatively restricted in its ability to undergo change. Initial increases in ICP are buffered by a translocation of CSF to the spinal subarachnoid space and increased reabsorption of CSF. The compensatory ability of the blood compartment is limited by the small amount of blood that is in the cerebral circulation. The cerebral blood vessels contain less than 10% of the intracranial volume, most of which is contained in the low-pressure venous system. As the volume-buffering capacity of this compartment becomes exhausted, venous pressure increases and cerebral blood volume and ICP rise. In addition, cerebral blood flow is highly controlled by autoregulatory mechanisms, which affect its compensatory capacity. Conditions such as ischemia and elevated carbon dioxide (PCO₂) levels produce a compensatory vasodilation of the cerebral blood vessels.

The impact of increases in blood, brain tissue, or CSF volumes on ICP varies among individuals and depends on the amount of increase that occurs, the effectiveness of compensatory mechanisms, and the compliance of brain tissue. Compliance represents the ratio of change in volume to the resulting change in pressure.⁶ As depicted in Figure 37-2 even small changes in volume produce large changes in ICP once the compensatory mechanisms have been exceeded.

The cerebral perfusion pressure (CPP), which represents the difference between the mean arterial blood pressure (MABP) and the ICP, is the pressure perfusing the brain.7 CPP is determined by the pressure gradient between the internal carotid artery and the subarachnoid veins. The MABP and ICP are monitored frequently in persons with brain conditions that increase ICP and impair brain perfusion. Normal CPP ranges from 70 to 100 mm Hg. Brain ischemia develops at levels



■ FIGURE 37-2 ■ Pressure-volume curve. From point A to just before B, the ICP remains constant although there is an addition of volume (compliance is high). At point B, even though the ICP is within normal limits, compliance begins to change, as evidenced by the slight rise in ICP. From points B to C, the ICP rises with an increase in volume (low compliance). From points C to D, ICP rises significantly with each minute increase in volume (compliance is lost). (Hickey J.V. [2003]. *Neurological and neurosurgical nursing* [5th ed., p. 286]. Philadelphia: Lippincott Williams & Wilkins)

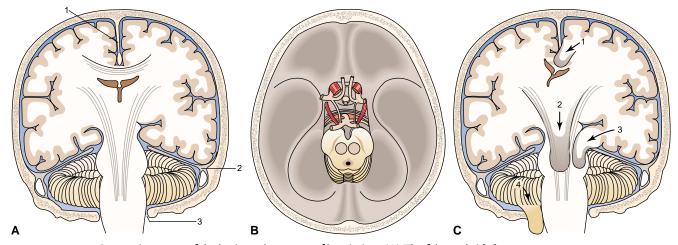
below 50 to 70 mm Hg.⁶ When the ICP approaches or exceeds the MABP, tissue perfusion becomes inadequate, cellular hypoxia results, and if the elevated pressure is maintained, neuronal death may occur. The highly specialized cortical neurons are the most sensitive to oxygen deficit; a decrease in the level of consciousness is one of the earliest and most reliable signs of increased ICP. One of the late reflexes seen with a marked increase in ICP is the CNS ischemic response, which is triggered by ischemia of the vasomotor center in the brain stem. Neurons in the vasomotor center respond directly to ischemia by producing a marked increase in MABP, sometimes to levels as high as 270 mm Hg, accompanied by a widening of the pulse pressure and reflex slowing of the heart rate. These three signs, sometimes called the *Cushing reflex*, are important but late indicators of increased ICP.² The ischemic reflex represents a "last-ditch" effort by the nervous system to maintain the cerebral circulation.

Brain Herniation

The brain is protected by the nonexpandable skull and supporting septa, the falx cerebri and the tentorium cerebelli, that divide the intracranial cavity into *fossae* or compartments that normally protect against excessive movement. The falx cerebri is a sickle-shaped septum that separates the two hemispheres. The tentorium is a tentlike structure, higher in the center than at the sides of the skull, which separates the occipital lobes of the brain from the cerebellum and much of the brain stem (Fig. 37-3A). It creates the area above the tentorium, the supratentorial space, and the area below the tentorium, the infratentorial space. Extending posteriorly into the center of the tentorium is a large semicircular opening called the *incisura* or *tentorial notch*. The brain stem, blood vessels, and accompanying nerves pass through the incisura (see Fig. 37-3B).

Brain herniation is displacement of brain tissue under the falx cerebri or through the incisura of the tentorium cerebelli. It occurs when the presence of cerebral edema or a mass results in shifting or herniation of brain tissue from a compartment of higher pressure to one of lower pressure. The different types of herniation syndromes are based on the area of the brain that has herniated and the structure under which it has been pushed (see Fig. 37-3C). They commonly are divided into two broad categories, supratentorial and infratentorial, based on whether they are located above or below the tentorium.

There are three major patterns of supratentorial herniation: cingulate or across the falx cerebri, uncal or lateral, and trans-



■ FIGURE 37-3 ■ Supporting septa of the brain and patterns of herniation. (A) The falx cerebri [1], tentorium cerebelli [2], foramen magnum [3]. (B) The location of the insicura or tentorial notch in relation to the cerebral arteries and oculomotor nerve. (C) Herniation of the cingulate gyrus under the falx cerebri [1], central or transtentorial herniation [2], herniation of the temporal lobe into the tentorial notch [3], and infratentorial herniation of the cerebellar tonsils [4]. (Courtesy of Carole Hilmer, C.M.I.)

tentorial or central.6 Cingulate herniation involves displacement of the cingulate gyrus and hemisphere beneath the sharp edges of the falx cerebri to the opposite side of the brain. Uncal herniation occurs when a lateral mass pushes the brain tissue centrally and forces the medial aspect of the temporal lobe, which contains the uncus and hippocampal gyrus, under the edge of the tentorial incisura, into the posterior fossa. Transtentorial or central herniation involves the downward displacement of the cerebral hemispheres, basal ganglia, diencephalon, and midbrain through the tentorial incisura.

Each supratentorial herniation syndrome has distinguishing features in the early phases, but as the forced downward displacement on the pons and medulla continues, clinical signs become similar (see Table 37-2). Any of the supratentorial herniation syndromes can compress vascular and CSF flow, which can further complicate the neurologic manifestations of brain lesions. The progressive downward displacement from any of the supraventricular herniation syndromes can result in brain stem herniation, in which the medulla herniates into the foramen magnum. Death is immediate, caused by compression of cardiorespiratory centers in the medulla.

Infratentorial herniation results from increased pressure in the infratentorial compartment. Herniation may occur superiorly (upward) through the tentorial incisura or inferiorly (downward) through the foramen magnum. Upward displacement of brain tissue can cause blockage of the aqueduct of Sylvius and lead to hydrocephalus and coma. Downward displacement of the midbrain through the tentorial notch or of the cerebellar tonsils through the foramen magnum can interfere with medullary functioning and cause cardiac or respiratory arrest. In cases of pre-existing ICP, herniation may occur when the pressure is released from below, such as in a lumbar puncture.

Hydrocephalus

Enlargement of the CSF compartment occurs with hydrocephalus, which is defined as an abnormal increase in CSF volume in any part or all of the ventricular system (see Chapter 36, Fig. 36-25). There are two types of hydrocephalus: noncom-

TABLE 37-2Key Structures andClinical Signs of Cingulate,Transtentorial, and Uncal Herniation			
Herniation Syndrome	Key Structures Involved	Key Clinical Signs	
Cingulate Transtentorial	Anterior cerebral artery Reticular activating system Corticospinal tract	Leg weakness Altered level of consciousness Decorticate posturing Rostral–caudal deterioration	
Uncal	Cerebral peduncle Oculomotor nerve Posterior cerebral artery Cerebellar tonsil Respiratory center	Hemiparesis Pupil dilatation Visual field loss Respiratory arrest	

municating and communicating. Hydrocephalus ex vacuo refers to dilation of the ventricular system and a compensatory increase in CSF volume secondary to a loss of brain tissue. It is commonly associated with other evidence of brain atrophy.^{1,8}

Noncommunicating or obstructive hydrocephalus occurs when obstruction in the ventricular system prevents the CSF from reaching the arachnoid villi. CSF flow can be obstructed by congenital malformations, tumors encroaching on the ventricular system, inflammation, or hemorrhage.8

Communicating hydrocephalus results from impaired reabsorption of CSF from the arachnoid villi into the venous system. Decreased absorption can result from a block in the CSF pathway to the arachnoid villi or a failure of the villi to transfer the CSF to the venous system. It can occur if too few villi are formed, if postinfective (meningitis) scarring occludes them, or if the villi become obstructed with fragments of blood or infectious debris. Normal-pressure hydrocephalus is an important type of communicating hydrocephalus seen in older adults. It is accompanied by ventricular enlargement with compression of cerebral tissue but normal CSF pressure.

Similar pathologic patterns occur with noncommunicating and communicating types of hydrocephalus. The cerebral hemispheres become enlarged, and the ventricular system is dilated behind the point of obstruction. The gyri on the surface of the brain tend to become less prominent as the sulci are compressed and the white matter is reduced in volume. The presence and extent of the increased ICP is determined by the amount of fluid accumulation and the type of hydrocephalus, the age at onset, and the rapidity and degree of pressure rise. Acute-onset hydrocephalus is usually marked by symptoms of increased ICP, including headache and vomiting, followed by edema of the optic disk (papilledema). If the obstruction is not relieved, mental deterioration eventually occurs. Slowly developing hydrocephalus is less likely to produce an increase in ICP, but it may produce deficits such as progressive dementia and gait changes.

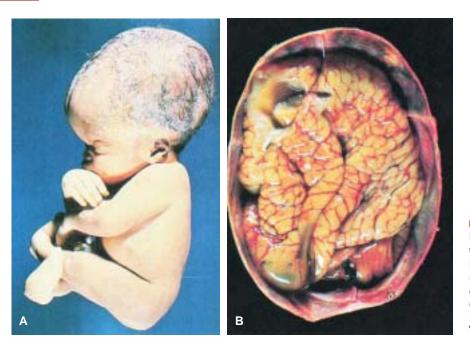
Computed tomographic (CT) scans are used to diagnose all types of hydrocephalus. Treatment depends on the cause of the disorder. In noncommunicating hydrocephalus, shunting procedures are used to provide an alternative route for return of CSF to the circulation. Treatment for communicating hydrocephalus includes attempts to clear the arachnoid villi of exudate; if this is unsuccessful, surgical shunting may be required.⁶

When hydrocephalus develops in utero or before the cranial sutures have fused in infancy, the ventricles expand beyond the point of obstruction, the cranial sutures separate, the head expands, and there is bulging of the fontanels (Fig. 37-4).9 Because the skull is able to expand, signs of increased ICP usually are absent. Surgical placement of a shunt (e.g., a ventriculoperitoneal shunt) that allows for diversion of excess CSF fluid is often used to prevent extreme enlargement of the head. The major complication of shunts is bacterial infection. Children with hydrocephalus are at risk for developmental disorders. Visual problems, including strabismus, visual field defects, and atrophy of the optic nerve, are common.9

Traumatic Head and Brain Injury

The term *head injury* is used to describe all structural damage to the head and has become synonymous with brain injury.^{10,11} In the United States, head injury is the leading cause of death

672 Unit Ten: Alterations in the Nervous System



■ FIGURE 37-4 ■ Congenital hydrocephalus. (A) Hydrocephalus occurring before the fusion of the cranial sutures causes pronounced enlargement of the head. (B) Removal of the calvarium demonstrates an atrophic and collapsed cerebral cortex. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 1454]. Philadelphia: Lippincott Williams & Wilkins)

among persons younger than 24 years. The main causes of head injury are road accidents, falls, and assaults, and the most common cause of fatal head injuries is road accidents involving vehicles and pedestrians.¹²

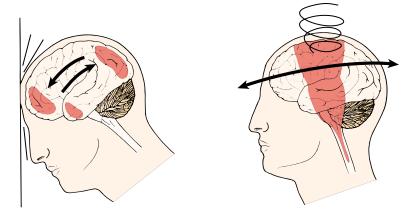
Skull Fractures

Although the skull affords protection for the tissues of the CNS, it also provides the potential for development of ischemic and traumatic brain injuries. This is because it cannot expand to accommodate the increase in volume that occurs when there is swelling or bleeding within its confines. The bony structures themselves can also cause injury to the nervous system. Fractures of the skull can compress sections of the nervous system, or they can splinter and cause penetrating wounds.

Skull fractures can be divided into three groups: simple, depressed, and basilar. A *simple* or *linear* skull fracture is a break in the continuity of bone. A *comminuted* skull fracture refers to a splintered or multiple fracture line. When bone fragments are embedded into the brain tissue, the fracture is said to be *depressed*. Radiologic examination usually is needed to confirm the presence and extent of a skull fracture.

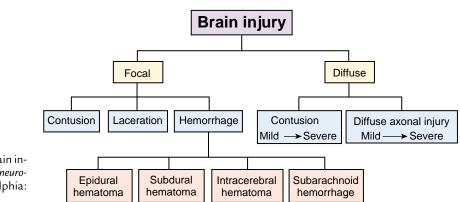
Traumatic Brain Injury

The skull and CSF provide protection for the brain, but they also can contribute to brain trauma at the time of head injury. The two mechanisms responsible for brain injury are acceleration-deceleration and concurrent rotational movement (Fig. 37-5). Because the brain floats freely in the CSF, blunt force to the head can cause the brain to accelerate in the skull and then abruptly decelerate on hitting the inner confines of the skull. At the time of impact to the skull, there is always some acceleration/ deceleration of the brain, whether the head is held in fixed or



■ FIGURE 37-5 ■ Mechanisms of (A) brain acceleration-deceleration injury and (B) rotational injury.

1 LINE



■ FIGURE 37-6 ■ Focal and generalized brain injuries. (Hickey J.V. [1996]. *Neurological and neurosurgical nursing* [4th ed., p. 393]. Philadelphia: Lippincott-Raven).

free position. This results in what are called *coup-contrecoup* injuries.⁶ The injury directly under the area of impact is called a *coup injury* and the injury sustained on the opposite pole of the brain, a *contrecoup injury*.

The effects of traumatic head injuries can be divided into two categories: primary or direct injuries, in which damage is caused by impact, and secondary injuries, in which damage results from the subsequent brain swelling, an intracranial hematoma, infection, or cerebral ischemia. Ischemia is considered to be the most common cause of secondary brain injury. It can cause the hypoxia and hypotension that occur during the resuscitation process or the impairment of regulatory mechanisms that maintain cerebral blood flow and oxygen supply.^{13,14}

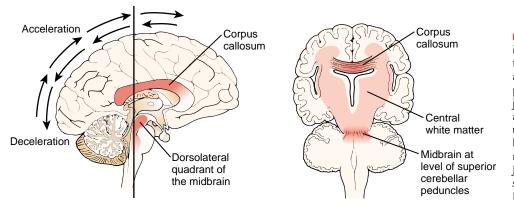
Even if there is no break in the skull, a blow to the head can cause severe and diffuse brain damage. Such closed injuries vary in severity and can be classified as focal or diffuse (Fig. 37-6). Diffuse injuries include concussion and diffuse axonal injury, and focal injuries, contusions, lacerations, and hemorrhage.

Concussion. The term *concussion* refers to a momentary interruption of brain function with or without loss of consciousness. In mild head injury, there may be momentary loss of consciousness without demonstrable neurologic symptoms or residual damage, except for possible residual amnesia. Microscopic changes usually can be detected in the neurons and supporting tissues within hours of injury. Although recovery usually takes place within 24 hours, mild symptoms, such as headache, irritability, insomnia, and poor concentration and memory, may persist for months. This is known as the *postconcussion syndrome*. Because these complaints are vague and subjective, they sometimes are regarded as being of psychological origin.

Diffuse Axonal Injury. Diffuse axonal injury is a primary injury with diffuse microscopic damage to axons in the cerebral hemisphere, corpus callosum, and brain stem. It is responsible for most cases of posttraumatic dementia and, in conjunction with hypoxic-ischemic injury, is the most common cause of persistent vegetative state. The lesions of diffuse axonal injury result from sudden deceleration and/or acceleration forces sufficient to stretch or, in extreme cases, tear nerve cell processes within the white matter of the brain (Fig. 37-7).

Contusion. A contusion is a bruise to the cortical surface of the brain caused by blunt head trauma.^{1,7} Contusions may be single or multiple and occur at any place where the brain comes in contact with the skull; they often occur when the moving head strikes a fixed object (Fig. 37-8). They are most common in the frontal poles, orbital surfaces of the frontal lobes, temporal poles, occipital poles, and posterior cerebellum.¹

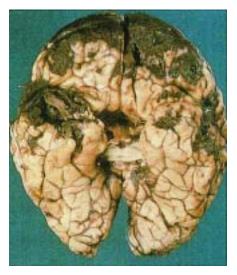
Contusions are generally the result of anteroposterior displacement, when the moving head strikes a fixed object. As the



■ FIGURE 37-7 ■ Diffuse axonal injury. Diffuse axonal injury results from acceleration-deceleration and shearing on the brain. Dependent upon the severity of injury, the areas of the brain most affected are the corpus callosum, the dorsolateral area of the midbrain, and the parasagittal white matter. Adapted from (Hickey J.V. [1996]. Neurological and neurosurgical nursing [4th ed., p. 386]. Philadelphia: Lippincott-Raven).

SHORT

Unit Ten: Alterations in the Nervous System



■ FIGURE 37-8 ■ Recent cerebral contusion. Multiple areas of hemorrhage mark the poles of the frontal and temporal lobes. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 1462]. Philadelphia: Lippincott Williams & Wilkins)

brain strikes the rough surface of the cranial vault, brain tissue, blood vessels, nerve tracts, and other structures are bruised and torn. The extent of brain damage that occurs is dependent upon the force causing the injury. If the force is minimal, the contusion is limited to the apex of the gyri. Greater forces destroy larger expanses of the cortex, creating deeper lesions that extend into the white matter or lacerate the cortex and initiate cortical and subcortical hemorrhages. Cerebral contusions, particularly those accompanied by tearing of the superficial layers of the brain, are an important cause of traumatic subarachnoid hemorrhage.

Contusions cause permanent damage to brain tissue.^{1,8} The bruised, necrotic tissue is phagocytized by macrophages, and scar tissue formed by astrocyte proliferation persists as a crater.

Hematomas. Hematomas result from vascular injury and bleeding. Depending on the anatomic position of the ruptured vessel, bleeding may involve the development of an epidural hematoma, the subdural hematoma, or an intracerebral hematoma (Fig. 37-9).

Epidural hematoma is one that develops between the inner table of the bones of the skull and the dura. It usually results from a tear in an artery, most often the middle meningeal, which is located under the thin temporal bone. Because bleeding is arterial in origin, rapid compression of the brain occurs from the expanding hematoma. Epidural hematoma is more common in a young person because the dura is not as firmly attached to the skull surface as it is in an older person. Typically, a person with an epidural hematoma presents with a history of head injury and a brief period of unconsciousness, followed by a lucid period in which consciousness is regained, followed by rapid progression to unconsciousness. The lucid interval does not always occur, but when it does, it is of great diagnostic value. With rapidly developing unconsciousness, there are focal symptoms related to the area of the brain involved. These symptoms can include ipsilateral (same side) pupil dilatation and contralateral (opposite side) hemiparesis. If the hematoma

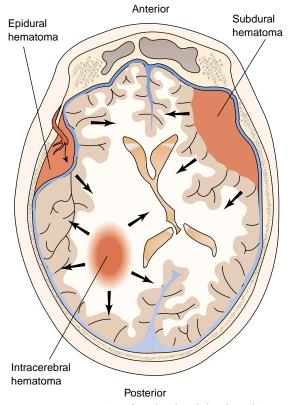


FIGURE 37-9 Location of epidural, subdural, and intracerebral hematomas.

is not removed, the condition progresses, with increased ICP, tentorial herniation, and death. However, prognosis is excellent if the hematoma is removed before loss of consciousness occurs.

A *subdural hematoma* develops in the area between the dura and the arachnoid (subdural space) and usually is the result of a tear in the small bridging veins that connect veins on the surface of the cortex to dural sinuses. These veins are readily snapped in head injury when the brain moves suddenly in relation to the skull. A subdural hematoma develops more slowly than an epidural hematoma because the tear is in the venous, rather than the arterial, system.

Subdural hematomas are classified as acute, subacute, or chronic. Symptoms of an acute subdural hematoma are seen within 24 hours of the injury. Acute subdural hematomas progress rapidly and are associated with a high mortality rate because of the severe secondary injuries related to edema and uncontrolled rise in ICP. Subacute hematomas do not produce symptoms until 2 to 10 days after injury. There may be a period of improvement in the level of consciousness and neurologic symptoms, followed by deterioration if the hematoma is not removed.

The symptoms of chronic subdural hematomas may not arise until several weeks after the injury, so much later that the person may not remember having had a head injury. This is especially true of the older person with fragile vessels whose brain has shrunk away from the dura. Seepage of blood into the subdural space may occur slowly. Because the blood in the subdural space is not absorbed, fibroblastic activity begins,



■ FIGURE 37-10 ■ Subdural hematoma. A chronic subdural hematoma is encapsulated by an outer membrane, evidenced as a narrow brown layer beneath the white dura. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 1460]. Philadelphia: Lippincott Williams & Wilkins)

and the hematoma becomes encapsulated⁸ (Fig. 37-10). Within this encapsulated area, the blood cells are slowly lysed, and a fluid with a high osmotic pressure is formed. This creates an osmotic gradient, with fluid from the surrounding subarachnoid space being pulled into the hematoma, causing the mass to increase in size and exert pressure on the surrounding intracranial contents. In some instances, the clinical picture is less defined, with the most prominent symptom being a decreasing level of consciousness indicated by drowsiness, confusion, and apathy. The person also may have headache.

Traumatic Intracerebral Hematomas. Traumatic intracerebral hematomas may be single or multiple. They can occur in any lobe of the brain but are most common in the frontal or temporal lobes. They may occur in association the severe motion that the brain undergoes during head injury, or a contusion can coalesce into a hematoma. Intracerebral hematomas occur more frequently in older persons and alcoholics whose brain vessels are more friable.

The signs and symptoms produced by an intracerebral hematoma depend on its size and location within the brain. Signs of increased ICP can be manifested if the hematoma is large and encroaching on vital structures. A hematoma in the temporal lobe can be dangerous because of the potential for lateral herniation.

Manifestations of Global Brain Injury

Global brain injury, whether caused by head trauma, stroke, or other pathologies, is manifested by alterations in sensory and motor function and by changes in the level of consciousness. In contrast to focal injury, which causes alterations in sensory function (Chapter 39) or motor function (Chapter 37), global injury tends to result in altered levels of consciousness. Severe injury that seriously compromises brain function may result in brain death.

The cerebral hemispheres are the most susceptible to damage, and the most common sign of brain dysfunction is altered level of consciousness and change in behavior. As the brain structures in the diencephalon, midbrain, pons, and medulla are affected, additional respiratory; pupillary and eye movement reflexes, and motor signs become evident (see Table 37-3). Hemodynamic and respiratory instability are the last signs to occur because their regulatory centers are located low in the medulla.

In progressive brain deterioration, the person's neurologic capabilities appear to deteriorate in stepwise fashion. Similarly, as neurologic function returns, there appears to be stepwise progress to higher levels of consciousness. Deterioration of brain function from supratentorial lesions tends to follow a rostral-to-caudal stepwise progression, which is observed as the brain initially compensates for injury and subsequently decompensates with loss of autoregulation and cerebral perfusion. Infratentorial (brain stem) lesions may lead to an early, sometimes abrupt disturbance in consciousness without any orderly rostrocaudal progression of neurologic signs.

Anatomic and Physiologic Basis of Consciousness

Consciousness is the state of awareness of self and the environment and of being able to become oriented to new stimuli.⁶ It has traditionally been divided into two components: (1) arousal and wakefulness and (2) content and cognition. The content and cognition aspects of consciousness are determined by a functioning cerebral cortex. Arousal and wakefulness requires the concurrent functioning of both cerebral hemispheres and an intact reticular activating system (RAS) in the brain stem.

Reticular Activating System

The reticular formation is a diffuse, primitive system of interlacing nerve cells and fibers in the brain stem that receive input from multiple sensory pathways (Fig. 37-11). Anatomically, the reticular formation constitutes the central core of the brain

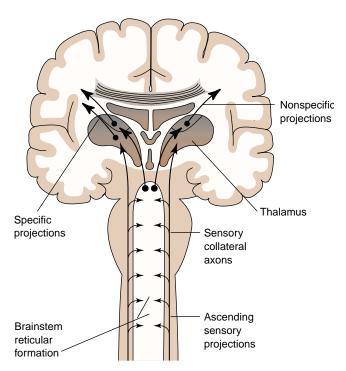
TABLE 37-3	Key Signs	in Rostral-to-Caudal Progression of Brain Lesions
Level of Brain	Injury	Key Clinical Signs
Diencephalon		Impaired consciousness (see Table 37-4); small, reactive pupils; intact oculocephalic re- flex; decorticate posturing; Cheyne-Stokes respirations
Midbrain		Coma, fixed, midsize pupils; impaired oculocephalic reflex; neurogenic hyperventilation; decerebrate posturing
Pons		Coma, fixed, irregular pupils; dysconjugate gaze; impaired cold caloric stimulation; loss of corneal reflex; hemiparesis/quadriparesis; decerebrate posturing; apneustic respirations
Medulla		Coma, fixed pupils, flaccidity, loss of gag and cough reflexes, ataxic/apneic respirations

KEY CONCEPTS

BRAIN INJURY AND LEVELS OF CONSCIOUSNESS

- Consciousness is a global function that depends on a diffuse neural network that includes activity of the reticular activating system (RAS) and both cerebral hemispheres.
- Impaired consciousness implies diffuse brain injury to the RAS at any level (medulla through thalamus) or both cerebral hemispheres simultaneously.
- In contrast, local brain injury causes focal neurologic deficit but does not disrupt consciousness.

stem, extending from the medulla through the pons to the midbrain, which is continuous caudally with the spinal cord and rostrally with the subthalamus, the hypothalamus, and the thalamus.¹⁵ Fibers from the RAS also project to the autonomic nervous system and motor systems. The hypothalamus plays a predominant role in maintaining homeostasis through integration of somatic, visceral, and endocrine functions. Inputs from the reticular formation, vestibulospinal projections, and other motor systems are integrated to provide a continuously



■ FIGURE 37-11 ■ The brain stem reticular formation and reticular activating system. Ascending sensory tracts send axon collateral fibers to the reticular formation. These give rise to fibers synapsing in the nonspecific nuclei of the thalamus. From there the nonspecific thalamic projections influence widespread areas of the cerebral cortex and limbic system. (Rhoades R.A., Tanner G.A. [1996]. *Medical physiology*. Boston: Little, Brown) adapting background of muscle tone and posture to facilitate voluntary motor actions. Reticular formation neurons that function in regulation of cardiovascular, respiratory, and other visceral functions are intermingled with those that maintain other reticular formation functions.

Ascending fibers of the reticular formation, known as the *ascending RAS*, transmit activating information to all parts of the cerebral cortex. The flow of information in the ascending RAS activates the hypothalamic and limbic structures that regulate emotional and behavioral responses such as those that occur in response to pain and loud noises, and they exert facilitatory effects on cortical neurons. Without cortical activation, a person is less able to detect specific stimuli, and the level of consciousness is reduced. The pathways for the ascending RAS travel through the midbrain, and lesions of the midbrain can interrupt RAS activity, leading to altered levels of consciousness and coma.

Any deficit in level of consciousness, from mild confusion to stupor or coma, indicates injury to either the RAS or to both cerebral hemispheres concurrently. For example, consciousness may decline because of severe systemic metabolic derangements that affect both hemispheres, or from head trauma causing shear injuries to white matter of both the RAS and the cerebral hemispheres. Brain injuries that affect a hemisphere unilaterally and also spare the RAS, such as cerebral infarction, usually do not cause impaired consciousness.

Levels of Consciousness

Levels of consciousness reflect an orientation to person, place, and time. A fully conscious person is totally aware of her or his surroundings.⁶ Levels of consciousness exist on a continuum that includes consciousness, confusion, delirium, obtundation, stupor, and coma (Table 37-4).

The earliest signs of diminution in level of consciousness are inattention, mild confusion, disorientation, and blunted responsiveness. With further deterioration, the delirious person becomes markedly inattentive and variably lethargic or agitated. The person may progress to become obtunded and may respond only to vigorous or noxious stimuli, such as shaking.

Because of its simplicity of application, the Glasgow Coma Scale has gained almost universal acceptance as a method for assessing the level of consciousness in persons with brain injury^{6,13,14} (Table 37-5). Numbered scores are given to responses of eye opening, verbal utterances, and motor responses. The total score is the sum of the best response in each category.

Other Manifestations of Deteriorating Brain Function

Additional elements in the initial neurological evaluation of a person with brain injury include checking for abnormalities in the size of the pupils and their reaction to light, weakness and asymmetry of motor function, and evidence of decortication or decerebration posturing.

Pupillary Reflexes and Eye Movements. Although the pupils may initially respond briskly to light, they become unreactive and dilated as brain function deteriorates. A bilateral loss of the pupillary light response is indicative of lesions of the brain stem. A unilateral loss of the pupillary light response may be caused by a lesion of the optic or oculomotor pathways. The oculocephalic reflex (doll's-head eye movement) can be used to determine if the brain stem centers for eye movement are intact (Fig. 37-12). If the oculocephalic reflex is inconclusive, and

TABLE 37-4	Descendin	g Levels of Consciousness and Their Characteristics
Level of Consc	iousness	Characteristics
Confusion		Disturbance of consciousness characterized by impaired ability to think clearly, and to perceive, respond to, and remember current stimuli; also disorientation
Delirium		State of disturbed consciousness with motor restlessness, transient hallucinations, disorientation, and sometimes delusions
Obtundation		Disorder of decreased alertness with associated psychomotor retardation
Stupor		A state in which the person is not unconscious but exhibits little or no spontaneous activity
Coma		A state of being unarousable and unresponsive to external stimuli or internal needs; often deter- mined by the Glasgow Coma Scale

(Data from Bates D. [1993]. The management of medical coma. Journal of Neurology, Neurosurgery, and Psychiatry 56, 590)

if there are no contraindications, the oculovestibular (*i.e.*, cold caloric test, in which cold water is instilled into the ear canal) may be used to elicit nystagmus (see Chapter 40).

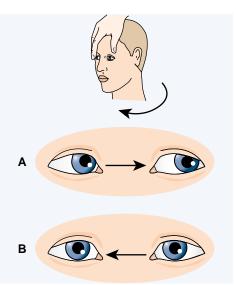
Decorticate and Decerebrate Posturing. With the early onset of unconsciousness there is some combative movement and purposeful movement in response to pain. As coma progresses, noxious stimuli can initiate rigidity and abnormal postures if the motor tracts are interrupted at specific levels. These abnormal postures are called *decortication* and *decerebration*. Decorticate (flexion) posturing is characterized by flexion of the arms, wrists, and fingers, with abduction of the upper extremities, internal rotation, and plantar flexion of the lower extremities (Fig. 37-13A). Decorticate posturing results from lesions of the cerebral hemisphere or internal capsule. Decerebrate (extensor) posturing results from increased muscle excitability (see Fig. 37-13B). It is characterized by rigidity of the arms with palms of the hands turned away from the body and with stiffly extended legs with plantar flexion of the feet. This response occurs when lesions of the diencephalon extend to involve the

TABLE 37-5 The Glasgow Co	ma Scale
Test	Score*
Eye Opening (E)	
Spontaneous	4
To call	3
To pain	2
None	1
Motor Response (M)	
Obeys commands	6
Localizes pain	5
Normal flexion (withdrawal)	4
Abnormal flexion (decorticate)	3
Extension (decerebrate)	2
None (flaccid)	1
Verbal Response (V)	
Oriented	5
Confused conversation	4
Inappropriate words	3
Incomprehensible sounds	2
None	1

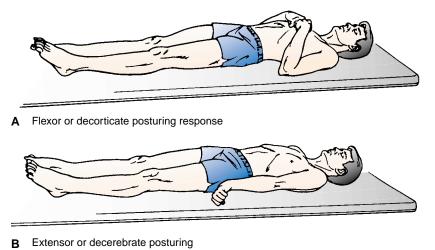
*GCS Score = E + M + V. Best possible score = 15; worst possible score = 3.

midbrain and upper brain stem. Both decerebrate and decorticate posturing are associated with poor prognosis.

Respiratory Responses. Early respiratory changes include yawning and sighing, with progression to Cheyne-Stokes breathing. With progression continuing to the midbrain, respirations change to neurogenic hyperventilation, in which the frequency of respirations may exceed 40 breaths per minute because of uninhibited stimulation of inspiratory and expiratory centers. With medullary involvement, respirations become ataxic (*i.e.*, totally uncoordinated and irregular). Apnea may occur because of a lack of responsiveness to carbon dioxide stimulation. Complete ventilatory assistance is often required at this point.



■ FIGURE 37-12 ■ The *doll's-head eye response* demonstrates the always-present vestibular static reflexes without forebrain interference or suppression. Severe damage to the forebrain or to the brain stem rostral to the pons often results in loss of rostral control of these static vestibular reflexes. If the person's head is moved from side to side or up and down, the eyes will move in conjugate gaze to the opposite side (**A**), much like those of a doll with counterweighted eyes. If the doll's-head phenomenon is observed, brain stem function at the level of the pons is considered intact (in a comatose person). In the unconscious person without intact brain stem function and vestibular static reflexes, the eyes stay in midposition (fixed) or turn in the same direction (**B**) as the head is turned.



■ FIGURE 37-13 ■ Abnormal posturing. (A) Decorticate rigidity. In decorticate rigidity, the upper arms are held at the sides, with elbows, wrists, and fingers flexed. The legs are extended and internally rotated. The feet are plantar flexed. (B) Decerebrate rigidity. In decerebrate rigidity, the jaws are clenched and neck extended. The arms are adducted and stiffly extended at the elbows with the forearms pronated, wrists and fingers flexed. (From Fuller J., Schaller-Ayers J. [1994]. *Health assessment: A nursing approach* [2nd ed.]. Philadelphia: J.B. Lippincott)

Brain Death

Brain death is defined as the irreversible loss of function of the brain, including the brain stem.¹⁶ Irreversibility implies that brain death cannot be reversed. Some conditions such as drug and metabolic intoxication can cause cessation of brain functions that is completely reversible, even when they produce clinical cessation of brain functions and EEG silence. This possibility needs to be excluded before declaring that a person is brain dead.

With advances in scientific knowledge and technology that have provided the means for artificially maintaining ventilatory and circulatory function, the definition of death has had to be continually re-examined. In 1995, the Quality of Standards Subcommittee of the American Academy of Neurology published the clinical parameters for determining brain death and procedures for testing persons older than 18 years.¹⁷ According to these parameters, "brain death is the absence of clinical brain function when the proximate cause is known and demonstrably irreversible."17 Clinical examination must disclose at least the absence of responsiveness, brain stem reflexes, and respiratory effort. Brain death is a clinical diagnosis, and a repeat evaluation at least 6 hours later is recommended.¹⁷ Longer periods of observation of absent brain activity are required in cases of drug overdose (e.g., barbiturates, other CNS depressants), drug toxicity (e.g., neuromuscular blocking drugs, aminoglycoside antibiotics), neuromuscular diseases such as myasthenia gravis, hypothermia, and shock. Medical circumstances may require use of confirmatory tests.

Persistent Vegetative State

Advances in the care of brain-injured persons during the past several decades have resulted in survival of many persons who previously would have died. Unfortunately, some of these persons remain in what often is called a *persistent vegetative state*. The vegetative state is characterized by loss of all cognitive functions and the unawareness of self and surroundings. Reflex and vegetative functions remain.¹⁸ Persons in the vegetative state must be fed and require full nursing care.

The criteria for diagnosis of vegetative state include the absence of awareness of self and environment and an inability to interact with others; the absence of sustained or reproducible voluntary behavioral responses; lack of language comprehension; sufficiently preserved hypothalamic and brain stem function to maintain life; bowel and bladder incontinence; and variably preserved cranial nerve (*e.g.*, pupillary, gag) and spinal cord reflexes.¹⁹ The diagnosis of persistent vegetative state requires that the condition has continued for at least 1 month.

In summary, many of the agents that cause brain damage do so through common pathways, including hypoxia or ischemia, accumulation of excitatory neurotransmitters, increased ICP, and cerebral edema. Deprivation of oxygen (*i.e.*, hypoxia) or blood flow (*i.e.*, ischemia) can have deleterious effects on the brain structures. Ischemia can be focal, as in stroke, or global. Global ischemia occurs when blood flow is inadequate to meet the metabolic needs of the brain, as in cardiac arrest.

The term *head injury* is used to describe all structural damage to the head and has become synonymous with *brain injury*. The effects of traumatic head injuries can be divided into two categories: primary or secondary injuries. In secondary injuries, damage results from the subsequent brain swelling, intracranial hematomas, infection, cerebral hypoxia, and ischemia. Primary injuries result from direct impact. Even if there is no break in the skull, a blow to the head can cause severe and diffuse brain damage. Such closed injuries vary in severity and can be classified as focal or diffuse. Diffuse injuries include concussion and diffuse axonal injury. Focal injuries include contusion, laceration, and hemorrhage.

Brain injury is manifested by alterations in sensory and motor function and by changes in the level of consciousness. Consciousness is a state of awareness of self and environment. It exists on a normal continuum of wakefulness and sleep and a pathologic continuum of wakefulness and coma. In progressive brain injury, coma may follow a rostral-to-caudal progression with characteristic changes in levels of consciousness, respiratory activity, pupillary and oculovestibular reflexes, and muscle tone occurring as the diencephalon through the medulla are affected.

Brain death is defined as the irreversible loss of function of the brain, including that of the brain stem. Clinical examination must disclose at least the absence of responsiveness, brain stem reflexes, and respiratory effort. The vegetative state is characterized by loss of all cognitive functions and the unawareness of self and surroundings, while reflex and vegetative functions remain intact.

CEREBROVASCULAR DISEASE

Cerebrovascular disease encompasses a number of disorders involving vessels in the cerebral circulation. These disorders include stroke and transient ischemic attacks (TIAs), aneurysmal subarachnoid hemorrhage, and arteriovenous malformations.

Cerebral Circulation

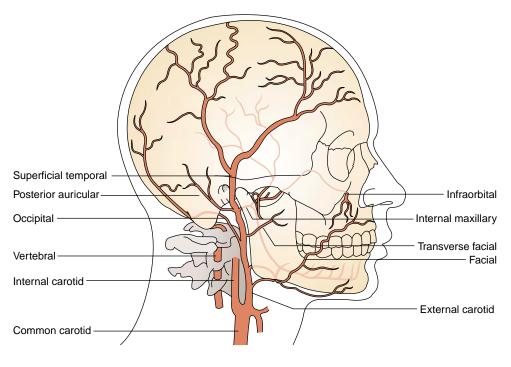
The blood flow to the brain is supplied by the two internal carotid arteries anteriorly and by the two vertebral arteries posteriorly (Fig. 37-14). The internal carotid artery, which provides the major blood supply to the brain, branches into several arteries: the ophthalmic artery, which supplies the eye and orbital structures; the posterior communicating artery, which forms part of the circle of Willis; and the anterior choroidal artery, which supplies the choroid plexus within the lateral ventricles of the brain. The internal carotid terminates by dividing into the anterior and middle cerebral arteries. The anterior cerebral arteries supply most of the medial and superior surfaces of the brain and the frontal lobe. The middle cerebral arteries supply the lateral surface of the brain, including the primary motor and sensory areas of the face and upper limbs, the optic radiations, and the speech area of the brain.

The two vertebral arteries unite to form the basilar artery. Branches of the basilar and vertebral arteries supply the medulla, pons, cerebellum, midbrain, and caudal part of the diencephalon. The basilar artery terminates by dividing into two posterior cerebral arteries that supply the remaining occipital and inferior regions of the temporal lobes and the thalamus. The posterior cerebral arteries also help to form the arterial circle of Willis, which connects the vertebral artery and the internal carotid arterial systems (Fig. 37-15). The union of these two systems provides alternate pathways for blood flow should one of the vessels become occluded.

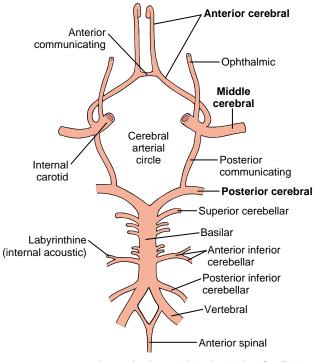
The cerebral blood is drained by two sets of veins that empty into the dural venous sinuses: the deep (great) cerebral venous system and the superficial venous system. The deep system is well protected, in contrast to the superficial cerebral veins that travel through the pia mater on the surface of the cerebral cortex. These vessels connect directly to the sagittal sinuses in the falx cerebri by way of bridging veins. They travel through the CSF-filled subarachnoid space and penetrate the arachnoid and then the dura to reach the dural venous sinuses. This system of sinuses returns blood to the heart primarily by way of the internal jugular veins. Alternate routes for venous flow also exist; for example, venous blood may exit through the emissary veins that pass through the skull and through veins that traverse various foramina to empty into extracranial veins.

Regulation of Cerebral Blood Flow

The blood flow to the brain is maintained at approximately 750 mL/minute or one sixth of the resting cardiac output.² The regulation of blood flow to the brain is controlled largely by autoregulatory mechanisms and by the sympathetic nervous system. The autoregulation of cerebral blood flow responds to the local metabolic needs of the brain tissue and is efficient within an MABP range of approximately 60 to 140 mm Hg.² If blood pressure falls below 60 mm Hg, cerebral blood flow becomes severely compromised, and if it rises above the upper limit of autoregulation, blood flow increases rapidly and overstretches the cerebral vessels. In persons with hypertension, this autoregulatory range shifts to a higher level.



■ FIGURE 37-14 ■ Branches of the right external carotid artery. The internal carotid artery ascends to the base of the brain. The right vertebral artery is also shown as it ascends through the transverse foramina of the cervical vertebrae.



■ FIGURE 37-15 ■ The cerebral arterial circle (circle of Willis).

Although total cerebral blood flow remains relatively stable throughout marked changes in cardiac output and arterial blood pressure, regional blood flow may change markedly in response to local changes in metabolism. At least three metabolic factors affect cerebral blood flow: carbon dioxide concentration, hydrogen ion concentration, and oxygen concentration. Increased carbon dioxide or increased hydrogen ion concentrations increase cerebral blood flow; decreased oxygen concentration also increases blood flow. Carbon dioxide, by way of the hydrogen ion concentration, provides a potent stimulus for control of cerebral blood flow-a doubling of the PCO₂ in the blood results in a doubling of cerebral blood flow. Other substances that alter the pH of the brain produce similar changes in cerebral blood flow. Because increased hydrogen ion concentration greatly depresses neural activity, the increase in blood flow is protective in that it washes the hydrogen ions and other acidic materials away from the brain tissue.2

In addition to the autoregulatory mechanisms that control blood flow in the deep cerebral vessels, the superficial and major cerebral blood vessels are innervated by the sympathetic nervous system. Under normal physiologic conditions, the sympathetic nervous system exerts little effect on superficial cerebral blood flow because local regulatory mechanisms are so powerful that they compensate almost entirely for the effects of sympathetic stimulation. However, when local mechanisms fail, sympathetic control of cerebral blood pressure becomes important.² For example, when the arterial pressure rises to very high levels during strenuous exercise or in other conditions, the sympathetic nervous system constricts the large and intermediate-size superficial blood vessels as a means of protecting the smaller, more easily damaged vessels. Sympathetic reflexes are believed to cause vasospasm in the intermediate and large arteries in some types of brain damage, such as that caused by rupture of a cerebral aneurysm.

Stroke (Brain Attack)

Stroke is an acute focal neurologic deficit from a vascular disorder that injures brain tissue. Stroke remains one of the leading causes of mortality and morbidity in the United States. Each year, 600,000 Americans are afflicted with stroke, and approximately 167,000 of these persons die, and many survivors of stroke are left with at least some degree of neurologic impairment.²⁰ The term *brain attack* has been promoted to highlight that time-dependent tissue damage occurs and to raise awareness of the need for rapid emergency treatment, similar to that with heart attack.

There are two main types of strokes: ischemic stroke and hemorrhagic stroke. Ischemic strokes are caused by an interruption of blood flow in a cerebral vessel and are the most common type of stroke, accounting for 70% to 80% of all strokes. The less common hemorrhagic strokes, which are caused by bleeding into brain tissue, are associated with a much higher fatality rate than are ischemic strokes.

Among the major risk factors for stroke are age, gender, race, heart disease, hypertension, high cholesterol levels, cigarette smoking, prior stroke, and diabetes mellitus.^{20,21} Other risk factors include sickle cell disease, polycythemia, blood dyscrasias, excess alcohol use, cocaine and illicit drug use, obesity, and sedentary lifestyle. The incidence of stroke increases with age, with a 1% per year increased risk for persons 65 to 74 years of age; the incidence of stroke is approximately 19% greater in men than women; and African Americans have a 60% greater risk of death and disability from stroke than do whites.²⁰ Heart disease, particularly atrial fibrillation and other conditions that predispose to clot formation on the wall of the heart or valve leaflets or to paradoxical embolism through right-to-left shunting, predisposes to cardioembolic stroke. Polycythemia, sickle

KEY CONCEPTS

STROKE/BRAIN ATTACK

- Stroke is an acute focal neurologic deficit from an interruption of blood flow in a cerebral vessel (ischemic stroke, the most common type) due to thrombi or emboli or to bleeding into the brain tissue (hemorrhagic stroke).
- During the evolution of an ischemic stroke, there usually is a central core of dead or dying cells surrounded by an ischemic band of minimally perfused cells called a *penumbra*. Whether the cells of the penumbra continue to survive depends on the successful timely return of adequate circulation.
- The realization that there is a window of opportunity during which ischemic but viable brain tissue can be salvaged has led to the use of thrombolytic agents in the early treatment of ischemic stroke.

cell disease (during sickle cell crisis), and blood disorders predispose to clot formation in the cerebral vessels. Alcohol abuse can contribute to stroke in several ways: induction of cardiac arrhythmias and defects in ventricular wall motion that lead to cerebral embolism, induction of hypertension, enhancement of blood coagulation disorders, and reduction of cerebral blood flow.²² Another cause of stroke is cocaine. Cocaine use causes both ischemic and hemorrhagic strokes by inducing vasospasm, enhanced platelet activity, and increased blood pressure, heart rate, body temperature, and metabolic rate. Cocaine stroke victims range in age from newborn (*i.e.*, from maternal cocaine use) to old age.²³

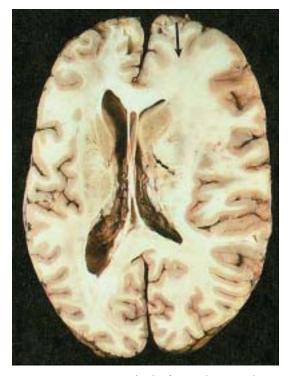
Elimination or control of risk factors for cerebrovascular disease (*e.g.*, use of tobacco, control of blood lipids and blood sugar, reduction of hypertension) offers the best opportunity to prevent cerebral ischemia from cerebral atherosclerosis. Early detection and treatment offer significant advantages over waiting until a serious event has occurred.

Ischemic Stroke

Ischemic strokes are caused by local interruption of blood flow caused by thrombosis or emboli. A common classification system identifies five stroke subtypes and their frequency: 20% large artery atherosclerotic disease (both thrombosis and plaque emboli); 25% small vessel or penetrating artery disease (*lacunar stroke*); 20% cardiogenic embolism; 30% cryptogenic stroke (undetermined cause); and 5% other, unusual causes²⁴ (*i.e.*, migraine, dissection, coagulopathy).

During the evolution of an ischemic stroke, there usually is a central core of dead or infarcted tissue, surrounded by an ischemic band or area of minimally perfused tissue called the *penumbra* (*i.e.*, halo). Brain cells of the penumbra receive marginal blood flow, and their metabolic activities are impaired; although the area undergoes an "electrical failure," the structural integrity of the brain cells is maintained.²⁵ Whether the cells of the penumbra continue to survive depends on the successful timely return of adequate circulation, the volume of toxic products released by the neighboring dying cells, the degree of cerebral edema, and alterations in local blood flow. If the toxic products result in additional death of cells in the penumbra, the core of infarcted tissue enlarges, and the volume of surrounding ischemic tissue increases (Fig. 37-16).

Large Vessel (Thrombotic) Stroke. Cerebral atherosclerosis is the most common cause of ischemic strokes. In the cerebral circulation, atherosclerotic plaques are found most commonly at arterial bifurcations. Common sites of plaque formation include larger vessels of the brain, notably the origins of the internal carotid and vertebral arteries, and junctions of the basilar and vertebral arteries (Fig. 37-17). Cerebral infarction can result from an acute local thrombosis and occlusion at the site of chronic atherosclerosis, with or without embolization of the plaque material distally, or from critical perfusion failure distal to a stenosis. These infarcts often affect the cortex, causing aphasia or neglect, visual field defects, or the retina. In most cases of stroke, a single cerebral artery and its territories are affected. Usually, thrombotic strokes are seen in older persons and frequently are accompanied by evidence of atherosclerotic heart or peripheral arterial disease. The thrombotic stroke is not associated with activity and may occur in a person at rest.



■ FIGURE 37-16 ■ Recent cerebral infarct. A horizontal section of the brain shows expansion and softening in the distribution of the right middle cerebral artery. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 1472]. Philadelphia: Lippincott Williams & Wilkins)

Small Vessel Stroke (Lacunar Infarct). Lacunar infarcts are small (1.5 to 2.0 cm) to very small (3 to 4 mm) infarcts located in the deeper, noncortical parts of the brain or in the brain stem. They are found in the territory of single deep penetrating arteries supplying the internal capsule, basal ganglia, or brain stem. They result from occlusion of the smaller branches of large cerebral arteries, commonly the middle cerebral and posterior cerebral arteries and less commonly the anterior cerebral, vertebral, or basilar arteries. In the process of healing, lacunar infarcts leave behind small cavities, or lacunae (lakes). Six basic causes of lacunar infarcts have been proposed: embolism, hypertension, small vessel occlusive disease, hematologic abnormalities, small intracerebral hemorrhages, and vasospasm. Because of their size and location, lacunar infarcts usually do not cause cortical deficits such as aphasia or apraxia. Instead, they produce classic recognizable "lacunar syndromes," such as pure motor hemiplegia, pure sensory hemiplegia, and dysarthria with the clumsy hand syndrome. Because CT scans are not sensitive enough to detect these tiny infarcts, diagnosis used to depend on clinical features alone. The use of magnetic resonance imaging (MRI) has allowed frequent visualization of small vessel infarcts and is obligatory to confirm such a lesion.

Cardiogenic Embolic Stroke. An embolic stroke is caused by a moving blood clot that travels from its origin to the brain. Various cardiac conditions predispose to formation of emboli that produce embolic stroke, including rheumatic heart disease, atrial fibrillation, recent myocardial infarction, ventricular aneurysm, mobile aortic arch atheroma, and bacterial endocarditis. Although most cerebral emboli originate from a



thrombus in the left heart, they also may originate in an atherosclerotic plaque in the carotid arteries.

Embolic strokes usually affect the larger proximal cerebral vessels, often lodging at bifurcations. The most common site is the middle cerebral artery, probably because it offers the path of least resistance, reflecting the large territory of this vessel and its position as the terminus of the carotid artery. Embolic stroke usually has a sudden onset with immediate maximum deficit.

Hemorrhagic Stroke

The most frequently fatal stroke is a spontaneous hemorrhage into the brain substance.²⁶ With rupture of a blood vessel, hemorrhage into the brain tissue occurs, resulting in edema, compression of the brain contents, or spasm of the adjacent blood vessels. The most common predisposing factors are advancing age and hypertension. Other causes of hemorrhage are aneurysm, trauma, erosion of the vessels by tumors, arteriovenous malformations, coagulopathies, vasculitis, and drugs.

A cerebral hemorrhage occurs suddenly, usually when the person is active. Vomiting commonly occurs at the onset, and headache sometimes occurs. Focal symptoms depend on which vessel is involved. In the most common situation, hemorrhage into the basal ganglia results in contralateral hemiplegia, with initial flaccidity progressing to spasticity. The hemorrhage and resultant edema exert great pressure on the brain substance, and the clinical course progresses rapidly to coma and frequently to death.

Transient Ischemic Attacks

A transient ischemic attack (TIA) is characterized by a focal ischemic cerebral neurologic deficit that lasts less than 24 hours (usually less than 1 to 2 hours). A TIA or "ministroke" is equivalent to "brain angina" and reflects a temporary disturbance in focal cerebral blood flow, which reverses before infarction occurs, analogous to angina in relation to heart attack. The causes of TIAs are the same as those of ischemic stroke and include atherosclerotic disease of cerebral vessels and emboli.

TIAs are important because they may provide warning of impending stroke. In fact, the risk of stroke after a TIA is similar to the risk after a first stroke and is maximal immediately after the event: 4% to 8% risk of stroke within 1 month, 12% to 13% risk during the first year, and 24% to 29% risk during the next 5 years.²⁷ Diagnosis of TIA before a stroke may permit surgical or medical intervention that prevents an eventual stroke and the associated neurologic deficits.²⁶

Acute Manifestations of Stroke

The specific manifestations of stroke or TIA are determined by the cerebral artery that is affected, by the area of brain tissue that is supplied by that vessel, and by the adequacy of the collateral circulation. Symptoms of stroke/TIA always are sudden in onset and focal and are usually one-sided. The most common symptom is unilateral weakness of the face and arm or less commonly of the leg. Other frequent stroke symptoms are unilateral numbness, vision loss in one eye or to one side (hemianopia), language disturbance (aphasia) or slurred speech (dysarthria), and sudden loss of balance or ataxia. In the event of TIA, symptoms rapidly resolve spontaneously, although the underlying mechanisms are the same as for stroke. The specific stroke signs depend on the specific vascular territory compromised (Table 37-6). Discrete subsets of these vascular syndromes usually occur, depending on which branches of the involved artery are blocked.

Diagnosis and Treatment

Diagnosis. Accurate diagnosis of stroke is based on a complete history and thorough physical and neurologic examination. A careful history, including documentation of previous TIAs, the time of onset and pattern and rapidity of system progression, the specific focal symptoms (to determine the likely vascular territory), and the existence of any coexisting diseases, can help to determine the type of stroke that is involved. The diagnostic evaluation should aim to determine the presence of hemorrhage or ischemia, identify the stroke or TIA mechanism (large vessel or small vessel atherothrombotic, cardioembolic, other or cryptogenic, hemorrhagic), characterize the severity of clinical deficits, and unmask the presence of risk factors.

Imaging studies document the brain infarction and the anatomy and pathology of the related blood vessels. CT scans and MRI have become essential tools in diagnosing stroke, differentiating cerebral hemorrhage from ischemia, and excluding intracranial lesions that mimic stroke clinically. CT scans are a necessary screening tool in the acute setting for rapid identification of hemorrhage but are insensitive to ischemia within 24 hours and to any brain stem or small infarcts. MRI is supe-

TABLE 37-6	Signs and Symptoms of Stroke by	Involved Cerebral Artery
Cerebral Artery	Brain Area Involved	Signs and Symptoms*
Anterior cerebral	Infarction of the medial aspect of one frontal lobe if lesion is distal to communicating artery; bilateral frontal infarction if flow in other anterior cerebral artery is inadequate	Paralysis of contralateral foot or leg; impaired gait; paresis of con- tralateral arm; contralateral sensory loss over toes, foot, and leg; problems making decisions or performing acts voluntarily; lack of spontaneity, easily distracted; slowness of thought; aphasia depends on the hemisphere involved; urinary incontinence; cognitive and affective disorders
Middle cerebral	Massive infarction of most of lateral hemi- sphere and deeper structures of the frontal, parietal, and temporal lobes; internal capsule; basal ganglia	Contralateral hemiplegia (face and arm); contralateral sensory impairment; aphasia; homonymous hemianopia; altered con- sciousness (confusion to coma); inability to turn eyes toward paralyzed side; denial of paralyzed side or limb (hemiatten- tion); possible acalculia, alexia, finger agnosia, and left–right confusion; vasomotor paresis and instability
Posterior cerebral	Occipital lobe; anterior and medial portion of temporal lobe	Homonymous hemianopia and other visual defects such as color blindness, loss of central vision, and visual hallucinations; mem- ory deficits, perseveration (repeated performance of same verbal or motor response)
	Thalamus involvement	Loss of all sensory modalities; spontaneous pain; intentional tremor; mild hemiparesis; aphasia
	Cerebral peduncle involvement	Oculomotor nerve palsy with contralateral hemiplegia
Basilar and vertebra	l Cerebellum and brain stem	Visual disturbance such as diplopia, dystaxia, vertigo, dysphagia, dysphonia

*Depend on hemisphere involved and adequacy of collaterals.

rior for imaging ischemic lesions in all territories. Arteriography can demonstrate the site of the vascular abnormality and afford visualization of most intracranial vascular areas. The introduction of several Doppler ultrasonographic techniques has facilitated the noninvasive evaluation of the cerebral circulation, especially for detection of carotid stenosis.

Treatment. The treatment of acute ischemic stroke has changed markedly since the early 1990s, with an emphasis on salvaging brain tissue and minimizing long-term disability. The realization that there is a window of opportunity during which ischemic but viable brain tissue can be salvaged has led to the use of thrombolytic (clot disrupting) agents in the early treatment of ischemic stroke.²⁴ Although the results of emergent treatment of hemorrhagic stroke have been less dramatic, continued efforts to reduce disability have been promising.

A subcommittee of the Stroke Council of the American Heart Association has developed guidelines for the use of thrombolytic therapy for acute stroke.^{28,29} The major risk of treatment with thrombolytic agents is intracranial hemorrhage of the infarcted brain. A number of conditions, including use of oral anticoagulant medications, a history of gastrointestinal bleeding, recent myocardial infarction, previous stroke or head injury within 3 months, surgery within the past 14 days, and a blood pressure greater than 200/120 mm Hg, are considered contraindications to thrombolytic therapy.²⁹

The successful treatment of stroke depends on education of the public, paramedics, and health care professionals in emergency care facilities about the need for early diagnosis and treatment. As with heart attack, the message should be "do not wait to decide if the symptoms subside but seek immediate treatment." Effective medical and surgical procedures may preserve brain function and prevent disability. During the acute phase, proper positioning and range-of-motion exercises are essential. Early rehabilitation efforts include all members of the rehabilitation team—physician, nurse, speech therapist, physical therapist, and occupational therapist—and the family.

Stroke and cerebrovascular disorders often cause long-term disabilities, including motor and sensory deficits, language and speech problems, and a condition called the *hemineglect syndrome* (denial of one half of the body and environment on that side of the body). Longer-term treatment is aimed at preventing complications and recurrent stroke and promoting the fullest possible recovery of function.

Aneurysmal Subarachnoid Hemorrhage

An aneurysm is a bulge at the site of a localized weakness in the muscular wall of an arterial vessel. Most cerebral aneurysms are small saccular aneurysms called *berry aneurysms* (Fig. 37-18). They usually occur in the anterior circulation and are found at bifurcations and other junctions of vessels such as those in the circle of Willis (Fig. 37-19). They are thought to arise from a congenital defect in the media of the involved vessels, particularly at bifurcations. Persons with heritable connective tissue disorders such as autosomal dominant polycystic kidney disease, Ehlers-Danlos syndrome, neurofibromatosis type I, and Marfan's syndrome are at particular risk.¹ Other causes of cerebral aneurysms are atherosclerosis, hypertension, and bacterial infections.

Rupture of a cerebral aneurysm results in subarachnoid hemorrhage.^{30,31} The probability of rupture increases with the size of the aneurysm; aneurysms larger than 10 mm in diameter have a 50% chance of bleeding per year.¹ Rupture often occurs with acute increases in ICP. Of the various environmental factors that may predispose to aneurysmal subarachnoid hemorrhage, cigarette smoking and hypertension appear to constitute



■ FIGURE 37-18 ■ Berry aneurysm. A thin-walled aneurysm protrudes from the arterial bifurcation in the circle of Willis. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 1466]. Philadelphia: Lippincott Williams & Wilkins)

the greatest threat. The mortality and morbidity rates associated with aneurysmal subarachnoid hemorrhage are high.

The signs and symptoms of cerebral aneurysms can be divided into two phases: those presenting before rupture and bleeding and those presenting after rupture and bleeding. Most small aneurysms are asymptomatic; intact aneurysms frequently are found at autopsy as an incidental finding.¹ Large aneurysms may cause chronic headache, neurologic deficits, or both. Approximately 50% of persons with subarachnoid hemorrhage have a history of atypical headaches occurring days to weeks before the onset of hemorrhage, suggesting the presence of a small leak.^{30,31} These headaches are characterized by sudden onset and often are accompanied by nausea, vomiting, and dizziness. Persons with these symptoms may be mistakenly diagnosed as having tension or migraine headaches.

The onset of subarachnoid aneurysmal rupture often is heralded by a sudden and severe headache, described as "the worst headache of my life." If the bleeding is severe, the headache may be accompanied by collapse and loss of consciousness. Vomiting may accompany the presenting symptoms. Other

■ FIGURE 37-19 ■ Common sites of berry aneurysms.

manifestations include signs of meningeal irritation such as nuchal rigidity (neck stiffness) and photophobia (light intolerance); cranial nerve deficits, especially cranial nerve II, and sometimes III and IV (diplopia and blurred vision); stroke syndrome (motor and sensory deficits); cerebral edema and increased ICP; and pituitary dysfunction (diabetes insipidus and hyponatremia).

The complications of aneurysmal rupture include rebleeding, vasospasm with cerebral ischemia, hydrocephalus, hypothalamic dysfunction, and seizure activity. Rebleeding and vasospasm are the most severe and most difficult to treat. Rebleeding, which has its highest incidence on the first day after the initial rupture, results in further and usually catastrophic neurologic deficits.

Vasospasm is a dreaded complication of aneurysmal rupture. The condition is difficult to treat and is associated with a high incidence of morbidity and mortality. Although the description of aneurysm-associated vasospasm is relatively uniform, its proposed mechanisms are a matter of controversy. Usually, the condition develops within 3 to 10 days (peak, 7 days) after aneurysm rupture and involves a focal narrowing of the cerebral artery or arteries that can be visualized on arteriography. The neurologic status gradually deteriorates as blood supply to the brain in the region of the spasm is decreased; this usually can be differentiated from the rapid deterioration seen in rebleeding.

Another complication of aneurysm rupture is the development of hydrocephalus. It is caused by plugging of the arachnoid villi with products from lysis of blood in the subarachnoid space.

The diagnosis of subarachnoid hemorrhage and intracranial aneurysms is made by clinical presentation, CT scan, lumbar puncture, and angiography. Lumbar puncture may be used to detect blood in the CSF, but the procedure carries with it the risk of rebleeding and brain herniation.

The course of treatment after aneurysm rupture depends on the extent of neurologic deficit. Persons with less severe deficits, with or without headache and no neurologic deficits, may undergo cerebral arteriography and early surgery, usually within 24 to 72 hours. A procedure involving craniotomy and clipping often is used. In this procedure, a specially designed silver clip is inserted and tightened around the neck of the aneurysm. This procedure offers protection from rebleeding and may permit removal of the hematoma. Some persons with subarachnoid hemorrhage are managed medically for 10 days or more in an attempt to improve their clinical status before surgery. The use of endovascular techniques such as balloon embolization and platinum coil electrothrombosis is evolving.

Arteriovenous Malformations

Arteriovenous malformations are a complex tangle of abnormal arteries and veins linked by one or more fistulas (Fig. 37-20).³² These vascular networks lack a capillary bed, and the small arteries have a deficient muscularis layer. Arteriovenous malformations are thought to arise from failure in development of the capillary network in the embryonic brain. As the child's brain grows, the malformation acquires additional arterial contributions that enlarge to form a tangled collection of thin-walled vessels that shunt blood directly from the arterial to the venous circulation. Arteriovenous malformations typically present before 40 years of age and affect men and



■ FIGURE 37-20 ■ Arteriovenous malformation. Abnormal blood vessels replace the cortical gray matter and extend deeply into the underlying white matter. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 1466]. Philadelphia: Lippincott Williams & Wilkins)

women equally. Rupture of vessels in the malformation accounts for approximately 2% of all strokes.³²

The hemodynamic effects of arteriovenous malformations are twofold. First, blood is shunted from the high-pressure arterial system to the low-pressure venous system without the buffering advantage of the capillary network. The draining venous channels are exposed to high levels of pressure, predisposing them to rupture and hemorrhage. Second, impaired perfusion affects the cerebral tissue adjacent to the arteriovenous malformation. The elevated arterial and venous pressures and lack of a capillary circulation impair cerebral perfusion by producing a high-flow situation that diverts blood away from the surrounding tissue. Clinically, this is evidenced by slowly progressive neurologic deficits.

The major clinical manifestations of arteriovenous malformations are hemorrhage, seizures, headache, and progressive neurologic deficits. Headaches often are severe, and persons with the disorder may describe them as being throbbing and synchronous with their heart beat. Other, less common symptoms include visual symptoms (*i.e.*, diplopia and hemianopia), hemiparesis, mental deterioration, and speech deficits. Learning disorders have been documented in 66% of adults with arteriovenous malformations.³²

Definitive diagnosis often is obtained through cerebral angiography. Treatment methods include surgical excision, endovascular occlusion, and radiation therapy. Because of the nature of the malformation, each of these methods is accompanied by some risk of complications.

In summary, a stroke, or "brain attack," is an acute focal neurologic deficit caused by a vascular disorder that injures brain tissue. It is the third leading cause of death in the United States and a major cause of disability. There are two main types of stroke: ischemic and hemorrhagic. Ischemic stroke, which is the most common type, is caused by cerebrovascular obstruction by a thrombus or emboli. Hemorrhagic stroke, which is associated with greater morbidity and mortality, is caused by the rupture of a blood vessel and bleeding into the brain. The acute manifestations of stroke depend on the location of the blood vessel that is involved and can include motor, sensory, language, speech, and cognitive disorders. Early diagnosis and treatment with thrombolytic agents has improved the outlook for many persons with ischemic stroke.

A subarachnoid hemorrhage involves bleeding into the subarachnoid space. Most subarachnoid hemorrhages are the result of a ruptured cerebral aneurysm. Presenting symptoms include headache, nuchal rigidity, photophobia, and nausea. Complications include rebleeding, vasospasm, and hydrocephalus. Arteriovenous malformations are congenital abnormal communications between arterial and venous channels that result from failure in the development of the capillary network in the embryonic brain. The vessels in the arteriovenous malformations may enlarge to form a spaceoccupying lesion, become weak and predispose to bleeding, and divert blood away from other parts of the brain; they can cause brain hemorrhage, seizures, headache, and other neurologic deficits.

INFECTIONS AND NEOPLASMS

Infections

Infections of the CNS may be classified according to the structure involved: the meninges, (meningitis); the brain parenchyma (encephalitis); the spinal cord (myelitis); and the brain and spinal cord (encephalomyelitis). They also may be classified by the type of invading organism: bacterial, viral, or other. In general, the pathogens enter the CNS through the bloodstream by crossing the blood-brain barrier or by direct invasion through skull fracture, a bullet hole, or rarely, by contamination during surgery or lumbar puncture.

Meningitis

Meningitis is an inflammation of the pia mater, the arachnoid, and the CSF-filled subarachnoid space. Inflammation spreads rapidly because of CSF circulation around the brain and spinal cord. The inflammation usually is caused by an infection, but chemical meningitis can occur. There are two types of acute infectious meningitis: acute pyogenic meningitis (usually bacterial) and acute lymphocytic (usually viral) meningitis.¹

Bacterial Meningitis. Most cases of bacterial meningitis are caused by *Streptococcus pneumoniae, Haemophilus influenzae,* or *Neisseria meningitidis* (the meningococcus). The incidence of *H. influenzae,* which is the common cause of meningitis in children younger than 5 years, has declined dramatically during recent years because of vaccination against *H. influenza.* Epidemics of meningococcal meningitis occur in settings such as the military, where the recruits must reside in close contact. The very young and the very old are at highest risk for pneumococcal meningitis. Risk factors associated with contracting meningitis include head trauma with basilar skull fractures, otitis media, sinusitis or mastoiditis, neurosurgery, dermal sinus tracts, systemic sepsis, or immunocompromise.

In the pathophysiology of bacterial meningitis, the bacterial organisms replicate and undergo lysis in the CSF, releasing endotoxins or cell wall fragments. These substances initiate the release of inflammatory mediators, which set the stage for a complex but coordinated sequence of events by which neutrophils bind to and damage the endothelial cells of the blood-brain barrier, permitting fluid to move across the capillary wall. This allows pathogens, neutrophils, and albumin to move across the endothelial wall into the CSF. As the pathogens enter the subarachnoid space, they cause inflammation, characterized by a cloudy, purulent exudate. Thrombophlebitis of the bridging veins and dural sinuses may develop, followed by congestion and infarction in the surrounding tissues. Ultimately, the meninges thicken, and adhesions form. These adhesions may impinge on the cranial nerves, giving rise to cranial nerve palsies, or may impair the outflow of CSF, causing hydrocephalus.

The most common symptoms of acute bacterial meningitis are fever and chills; headache; stiff neck; back, abdominal, and extremity pains; and nausea and vomiting. Other signs include seizures, cranial nerve palsies, and focal cerebral signs.³³ A petechial rash is found in most persons with meningococcal meningitis. These petechiae vary from pinhead size to large ecchymoses or even areas of skin gangrene that slough if the person survives. Other types of meningitis also may produce a petechial rash. Persons infected with *H. influenzae* or *S. pneumoniae* may present with difficulty in arousal and seizures, whereas those with *N. meningitidis* infection may present with delirium or coma.³³ The development of brain edema, hydrocephalus, or increased cerebral blood flow can increase ICP.

Meningeal signs (*e.g.*, photophobia and nuchal rigidity), such as those seen in subarachnoid hemorrhage, also may be present. Two assessment techniques can help determine whether meningeal irritation is present. *Kernig's sign* is resistance to extension of the leg while the person is lying with the hip flexed at a right angle. *Brudzinski's sign* is elicited when flexion of the neck results in flexion of the hip and knee. These postures are caused by stretching of the inflamed meninges from the lumbar level to the head. Stretching of the inflamed meninges is extremely painful, producing resistance to stretching. Cranial nerve damage (especially the eighth nerve, with resulting deafness) and hydrocephalus may occur as complications of pyogenic meningitis.

The diagnosis of bacterial meningitis is based on the history and physical examination, along with laboratory data. Lumbar puncture (*i.e.*, spinal tap) findings, which is necessary for accurate diagnosis, includes a cloudy and purulent CSF under increased pressure. The CSF typically contains large numbers of polymorphonuclear neutrophils (up to 90,000/mm³), increased protein content, and reduced sugar content. Bacteria can be seen on smears and can easily be cultured with appropriate media.

Treatment includes antibiotics and corticosteroids. Optimal antibiotic treatment requires that the drug have a bactericidal effect in the CSF. Because bactericidal therapy often results in rapid lysis of the pathogen, treatment can promote the release of biologically active cell wall products into the CSF. The release of these cell wall products can increase the production of inflammatory mediators that have the potential for exacerbating the abnormalities of the blood-brain barrier and the inflammatory process. Because of evidence linking the inflammatory mediators to the pathogenesis of bacterial meningitis, adjunctive corticosteroid therapy usually is administered with or just before the first dose of antibiotics in infants and children. The adjunctive use of corticosteroid therapy in adults is a matter of controversy.

Persons who have been exposed to someone with meningococcal meningitis should be treated prophylactically with antibiotics.³⁴ Effective polysaccharide vaccines are available to protect against meningococcal groups A, C, Y, and W-135.³⁵ These vaccines are recommended for military recruits and college students, who are at increased risk of invasive meningo-coccal disease.

Viral Meningitis. Viral meningitis manifests in much the same way as bacterial meningitis, but the course is less severe, and the CSF findings are markedly different. There are lymphocytes in the fluid, rather than polymorphonuclear cells, the protein content is only moderately elevated, and the sugar content usually is normal. The acute viral meningitides are self-limited and usually require only symptomatic treatment. Viral meningitis can be caused by many different viruses, including mumps, coxsackievirus, Epstein-Barr virus, and herpes simplex type 2. In many cases, the virus cannot be identified.

Encephalitis

Encephalitis represents a generalized infection of the parenchyma of the brain or spinal cord. It usually is caused by a virus, but it also may be caused by bacteria, fungi, and other organisms. The nervous system is subjected to invasion by many viruses, such as arbovirus, poliovirus, and rabies virus. The mode of transmission may be the bite of a mosquito (arbovirus), a rabid animal (rabies virus), or ingestion (poliovirus). A common cause of encephalitis in the United States is herpes simplex virus. Less common causes of encephalitis are toxic substances such as ingested lead and vaccines for measles and mumps. Encephalitis caused by human immunodeficiency virus infection is discussed in Chapter 10.

The pathologic picture of encephalitis includes local necrotizing hemorrhage, which ultimately becomes generalized, with prominent edema. There is progressive degeneration of nerve cell bodies. The histologic picture, although rather general, demonstrates some specific characteristics. For example, the poliovirus selectively destroys the cells of the anterior horn of the spinal cord.

Encephalitis, like meningitis, is characterized by fever, headache, and nuchal rigidity. Patients experience a wide range of neurologic disturbances, such as lethargy, disorientation, seizures, focal paralysis, delirium, and coma. Diagnosis of encephalitis is made by clinical history and presenting symptoms, in addition to traditional CSF studies.

Neoplasms

For most neoplasms, the term *malignant* is used to describe the tumor's lack of cell differentiation, its invasive nature, and its ability to metastasize. However, in the brain even a welldifferentiated and histologically benign tumor may grow and cause death because of its location.

Brain cancer accounts for 2% of all cancer deaths. The American Cancer Society estimates that there were 17,200 new cases and more than 13,100 deaths of brain and CNS cancers in 2001.³⁶ Metastasis to the brain from other sites is even more common.³⁷ In children, brain tumors are second only to leukemia as a cause of death from cancer, and they kill approximately 1600 children and young adults annually.

Although a number of chemical and viral agents can cause brain tumors in laboratory animals, there is no evidence that these agents directly cause brain cancer in humans. Cranial irradiation and exposure to some chemicals may lead to an increased incidence of astrocytomas and meningiomas. There may also be a hereditary factor. Childhood tumors are considered to be developmental in origin.

Types of Tumors

Brain tumors can be divided into three types: primary intracranial tumors of CNS tissue (*e.g.*, neuroglial tumors [gliomas]), primary intracranial tumors that originate in the skull cavity but are not derived from the brain tissue itself (*e.g.*, meninges, pituitary gland, pineal gland), and metastatic tumors.

Primary Neuroglial Tumors. Primary intracranial neoplasms of CNS origin can be classified according to the site of origin and histologic type. They include astrocytomas, oligodendrogliomas, and ependymomas.

Collectively, the neoplasms of astrocyte origin are the most common type of primary brain tumor in the adult. Astrocytomas can be subdivided into fibrillary (filtrating) astrocytic neoplasms and pilocytic astrocytomas.1 Fibrillary or diffuse astrocytomas account for 80% of adult primary brain tumors. They are most commonly seen in adults, but may occur at any age. Although they usually are found in the cerebral hemispheres, they also can occur in the cerebellum, brain stem, or spinal cord. These tumors are subdivided into histologic grades based on their degree of differentiation. The Word Health Organization grading system divides these tumors into three types: (1) the well differentiated lesions, designated astrocytomas; (2) intermediate-grade tumors, designated anaplastic astrocytomas; and (3) the most aggressive lesions, designated glioblastoma multiforme. Astrocytomas have a marked tendency to become more anaplastic with time, so a tumor beginning as an astrocytoma may develop into a glioblastoma. Pilocytic astrocytomas are distinguished from other astrocytomas by their cellular appearance and their benign behavior. Typically, they occur in children and young adults and usually are located in the cerebellum, but they also can be found in the floor and walls of the third ventricle, the optic chiasm and nerves, and occasionally in the cerebral hemispheres.

Oligodendrogliomas comprise approximately 5% to 15% of glial tumors. They are most common in middle life and are found in the cerebral hemispheres.¹ In general, persons with oligodendrogliomas have a better prognosis than do persons with astrocytomas.

Ependymomas are derived from the single layer of epithelium that lines the ventricles and spinal canal. Although they can occur at any age, they are most likely to occur in the first 2 decades of life.¹

Meningiomas. Meningiomas develop from the meningothelial cells of the arachnoid and are outside the brain. They usually have their onset in the middle or later years of life. They are slow-growing, well-circumscribed, and often highly vascular tumors. They usually are benign, and complete removal is possible if the tumor does not involve vital structures.

Metastatic Tumors. The brain is a common site for metastatic tumors. They occur primarily in older persons, paralleling the increase in solid tumors that occurs with increasing age.¹ Hematopoietic tumors, such as lymphomas and leukemias, which occur in children as well as adults, may also spread to the CNS. Among the tumors that commonly metastasize to the brain are lung and breast tumors and malignant melano-

mas. In some cases, particularly in carcinomas arising in the lung cancer, neurologic symptoms may provide the first evidence of cancer.

Clinical Course

Brain tumors can produce either focal or generalized disorders of brain function.³⁷ Focal disturbances result from brain compression, tumor infiltration, disturbances in blood flow, and cerebral edema. Cysts may form in tumors and contribute to brain compression. Generalized disruption of brain function usually reflects an increase in ICP. Because the volume of the intracranial cavity is fixed, brain tumors can cause a generalized increase in ICP when they reach sufficient size. Tumors can obstruct the flow of CSF in the ventricular cavities and produce hydrocephalic dilatation of the proximal ventricles and atrophy of the cerebral hemispheres.

The clinical manifestations of brain tumors depend on the size and location of the tumor. Generalized signs and symptoms include headache, nausea, vomiting, and papilledema. The brain itself is insensitive to pain. The headache that accompanies brain tumors results from compression or distortion of pain-sensitive dural or vascular structures. It may be felt on the same side of the head as the tumor but more commonly is diffuse. In the early stages, the headache, which is caused by irritation, compression, and traction on the dural sinuses or blood vessels, is mild and occurs in the morning when the person awakens.³⁷ It usually disappears after the person has been up for a short time. The headache becomes more constant as the tumor enlarges and often is worsened by coughing, bending, or sudden movements of the head.

Vomiting occurs with or without preceding nausea and is a common symptom of increased ICP and brain stem compression. Direct stimulation of the vomiting center, which is located in the medulla, may contribute to the vomiting that occurs with brain tumors. The vomiting may be projectile. Vomiting caused by brain tumor usually is unrelated to meals and often is associated with headache. Papilledema (edema of the optic disk) results from increased ICP and obstruction of the CSF pathways. It is associated with decreased visual acuity, diplopia, and deficits in the visual fields. Visual defects associated with papilledema often are the reason that persons with brain tumor seek medical care.

Personality and mental changes are common with brain tumors. Persons with brain tumors often are irritable initially and later become quiet and apathetic. They may become forgetful, seem preoccupied, and appear to be psychologically depressed. Because of the mental changes, a psychiatric consultation may be sought before a diagnosis of brain tumor is made.

Focal signs and symptoms are determined by the location of the tumor. Tumors arising in the frontal lobe may grow to large size, increase the ICP, and cause signs of generalized brain dysfunction before focal signs are recognized. Tumors that impinge on the visual system cause visual loss or visual field defects long before generalized signs develop. Certain areas of the brain have a relatively low threshold for seizure activity; tumors arising in relatively silent areas of the brain may produce focal epileptogenic discharges. Temporal lobe tumors often produce seizures as their first symptom. Hallucinations of smell or hearing and déjà vu phenomena are common focal manifestations of temporal lobe tumors. Brain stem tumors commonly produce upper and lower motoneuron signs, such as weakness of facial muscles and ocular palsies that occur with or without involvement of sensory or long motor tracts. Cerebellar tumors often cause ataxia of gait.

Diagnosis and Treatment. Diagnostic procedures for brain tumor include physical and neurologic examinations, visual field and funduscopic examination, CT scans and MRI, skull x-ray films, technetium pertechnetate brain scans, electroencephalography, and cerebral angiography. Physical examination is used to assess motor and sensory function. Because the visual pathways travel through many areas of the cerebral lobes, detection of visual field defects can provide information about the location of tumors. Although CT scanning is used as a screening test, MRI scans are more sensitive than CT scans for mass lesions and can be diagnostic when a clinically suspected tumor is not detected by CT scanning. Skull x-ray films are used to detect calcified areas in a neoplasm or erosion of skull structures caused by tumors. Cerebral angiography can be used to locate a tumor and visualize its vascular supply, information that is important when planning surgery.

The three general methods for treatment of brain tumors are surgery, irradiation, and chemotherapy. Surgery is part of the initial management of virtually all brain tumors; it establishes the diagnosis and achieves tumor removal in many cases. The use of chemotherapy for brain tumors is somewhat limited by the blood-brain barrier. Chemotherapeutic agents can be administered intravenously, intra-arterially, intrathecally (*i.e.*, into the spinal canal), or intraventricularly.

In summary, infections of the CNS may be classified according to the structures involved (*e.g.*, meningitis, encephalitis) or the type of organism causing the infection. The damage caused by infection may predispose to hydrocephalus, seizures, or other neurologic defects.

Brain tumors account for 2% of all cancer deaths and are the second most common type of cancer in children. Primary brain tumors can arise from any structure in the cranial cavity. Most begin in brain tissue, but the pituitary, the pineal region, and the meninges also are sites of tumor development. Brain tumors cause generalized or focal disturbances in brain function. Generalized disruption of brain function usually reflects an increase in ICP. General symptoms such as headache, nausea, vomiting, mental changes, and papilledema usually reflect an increase in ICP. Focal symptoms, such disorders of motor function, result from brain compression, tumor infiltration, disturbances in blood flow, and cerebral edema.

SEIZURE DISORDERS

Seizures, sometimes called *convulsions*, are paroxysmal motor, sensory, or cognitive manifestations of spontaneous, abnormally synchronous discharges of collections of neurons in the cerebral cortex. Approximately 2 million persons in the United States experience recurrent seizures.³⁸ Seizure activity is the most common disorder encountered in pediatric neurology, and among adults, its incidence is exceeded only by cerebrovascular disorders. In most persons, the first seizure episode occurs before 20 years of age. After 20 years of age, a seizure is caused most often by a structural change, trauma, tumor, or stroke.

KEY CONCEPTS

SEIZURES

- Seizures are paroxysmal motor, sensory, or cognitive manifestations of spontaneous, abnormally synchronous electrical discharges from collections of neurons in the cerebral cortex that are thought to result directly or indirectly from changes in excitability of single neurons or groups of neurons.
- Partial seizures originate in a small group of neurons in one hemisphere with secondary spread of seizure activity to other parts of the brain. Simple partial seizures usually are confined to one hemisphere and do not involve loss of consciousness. Complex partial seizures begin in a localized area, spread to both hemispheres, and involve impairment of consciousness.
- Generalized seizures show simultaneous disruption of normal brain activity in both hemispheres from the onset. They include unconsciousness and varying bilateral degrees of symmetric motor responses with evidence of localization to one hemisphere. Absence seizures are generalized nonconvulsive seizure events that are expressed mainly by brief periods of unconsciousness. Tonic-clonic seizures involve unconsciousness along with both tonic and clonic muscle contractions.

A seizure represents the clinical manifestations of an abnormal, uncontrolled electrical discharge from a group of neurons. This uncontrolled neuronal activity causes signs and symptoms that vary according to the location of the originating focus of seizure activity, involvement of surrounding neurons, and spread to other parts of the brain. These signs and symptoms can include strange sensations and perceptions (*e.g.*, hallucinations), unusual or repetitive muscle movements, autonomic visceral activity, and the onset of a confusional state or loss of consciousness.

Many theories have been proposed to explain the cause of the abnormal brain electrical activity that occurs with seizures. Seizures may be caused by alterations in cell membrane permeability or distribution of ions across the neuronal cell membranes. Another cause may be decreased inhibition of cortical or thalamic neuronal activity or structural changes that alter the excitability of neurons. Neurotransmitter imbalances such as an acetylcholine excess or γ -aminobutyric acid (GABA, an inhibitory neurotransmitter) deficiency have been proposed as causes.

Provoked and Unprovoked Seizures

Clinically, seizures may be categorized as unprovoked (primary or idiopathic) or provoked (secondary or acute symptomatic).³⁹ Provoked or symptomatic seizures include febrile seizures, seizures precipitated by systemic metabolic conditions, and those that follow a primary insult to the CNS. Unprovoked or idiopathic seizures are those for which no identifiable cause can be determined.

The most common subgroup of seizures under the category of provoked seizures is that of febrile seizures in children.⁴⁰ They are associated with a high fever, usually with a temperature higher than 104°F. Seizures can also be precipitated by systemic or metabolic disturbances, and primary CNS insults also fall into the category of provoked seizures. Transient systemic metabolic disturbances may precipitate seizures. Examples include electrolyte imbalances, hypoglycemia, hypoxia, hypocalcemia, and alkalosis. Toxemia of pregnancy, water intoxication, uremia, and CNS infections such as meningitis also may precipitate a seizure. The rapid withdrawal of sedative-hypnotic drugs, such as alcohol or barbiturates, is another cause of seizures. Approximately 5% to 10% of those who sustain a CNS insult, such as occurs with cerebral bleeding, edema, or neuronal damage, experience a seizure. Treatment of the immediate cause of these seizures often results in their disappearance.

Multiple episodes or frequent recurrences of apparently unprovoked seizures are considered a seizure disorder. The terms *seizure disorder* and *epileptic syndrome* often are used interchangeably. However, the term *seizure disorder* is often preferred because of the negative connotations associated with the term *epilepsy*. A seizure disorder can be defined as a syndrome in which there is a tendency to have recurrent, paroxysmal seizure activity without evidence of a reversible metabolic cause. It is a chronic condition for which long-term medication may be appropriate.

Classification

The International Classification of Epileptic Seizures is based on symptoms during the seizure and EEG activity. It divides seizures into two broad categories: partial seizures, in which the seizure begins in a specific or focal area of one cerebral hemisphere, and generalized seizures, which begin simultaneously in both cerebral hemispheres⁴¹ (Chart 37-1).

Partial Seizures

Partial or focal seizures are the most common type of seizure among newly diagnosed cases in all persons older than 10 years of age. Partial seizures can be subdivided into three major groups: simple partial (consciousness is not impaired), complex partial (impairment of consciousness), and secondarily generalized partial seizures.

Simple Partial Seizures. Simple partial seizures usually involve only one hemisphere and are not accompanied by loss of consciousness or responsiveness. These seizures also have been referred to as *elementary partial seizures, partial seizures with elementary symptoms,* or *focal seizures.* Simple partial seizures are classified according to motor signs, sensory symptoms, autonomic manifestations, and psychic symptoms.

The observed clinical signs and symptoms depend on the area of the brain where the abnormal neuronal discharge is taking place. If the motor area of the brain is involved, the earliest symptom is motor movement corresponding to the location of onset on the contralateral side of the brain. The motor movement may remain localized or may spread to other cortical areas, with sequential involvement of body parts in an epileptic-type

CHART 37-1 Classification of Epileptic Seizures

Partial Seizures

Simple partial seizures (no impairment of consciousness)
With motor symptoms
With sensory symptoms
With autonomic signs
With psychic symptoms
Complex partial seizures (impairment of consciousness)
Simple partial onset followed by impaired consciousness
Impairment of consciousness at onset
Partial seizures evolving to secondarily generalized seizures
Simple partial leading to generalized seizures
Complex partial leading to generalized seizures
Unclassified Seizures
Unclassified Seizures Classification not possible because of inadequate or
Unclassified Seizures Classification not possible because of inadequate or incomplete data Generalized Seizures
Unclassified Seizures Classification not possible because of inadequate or incomplete data
Unclassified Seizures Classification not possible because of inadequate or incomplete data Generalized Seizures Absence seizures (typical or atypical)
Unclassified Seizures Classification not possible because of inadequate or incomplete data Generalized Seizures Absence seizures (typical or atypical) Atonic seizures
Unclassified Seizures Classification not possible because of inadequate or incomplete data Generalized Seizures Absence seizures (typical or atypical) Atonic seizures Myoclonic seizures

International League Against Epilepsy [1981]. Proposal for revised clinical and electroencephalographic classification of epileptic seizures. *Epilepsia* 22, 489)

"march," known as a *Jacksonian seizure*. If the sensory portion of the brain is involved, there may be no observable clinical manifestations. Sensory symptoms correlating with the location of seizure activity on the contralateral side of the brain may involve somatic sensory disturbance (*e.g.*, tingling and crawling sensations) or special sensory disturbance (*i.e.*, visual, auditory, gustatory, or olfactory phenomena). When abnormal cortical discharge stimulates the autonomic nervous system, flushing, tachycardia, diaphoresis, hypotension or hypertension, or pupillary changes may be evident.

The term *prodrome* or *aura* traditionally has meant a sensory warning sign of impending seizure activity that affected persons could recognize or describe because they were conscious. Simple partial seizures may progress to complex partial seizures or generalized tonic-clonic seizures that result in unconsciousness.

Complex Partial Seizures. Complex partial seizures involve impairment of consciousness and often arise from the temporal lobe. The seizure begins in a localized area of the brain but may progress rapidly to involve both hemispheres. These seizures also may be referred to as *temporal lobe seizures* or *psychomotor seizures*.

Complex partial seizures often are accompanied by automatisms. Automatisms are repetitive, nonpurposeful activity, such as lip smacking, grimacing, patting, or rubbing clothing. Confusion during the postictal state (after a seizure) is common. Hallucinations and illusional experiences such as *déjà vu* (familiarity with unfamiliar events or environments) or *jamais vu* (unfamiliarity with a known environment) have been reported. There may be overwhelming fear, uncontrolled forced thinking or a flood of ideas, and feelings of detachment and depersonalization.

Secondarily Generalized Partial Seizures. These seizures are focal at onset but then become generalized as the seizure activity spreads, involving deeper structures of the brain, such as the thalamus or the reticular formation. Discharges spread to both hemispheres, resulting in progression to tonic-clonic seizure activity. These seizures may start as simple or complex partial seizures and may be preceded by an aura. The aura, often a stereotyped peculiar sensation that precedes the seizure, is the result of partial seizure activity. A history of an aura is clinically useful to identify the seizure as partial and not generalized in onset. However, absence of an aura does not reliably exclude a focal onset because many partial seizures generalize too rapidly to generate an aura.

Generalized Seizures

Generalized seizures begin with initial involvement of both hemispheres. Generalized-onset seizures are the most common type in young children. These seizures are classified as primary or generalized when clinical signs, symptoms, and supporting EEG changes indicate involvement of both hemispheres at onset. The clinical symptoms include unconsciousness and involve varying degrees of symmetric bilateral motor responses without evidence of localization to one hemisphere.

These seizures are divided into four broad categories: absence seizures (typical and atypical), atonic seizures, myoclonic seizures, and major motor (tonic-clonic) seizures.⁴¹

Absence Seizures. Absence seizures, formerly referred to as petit mal seizures, are generalized, nonconvulsive epileptic events and are expressed mainly as disturbances in consciousness. Absence seizures typically occur only in children and cease in adulthood or evolve to generalized motor seizures. Children may present with a history of school failure that predates the first evidence of seizure episodes. Although typical absence seizures have been characterized as a blank stare, motionlessness, and unresponsiveness, motion occurs in many cases of typical absence seizures. This motion may take the form of automatisms such as lip smacking, mild clonic motion (usually in the eyelids), increased or decreased postural tone, and autonomic phenomena. There often is a brief loss of contact with the environment. The seizure usually lasts only a few seconds, and then the child is able to resume normal activity immediately. The manifestations often are so subtle that they may pass unnoticed.

Atypical absence seizures are similar to typical absence seizures except for greater alterations in muscle tone and less abrupt onset and cessation. In practice, it is difficult to distinguish typical from atypical absence seizures without benefit of supporting EEG findings. Because automatisms and unresponsiveness are common to complex partial seizures, the latter may be mistakenly labeled as absence seizures. However, it is important to distinguish between the two types of seizures because the drug treatment is different. Medications that are effective for partial seizures may increase the frequency of absence seizures.

Atonic Seizures. In atonic or akinetic seizures, there is a sudden, split-second loss of muscle tone leading to slackening of the jaw, drooping of the limbs, or falling to the ground. These seizures also are known as *drop attacks*.

Myoclonic Seizures. Myoclonic seizures involve brief involuntary muscle contractions induced by stimuli of cerebral origin. A myoclonic seizure involves bilateral jerking of muscles, generalized or confined to the face, trunk, or one or more extremities. Tonic seizures are characterized by a rigid, violent contraction of the muscles, fixing the limbs in a strained position. Clonic seizures consist of repeated contractions and relaxations of the major muscle groups.

Tonic-Clonic Seizures. Tonic-clonic seizures, formerly called *grand mal seizures*, are the most common major motor seizure. Frequently, a person has a vague warning (probably a simple partial seizure) and experiences a sharp tonic contraction of the muscles with extension of the extremities and immediate loss of consciousness. Incontinence of bladder and bowel is common. Cyanosis may occur from contraction of airway and respiratory muscles. The tonic phase is followed by the clonic phase, which involves rhythmic bilateral contraction and relaxation of the extremities. At the end of the clonic phase, the person remains unconscious until the RAS begins to function again. This is called the *postictal phase*. The tonic-clonic phases last approximately 60 to 90 seconds.

Unclassified Seizures

Unclassified seizures are those that cannot be placed in one of the previous categories. These seizures are observed in the neonatal and infancy periods. Determination of whether the seizure is focal or generalized is not possible. Unclassified seizures are difficult to control with medication.

Diagnosis and Treatment

The diagnosis of seizure disorders is based on a thorough history and neurologic examination, including a full description of the seizure. The physical examination and laboratory studies help exclude any metabolic disease (*e.g.*, hyponatremia) that could precipitate seizures. Skull radiographs and CT or MRI scans are used to identify structural defects. One of the most useful diagnostic tests is the EEG, which is used to record changes in the brain's electrical activity. It is used to support the clinical diagnosis of epilepsy, to provide a guide for prognosis, and to assist in classifying the seizure disorder.

The first rule of treatment is to protect the person from injury during a seizure, preserve brain function by aborting or preventing seizure activity, and treat any underlying disease. Persons with epilepsy should be advised to avoid situations that could be dangerous or life threatening if seizures occur.

Anticonvulsant Medications

Since the late 1970s, the therapy for epilepsy has changed drastically because of an improved classification system, the ability to measure serum anticonvulsant levels, and the availability of potent new anticonvulsant drugs. With proper drug management, 60% to 80% of persons with epilepsy can obtain good seizure control. More than 20 drugs are available in the United States for the treatment of epilepsy.⁴² Antiseizure drugs act mainly by suppressing repetitive firing of isolated neurons that act as epileptogenic foci for seizure activity or by inhibiting the transmission of electrical impulses involved in seizure activity. Because of their selective mechanisms of action, different drugs are used to treat the different types of seizures. For example, ethosuximide is the drug of choice for absence seizures, but it is not effective for tonic-clonic seizures that progress from partial seizures.

The goal of pharmacologic treatment is to bring the seizures under control with the least possible disruption in lifestyle and minimum side effects from medication. When possible, a single drug should be used. Monotherapy eliminates drug interactions and additive side effects. Determining the proper dose of the anticonvulsant drug is often a long and tedious process, which can be very frustrating for the person with epilepsy. Often blood tests are used to determine that the blood concentration is within the therapeutic range. Consistency in taking the medication is essential. Anticonvulsant drug use never should be discontinued abruptly. Special consideration is needed when a person taking an anticonvulsant medication becomes ill and must take additional medications. Some drugs act synergistically, and others interfere with the actions of anticonvulsant medications. This situation needs to be carefully monitored to avoid overmedication or interference with successful seizure control.

Women of childbearing age require special consideration concerning fertility, contraception, and pregnancy. Many of the drugs interact with oral contraceptives; some affect hormone function or decrease fertility. All such women should be advised to take folic acid supplementation. For women with epilepsy who become pregnant, antiseizure drugs increase the risk of congenital abnormalities and other perinatal complications.

Surgical Therapy

Surgical treatment may be an option for persons with epilepsy that is refractory to drug treatment.⁴³ With the use of modern neuroimaging and surgical techniques, a single epileptogenic lesion can be identified and removed without leaving a neurologic deficit.

Generalized Convulsive Status Epilepticus

Seizures that do not stop spontaneously or occur in succession without recovery are called *status epilepticus*. There are as many types of status epilepticus as there are types of seizures. Tonicclonic status epilepticus is a medical emergency and, if not promptly treated, may lead to respiratory failure and death.

The disorder occurs most frequently in the young and old. Morbidity and mortality rates are highest in elderly persons and persons with acute symptomatic seizures, such as those related to anoxia or cerebral infarction.⁴⁴ Approximately one third of patients have no history of a seizure disorder, and in another one third, status epilepticus occurs as an initial manifestation of epilepsy.⁴⁴

Treatment consists of appropriate life-support measures. If status epilepticus is caused by neurologic or systemic disease, the cause needs to be identified and treated immediately. Medications are given to control seizure activity.

In summary, seizures are caused by spontaneous, uncontrolled, paroxysmal, transitory discharges from cortical centers in the brain. Seizures may occur as a reversible symptom of another disease condition or as a recurrent condition called epilepsy. Epileptic seizures are classified as partial or generalized seizures. Partial seizures have evidence of local onset, beginning in one hemisphere. They include simple partial seizures, in which consciousness is not lost, and complex partial seizures, which begin in one hemisphere but progress to involve both. Because consciousness is not lost, an aura is now considered to be part of a simple partial seizure. Generalized seizures involve both hemispheres and include unconsciousness and rapidly occurring, widespread, bilateral symmetric motor responses. They include minor motor seizures, such as absence and akinetic seizures, and major motor or grand mal seizures.

DEMENTIAS

Dementia is a syndrome of intellectual deterioration severe enough to interfere with occupational or social performance. It may involve disturbances in memory, language use, perception, and motor skills and may interrupt the ability to learn necessary skills, solve problems, think abstractly, and make judgments. The dementias include Alzheimer's disease, multiinfarct dementia, Pick's disease, Creutzfeldt-Jakob disease, Wernicke-Korsakoff syndrome, and Huntington's chorea.

Depression is the most common treatable illness that may masquerade as dementia, and it must be excluded when a diagnosis of dementia is considered. This is important because cognitive functioning usually returns to baseline levels after depression is treated.

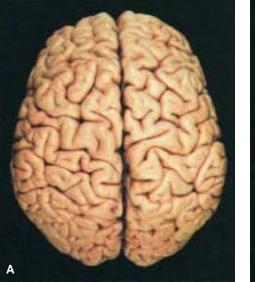
Alzheimer's Disease

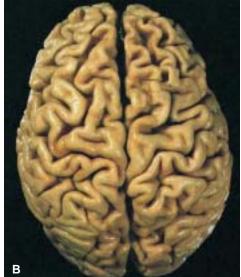
Dementia of the Alzheimer's type occurs in middle or late life and accounts for 50% to 70% of all cases of dementia. The disorder affects approximately 4 million Americans and may be the fourth leading cause of death in the United States.⁴⁵ The risk of Alzheimer's disease increases with age, and it occurs in nearly half of persons 85 years of age and older.

Pathophysiology

Alzheimer's disease is characterized by cortical atrophy and loss of neurons, particularly in the parietal and temporal lobes (Fig. 37-21). With significant atrophy, there is ventricular enlargement (*i.e.*, hydrocephalus) from the loss of brain tissue. Neurochemically, Alzheimer's disease has been associated with a decrease in the level of choline acetyltransferase activity in the cortex and hippocampus. This enzyme is required for the synthesis of acetylcholine, a neurotransmitter that is associated with memory. The reduction in choline acetyltransferase is quantitatively related to the numbers of neuritic plaques and severity of dementia.

The major microscopic features of Alzheimer's disease are the presence of amyloid-containing neuritic plaques and neurofibrillary tangles.⁸ The neurofibrillary tangles, which





■ FIGURE 37-21 ■ Alzheimer's disease. (A) Normal brain. (B) The brain of a patient with Alzheimer's disease shows cortical atrophy, characterized by slender gyri and prominent sulci. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 1511]. Philadelphia: Lippincott Williams & Wilkins)

are found in the cytoplasm of abnormal neurons, consist of fibrous proteins that are wound around each other in a helical fashion. These tangles are resistant to chemical or enzymatic breakdown, and they persist in brain tissue long after the neuron in which they arose has died and disappeared. The senile plaques are patches or flat areas composed of clusters of degenerating nerve terminals arranged around a central core of β -amyloid (A β).⁸ These plaques are found in areas of the cerebral cortex that are linked to intellectual function. Aß is a fragment of a much larger membrane-spanning amyloid precursor protein (APP). The function of APP is unclear, but it appears to be associated with the cytoskeleton of nerve fibers. Normally, the degradation of APP involves enzymatic cleavage, with formation of soluble nonpathogenic fragments. In Alzheimer's disease, cleavage of the APP molecule results in the formation of the less soluble A β peptide, which tends to aggregate into the amyloid fibrils found in the senile plaques.8

Some plaques and tangles can be found in the brains of older persons who do not show cognitive impairment. The number and distribution of the plaques and tangles appear to contribute to the intellectual deterioration that occurs with Alzheimer's disease. In persons with the disease, the plaques and tangles are found throughout the neocortex and in the hippocampus and amygdala, with relative sparing of the primary sensory cortex.¹ Hippocampal function in particular may be compromised by the pathologic changes that occur in Alzheimer's disease. The hippocampus is crucial to information processing, acquisition of new memories, and retrieval of old memories. The development of neurofibrillary tangles in the entorhinal cortex and superior portion of the hippocampal gyrus interferes with cortical input and output, thereby isolating the hippocampus from the remainder of the cortex and rendering it functionless.

It is likely that Alzheimer's disease is caused by several factors that interact differently in different persons. Progress on the genetics of inherited early-onset Alzheimer's disease shows

that mutations in at least three genes—the APP gene on chromosome 21; presenilin-1 (PS1), a gene on chromosome 14; and presenilin-2 (PS2), a gene on chromosome 1-can cause Alzheimer's disease in certain families.^{8,46,47} The APP gene is associated with an autosomal dominant form of early-onset Alzheimer's disease and can be tested clinically. Persons with Down's syndrome (trisomy 21) experience the pathologic changes of Alzheimer's disease and a comparable decline in cognitive functioning at a relatively young age. Virtually all persons with Down's syndrome who survive past 50 years experience the full-blown pathologic features of dementia. Because the APP gene is located on chromosome 21, it is thought that the additional dosage of the gene product in trisomy 21 predisposes to accumulation of BAP.8 There is some indication that PS1 and PS2 are mutant proteins that alter the processing of APP.⁸ A fourth gene, an allele of the apolipoprotein E gene, APOE e4, has been identified as a risk factor for late-onset Alzheimer's disease.

Clinical Course

Alzheimer's-type dementia follows an insidious and progressive course. The hallmark symptoms are loss of short-term memory and a denial of such memory loss, with eventual disorientation, impaired abstract thinking, apraxias, and changes in personality and affect. Three stages of Alzheimer's dementia have been identified, each characterized by progressive degenerative changes.

The *first stage*, which may last for 2 to 4 years, is characterized by short-term memory loss that often is difficult to differentiate from the normal forgetfulness that occurs in the elderly, and usually is reported by caregivers and denied by the patient. Although most elderly have trouble retrieving from memory incidental information and proper names, persons with Alzheimer's disease randomly forget important and unimportant details. They forget where things are placed, get lost easily, and have trouble remembering appointments and performing novel tasks. Mild changes in personality, such as lack of spontaneity, social withdrawal, and loss of a previous sense of humor, occur during this stage.

As the disease progresses, the person with Alzheimer's disease enters the second or confusional stage of dementia. This stage may last several years and is marked by a more global impairment of cognitive functioning. During this stage, there are changes in higher cortical functioning needed for language, spatial relationships, and problem solving. Depression may occur in persons who are aware of their deficits. There is extreme confusion, disorientation, lack of insight, and inability to carry out the activities of daily living. Personal hygiene is neglected, and language becomes impaired because of difficulty in remembering and retrieving words. Wandering, especially in the late afternoon or early evening, becomes a problem. The sundown syndrome, which is characterized by confusion, restlessness, agitation, and wandering, may become a daily occurrence late in the afternoon. Some persons may become hostile and abusive toward family members. Persons who enter this stage become unable to live alone and should be assisted in making decisions about supervised placement with family members or friends or in a community-based facility.

Stage 3 is the terminal stage. It usually is relatively short (1 to 2 years) compared with the other stages, but it has been known to last for as long as 10 years.48 The person becomes incontinent, apathetic, and unable to recognize family or friends. It usually is during this stage that the person is institutionalized.

Diagnosis and Treatment

Alzheimer's disease is essentially a diagnosis of exclusion. There are no peripheral biochemical markers or tests for the disease. The diagnosis can be confirmed only by microscopic examination of tissue obtained from a cerebral biopsy or at autopsy. The diagnosis is based on clinical findings. A diagnosis of Alzheimer's disease requires the presence of dementia established by clinical examination, mental status tests, and the absence of systemic or brain disorders that could account for the memory or cognitive deficits.^{48,49} Brain imaging, CT scan, or MRI is done to exclude other brain disease. Metabolic screening should be done for known reversible causes of dementia, such as vitamin B₁₂ deficiency, thyroid dysfunction, and electrolyte imbalance.

There is no curative treatment for Alzheimer's dementia. Drugs are used primarily to slow the progression and to control depression, agitation, or sleep disorders. Two major goals of care are maintaining the person's socialization and providing support for the family. Day care and respite centers are available in many areas to provide relief for caregivers and appropriate stimulation for the patient.

Although there is no current drug therapy that is curative for Alzheimer's disease, some show promise in terms of slowing the progress of the disease. Several drugs have been shown to be effective in slowing the progression of the disease by potentiating the available acetylcholine. The drugs-tacrine, donepezil, rivastigmine, and galantamine-inhibit acetylcholinesterase, preventing the metabolism of endogenous acetylcholine. Thus far, such therapy has not halted disease progression, but it can establish a meaningful plateau in decline.⁵⁰ There also is interest in the use of agents such as antioxidants (e.g., vitamin E, ginkgo) and anti-inflammatory agents to prevent or delay the onset of the disease.

Other Types of Dementia

Vascular Dementia

Dementia associated with cerebrovascular disease does not result directly from atherosclerosis, but rather is caused by multiple infarctions throughout the brain, thus the name vascular or multi-infarct dementia. Approximately 20% to 25% of dementias are vascular in origin, and the incidence is closely associated with hypertension. Other contributing factors are arrhythmias, myocardial infarction, peripheral vascular disease, diabetes mellitus, and smoking. The usual onset is between the ages of 55 and 70 years. The disease differs from Alzheimer's dementia in its presentation and tissue abnormalities. The onset may be gradual or abrupt, the course usually is a stepwise progression, and there should be focal neurologic symptoms related to local areas of infarction.

Pick's Disease

Pick's disease is a rare form of dementia characterized by atrophy of the frontal and temporal areas of the brain. The neurons in the affected areas contain cytoplasmic inclusions called *Pick* bodies. The average age at onset of Pick's disease is 38 years. The disease is more common in women than men. Behavioral manifestations may be noticed earlier than memory deficits, taking the form of a striking absence of concern and care, a loss of initiative, echolalia (i.e., automatic repetition of anything said to the person), hypotonia, and incontinence. The course of the disease is relentless, with death ensuing within 2 to 10 years. The immediate cause of death usually is infection.

Creutzfeldt-lakob Disease

Creutzfeldt-Jakob disease is a rare transmissible form of dementia thought to be caused by an infective protein agent called a prion.⁵¹ Similar diseases occur in animals, including scrapie in sheep and goats, and bovine spongiform encephalitis (BSE; mad cow disease) in cows. The pathogen is resistant to chemical and physical methods commonly used for sterilizing medical and surgical equipment. The disease reportedly has been transmitted through corneal transplants and human growth hormone obtained from cadavers.

Creutzfeldt-Jakob disease causes degeneration of the pyramidal and extrapyramidal systems and is distinguished most readily by its rapid course. Affected persons usually are demented within 6 months of onset. The disease is uniformly fatal, with death often occurring within months, although a few persons may survive for several years.¹ The early symptoms consist of abnormalities in personality and visual-spatial coordination. Extreme dementia, insomnia, and ataxia follow as the disease progresses.51

Wernicke-Korsakoff Syndrome

Wernicke-Korsakoff syndrome results from chronic alcoholism. Wernicke's disease is characterized by acute weakness and paralysis of the extraocular muscles, nystagmus, ataxia, and confusion. The affected person also may have signs of peripheral neuropathy. The person has an unsteady gait and reports diplopia. There may be signs attributable to alcohol withdrawal, such as delirium, confusion, and hallucinations. This disorder is caused by a deficiency of thiamine (vitamin B_1), and many of the symptoms are reversed when nutrition is improved with supplemental thiamine.

The Korsakoff component of the syndrome involves the chronic phase with severe impairment of recent memory. There often is difficulty in dealing with abstractions, and the person's capacity to learn is defective. Confabulation (*i.e.*, recitation of imaginary experiences to fill in gaps in memory) probably is the most distinctive feature of the disease. Polyneuritis also is common. Unlike Wernicke's disease, Korsakoff's psychosis does not improve significantly with treatment.

Huntington's Disease

Huntington's disease is a rare hereditary disorder characterized by chronic progressive chorea, psychological changes, and dementia. The disease is inherited as an autosomal dominant disorder with complete penetrance, meaning that anyone inheriting the gene will eventually experience the disease. The age of onset most commonly is in the fourth and fifth decades, often after affected persons have passed the gene on to their children.¹ Juvenile- or early-onset cases can occur and are more likely to be associated with inheritance from the father than from the mother.

The responsible gene (which encodes a protein called *huntingtin*) has located on chromosome 4.¹ The presence of the mutant huntingtin gene leads to localized death of brain cells. The first and most severely affected neurons are those of the basal ganglia. There is symmetric atrophy of caudate nucleus and lesser involvement of the putamen. There also is atrophy of the frontal cortex. Although the exact mechanisms whereby the mutant gene produces its effects is unclear, it is likely that it causes cell loss by some combination of activation of apoptotic pathways and impairment of normal metabolic processes in susceptible neurons.

Depression and personality changes are the most common early psychological manifestations; memory loss often is accompanied by impulsive behavior, moodiness, antisocial behavior, and a tendency toward emotional outbursts.⁵² Other early signs of the disease are lack of initiative, loss of spontaneity, and inability to concentrate. Fidgeting or restlessness may represent early signs of dyskinesia, followed by choreiform and some dystonic posturing. Eventually, progressive rigidity and akinesia (rather than chorea) develop in association with dementia.

There is no cure for Huntington's disease. The treatment is largely symptomatic. Drugs may be used to treat the dyskinesias and behavioral disturbances. The discovery of a marker probe for the gene locus has enabled testing that can predict whether a person will experience the disease.

In summary, cognitive disorders can be caused by any disorder that permanently damages large cortical or subcortical areas of the hemispheres. The most common cause of dementia is Alzheimer's disease, which is a major health problem among the elderly. It is characterized by cortical atrophy and loss of neurons, the presence of neuritic plaques, granulovacuolar degeneration, and cerebrovascular deposits of amyloid. The disease follows an insidious and progressive course that begins with memory impairment and terminates in an inability to recognize family or friends and the loss of control over bodily functions. Multi-infarct dementia is associated with vascular disease and Pick's disease with atrophy of the frontal and temporal lobes. Creutzfeldt-Jakob disease is a rare transmissible form of dementia. Wernicke-Korsakoff syndrome results from chronic alcoholism. Huntington's disease is a rare hereditary disorder characterized by chronic and progressive chorea, psychological change, and dementia.

REVIEW QUESTIONS

Differentiate cerebral hypoxia from ischemia and focal from global ischemia.

Compare cytotoxic and vasogenic cerebral edema in terms of pathophysiology and distribution with brain tissue.

Characterize the role of excitatory amino acids as a common pathway for neurologic disorders.

Compare the brain damage associated with concussion, diffuse axonal injury, contusion, and intracerebral hemorrhage and hematoma.

State the determinants of intracranial pressure and describe compensatory mechanisms used to prevent large changes in intracranial pressure when there are changes in brain, blood, and cerebrospinal fluid volumes.

Compare the causes of communicating and noncommunicating hydrocephalus.

Define consciousness and trace the rostral-to-caudal progression of unconsciousness in terms of arousal and cognition, pupillary changes, muscle tone and motor function, and respiration.

Compare the pathologies of ischemic and hemorrhagic stroke.

Compare the pathology, manifestations, and outcomes associated with meningitis and encephalitis.

List the major categories of brain tumors and interpret the meaning of *benign* and *malignant* as related to brain tumors.

Differentiate between the origin of seizure activity in partial and generalized forms of epilepsy and compare the manifestations of simple partial seizures with those of complex partial seizures and major and minor motor seizures.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- Kumar V., Cotran R.S., Robbins S.L. (2003). Basic pathology (7th ed., pp. 810–849). Philadelphia: W.B. Saunders.
- Guyton A.C., Hall J.E. (2001). *Textbook of medical physiology* (10th ed., pp. 192, 671–722). Philadelphia: W.B. Saunders.
- Meyer F.B. (1992). Brain metabolism, blood flow, and ischemic thresholds. In Awad I.A. (Ed.), *Neurosurgical topics: Cerebrovascular occlusive disease and brain ischemia* (pp. 1–24). Cleveland: American Association of Neurological Surgeons.

695 **Chapter 37: Alterations in Brain Function**

- 4. Richmond T.S. (1997). Cerebral resuscitation after global brain ischemia: Linking research to practice. AACN Clinical Issues 8, 171-181.
- 5. Lipton S.A., Rosenberg P.A. (1994). Excitatory amino acids as a final common pathway in neurologic disorders. New England Journal of Medicine 330, 613-622.
- 6. Hickey J.V. (1996). The clinical practice of neurological and neurosurgical nursing (4th ed., pp. 295-327, 569-584). Philadelphia: Lippincott-Raven
- 7. Lang E.W., Chestnut R.M. (1995). Intracranial pressure and cerebral perfusion pressure in severe head injury. New Horizons 3, 400-409.
- 8. Rubin E., Farber J.L. (1999). Pathology (3rd ed., pp. 1470-1473, 1509-1512). Philadelphia: Lippincott Williams & Wilkins.
- 9. Behrman R.E., Kliegman R.M., Jenson H.B. (2000). Nelson textbook of pediatrics (16th ed., pp. 1812-1813). Philadelphia: W.B. Saunders.
- 10. Ghajar J. (2000). Traumatic brain injury. Lancet 356, 923-929.
- 11. White R.J., Likavec M.J. (1992). The diagnosis and initial management of head injury. New England Journal of Medicine 327, 1507–1511.
- 12. Jennett B. (1996). Epidemiology of head injury. Journal of Neurology, Neurosurgery, and Psychiatry 60, 362-369.
- 13. Chestnut R.M. (1995). Secondary brain insults after head injury: Clinical perspectives. New Horizons 3, 366-375.
- 14. Teasdale G.M. (1995). Head injury. Journal of Neurology, Neurosurgery, and Psychiatry 58, 526-539.
- 15. Conn P.M. (1995). Neuroscience in medicine (pp. 232-235). Philadelphia: J.B. Lippincott.
- 16. Wijdicks E.F.M. (2001). The diagnosis of brain death. New England Journal of Medicine 344, 1215-1221.
- 17. Quality Standards Subcommittee of American Academy of Neurology. (1995). Practice parameters for determining brain death in adults. Neurology 45, 1012-1014.
- 18. Celesia G.G. (1993). Persistent vegetative state. Neurology 43, 1457-1458.
- 19. Quality Standards Subcommittee of American Academy of Neurology. (1995). Practice parameters: Assessment and management of patients with persistent vegetative state. Neurology 45, 1015–1018.
- 20. American Heart Association. (2002). 2002 heart and stroke statistical update. American Stroke Association. American Heart Association.
- 21. Stroke Council of the American Heart Association, Goldstein L.B. (Chairman). (2001). Primary prevention of ischemic stroke. Circulation 101, 161-182.
- 22. Gorelick P.B. (1987). Alcohol and stroke. Current Concepts in Cerebrovascular Disease 21 (5), 21.
- 23. Blank-Reid C. (1996). How to have a stroke at an early age: The effects of crack, cocaine and other illicit drugs. Journal of Neuroscience Nursing 28 (1), 19-27.
- 24. Albers W.A. (Chair). (1998). Antithrombotic and thrombolytic therapy for ischemic stroke. Chest 114, 683S-698S.
- 25. Zambramski J.M., Anson J.A. (1992). Diagnostic evaluation of ischemic cerebrovascular disease. In Awad I.A. (Ed.), Neurosurgical topics: Cerebrovascular occlusive disease and brain ischemia (pp. 73–101). Cleveland: American Association of Neurological Surgeons.
- 26. Qureshi A.I., Tuhrim S., Broderick J.P., et al. (2001). Spontaneous intracerebral hemorrhage. New England Journal of Medicine 344, 1450-1460
- 27. Gregory W. (Chair, Ad Hoc Committee on Guidelines for Management of Transient Ischemic Attacks, Stroke Council, American Heart Association). (1999). Supplement to the guidelines for transient ischemic attacks. Stroke 30, 2502-2511.
- 28. Adams H.P. (Chair). (1994). Guidelines for the management of patients with acute ischemic stroke: A statement for healthcare professionals from a Special Writing Group of the Stroke Council, American Heart Association. Stroke 25, 1901-1914.

- 29. Adams H.P. (Chair). (1996). Guidelines for thrombolytic therapy of acute stroke: A supplement to the guidelines for the management of patients with acute ischemic stroke: A statement for healthcare professionals from the Special Writing Group of the Stroke Council, American Heart Association. Circulation 94, 1167-1174.
- 30. Schievink W.I. (1997). Intracranial aneurysms. New England Journal of Medicine 336, 28-39.
- 31. Mayberg M.R. (Chair). (1994). Guidelines for the management of aneurysmal subarachnoid hemorrhage: A statement for healthcare professionals from a Special Writing Group of the Stroke Council, American Heart Association. Stroke 25, 2315–2327.
- 32. Arteriovenous Malformations Study Group. (1999). Arteriovenous malformations of the brain in adults. New England Journal of Medicine 340, 1812-1818.
- 33. Tunkel A.R., Scheld W.M. (1997). Issues in management of bacterial meningitis. American Family Physician 56, 1355-1365.
- 34. Quagliarello V.J., Scheld W.M. (1997). Treatment of bacterial meningitis. New England Journal of Medicine 336, 708-716.
- 35. Mehta N., Levin M. (2000). Management and prevention of meningococcal disease. Hospital Practice 35 (8), 75-86.
- 36. American Cancer Society. (2001). Brain and spinal cord cancers in adults. [On-line]. Available: http://www.cancer.org
- 37. DeAngelo L.M. (2001). Brain tumors. New England Journal of Medicine 344, 114-123.
- 38. Browne T.R., Holmes G.L. (2001). Epilepsy. New England Journal of Medicine 344, 1145-1151.
- 39. Mosewich R.K., So E.L. (1996). The clinical approach to classification of seizures and epileptic syndromes. Mayo Clinic Proceedings 71, 405-441.
- 40. Haslam R.H. (2000). The nervous system. In Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.), Nelson textbook of pediatrics (16th ed., pp. 1813-1829). Philadelphia: W.B. Saunders.
- 41. Commission on Classification and Terminology of the International League Against Epilepsy. (1989). Proposal for revised classification of epilepsies and epileptic syndromes. Epilepsia 30, 389-399.
- 42. Dichter M.A., Brodie M.J. (1996). New antiepileptic drugs. New England Journal of Medicine 334, 1583–1589.
- 43. Engel J. (1996). Surgery for seizures. New England Journal of Medicine 334, 647-652.
- 44. Cascino G.D. (1996). Generalized convulsive status epilepticus. New England Journal of Medicine 71, 787–792.
- 45. Morrison-Borgorad M., Phelps C., Buckholtz N. (1996). Alzheimer disease research comes of age. Journal of the American Medical Association 277, 837-840.
- 46. van Duijn C.M. (1996). Epidemiology of the dementias: Recent developments and new approaches. Journal of Neurology, Neurosurgery, and Psychiatry 60, 478-488.
- 47. Lendon C.L., Ashall F., Goate A.M. (1996). Exploring the etiology of Alzheimer's disease using molecular genetics. Journal of the American Medical Association 277, 825-831.
- 48. U.S. Department of Health and Human Services. (1996). Recognition and initial assessment of Alzheimer's disease and related disorders. AHCPR publication no. 97-0702. Washington, DC: Public Health Service, Agency for Health Care Policy and Research.
- 49. Morris J.C. (1997). Alzheimer's disease: A review of clinical assessment and management issues. Geriatrics 52 (Suppl. 2), S22-S25.
- 50. Mayeux R., Sano M. (1999). Treatment of Alzheimer's disease. New England Journal of Medicine 341, 1670-1679.
- 51. Prusiner S.B. (2001). Shattuck lecture: Neurodegenerative diseases and prions. New England Journal of Medicine 344, 1516-1526.
- 52. Martin J., Gusella J. (1987). Huntington's disease: Pathogenesis and management. New England Journal of Medicine 315, 1267-1276.

CHAPTER 38

Alterations in Neuromuscular Function

The Organization and Control of Motor Function The Motor Unit **The Motor Cortex Spinal Reflexes** Myotatic or Stretch Reflex **Crossed-Extensor Reflex** Disorders of Muscle Tone and Movement **Disorders of Muscle Tone Disorders of Muscle Movement Skeletal Muscle and Peripheral Nerve Disorders Skeletal Muscle Disorders Muscle Atrophy Muscular Dystrophy Neuromuscular Junction Disorders** Myasthenia Gravis **Peripheral Nerve Disorders** Peripheral Nerve Injury and Repair **Peripheral Neuropathies Back Pain and Intervertebral Disk Injury Back Pain** Herniated Intervertebral Disk **Basal Ganglia and Cerebellum Disorders** The Basal Ganglia Movement Disorders Parkinson's Disease The Cerebellum **Cerebellar Dysfunction Upper and Lower Motoneuron Disorders Amyotrophic Lateral Sclerosis Clinical Course Multiple Sclerosis Clinical Course** Diagnosis Treatment

Spinal Cord Injury Injury to the Vertebral Column Types and Classification of Spinal Cord Injury Primary and Secondary Injuries Classification Acute Spinal Cord Injury Disruption of Functional Abilities Somatosensory and Skeletal Muscle Function Autonomic Nervous System Function Other Functions

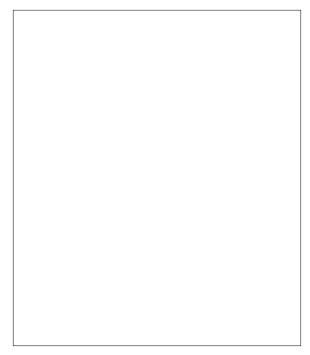
ffective motor function requires that muscles move and that the mechanics of their movement be programmed in a manner that provides for smooth and coordinated movement. In some cases, purposeless and disruptive movements can be almost as disabling as relative or complete absence of movement.

THE ORGANIZATION AND CONTROL OF MOTOR FUNCTION

As with other parts of the nervous system, the motor systems are organized in functional hierarchy, with each concerned with levels of function (Fig. 38-1). The highest level of function, which occurs at the level of the frontal cortex, is concerned with the purpose and planning of the movement.¹ The lowest level of the hierarchy occurs at the level of the spinal cord, which contains the basic reflex circuitry needed to coordinate the function of the motor units involved in the planned movement. Several anatomically distinct pathways project in parallel to the spinal cord from the higher motor centers. Above the spinal cord is the brain stem, and above the brain stem is the cerebellum and basal ganglia, structures that modulate the actions of the brain stem systems. Overseeing these supraspinal structures are the motor centers in the cerebral cortex.

The Motor Unit

The major effects of the elaborate processing of movement information that takes place in the brain has to do with contraction of skeletal muscles. The neurons that control skeletal



■ FIGURE 38-1 ■ The motor control system. The final common pathway transmits all central nervous system commands to the skeletal muscles. This path is influenced by sensory input from the muscle spindles and tendon organs (*dashed lines*) and descending signals from the cerebral cortex and brain stem. The cerebellum and basal ganglia influence the motor function indirectly, using brain stem and cortical pathways.

muscle contraction are referred to as *motoneurons* or sometimes as *alpha motoneurons*. A motor unit consists of one motoneuron and the group of muscle fibers it innervates in a skeletal muscle. The motoneurons supplying a motor unit are located in the ventral horn of the spinal cord and are called *lower motoneurons* (LMNs) (Fig. 38-2). The synapse between a LMN and the muscle fibers of a motor unit is called the *neuromuscular junction*. Upper motoneurons (UMNs), which exert control over LMNs, project from the motor strip in the cerebral cortex to the ventral horn of the spinal cord and are fully contained within the central nervous system (CNS).

Axons of the LMNs exit the spinal cord at each segment to innervate skeletal muscle cells, including those of the limbs, back, abdomen, and chest. Each LMN undergoes multiple branching, making it possible for a single LMN to innervate 10 to 2000 muscle cells. In general, large muscles—those containing hundreds or thousands of muscle cells and providing gross motor movement—have large motor units. This sharply contrasts with those that control the hand, tongue, and eye movements, for which the motor units are small and permit very discrete control.

In addition to the output from LMNs that innervate the motor unit, the body uses information from a vast array of sensory input to ensure the generation of correct patterns of muscle activity. Much of this information goes to spinal cord reflexes that control muscle tone and coordinate the movement of the extensor and flexor muscles used in walking and other motor activities.



The Motor Cortex

Delicate, skillful, intentional movement of distal and especially flexor muscles of the limbs and the speech apparatus is initiated and controlled from the motor cortex located in the posterior part of the frontal lobe. It consists of the primary, premotor, and supplementary motor cortex² (Fig. 38-3). These areas receive information from the thalamus and the somesthetic (sensory) cortex and indirectly from the cerebellum and basal ganglia.

KEY CONCEPTS

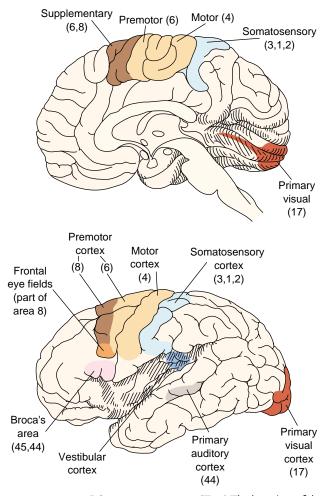
MOTOR SYSTEMS

- Motor systems require upper motoneurons (UMNs) that project from the motor cortex to the brain stem or spinal cord, where they directly or indirectly innervate the lower motoneurons (LMNs) of the contracting muscles; sensory feedback from the involved muscles that is continuously relayed to the cerebellum, basal ganglia, and sensory cortex; and a functioning neuromuscular junction that links nervous system activity with muscle contraction.
- Input from the basal ganglia and cerebellum provides the background for the more crude, supportive movement patterns.
- The efficiency of the movement by the motor system depends on a background of muscle tone provided by the stretch reflex and vestibular system input to maintain stable postural support.

The primary motor cortex (area 4), also called the motor strip, is located on the rostral surface and adjacent portions of the central sulcus. The primary motor cortex controls discrete muscle movement sequences and is the first level of descending control for precise movements. The neurons in the primary motor cortex are arranged in a somatotopic array or distorted map of the body called the *motor homunculus* (Fig. 38-4).³ The body parts that require the greatest dexterity have the largest cortical areas devoted to them. More than one half of the primary motor cortex is concerned with controlling the muscles of the hands, of facial expression, and of speech.²

The premotor cortex (areas 6 and 8), which is located just anterior to the primary motor cortex, sends some fibers into the corticospinal tract but mainly innervates the primary motor strip. A movement pattern to accomplish a particular objective, such as throwing a ball or picking up a fork, is programmed by the prefrontal association cortex and associated thalamic nuclei. The *supplementary motor cortex*, which contains representations of all parts of the body, is located on the medial surface of the hemisphere (areas 6 and 8) in the premotor region. It is intimately involved in the performance of complex, skillful movements that involve both sides of the body.

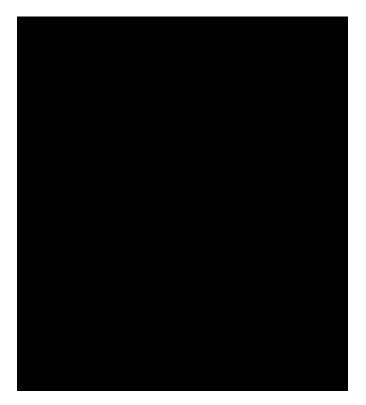
The primary motor cortex contains many layers of pyramidshaped output neurons that project to the same side of the cortex (*i.e.*, premotor and somesthetic areas), project to the opposite side of the cortex, or descend to subcortical structures such as the basal ganglia and thalamus. The large pyramidal cells located in the fifth layer project to the brain stem and spinal cord. The axons of these UMNs project through the subcortical white matter and internal capsule to the deep surface of the brain stem, through the ventral bulge of the pons, and to the ventral surface of the medulla, where they form a ridge or pyramid (see Fig. 38-1). At the junction between the medulla and cervical spinal cord, 80% or more of the UMN axons cross the midline to form the lateral corticospinal tract in the lateral white matter of the spinal cord. This tract extends throughout the spinal



■ FIGURE 38-3 ■ Primary motor cortex. (**Top**) The location of the primary, premotor, and supplementary cortex on the medial surface of the brain. (**Bottom**) The location of the primary and premotor cortex on the lateral surface of the brain. (Courtesy of Carole Russell Hilmer, C.M.I.)

cord, with roughly 50% of the fibers terminating in the cervical segments, 20% in the thoracic segments, and 30% in the lumbosacral segments. Most of the remaining uncrossed fibers travel down the ventral column of the cord, mainly to cervical levels, where they cross and innervate contralateral LMNs.

By convention, motor tracts have been classified as belonging to one of two motor systems: the pyramidal and extrapyramidal systems. According to this classification system, the pyramidal system consists of the motor pathways originating in the motor cortex and terminating in the corticobulbar fibers in the brain stem and the corticospinal fibers in the spinal cord. The corticospinal fibers traverse the ventral surface of the medulla in a bundle called the pyramid before decussating or crossing to the opposite side of the brain at the medulla-spinal cord junction, thus the name pyramidal system. Other fibers from the cortex and basal ganglia also project to the brain stem reticular formation and reticulospinal systems, following a more ancient pathway to LMNs of proximal and extensor muscles. These fibers do not decussate in the pyramids, thus the name extrapyramidal system. Disorders of the pyramidal tracts (e.g., stroke) are characterized by spasticity and paralysis and those affecting



the extrapyramidal tracts (e.g., Parkinson's disease) by involuntary movements, muscle rigidity, and immobility without paralysis. This classification is no longer used extensively. As increased knowledge regarding motor pathways has emerged, it has become evident that the extrapyramidal and pyramidal systems are extensively interconnected and cooperate in the control of movement.1

Spinal Reflexes

Reflexes are coordinated, involuntary motor responses initiated by a stimulus applied to peripheral receptors. Some reflexes, such as the flexion-withdrawal reflex, initiate movements to avoid hazardous situations; whereas others, such as the stretch or crossed-extensor reflex, serve to integrate motor movements so they function in a coordinated manner. The anatomic basis of a reflex consists of (1) an afferent neuron, (2) the connection or synapse with CNS interneurons that communicate with the effector neuron, and (3) the effector neuron that innervates a muscle. Reflexes are essentially "wired into" the CNS so that they are always ready to function; with training, most reflexes can be modulated to become parts of more complicated movements. A reflex may involve neurons in a single cord segment (*i.e.*, segmental reflexes), several or many segments (*i.e.*, intersegmental reflexes), or structures in the brain (i.e., suprasegmental reflexes).

Myotatic or Stretch Reflex

The myotatic ("myo" from the Greek for "muscle," "tatic" from the Greek for "stretch") or stretch reflex controls muscle tone and helps maintain posture. Stretch reflexes can be evoked in many muscles throughout the body and are routinely tested (e.g., knee-jerk reflex) during the clinical examination for the diagnosis of neurologic conditions. Disorders of muscle tone

caused by dysregulated function of the stretch reflex are seen in persons with conditions such as spinal cord injury and stroke.

The myotatic reflex uses specialized afferent sensory endings in skeletal muscles and tendons to relay information regarding the sense of body position, movement, and muscle tone to the CNS. Information from these sensory afferents is relayed to the cerebellum and cerebral cortex and is experienced as the sense of body movement and position (proprio*ception*). To provide this information, the muscles and their tendons are supplied with two types of stretch receptors: muscle spindle receptors and Golgi tendon organs (Fig. 38-5A). The muscle spindles, which are distributed throughout the belly of a muscle, provide information about muscle length and rate of stretch. The Golgi tendon organs are found in muscle tendons and transmit information about muscle tension or force of contraction at the junction of the muscle and the tendon that attaches to bone. A likely role of the tendon organs is to equalize the contractile forces of the separate muscle groups, spreading the load over all the fibers to prevent the local muscle damage that might occur when small numbers of fibers are overloaded.

The muscle spindles consist of a group of specialized miniature skeletal muscle fibers called intrafusal fibers that are encased in a connective tissue capsule and attached to muscle fibers (*i.e.*, extrafusal fibers) of a skeletal muscle (Fig. 38-5A). An afferent sensory neuron, which spirals around the intrafusal fibers, transmits information to the spinal cord. The extrafusal fibers and the intrafusal fibers are innervated by motoneurons that reside in the ventral horns of the spinal cord. Extrafusal fibers are innervated by large alpha motoneurons that produce contraction of the muscle. The intrafusal fibers are innervated by gamma motoneurons that adjust the length of the intrafusal fibers to match that of the extrafusal fibers.

The intrafusal muscle fibers function as stretch receptors. When a skeletal muscle is stretched, the spindle and its intrafusal fibers are stretched, resulting in increased firing of its afferent fibers. The increased firing of the afferent neurons synaptically depolarizes the alpha motoneuron fibers. This causes the extrafusal muscle fibers to contract, thereby shortening the muscle. The knee-jerk reflex that occurs when the knee is tapped with a reflex hammer tests for the intactness of the myotatic reflex arc in the quadriceps muscle (Fig. 38-5B).

Axons of the spindle afferent neurons enter the spinal cord through the several branches of the dorsal root. Some branches end in the segment of entry, and others ascend in the dorsal column of the cord to the medulla of the brain stem. Segmental branches make connections, along with other branches, that pass directly to the anterior gray matter of the spinal cord and establish monosynaptic contact with each of the LMNs that have motor units in the muscle containing the spindle receptor. This produces an opposing muscle contraction. Another segmental branch of the same afferent neuron innervates an internuncial neuron that is inhibitory to motor units of antagonistic muscle groups. Inhibition of these muscle units helps in opposing muscle stretch. Branches of the afferent axon also ascend into the dorsal horn of the adjacent segments, influencing intersegmental reflex function. Ascending fibers from the stretch reflex ultimately provide information about muscle length to the cerebellum and cerebral cortex.

The role of afferent spindle fibers is to inform the CNS of the status of muscle length. When a skeletal muscle lengthens or

■ FIGURE 38-5 ■ (A) Spinal cord innervation of the muscle spindles. Cell bodies from both the alpha motoneuron that innervates the extrafusal muscle fiber and the gamma motoneuron that innervates the intrafusal fiber reside in the ventral horn of the same segment of the spinal cord. (B) The knee-jerk reflex. Stretching of the extrafusal fiber by tapping with a reflex hammer leads to lengthening of the intrafusal fiber and increased firing of the type la afferent fiber. Impulses from the la fiber enter the dorsal horn of the spinal cord and make monosynaptic contact with the ventral horn alpha motoneuron supplying the extrafusal fibers in the quadriceps muscle. The resultant contraction (shortening) of the guadriceps muscle is responsible for the knee-jerk response.

shortens against tension, a feedback mechanism needs to be available for readjustment such that the spindle apparatus remains sensitive to moment-to-moment changes in muscle stretch, even while changes in muscle length are occurring. This is accomplished by the gamma motoneurons that adjust spindle fiber length to match the length of the extrafusal muscle fiber. Descending fibers of motor pathways synapse with and simultaneously activate both alpha and gamma motoneurons so that the sensitivity of the spindle fibers is coordinated with muscle movement.

Central control over the gamma LMN mechanism permits increases or decreases in muscle tone in anticipation of changes in the muscle force required to oppose ongoing conditions, such as when weight is about to be lifted. The CNS, through its coordinated control of the muscle's alpha LMNs and the spindle's gamma LMNs, can suppress the stretch reflex. This occurs during centrally programmed movements, such as pitching a baseball, permitting the muscle to produce its greatest range of motion. Without this programmed adjustability of the stretch reflex, any movement is immediately opposed and prevented.

Crossed-Extensor Reflex

The crossed-extensor reflex, in which the limb on one side of the body extends as the limb on the other side relaxes, provides the basis for postural stability during walking (Fig. 38-6).⁴ For example, when the crossed-extensor reflex produces relaxation of antigravity muscles (with flexion) of one leg as we walk, the contralateral component produces contraction and extension of the opposite leg. Intersegmental connections of the crossed-extensor reflex between the lumbar and cervical spinal segments also accounts for the swinging of the arms that accompanies walking.

Disorders of Muscle Tone and Movement

Disorders of Muscle Tone

In the muscles that are supporting body weight, the stretch reflex operates continuously, producing a continuous resistance to passive stretch called *muscle tone*. Muscle tone is evidenced by the resistance to passive movement around a joint. Disorders of skeletal muscle tone are characteristic of many nervous system pathologies. Any interruption of the myotatic reflex circuit by peripheral nerve injury, pathology of the neuromuscular junction and of skeletal muscle fibers, damage to the corticospinal system, or injury to the spinal cord or spinal nerve root results in disturbance of muscle tone. Muscle tone may be



Chapter 38: Alterations in Neuromuscular Function

701

described as less than normal (hypotonia), absent (flaccidity), or excessive (hypertonia, rigidity, spasticity, or tetany).

Reduced excitability of the stretch reflex results in decreased muscle tone, or *hypotonia*, ranging from postural weakness to total flaccid paralysis. It can result from decreased function of the descending facilitatory systems controlling the gamma LMNs that innervate the muscle or damage to the stretch reflex or peripheral nerves innervating the muscle.

Hypertonia, or spasticity, is an abnormal increase in muscle tone. It can result from increased excitation or loss of inhibition of the spindle's gamma LMNs or changes in the segmental spinal cord circuitry controlling the stretch reflex. It is characterized by hyperactive tendon reflexes and an increase in resistance to rapid muscle stretch. Spasticity commonly occurs with UMN lesions such as those that exist after spinal shock in persons with spinal cord injury.

Rigidity is a greatly increased resistance to movements in all directions. It is caused by increased activation of the alpha LMNs innervating the extrafusal muscle fibers and does not depend on the dorsal root innervation of the intrafusal spindle fibers. It is seen in conditions, such as Parkinson's disease, in which descending CNS inhibition of alpha LMNs is impaired.

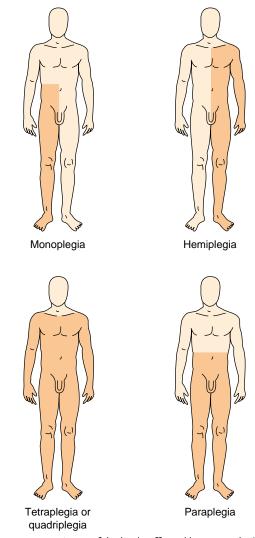
Clonus is the rhythmic contraction and alternate relaxation of a limb that is caused by suddenly stretching a muscle and gently maintaining it in the stretched position. It is seen in the hypertonia of spasticity associated with UMN lesions, such as spinal cord injury. It is caused by an oscillating stimulation of the muscle spindles that occurs when the spindle fibers are activated by an initial muscle stretch. This results in reflex contraction of the muscle and unloading of the spindle fibers with decreased afferent activity. The reduced spindle fiber activity causes the muscle to relax, which causes the spindle fiber to stretch again, and the cycle starts over.

Disorders of Muscle Movement

The suffix *plegia* comes from the Greek word for a blow, a stroke, or paralysis. The terms that are used to describe the extent and anatomic location of motor damage include *paralysis*, meaning loss of movement, and *paresis*, implying weakness or incomplete loss of muscle function. *Monoparesis* or *monoplegia* results from the destruction of pyramidal UMN innervation of one limb; *hemiparesis* or *hemiplegia*, both limbs on one side; *diparesis* or *diplegia* or *paraparesis* or *paraplegia*, both upper or lower limbs; and *tetraparesis* or *tetraplegia*, also called *quadriparesis* or *quadriplegia*, all four limbs⁵ (Fig. 38-7). Paresis or paralysis can be further designated as of UMN or LMN origin.

Upper Motoneuron Lesions. A UMN lesion can involve any part of the CNS: the motor cortex, the internal capsule, or other brain structures through which the corticospinal or corticobulbar tracts descend, or the spinal cord. When the lesion is at or above the level of the pyramids, paralysis affects structures on the opposite side of the body. In UMN disorders involving injury to the L1 level or above, there is an immediate, profound weakness and loss of fine, skilled voluntary lower limb movement, reduced bowel and bladder control, and diminished sexual functioning, followed by an exaggeration of muscle tone. With UMN damage above C7, upper limb movement also is affected (see section on spinal cord injury).

With UMN lesions, the LMN spinal reflexes remain intact, but communication and control from higher brain centers are



■ **FIGURE 38-7** ■ Areas of the body affected by monoplegia, hemiplegia, tetraplegia or quadriplegia, and paraplegia. The *shaded area* shows the extent of motor and sensory loss.

lost. Descending excitatory influences from the pyramidal system and some descending inhibitory influences from other cortical regions are lost after injury, resulting in immediate weakness accompanied by the loss of control of delicate, skilled movements. After several weeks, this weakness becomes converted to hypertonicity or spasticity, which is manifested by an initial increased resistance (stiffness) to the passive movement of a joint at the extremes of range of motion followed by a sudden or gradual release of resistance. The spasticity often is greatest in the flexor muscles of the upper limbs and extensor muscles of the lower limbs. Sometimes, a lesion of the pyramidal tract is less severe and results in a relatively minor degree of weakness. In this case, the finer and more skilled movements are most severely impaired.

Lower Motoneuron Lesions. In contrast to UMN lesions, in which the spinal reflexes remain intact, LMN disorders disrupt communication between the muscle and all neural input from spinal cord reflexes, including the stretch reflex, which maintains muscle tone.

KEY CONCEPTS

SPASTIC VERSUS FLACCID PARALYSIS

- Afferent input from stretch receptors located in muscles and joints is incorporated into spinal cord reflexes that control muscle tone. The activity of the spinal cord reflexes that control muscle tone is constantly monitored and regulated by input from higher brain centers.
- Upper motoneuron lesions that interrupt communication between the spinal cord reflexes and higher brain centers result in unregulated reflex activity, increased muscle tone, and spastic paralysis.
- Lower motoneuron lesions that interrupt communication between the muscle and the spinal cord reflex result in a loss of reflex activity, decreased or absent muscle tone, and flaccid paralysis.

Infection or irritation of the cell body of the LMN or its axon can lead to hyperexcitability, which causes spontaneous contractions of the muscle units. These can be observed as twitching and squirming movements on the muscle surface, a condition called *fasciculations*. Toxic agents, such as the tetanus toxin, produce extreme hyperexcitability of the LMN, which results in continuous firing at maximum rate. The resultant sustained contraction of the muscles is called *tetany*. Tetany of muscles on both sides of a joint produces immobility or tetanic paralysis. When a virus, such as the poliomyelitis virus, attacks an LMN, it first irritates the LMN, causing fasciculations to occur. These fasciculations often are followed by death of LMNs. Weakness and severe muscle wasting or denervation atrophy result. If muscles are totally denervated, total weakness and total loss of reflexes, called *flaccid paralysis*, occurs.

With complete LMN lesions, the muscles of the affected limbs, bowel, bladder, and genital areas become atonic, and it is impossible to elicit contraction by stretching the tendons. One of the outstanding features of LMN lesions is the profound development of muscle atrophy. Damage to an LMN with or without spinal cord damage, often called *peripheral nerve injury*, may occur at any level of the spinal cord. For example, a C7 peripheral nerve injury leads to LMN hand weakness only. All segments below the level of injury that have intact LMNs manifest UMN signs. Usually, injury to the spinal cord at the T12 level or below results in LMN injury and flaccid paralysis to all areas below the level of injury. This occurs because the spinal cord ends at the T12 to L1 level, and from this level, the spinal roots of the LMNs continue caudally in the vertebral canal as part of the cauda equina.

In summary, motor function involves the neuromuscular unit, spinal cord circuitry, brain stem neurons, the cerebellum, the basal ganglia, and the motor cortex. A motor unit consists of one LMN and the group of muscle fibers it innervates in the muscle. Delicate, skillful, intentional movement of distal and especially flexor muscles of the limbs and the speech apparatus is initiated and controlled from the motor cortex located in the posterior frontal lobe. It consists of the primary, premotor, and supplementary motor cortex. These areas receive information from the thalamus and somesthetic cortex and, indirectly, from the cerebellum and basal ganglia. The UMNs in the motor cortex send their axons through the subcortical white matter and internal capsule and the deep surface of the brain stem to the ventral surface to the opposite side of the medulla, where they form a pyramid before crossing the midline to form the lateral corticospinal tract in the spinal cord.

Alterations in musculoskeletal function include muscle tone and movement. Muscle tone is maintained through the combined function of the muscle spindle system and the CNS centers that monitor and buffer UMN innervation of the LMNs. Hypotonia is a condition of less-than-normal muscle tone, and hypertonia or spasticity is a condition of excessive tone. Paresis refers to weakness in muscle function, and paralysis refers to a loss of muscle movement. UMN lesions produce spastic paralysis, and LMN lesions produce flaccid paralysis.

SKELETAL MUSCLE AND PERIPHERAL NERVE DISORDERS

Skeletal Muscle Disorders

Muscle Atrophy

Atrophy describes a decrease in muscle mass. Maintenance of muscle strength requires relatively frequent movements against resistance. When a normally innervated muscle is not used for long periods, the muscle cells shrink in diameter, and although the muscle cells do not die, they lose much of their contractile protein and become weakened. This is called *disuse atrophy*, and it occurs with conditions such as bed rest, application of a cast for fracture healing, and chronic illness.

The most extreme examples of muscle atrophy are found in persons with disorders, such as spinal cord injury, that deprive muscles of their innervation. This form is called *denervation atrophy*. Denervated muscles lose more than half their original bulk within 2 to 3 months.

Muscular Dystrophy

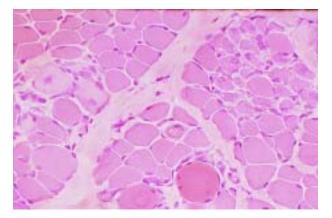
The term dystrophy refers to abnormal growth. *Muscular dystrophy* is a term applied to a number of genetic disorders that produce progressive deterioration of skeletal muscles because of mixed muscle cell hypertrophy, atrophy, and necrosis. They are primary diseases of muscle tissue and probably do not involve the nervous system. As the muscle undergoes necrosis, fat and connective tissue replace the muscle fibers, which increases muscle size and results in muscle weakness. The increase in muscle size resulting from connective tissue infiltration is called *pseudohypertrophy*. The muscle weakness is insidious in onset but continually progressive, varying with the type of disorder.

The most common form of muscular dystrophy is *Duchenne muscular dystrophy*, which has an incidence of approximately 3 cases per 100,000 male children. The disorder is inherited as a recessive single-gene defect on the X chromosome and is transmitted from the mother to her male offspring (see Chapter 4).^{6,7} Despite the X-linked inheritance pattern, about 30% of cases are new mutations and the mother is not the carrier.⁸ Another form of muscular dystrophy, *Becker muscular dystrophy*, is similarly X-linked but manifests later in childhood or adolescence and has a slower course.

The Duchenne muscular dystrophy mutation results in a defective form of a very large protein associated with the muscle cell membrane, called *dystrophin*, which fails to provide the normal attachment site for the contractile proteins. As a result there is necrosis of muscle fibers, a continuous effort at repair and regeneration, and progressive necrosis (Fig. 38-8).⁷

Clinical Course. The postural muscles of the hip and shoulder are affected first in the Duchenne type muscular dystrophy, and the child usually has no problems until approximately 3 years of age, when frequent falling begins to occur. Imbalances between agonist and antagonist muscles lead to abnormal postures and the development of contractures and joint immobility. Scoliosis is common. Wheelchairs usually are needed at approximately 7 to 12 years of age.⁸ The function of the distal muscles usually is preserved well enough that the child can continue to use eating utensils and a computer keyboard. The function of the extraocular nerves also is well preserved, as is the function of the muscles controlling urination and defecation. Incontinence is an uncommon and late event. Respiratory muscle involvement results in weak and ineffective cough, frequent respiratory infections, and decreasing respiratory reserve. Cardiomyopathy caused by involvement of cardiac muscle is a common feature of the disease. However, the severity of cardiac involvement does not necessarily correlate with skeletal muscle weakness. Some patients die at an early age of severe cardiomyopathy, whereas others maintain adequate cardiac function until the terminal stages of the disease. Death from respiratory and cardiac muscle involvement usually occurs in young adulthood.

Observation of the child's voluntary movement and a complete family history provide important diagnostic data for the disease. Serum levels of the enzyme creatine kinase, which



■ FIGURE 38-8 ■ Duchenne muscular dystrophy: hematoxylin and eosin stain. A section of the vastus lateralis muscle shows necrotic muscle fibers, some of them invaded by macrophages. The endomysial septa are thickened, indicating fibrosis. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 1422]. Philadelphia: Lippincott Williams & Wilkins)

leaks out of damaged muscle fibers, can be used to confirm the diagnosis. Muscle biopsy, which shows a mixture of muscle cell degeneration and regeneration and reveals fat and scar tissue replacement, is diagnostic of the disorder. Echocardiography, electrocardiography, and chest radiography are used to assess cardiac function. A specific molecular genetic diagnosis is possible by demonstrating the defective dystrophin in muscle biopsy tissue or by DNA analysis from the peripheral blood. The same methods of DNA analysis may be used on blood samples to establish carrier status in female relatives at risk, such as sisters and cousins. Prenatal diagnosis is possible as early as 12 weeks' gestation by sampling chorionic villi for DNA analysis (see Chapter 4).⁸

Management of the disease is directed toward maintaining ambulation and preventing deformities. Passive stretching, correct or counter posturing, and splints help to prevent deformities. Precautions should be taken to avoid respiratory infections. Although there have been exciting advances in identifying the gene and gene product involved in Duchenne muscular dystrophy, there is no known cure.

Neuromuscular Junction Disorders

The neuromuscular junction serves as a synapse between a motor neuron and a skeletal muscle fiber. It consists of the axon terminals of a motor neuron and a specialized region of the muscle membrane called the end-plate. The transmission of impulses at the neuromuscular junction is mediated by the release of the neurotransmitter acetylcholine from the axon terminals. Acetylcholine binds to specific receptors in the end-plate region of the muscle fiber surface to cause muscle contraction (Fig. 38-9). Acetylcholine is active in the neuromuscular junction only for a brief period, during which an action potential is generated in the innervated muscle cell. Some of the transmitter diffuses out of the synapse, and the remaining transmitter is rapidly inactivated by an enzyme called acetylcholinesterase. The rapid inactivation of acetylcholine allows repeated muscle contractions and gradations of contractile force.

A number of drugs and agents can alter neuromuscular function by changing the release, inactivation, or receptor binding of acetylcholine.⁹ Curare acts on the postjunctional membrane of the motor end-plate to prevent the depolarizing effect of the neurotransmitter. Blocking of neuromuscular transmission by curare-type drugs is used during many types of surgical procedures to facilitate relaxation of involved musculature. Drugs such as physostigmine and neostigmine inhibit the action of acetylcholinesterase and allow acetylcholine released from the motoneuron to accumulate. These drugs are used in the treatment of myasthenia gravis.

Toxins from the botulism organism (*Clostridium botulinum*) produce paralysis by blocking acetylcholine release. Spores from the botulism organism may be found in soil-grown foods that are not properly cooked. A pharmacologic preparation of the botulism toxin (botulism toxin type A [Botox]) has become available for use in treating eyelid and eye movement disorders such as blepharospasm and strabismus. It also is used for treatment of spasmodic torticollis, spasmodic dysphonias (laryngeal dystonia), and other dystonias. The drug is injected into the target muscle using the electrical activity recorded from the tip of a special electromyographic injection needle to guide the

Dr Murtadha Alshareifi e-Library

■ FIGURE 38-9 ■ Neuromuscular junction. (A) Acetylcholine (ACh) released from the motoneurons in the myoneural junction crosses the synaptic space to reach receptors that are concentrated in the folds of the end-plate of the muscle fiber. Once released, ACh is rapidly broken down by the enzyme acetylcholinesterase (AChesterase). (B) Decrease in ACh receptors in myasthenia gravis.

injection. The treatment is not permanent and usually needs to be repeated approximately every 3 months.

The organophosphates (e.g., malathion, parathion) that are used in some insecticides bind acetylcholinesterase to prevent the breakdown of acetylcholine. They produce excessive and prolonged acetylcholine action with a depolarization block of cholinergic receptors, including those of the neuromuscular junction.9 The organophosphates are well absorbed from the skin, lungs, gut, and conjunctiva of the eye, making them particularly effective as insecticides but also potentially dangerous to humans. Malathion and certain other organophosphates are rapidly metabolized to inactive products in humans and are considered safe for sale to the general public. The sale of other insecticides, such as parathion, which is not effectively metabolized to inactive products, has been banned. Other organophosphate compounds (e.g., soman) were developed as "nerve gases"; if absorbed in high enough concentrations, they have lethal effects from depolarization block and loss of respiratory muscle function.

Myasthenia Gravis

Myasthenia gravis is a disorder of transmission at the neuromuscular junction that affects communication between the motoneuron and the innervated muscle cell. The disease may occur at any age, but the peak incidence occurs between 20 and 30 years of age and affects women more often than men. A smaller second peak occurs in later life and affects men more often than women.

Now recognized as an autoimmune disease, the disorder is caused by antibody-mediated loss of acetylcholine receptors in the neuromuscular junction (Fig. 38–9).^{6,10} Although the exact mechanism that triggers the autoimmune response is unclear, it is thought to be related to abnormal T-lymphocyte function. Approximately two thirds of persons with myasthenia gravis also have thymic abnormalities, such as a thymoma (*i.e.*, thymus tumor) or thymic hyperplasia.^{6,10}

In persons with myasthenia gravis who have fewer acetylcholine receptors in the postsynaptic membrane, each release of acetylcholine from the presynaptic membrane results in a lower-amplitude end-plate potential. This results in both muscle weakness and fatigability with sustained effort. Most commonly affected are the eye and preorbital muscles. Either ptosis caused by eyelid weakness or diplopia caused by weakness of the extraocular muscles is an initial symptom in approximately 50% of persons with the disease. The disease may progress from ocular muscle weakness to generalized weakness, including respiratory muscle weakness. Chewing and swallowing may be difficult, and persons with the disease often choose to eat soft foods and cereals, rather than meats and hard fruit. Weakness in limb movement usually is more pronounced in proximal than in distal parts of the extremity, so climbing stairs and lifting objects are difficult. As the disease progresses, the muscles of the lower face are affected, causing speech impairment. When this happens, the person often supports the chin with one hand to assist in speaking. In most persons, symptoms are least evident when arising in the morning, but they grow worse with effort and as the day proceeds.

Persons with myasthenia gravis may experience a sudden exacerbation of symptoms and weakness known as *myasthenia crisis*.¹¹ Myasthenia crisis occurs when muscle weakness becomes severe enough to compromise ventilation to the extent that ventilatory support and airway protection are needed. This usually occurs during a period of stress, such as infection, emotional upset, pregnancy, alcohol ingestion, cold, or after surgery. It also can result from inadequate or excessive doses of the anticholinesterase drugs used in treatment of the disorder.

Diagnosis and Treatment. The diagnosis of myasthenia gravis is based on history and physical examination, the anticholinesterase test, nerve stimulation studies, and an assay for acetylcholine receptor antibodies. The anticholinesterase test uses a drug that inhibits acetylcholinesterase, the enzyme that breaks down acetylcholine. When weakness is caused by myasthenia gravis, a dramatic transitory improvement in muscle function occurs when the drug used for the test is administered. Electrophysiologic studies can be done to demonstrate a muscle response to repetitive stimulation of motor nerves. An immunoassay test can be used to detect the presence of acetylcholine receptor antibodies circulating in the blood.

Treatment methods include the use of pharmacologic agents; immunosuppressive therapy, including corticosteroid drugs; management of myasthenic crisis; thymectomy; and plasmapheresis or intravenous immunoglobulin.¹⁰ Pharmacologic treatment with reversible anticholinesterase drugs inhibits the

Chapter 38: Alterations in Neuromuscular Function

breakdown of acetylcholine at the neuromuscular junction by acetylcholinesterase. Corticosteroid drugs, which suppress the immune response, are used in cases of a poor response to anticholinesterase drugs and thymectomy. Immunosuppressant drugs (*e.g.*, azathioprine, cyclosporine) also may be used, often in combination with plasmapheresis.

Plasmapheresis removes antibodies from the circulation and provides short-term clinical improvement. It is used primarily to stabilize the condition of persons in myasthenic crisis or for short-term treatment in persons undergoing thymectomy. Intravenous immunoglobulin also produces improvement in persons with myasthenia gravis. Although the effect is temporary, it may last for weeks to months. The indications for its use are similar to those for plasmapheresis.

Thymectomy, or surgical removal of the thymus, may be used as a treatment for myasthenia gravis. Because the mechanism whereby surgery exerts its effect is unknown, the treatment is a matter of controversy. Thymectomy is performed in persons with thymoma, regardless of age, and in persons 50 to 60 years of age or older with recent onset of moderate disease.

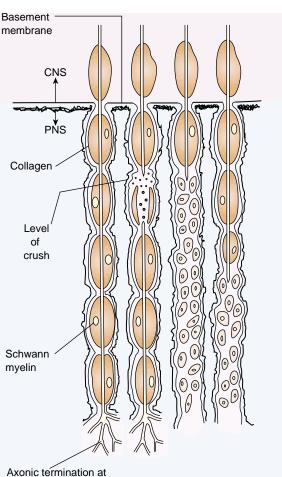
Peripheral Nerve Disorders

The peripheral nervous system consists of the motor and sensory branches of the cranial and spinal nerves, the peripheral parts of the autonomic nervous system (ANS), and peripheral ganglia. Unlike the nerves of the CNS, peripheral nerves are fairly strong and resilient. They contain a series of connective tissue sheaths that enclose their nerve fibers. An outer fibrous sheath called the *epineurium* surrounds the medium to large nerves; inside, a sheath called the *perineurium* invests each bundle of nerve fibers, and within each bundle, a delicate sheath of connective tissue known as the *endoneurium* surrounds each nerve fiber (see Chapter 36, Fig. 36-3). Small peripheral nerves lack the epineurial covering. In its endoneurial sheath, each nerve fiber is invested by a segmented sheath of Schwann cells. The Schwann cells produce the myelin sheath that surrounds the peripheral nerves.

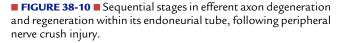
Peripheral Nerve Injury and Repair

Neurons exemplify the general principle that the more specialized the function of a cell type, the less able it is to regenerate. Although the entire neuron cannot be replaced, it often is possible for the dendritic and axonal cell processes to regenerate as long as the cell body remains viable.

When a peripheral nerve is destroyed by a crushing force or by a cut that penetrates the nerve, the portion of the nerve fiber that is separated from the cell body rapidly undergoes degenerative changes, whereas the central stump and cell body of the nerve often are able to survive. Because the cell body synthesizes the material required for nourishing and maintaining the axon, it is likely that the loss of these materials results in the degeneration of the separated portion of the nerve fibers. In crushing injuries in which the endoneurial tube remains intact, the outgrowing fiber will grow back down this tube to the structure that was originally innervated by the neuron (Fig. 38-10). However, it can take weeks or months for the regrowing fiber to reach its target organ and for communicative function to be re-established. More time is required for the Schwann cells to form new myelin segments and for the axon to recover its original diameter and conduction velocity. However, if the injury involves the severing of a nerve, the outgrowing branch must



neuromuscular junction



come in contact with its original endoneurial tube if it is to be reunited with its original target structure.

The successful regeneration of a nerve fiber in the peripheral nervous system depends on many factors. If a nerve fiber is destroyed relatively close to the neuronal cell body, the chances are that the nerve cell will die, and if it does, it will not be replaced. The degree of axonal regeneration that occurs after injury to a peripheral nerve also depends on the amount of scar tissue that develops at the site of injury and how quickly reinnervation occurs. If reinnervation occurs after the muscle cells have degenerated, no recovery is possible.

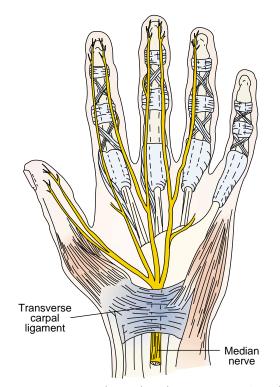
Perhaps the most difficult problem in the treatment of peripheral nerve injuries is the alignment of the proximal and distal endoneurial tubes so that a regenerating fiber can return down its former tube and innervate its former organ. Microscopic alignment of the cut edges during microsurgical repair results in improved success.

Peripheral Neuropathies

A peripheral neuropathy is any primary disorder of the peripheral nerves. The result usually is muscle weakness caused by LMN damage, with or without atrophy and sensory changes. The disorder can involve a single nerve (mononeuropathy) or multiple nerves (polyneuropathy). **Mononeuropathies.** Mononeuropathies usually are caused by localized conditions such as trauma, compression, or infections that affect a single spinal nerve, plexus, or peripheral nerve trunk. Fractured bones may lacerate or compress nerves; excessively tight tourniquets may injure nerves directly or produce ischemic injury; and infections such as herpes zoster may affect a single segmental afferent nerve distribution.

Carpal Tunnel Syndrome. Carpal tunnel syndrome is an example of a compression-type mononeuropathy that is relatively common. It is caused by compression of the median nerve as it travels with the flexor tendons through a canal made by the carpal bones and transverse carpal ligament (Fig. 38-11). The condition can be caused by a variety of conditions that produce a reduction in the capacity of the carpal tunnel (*i.e.*, bony or ligament changes) or an increase in the volume of the tunnel contents (i.e., inflammation of the tendons, synovial swelling, or tumors).¹² Carpal tunnel syndrome can be a feature of many systemic diseases such as rheumatoid arthritis, hyperthyroidism, acromegaly, and diabetes mellitus.¹² The condition can result from wrist injury; it can occur during pregnancy and use of birth control drugs; and it is seen in persons with repetitive use of the wrist (i.e., flexion-extension movements and stress associated with pinching and gripping motions).

Carpal tunnel syndrome is characterized by pain, paresthesia, and numbness of the thumb and first two and one-half digits of the hand; pain in the wrist and hand, which worsens at night; atrophy of abductor pollicis muscle; and weakness in precision grip. All of these abnormalities may contribute to clumsiness of fine motor activity.



■ FIGURE 38-11 ■ Carpal tunnel syndrome: compression of the median nerve by the transverse carpal ligament. (Courtesy Carole Russell Hilmer, C.M.I.)

Diagnosis usually is based on sensory disturbances confined to median nerve distribution. Electromyography and nerve conduction studies often are done to confirm the diagnosis and exclude other causes of the disorder.

Treatment includes avoidance of use, splinting, and antiinflammatory medications. Measures to decrease the causative repetitive movements should be initiated. Splints may be confined to nighttime use. When splinting is ineffective, corticosteroids may be injected into the carpal tunnel to reduce inflammation and swelling. Surgical intervention consists of operative division of the volar carpal ligaments as a means of relieving pressure on the medial nerve.

Polyneuropathies. Polyneuropathies involve demyelination or axonal degeneration of multiple peripheral nerves that leads to symmetric sensory, motor, or mixed sensorimotor deficits. Typically, the longest axons are involved first, with symptoms beginning in the distal part of the extremities. If the autonomic nervous system is involved, there may be postural hypotension, constipation, and impotence. Polyneuropathies can result from immune mechanisms (*e.g.*, Guillain-Barré syndrome), toxic agents (*e.g.*, arsenic polyneuropathy, lead polyneuropathy, alcoholic polyneuropathy), and metabolic diseases (*e.g.*, diabetes mellitus, uremia). Different causes tend to affect axons of different diameters and to affect sensory, motor, or autonomic neurons to different degrees.

Guillain-Barré Syndrome. Guillain-Barré syndrome is an acute inflammatory polyneuropathy. The annual incidence of Guillain-Barré syndrome is approximately 1 case per 50,000 persons, and it is more common with increasing age.¹³ Approximately 80% to 90% of persons with the disease achieve a spontaneous recovery.

The disorder involves an infiltration of mononuclear cells around the capillaries of the peripheral neurons, edema of the endoneurial compartment, and demyelination of ventral spinal roots. The cause of Guillain-Barré syndrome is unknown. Approximately two thirds of cases follow an infection that is seemingly mundane and often of viral origin.⁶ There is an association with preceding gastrointestinal tract infection with *Campylobacter jejuni*.¹⁴ A widely studied outbreak of the disorder followed the swine flu vaccination program of 1976 and 1977.¹⁵ It has been suggested that an altered immune response to peripheral nerve antigens contributes to the development of the disorder.

The disorder is characterized by progressive ascending muscle weakness of the limbs, producing a symmetric flaccid paralysis. Symptoms of paresthesia and numbness often accompany the loss of motor function. The rate of disease progression varies, and there may be disproportionate involvement of the upper or lower extremities. Paralysis may progress to involve the respiratory muscles; approximately 20% of persons with the disorder require ventilatory assistance.¹³ ANS involvement that causes postural hypotension, arrhythmias, facial flushing, abnormalities of sweating, and urinary retention is common.

Guillain-Barré syndrome usually presents as a medical emergency. There may be a rapid development of ventilatory failure and autonomic disturbances that threaten circulatory function. Treatment includes support of vital functions and prevention of complications such as skin breakdown and thrombophlebitis. Clinical trials have shown the effectiveness of plasma-

Chapter 38: Alterations in Neuromuscular Function

pheresis in decreasing morbidity and shortening the course of the disease. Treatment is most effective if initiated early in the course of the disease.

Back Pain and Intervertebral Disk Injury

Back Pain

Low back pain is an exceeding common health problem. The differential diagnosis is broad and includes muscle strain, primary spine disease (*e.g.*, disk herniation, degenerative arthritis [discussed in Chapter 43]), and systemic disease (*e.g.*, metastatic cancer). Although acute back problems commonly are attributed to a herniated intervertebral disk, most are caused by other, less serious conditions, such as muscle strain. It has been reported that 90% of persons with acute lower back problems of less than 3 months' duration experience spontaneous recovery.¹⁶ The diagnostic challenge is to identify those persons who require more extensive evaluation for more serious problems, including disk herniation.¹⁶

Treatment of back pain usually is conservative and consists of analgesic medications and education on how to protect the back. Pain relief usually can be provided using nonsteroidal anti-inflammatory drugs, although short-term use of opioid pain medications may be required for severe pain. Muscle relaxants may be used on a short-term basis. Bed rest, once the mainstay of conservative therapy, is now understood to be ineffective for acute back pain.¹⁷ Instruction in the correct mechanics for lifting and methods of protecting the back is important. Conditioning exercises of the trunk muscles, particularly the back extensors, may be recommended for persons with acute low back problems, particularly if the problem persists. Surgical treatment may be indicated when herniation is documented by some imaging procedure or in the presence of consistent pain or consistent neurologic deficit that has failed to respond to conservative therapy.

Herniated Intervertebral Disk

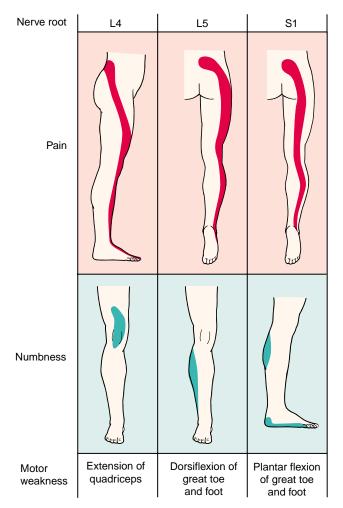
The intervertebral disk is considered the most critical component of the load-bearing structures of the spinal column. The intervertebral disk consists of a soft, gelatinous center called the *nucleus pulposus*, which is encircled by a strong, ringlike collar of fibrocartilage called the *annulus fibrosus*. The structural components of the disk make it capable of absorbing shock and changing shape while allowing movement. With dysfunction, the nucleus pulposus can be squeezed out of place and herniate through the annulus fibrosus, a condition referred to as a *herniated* or *slipped disk* (Fig. 38-12A and B).

The intervertebral disk can become dysfunctional because of trauma, the effects of aging, or degenerative disorders of the spine. Trauma accounts for 50% of disk herniations. It results from activities such as lifting while in the flexed position, slipping, falling on the buttocks or back, or suppressing a sneeze. With aging, the gelatinous center of the disk dries out and loses much of its elasticity, causing it to fray and tear. Degenerative processes such as osteoarthritis or ankylosing spondylitis predispose to malalignment of the vertebral column.

The cervical and lumbar regions are the most flexible area of the spine and are most often involved in disk herniations. Usually, herniation occurs at the lower levels of the lumbar spine, where the mass being supported and the bending of the vertebral column are greatest. Approximately 90% to 95% of lumbar herniations occur in the L4 or L5 to S1 regions. With herniations of the cervical spine, the most frequently involved levels are C6 to C7 and C5 to C6.⁵ Protrusion of the nucleus pulposus usually occurs posteriorly and toward the intervertebral foramen and its contained spinal nerve root, where the anulus fibrosus is relatively thin and poorly supported by either the posterior or anterior ligaments¹⁸ (see Fig. 38-12A).

The level at which a herniated disk occurs is important (Fig. 38-12C). When the injury occurs in the lumbar area, only the cauda equina is irritated or crushed. Because these elongated dorsal and ventral roots contain endoneurial tubes of connective tissue, regeneration of the nerve fibers is likely. However, several weeks or months are required for full recovery to occur because of the distance to the innervated muscle or skin of the lower limbs.

The signs and symptoms of a herniated disk are localized to the area of the body innervated by the nerve roots and include both motor and sensory manifestations (Fig. 38-13). Pain is the



■ FIGURE 38-13 ■ Dermatomes of the leg (L1 through S5) where pain and numbness would be experienced with spinal root irritation.

first and most common symptom of a herniated disk. The nerve roots of L4, L5, S1, S2, and S3 give rise to a syndrome of back pain that spreads down the back of the leg and over the sole of the foot. The pain is usually intensified with coughing, sneezing, straining, stooping, standing, and the jarring motions that occur during walking or riding. Slight motor weakness may occur, although major weakness is rare. The most common sensory deficits from spinal nerve root compression are paresthesias and numbness, particularly of the leg and foot. Knee and ankle reflexes also may be diminished or absent.

A herniated disk must be differentiated from other causes such as traumatic injury or fracture of the vertebral column, tumor, infection, cauda equina syndrome (see spinal cord injury), or other conditions that cause back pain.¹⁶ Diagnostic measures include history and physical examination. Neurologic assessment includes testing of muscle strength and reflexes. Other diagnostic methods include radiographs of the back, magnetic resonance imaging (MRI), myelography, and computed tomography (CT).¹⁹ Myelography, MRI, and CT usually are reserved for persons suspected of having more complex causes of back pain.

In summary, disorders of motor function include disorders of the skeletal muscle, the neuromuscular junction, and the peripheral nerves. *Muscular dystrophy* is a term used to describe a number of disorders that produce progressive deterioration of skeletal muscle. Muscle necrosis is followed by fat and connective tissue replacement. One form, Duchenne muscular dystrophy, is inherited as a X-linked trait and transmitted by the mother to her male offspring.

Myasthenia gravis is a disorder of the neuromuscular junction resulting from a deficiency of functional acetylcholine receptors, which causes weakness of the skeletal muscles. Because the disease affects the neuromuscular junction, there is no loss of sensory function. The most common manifestations are weakness of the eye muscles, with ptosis and diplopia; the jaw muscles, which make chewing and swallowing difficult; and proximal muscles and extremities, which make climbing stairs and lifting objects difficult.

Disorders of peripheral nerves include mononeuropathies and polyneuropathies. Mononeuropathies involve a single spinal nerve, plexus, or peripheral nerve trunk. Carpal tunnel syndrome, a mononeuropathy, is caused by compression of the medial nerve that passes through the carpal tunnel in the wrist. Polyneuropathies involve multiple peripheral nerves and produce symmetric sensory, motor, and mixed sensorimotor deficits. Guillain-Barré syndrome is a subacute polyneuropathy of uncertain origin. It causes progressive ascending motor, sensory, and ANS manifestations. Respiratory involvement may occur and necessitate mechanical ventilation.

Acute back pain is most commonly the result of conditions such as muscle strain, with treatment that focuses on measures to improve activity tolerance. A herniated intervertebral disk is characterized by protrusion of the nucleus pulposus into the spinal canal with irritation or compression of the nerve root. Usually, herniation occurs at the lower levels of the lumbar and sacral (L4 or L5 to S1) and cervical (C6 to C7 and C5 to C6) regions of the spine. The signs and symptoms of a herniated disk are localized to the area of the body innervated by the nerve roots and include pain and both motor and sensory manifestations.

BASAL GANGLIA AND CEREBELLUM DISORDERS

The Basal Ganglia

The basal ganglia are a group of deep, interrelated subcortical nuclei that play an essential role in control of movement. They receive indirect input from the cerebellum and from all sensory ganglia systems, including vision, and the motor cortex. The structural components of the basal ganglia include the caudate nucleus, putamen, and the globus pallidus in the forebrain. The caudate and putamen are collectively referred to as the neostriatum, and the putamen and the globus pallidus form a wedge-shaped region called the lentiform nucleus. Two other structures, the subthalamic nucleus of the diencephalon and the substantia nigra of the midbrain, are considered part of the basal ganglia (Fig. 38-14). The substantia nigra contains cells that use dopamine as a neurotransmitter and are rich in a black pigment called *melanin*. The high concentration of melanin gives the structure a black color, thus the name substantia nigra. The axons of the substantia nigra form the *nigrostriatal pathway*, which supplies dopamine to the striatum. The dopamine released from the substantia nigra regulates the overall excitability of the striatum and release of other neurotransmitters.

The basal ganglia have input structures that receive afferent information from outside structures, internal circuits that connect the various structures of the basal ganglia, and output structures that deliver information to other brain centers. The neostriatum represents the major input structure for the basal ganglia. Information coming from virtually all areas of the cortex and thalamus are projected to the neostriatum. The output areas of the basal ganglia have both ascending and descending components. The major ascending output is transmitted to thalamic nuclei, which process all incoming sensory information that is transmitted to the cerebral cortex. Descending output is directed to the midbrain, brain stem, and spinal cord. The output functions of the basal ganglia are mainly inhibitory.

The most is known about the inhibitory basal ganglia loop involved in modulating cortical motor control. This loop regulates release of stereotyped movement patterns that add efficiency and gracefulness to precise and delicate cortically



controlled movements. These movements include inherited patterns that add efficiency, balance, and gracefulness to motion, such as the swinging of the arms during walking and running, and the highly learned automatic postural and followthrough movements of throwing a ball or swinging a bat.

Chapter 38: Alterations in Neuromuscular Function

An additional modulating circuit involves a neostriatal inhibitory projection on the substantia nigra. The substantia nigra projects dopaminergic axons back on the neostriatum. A deficiency in the dopaminergic projection of this modulating circuit is implicated in Parkinson's syndrome. The function of the neostriatum also involves local cholinergic interneurons, and their destruction is thought to be related to the choreiform movements of Huntington's chorea, another basal gangliarelated syndrome (see Chapter 37).

Movement Disorders

Disorders of the basal ganglia comprise a complex group of motor disturbances characterized by involuntary movements, alterations in muscle tone, and disturbances in body posture. Unlike disorders of the motor cortex and corticospinal tract, lesions of the basal ganglia disrupt movement but do not cause paralysis.

Movement disorders associated with dysfunction of the basal ganglia include bradykinesis, hyperkinesis, and abnormal movements. Hyperfunction of the basal ganglia inhibitory loop result in excessive inhibition of cortical function, resulting in *bradykinesia* or *hypokinesis*. Reduced function of the basal ganglia loop results in *hyperkinesis*, or release of movement patterns at inappropriate times or sometimes continuously. Pathologically released patterns that often are disabling include rigidity and movement disorders. These movement patterns are not under cortical control and often are referred to as *involuntary movements*.

Involuntary movements include tremor, tics, choreiform movements, athetoid movements, and ballismus. These disorders are summarized in Table 38-1. Tremor is caused by involuntary, oscillating contractions of opposing muscle groups around a joint. It usually is fairly uniform in frequency and amplitude. Certain tremors are considered physiologic in that they are transitory and normally occur under conditions of increased muscle tone, as in highly emotional situations, or they may be related to muscle fatigue or reduced body temperature (*i.e.*, shivering). Toxic tremors are produced by hyperexcitability related to conditions such as thyrotoxicosis. The tremor of Parkinson's disease is caused by degenerative changes in the basal ganglia. Tics involve sudden and irregularly occurring contractions of whole muscles or major portions of a muscle. These are particularly evident in the muscles of the face but can occur elsewhere.

Choreiform movements are sudden, jerky, and irregular but are coordinated and graceful. They can involve the distal limb, face, tongue, or swallowing muscles. Choreiform movements, such as those seen in Huntington's disease (see Chapter 40), are accentuated by movement and by environmental stimulation; they often interfere with normal movement patterns. The word *chorea* originated from the Greek word meaning "to dance." There may be grimacing movements of the face; raising of the eyebrows; rolling of the eyes; and curling, protrusion, and withdrawal of the tongue. In the limbs, the movements largely are distal; there may movements that mimic piano playing with alternating extension and flexion of the fingers. The shoulders may be elevated and depressed or rotated. Movements

TABLE 38-1 Involuntary Movements Disorders Associated With Extrapyramidal Disorders

Movement Disorder Characteristics

Tremor	Rhythmic oscillating contractions or movements of whole muscles or major portions of a muscle. They can occur as resting tremors, which are prominent at rest and decrease or disappear with movement; intention tremors, which increase with activity and become worse when the target is reached; and postural tremors, which appear when the affected part is maintained in a stabilized position.
Tics	Irregularly occurring brief, repetitive, stereotyped, coordinated movements such as winking, grimacing, or shoulder shrugging
Chorea	Brief, rapid, jerky, and irregular movements that are coordinated and graceful. The face, head, and distal limbs are most commonly involved. They often interfere with normal movement patterns.
Athetosis	Continuous, slow, wormlike, twisting and turning motions of a limb or body that most commonly involve the face and distal extremities and are often associated with spasticity
Ballismus	Involve violent sweeping, flinging-type limb movements, especially on one side of the body (hemiballismus)
Dystonia	Abnormal maintenance of posture results from a twisting, turning motion of the limbs, neck, or trunk. Motions are similar to athetosis but involve larger portions of the body. They can result in grotesque and twisted postures.
Dyskinesias	Rhythmic, repetitive, bizarre movements that chiefly involve the face, mouth, jaw, or tongue, causing grimac- ing, pursing of the lips, protrusion of the tongue, opening and closing of the mouth, and deviations of the jaw. The limbs are affected less often.

(From Bates B. [1991]. A guide to physical examination and history taking [5th ed., pp. 554–556]. Philadelphia: J.B. Lippincott)

of the face or limbs may occur alone or, more commonly, in combination.

Athetoid movements are relatively continuous, wormlike twisting and turning motions of the joints of a limb or body, such as those seen in one form of cerebral palsy. These result from continuous and prolonged contraction of agonist and antagonistic muscle groups. The term *ballismus* originated from a Greek word meaning "to jump around." Ballistic movements are violent, sweeping, flinging motions, especially of the limbs on one side of the body (hemiballismus). They may occur as the result of a small stroke involving the subthalamic nucleus on the opposite side of the brain.

Dystonia refers to the abnormal maintenance of a posture resulting from a twisting, turning movement of the limbs, neck, or trunk. These postures often result from simultaneous contraction of agonist and antagonist muscles. Long-sustained simultaneous hypertonia across a joint can result in degenerative changes and permanent fixation in unusual postures. These effects can occur as a side effect of some antipsychotic medications. Spasmodic torticollis, the most common type of dystonia, affects the muscles of the neck and shoulder. The condition, which is caused by bilateral and simultaneous contraction of the neck and shoulder muscles, results in unilateral head turning or head extension, sometimes limiting rotation. Elevations of the shoulder commonly accompany the spasmodic movements of the head and neck. Immobility of the cervical vertebrae eventually can lead to degenerative fixation in the twisted posture. Torsional spasm involving the trunk also can occur.

Dyskinesias are rhythmic, repetitive, bizarre movements. They frequently involve the face, mouth, jaw, and tongue, causing grimacing, pursing of the lips, or protrusion of the tongue. The limbs are affected less often. Tardive dyskinesia is an untoward reaction that can develop with long-term use of some of the antipsychotic medications.

Parkinson's Disease

Parkinson's disease is a degenerative disorder of basal ganglia function that results in variable combinations of tremor, rigidity, and bradykinesia. The disorder is characterized by progressive destruction of the nigrostriatal pathway, with subsequent reduction in striatal concentrations of dopamine. As many as 1.5 million people in the United States are affected by the disease.²⁰ It usually begins after 50 years of age; most cases are diagnosed in the sixth and seventh decade of life.

The clinical syndrome arising from the degenerative changes in basal ganglia function often is referred to as parkinsonism. Parkinson's disease, the most common form of parkinsonism, is named after James Parkinson, a British physician who first described the disease in a paper he published in 1817 on the "shaking palsy."²¹ In Parkinson's disease, also known as idiopathic parkinsonism, dopamine depletion results from degeneration of the dopamine nigrostriatal system. Parkinsonism can also develop as a postencephalitic syndrome, as a side effect of therapy with antipsychotic drugs that block dopamine receptors, as a toxic reaction to a chemical agent, or as an outcome of severe carbon monoxide poisoning. Symptoms of parkinsonism also may accompany conditions such as cerebral vascular disease, brain tumors, repeated head trauma, or degenerative neurologic diseases that structurally damage the nigrostriatal pathway.

Drug-induced parkinsonism can follow the administration of antipsychotic drugs in high doses (*e.g.*, phenothiazines, butyrophenones). These drugs block dopamine receptors and dopamine output by the cells of the substantia nigra. Of interest in terms of research was the development of Parkinson's disease in several persons who had attempted to make a narcotic drug and instead synthesized a compound called MPTP (1-methyl-phenyl-2,3,6-tetrahydropyridine).²² This compound selectively destroys the dopaminergic neurons of the substantia nigra. This incident prompted investigations into the role of toxins that are produced by the body as a part of metabolic processes and those that enter the body from outside sources in the pathogenesis of Parkinson's disease. One theory is that the auto-oxidation of catecholamines such as dopamine during melanin synthesis has the potential for injuring neurons in the substantia nigra. Accordingly, the development of Parkinson's disease may be related to oxidative metabolites of this process and the inability of neurons to render these products harmless. MPTP is an inhibitor of the metabolic pathway that functions in the inactivation of these metabolites, suggesting that it may produce Parkinson's disease in a manner similar to the naturally occurring disease.⁶

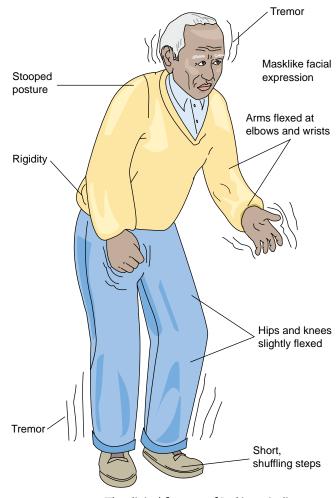
A recent discovery suggests that genetic susceptibility may play a role in the pathogenesis of early-onset (before 45 years of age) Parkinson's disease. A mutation in a gene called the *Parkin gene* has been identified in a high percentage of family members and persons with early-onset Parkinson's disease.²³

Clinical Course. The cardinal manifestations of Parkinson's disease are tremor, rigidity, and bradykinesia or slowness of movement.²⁴ Other advanced-stage parkinsonian manifestations are falls, fluctuations in motor function, neuropsychiatric disorders, and sleep disorders.

Tremor is the most visible manifestation of the disorder. The tremor affects the distal segments of the limbs, mainly the hands and feet; head, neck, face, lips, and tongue; or jaw. It is characterized by rhythmic, alternating flexion and contraction movements (four to six beats per minute) that resemble the motion of rolling a pill between the thumb and forefinger. The tremor usually is unilateral, occurs when the limb is supported and at rest, and disappears with movement and sleep. The tremor eventually progresses to involve both sides of the body. Although the most noticeable sign of Parkinson's disease, tremor usually is the least disabling because of the dampening effect of purposeful movement.

Rigidity is defined as resistance to movement of both flexors and extensors throughout the full range of motion. It is most evident during passive joint movement and involves jerky, cogwheel-type or ratchet-like movements that require considerable energy to perform. Flexion contractions may develop as a result of the rigidity. As with tremor, rigidity usually begins unilaterally but progresses to involve both sides of the body.

Bradykinesia is characterized by slowness in initiating and performing movements and difficulty in sudden, unexpected stopping of voluntary movements. Unconscious associative movements occur in a series of disconnected steps, rather than in a smooth, coordinated manner. This is the most disabling of the symptoms of Parkinson's disease. Persons with Parkinson's disease have difficulty initiating walking and difficulty turning. While walking, they may freeze in place and feel as if their feet are glued to the floor, especially when moving through a doorway or preparing to turn. When they walk, they lean forward to maintain their center of gravity and take small, shuffling steps without swinging their arms, and they have difficulty in changing their stride (Fig. 38-15). Loss of postural reflexes predisposes them to falling, often backward. Emotional and voluntary facial movements become limited and slow as the disease progresses, and facial expression becomes stiff and masklike. There is loss of the blinking reflex and a failure to express emotion. The tongue, palate, and throat muscles become



■ FIGURE 38-15 ■ The clinical features of Parkinson's disease.

rigid; the person may drool because of difficulty in moving the saliva to the back of the mouth and swallowing it. The speech becomes slow and monotonous, without modulation and poorly articulated.

Because the basal ganglia also influence the ANS, persons with Parkinson's disease often have excessive and uncontrolled sweating, sebaceous gland secretion, and salivation. Autonomic symptoms such as lacrimation, dysphagia, orthostatic hypotension, thermal regulation, constipation, impotence, and urinary incontinence may be present, especially late in the disease.

Dementia is an important feature associated with Parkinson's disease. It occurs in approximately 20% of persons with the disease and develops late in the course of the disease.²⁴ The mental state of some persons with Parkinson's disease may be indistinguishable from that seen in Alzheimer's disease. It has been suggested that many of the brain changes in both diseases may result from degeneration of acetylcholine-containing neurons.

Treatment. The approach to treatment of Parkinson's disease must be highly individualized. It includes nonpharmacologic, pharmacologic, and when indicated, surgical methods. Non-pharmacologic interventions offer group support, education,

daily exercise, and adequate nutrition. Botulism toxin injections may be used in the treatment of dystonias such as eyelid spasm and limb dystonias that frequently are associated with Parkinson's disease.²⁵

Pharmacologic treatment usually is determined by the severity of symptoms. Antiparkinson drugs act by increasing the functional ability of the underactive dopaminergic system, or they reduce the excessive influence of excitatory cholinergic neurons. They include drugs that increase dopamine levels (*e.g.*, levodopa), augment the release of dopamine (*e.g.*, amantadine), function as dopamine agonists or directly stimulate dopamine receptors (*e.g.*, bromocriptine, pergolide, pramipexole, ropinirole), or inhibit the metabolic breakdown of dopamine (*e.g.*, selegiline). Because dopamine transmission is disrupted in Parkinson's disease, there is a preponderance of cholinergic activity, which may be treated with anticholinergic drugs.

Surgical treatment includes thalamotomy or pallidectomy performed using stereotactic surgery. With these procedures, part of the thalamus or globus pallidum in the basal ganglia is destroyed using an electrical stimulator or supercooled tip of a metal probe (cryothalamotomy). Brain mapping is done during the surgery to identify and prevent injury to sensory and motor tracts. Surgery is generally confined to one side of the brain and is usually restricted to persons who have not experienced satisfactory response to drug therapy.

The Cerebellum

The cerebellum, or "little brain," constitutes only 10% of the total volume of the brain but contains more than half of all its neurons.¹ The cerebellum is responsible for smoothing the temporal and spatial aspects of rapid movement anywhere in the body. It does this by evaluating disparities between intention and action and by adjusting the operation of motor centers in the cortex while the movement is in progress as well as during repetitions of the same movement. The cerebellum also appears to have a role in learning both motor and cognitive tasks in which skilled responses are developed through repeated practice. It does not alter sensory thresholds or the strength of muscle contraction.

The cerebellum is located in the posterior fossa of the cranium superior to the pons (see Chapter 36, Fig. 36–19). It is separated from the cerebral hemispheres by a fold of dura mater, the tentorium cerebelli. The cerebellum consists of a small, unpaired median portion, called the *vermis*, and two large lateral masses, the *cerebellar hemispheres*. In contrast to the brain stem with its external white matter and internal gray nuclei, the cerebellum, like the cerebrum, has an outer cortex of gray matter overlying the white matter. Several masses of gray matter, called the *deep cerebellar nuclei*, border the roof of the fourth ventricle. Cells of the cerebellar cortex and the deep nuclei interact and axons from the latter send information to many regions, particularly to the motor cortex by means of a thalamic relay.

Synergistic (*i.e.*, temporal and spatial smoothing) functions of the cerebellum participate in all movements of limbs, trunk, head, larynx, and eyes, whether the movement is part of a voluntary movement or of a highly learned semiautomatic or automatic movement. During highly skilled movements, the motor cortex sends signals to the cerebellum, informing it about the movement that is to be performed. The cerebellum makes continuous adjustments, resulting in smoothness of movement, particularly during delicate maneuvers. Highly skillful movement requires extensive motor training, and considerable evidence suggests many of these learned movement patterns involve cerebellar circuits.

The cerebellum receives proprioceptor input from the vestibular system; feedback from the muscles, tendons, and joints; and indirect signals from the somesthetic, visual, and auditory systems that provide background information for ongoing movement. Sensory and motor information from a given area of the body is sent to the same area in the cerebellum. In this way, the cerebellum can assess continuously the status of each body part—position, rate of movement, and forces such as gravity that are opposing movement. The cerebellum compares what is actually happening with what is intended to happen. It then transmits the appropriate corrective signals back to the motor system, instructing it to increase or decrease the activity of the participating muscle groups so that smooth and accurate movements can be performed.

Cerebellar Dysfunction

Cerebellar dysfunction can be caused by injuries, global ischemia, occlusion of any of the cerebellar arteries, cerebral hemorrhage, and neoplastic lesions. Massive cerebellar infarction may lead to coma, tonsillar herniation, and death. Chronic alcoholism can lead to atrophy of the superior vermis of the cerebellulum.⁷

The signs of cerebellar dysfunction can be grouped into three classes: (1) vestibulocerebellar disorders, (2) cerebellar ataxia or decomposition of movement, and (3) cerebellar tremor. These disorders occur on the side of cerebellar damage. The abnormality of movement occurs whether the eyes are open or closed. Visual monitoring of movement cannot compensate for cerebellar defects.

Damage to the part of the cerebellum associated with the vestibular system leads to difficulty or inability to maintain a steady posture of the trunk, which normally requires constant readjusting movements. This is seen as an unsteadiness of the trunk, called *truncal ataxia*, and it can be so severe that standing is not possible. The ability to fix the eyes on a target also can be affected. Constant conjugate readjustment of eye position, called *nystagmus*, results and makes reading extremely difficult, especially when the eyes are deviated toward the side of cerebellar damage.

Cerebellar ataxia and tremor are different aspects of defects in the smooth, continuously correcting functions. Cerebellar dystaxia or, if severe, ataxia includes a decomposition of movement; each succeeding component of a complex movement occurs separately instead of being blended into a smoothly proceeding action. Because ethanol specifically affects cerebellar function, persons who are inebriated often walk with a staggering and unsteady gait. Rapid alternating movements such as supination-pronation-supination of the hands are jerky and performed slowly (dysdiadochokinesia). Reaching to touch a target breaks down into small sequential components, each going too far, followed by overcorrection. The finger moves jerkily toward the target, misses, corrects in the other direction, and misses again, until the target is finally reached. This is called *over-and-under reaching*. Tests of cerebellar function (walking a straight line and touching the nose with the index finger) are often used; these are the same tests used by law enforcement officials for screening persons suspected of driving while intoxicated.

Cerebellar tremor is a rhythmic back-and-forth movement of a finger or toe that worsens as the target is approached. The tremor results from the inability of the damaged cerebellar system to maintain ongoing fixation of a body part and to make smooth, continuous corrections in the trajectory of the movement; overcorrection occurs, first in one direction and then the other. Often, the tremor of an arm or leg can be detected during the beginning of an intended movement. The common term for cerebellar tremor is *intention tremor*. Cerebellar function as it relates to tremor can be assessed by asking a person to touch one heel to the opposite knee, to gently move the toes along the back of the opposite shin, or to move the hand so as to touch the nose with a finger.

Cerebellar function also can affect the motor skills of chewing and swallowing (dysphagia) and of speech (dysarthria). Normal speech requires smooth control of respiratory muscles and highly coordinated control of the laryngeal, lip, and tongue muscles. Cerebellar dysarthria is characterized by slow, slurred speech of continuously varying loudness. Rehabilitative efforts directed by speech therapists include learning to slow the rate of speech and to compensate as much as possible through the use of less-affected muscles.

In summary, alterations in coordination of muscle movements and abnormal muscle movements result from disorders of the basal ganglia and cerebellum. The basal ganglia organize basic movement patterns into more complex patterns and release them when commanded by the motor cortex, contributing gracefulness to cortically initiated and controlled skilled movements. Disorders of the basal ganglia are characterized by involuntary movements, alterations in muscle tone, and disturbances in posture. These disorders include tremor, tics, hemiballismus, chorea, athetosis, dystonias, and dyskinesias.

Parkinsonism, a disorder of the basal ganglia, is characterized by destruction of the nigrostriatal pathway, with a subsequent reduction in striatal concentrations of dopamine. This results in an imbalance between the inhibitory effects of dopaminergic basal ganglia functions and an increase in the excitatory cholinergic functions. The disorder is manifested by combinations of slowness of movement (*i.e.*, bradykinesia), increased muscle tonus and rigidity, rest tremor, gait disturbances, and impaired autonomic postural responses.

The cerebellum is responsible for smoothing the temporal and spatial aspects of rapid movement anywhere in the body. It does this by evaluating disparities between intention and action and by adjusting the operation of motor centers in the cortex while the movement is in progress as well as during repetitions of the same movement. The cerebellum also appears to have a role in learning both motor and cognitive tasks in which skilled responses are developed through repeated practice. Cerebellar disorders include vestibulocerebellar dysfunction, cerebellar ataxia, and cerebellar tremor.

UPPER AND LOWER MOTONEURON DISORDERS

Upper and Lower motoneurons innervate the skeletal muscles and are essential for motor function. The UMNs are fully contained within the CNS. They include the motor neurons arising in the motor areas of the cortex and their fibers as they project through the brain and descend in the spinal cord. Disorders that affect UMNs include multiple sclerosis and spinal cord injury (discussed later). Stroke, which is a common cause of UMN damage, is discussed in Chapter 37. Amyotrophic lateral sclerosis is a mixed UMN and LMN disorder.

Amyotrophic Lateral Sclerosis

Amyotrophic lateral sclerosis (ALS), also known as *Lou Gehrig's disease* after the famous New York Yankees baseball player, is a devastating neurologic disorder that selectively affects motor function. ALS is primarily a disorder of middle to late adulthood, affecting persons between 55 and 60 years of age, with men developing the disease nearly twice as often as women. The disease typically follows a progressive course, with a mean survival period of 2 to 5 years from the onset of symptoms.

ALS affects motoneurons in three locations: the anterior horn cells of the spinal cord; the motor nuclei of the brain stem, particularly the hypoglossal nuclei; and the UMNs of the cerebral cortex.²⁶ The fact that the disease is more extensive in the distal, rather than the proximal, parts of the affected tracts in the lower spinal cord suggests that affected neurons first undergo degeneration at their distal terminals and that the disease proceeds in a centripetal direction until ultimately the parent nerve cell dies. A remarkable feature of the disease is that the entire sensory system, the regulatory mechanisms of control and coordination of movement, and the intellect remain intact. The neurons for eye movement and the parasympathetic neurons in the sacral spinal cord also are spared.

Degeneration and loss of neurons in the primary motor cortex leads to loss of fibers within the corticospinal tract and lateral and anterior columns of the spinal cord.⁶ It is this fiber atrophy, called *amyotrophy*, that appears in the name of the disease. The loss of nerve fibers in lateral columns of the white matter of the spinal cord along with fibrillary gliosis imparts a firmness or sclerosis to this CNS tissue; the term *lateral sclerosis* designates these changes.

The cause of LMN and UMN destruction in ALS is uncertain. Five percent to 10% of cases are familial; the others are believed to be sporadic, with no family history of the disease. Recently, mutations to a gene encoding superoxide dismutase 1 (SOD1) was mapped to chromosome 21. This enzyme functions in the prevention of free radical formation (see Chapter 2). The mutation accounts for 20% of familial ALS, with the remaining 80% being caused by mutations in other genes.²⁶ Five percent of persons with sporadic ALS also have SOD1 mutations. Possible targets of SOD1-induced toxicity include the neurofilament proteins, which function in the axonal transport of molecules necessary for the maintenance of axons.²⁶ Another suggested mechanism of pathogenesis in ALS is exotoxic injury through activation of glutamate-gated ion channels, which are distinguished by their sensitivity to *N*-methyl-D-aspartic acid

(see Chapter 37). The possibility of glutamate excitotoxicity in the pathogenesis of ALS was suggested by the finding of increased glutamine levels in the cerebrospinal fluid of patients with sporadic ALS.²⁶ Although autoimmunity has been suggested as a cause of ALS, the disease does not respond to the immunosuppressant agents that normally are used in treatment of autoimmune disorders.

Clinical Course

The symptoms of ALS may be referable to UMN or LMN involvement. Manifestations of UMN lesions include weakness, spasticity or stiffness, and impaired fine motor control.^{26,27} Dysphagia (difficulty swallowing), dysarthria (impaired articulation of speech), and dysphonia (difficulty making the sounds of speech) may result from brain stem LMN involvement or from dysfunction of UMNs descending to the brain stem. Manifestations of LMN destruction include fasciculations, weakness, muscle atrophy, and hyporeflexia. Muscle cramping involving the distal legs often is an early symptom. The most common clinical presentation is slowly progressive weakness and atrophy in distal muscles of one upper extremity. This is followed by regional spread of clinical weakness, reflecting involvement of neighboring areas of the spinal cord. Eventually, UMNs and LMNs involving multiple limbs and the head are affected. In the more advanced stages, muscles of the palate, pharynx, tongue, neck, and shoulders become involved, causing impairment of chewing, swallowing, and speech. Dysphagia with recurrent aspiration and weakness of the respiratory muscles produces the most significant acute complications of the disease. Death usually results from involvement of cranial nerves and respiratory musculature.

Currently, there is no cure for ALS. Rehabilitation measures assist persons with the disorder to manage their disability, and respiratory and nutritional support allows persons with the disorder to survive longer than would otherwise have been the case. An antiglutamate drug, riluzole, is the only drug approved by the U.S. Food and Drug Administration (FDA) for treatment of ALS. The drug is designed to decrease glutamate accumulation and slow the progression of the disease.²⁶

Multiple Sclerosis

Multiple sclerosis (MS), a demyelinating disease of the CNS, is a major cause of neurologic disability among young and middleaged adults. Approximately two thirds of persons with MS experience their first symptoms between 20 and 40 years of age. In approximately 80% of the cases, the disease is characterized by exacerbations and remissions over many years in several different sites in the CNS.²⁸ Initially, there is normal or nearnormal neurologic function between exacerbations. As the disease progresses, there is less improvement between exacerbations and increasing neurologic dysfunction.

The prevalence of MS varies considerably around the world. The disease is more prevalent in the colder northern latitudes; it is more common in the northern Atlantic states, the Great Lakes region, and the Pacific Northwest than in the southern parts of the United States. Other high-incidence areas include northern Europe, Great Britain, southern Australia, and New Zealand.²⁹ The incidence among women is almost double that of men. Although MS is not directly inherited, there is a familial predisposition in some cases, suggesting a genetic influence on susceptibility. For example, there is evidence of a genetic linkage of MS susceptibility to the inherited major histocompatibility complex DR2 haplotype (Chapter 8).^{6,28}

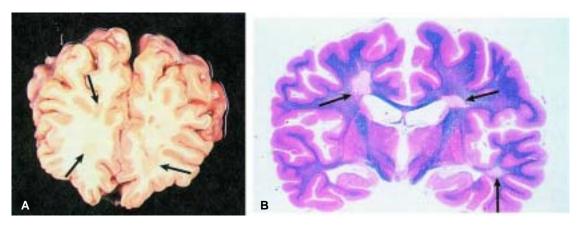
The pathophysiology of MS involves the demyelination of nerve fibers in the white matter of the brain, spinal cord, and optic nerve. In the CNS, myelin is formed by the oligodendrocytes, chiefly those lying among the nerve fibers in the white matter. The properties of the myelin sheath—high electrical resistance and low capacitance—permit it to function as an electrical insulator. Demyelinated nerve fibers display a variety of conduction abnormalities, ranging from decreased conduction velocity to conduction blocks.

The lesions of MS consist of hard, sharp-edged demyelinated or sclerotic patches that are macroscopically visible throughout the white matter of the CNS⁷ (Fig. 38-16). These lesions, which represent the end result of acute myelin breakdown, are called *plaques*. The lesions have a predilection for the optic nerves, periventricular white matter, brain stem, cerebellum, and spinal cord white matter.²⁸ In an active plaque, there is evidence of ongoing myelin breakdown. The sequence of myelin breakdown is not well understood, although it is known that the lesions contain small amounts of myelin basic proteins and increased amounts of proteolytic enzymes, macrophages, lymphocytes, and plasma cells. Oligodendrocytes are decreased in number and may be absent, especially in older lesions. Acute, subacute, and chronic lesions often are seen at multiple sites throughout the CNS.

The lesions of MS are generally thought to result from an immune-mediated inflammatory response that occurs in genetically susceptible individuals. The demyelination process in MS is marked by prominent lymphocytic invasion in the lesion. The infiltrate in plaques contains both CD8⁺ and CD4⁺ T cells as well as macrophages. Both macrophages and cytotoxic CD8⁺ T cells are thought to induce oligodendrocyte injury. There also is evidence of antibody-mediated damage involving myelin oligodendroglial protein.⁶ Magnetic resonance imaging has shown that the lesions of MS may occur in two stages: a first stage that involves the sequential development of small inflammatory lesions, and a second stage during which the lesions extend and consolidate and when demyelination and gliosis (scar formation) occur. It is not known whether the inflammatory process, present during the first stage, is directed against the myelin or against the oligodendrocytes that produce myelin. There is evidence that remyelination can occur in the CNS if the process that initiated the demyelination is halted before the oligodendrocyte dies.²⁸

Clinical Course

The interruption of neural conduction in the demyelinated nerves is manifested by a variety of symptoms, depending on the location and extent of the lesion. Areas commonly affected by MS are the optic nerve (visual field), corticobulbar tracts (speech and swallowing), corticospinal tracts (muscle strength), cerebellar tracts (gait and coordination), spinocerebellar tracts (balance), medial longitudinal fasciculus (conjugate gaze function of the extraocular eye muscles), and posterior cell columns of the spinal cord (position and vibratory sensation). Typically, an otherwise healthy person presents with an acute or subacute episode of paresthesias, optic neuritis (*i.e.*, visual clouding or loss of vision in part of the visual field with pain on movement of the globe), diplopia, or specific types of gaze paralysis.



■ FIGURE 38-16 ■ Multiple sclerosis. (A) In this unfixed brain, the plaques of multiple sclerosis in the white matter (*arrows*) assume the darker color of the cerebral cortex. (B) A coronal section of the brain from a patient with long-standing multiple sclerosis, which has been stained for myelin, shows discrete areas of demyelination (*arrows*) with characteristic involvement of the superior angles of the lateral ventricles. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 1497]. Philadelphia: Lippincott Williams & Wilkins)

Paresthesias are evidenced as numbness, tingling, a burning sensation, or pressure on the face or involved extremities; symptoms can range from annoying to severe. Pain from spasticity also may be a factor that can be aided by appropriate stretching exercises. Other common symptoms are abnormal gait, bladder and sexual dysfunction, vertigo, nystagmus, fatigue, and speech disturbance. These symptoms usually last for several days to weeks, and then completely or partially resolve. After a period of normal or relatively normal function, new symptoms appear. Psychological manifestations, such as mood swings, may represent an emotional reaction to the nature of the disease or, more likely, involvement of the white matter of the cerebral cortex. Depression, euphoria, inattentiveness, apathy, forgetfulness, and loss of memory may occur. Fatigue is one of the most common problems for persons with MS. It often is described as a generalized low-energy feeling not related to depression and different from weakness.

The course of the disease may fall into one of four categories: relapsing-remitting, secondary progressive, primary progressive, or progressive relapsing.^{31,32} The *relapsing-remitting* form of the disease is characterized by episodes of acute worsening with recovery and a stable course between relapses. *Secondary progressive disease* involves a gradual neurologic deterioration with or without superimposed acute relapses in a person with previous relapsing-remitting disease. *Primary progressive disease* is characterized by nearly continuous neurologic deterioration from onset of symptoms. The *progressive relapsing category* of disease involves gradual neurologic deterioration from the onset of symptoms but with subsequent superimposed relapses.

Diagnosis

The diagnosis of MS is based on established clinical, MRI, and laboratory findings in cerebrospinal fluid analysis. Examination of the cerebrospinal fluid is used for detecting signs of inflammation and immune responses. A large percentage of patients with MS have elevated immunoglobulin G (IgG) levels, and some have oligoclonal patterns (*i.e.*, discrete electrophoretic bands) even with normal IgG levels. Total protein or lymphocyte levels may be mildly elevated in the cerebrospinal fluid. These test results can be altered in a variety of inflammatory neurologic disorders and are not specific for MS. Advances in MRI have greatly simplified the diagnosis and evaluation of MS. MRI studies can detect the multiplicity of lesions even when CT scans appear normal. A computer-assisted method of MRI can measure lesion size. Many new areas of myelin abnormality are asymptomatic. Serial MRI studies can be done to detect asymptomatic lesions, monitor the progress of existing lesions, and evaluate the effectiveness of treatment. Electrophysiologic evaluations (*e.g.*, evoked potential studies) and CT scans may assist in the identification and documentation of lesions.

Treatment

Most treatment measures for MS are directed at modifying the course and managing the primary symptoms of the disease. The variety of symptoms, unpredictable course, and lack of specific diagnostic methods have made the evaluation and treatment of MS difficult. Persons who are minimally affected by the disorder require no specific treatment. The person should be encouraged to maintain as healthy a lifestyle as possible, including good nutrition and adequate rest and relaxation. Physical therapy may help maintain muscle tone. Every effort should be made to avoid excessive fatigue, physical deterioration, emotional stress, viral infections, and extremes of environmental temperature, which may precipitate an exacerbation of the disease.

The pharmacologic agents used in the treatment of MS fall into four categories: (1) those used to treat acute symptoms of the disease, (2) those used to modify the course of the disease, (3) those used to interrupt progressive disease, and (4) those used to treat the symptoms of the disorder.³¹ Corticosteroids are the mainstay of treatment for acute relapses of MS. These agents are thought to reduce the inflammation, improve nerve conduction, and have important immunologic effects. However, long-term administration does not appear to alter the course of the disease and can have harmful side effects. Adrenocorticotropic hormone (ACTH) also may be used in treatment of MS. Plasmapheresis has proved beneficial in some cases.

The agents used to modify the course of the disease include interferon beta and glatiramer acetate.³¹ Both agents have shown some benefit in reducing exacerbations in persons with relapsing-remitting MS. Interferon beta is a cytokine that acts as an immune enhancer. Glatiramer acetate is a synthetic polypeptide that simulates parts of the myelin basic protein. Although the exact mechanism of action is unknown, the drug seems to block myelin-damaging T cells by acting as a myelin decoy.

Progressive MS may be treated with immunosuppressive drugs such as methotrexate, cyclophosphamide, mitoxantrone, and cyclosporine. The drugs used to treat symptoms of the disease include muscle relaxants (*e.g.*, dantrolene, baclofen, diazepam), which are used to relieve spasticity; cholinergic drugs, which are used to treat bladder problems; and antidepressant drugs, which are used for depression.

In summary, UMNs and LMNs innervate the skeletal muscles and are essential for motor function. Amyotrophic lateral sclerosis is a mixed UMN and LMN disorder. Degeneration and loss of neurons in the UMNs in the primary motor cortex leads to loss of fibers within the corticospinal tract and lateral and anterior columns of the spinal cord. Disease of the LMNs leads to weakness, fasciculations, and atrophy of the affected muscle. A remarkable feature of the disease is that the entire sensory system and the intellect remain intact. The neurons for eye movement and the parasympathetic neurons in the sacral spinal cord also are spared. In the more advanced stages, muscles of the palate, pharynx, tongue, neck, and shoulders become involved, causing impairment of chewing, swallowing, and speech. Death usually results from involvement of cranial nerves and respiratory musculature.

MS is an example of a demyelinating disease of the CNS in which there is a slowly progressive breakdown of myelin and formation of plaques but sparing of the axis cylinder of the neuron. Interruption of neural conduction in MS is manifested by a variety of disabling signs and symptoms that depend on the neurons that are affected. The most common symptoms are paresthesias, optic neuritis, and motor weakness. The disease usually is characterized by exacerbations and remissions. Initially, near-normal function returns between exacerbations. The variety of symptoms, course of the disease, and lack of specific diagnostic tests make diagnosis and treatment of the disease difficult.

SPINAL CORD INJURY

Spinal cord injury (SCI) represents damage to the neural elements of the spinal cord. SCI is a disorder primarily of young adults, the most common cause being motor vehicle accidents, followed by falls, violence, sports injuries, and other types of injuries, including attempted suicide and occupational injuries. Of sports-related injuries, 66% are from diving.³³

Most spinal cord injuries involve damage to the vertebral column and or supporting ligaments as well as the spinal cord. Because of extensive tract systems that connect sensory afferent neurons and LMNs with high brain centers, spinal cord injuries commonly involve both sensory and motor function.

Injury to the Vertebral Column

Injuries to the vertebral column include fractures, dislocations, and subluxations. A fracture can occur at any part of the bony vertebrae, causing fragmentation of the bone. It most often involves the pedicle, lamina, or processes (*e.g.*, facets). Dislocation or subluxation (partial dislocation) injury causes the vertebral bodies to become displaced, with one overriding another and preventing correct alignment of the vertebral column. Damage to the ligaments or bony vertebrae may make the spine unstable. In an unstable spine, further unguarded movement of the spinal column can impinge on the spinal canal, causing compression or overstretching of neural tissue.

Most injuries result from some combination of compressive force or bending movement. Flexion injuries occur when forward bending of the spinal column exceeds the limits of normal movement. A typical flexion injury results when the head is struck from behind, such as in a fall with the back of the head as the point of impact. Extension injuries occur with excessive forced bending (*i.e.*, hyperextension) of the spine backward. A typical extension injury involves a fall in which the chin or face is the point of impact, causing hyperextension of the neck. Injuries of flexion and extension occur more commonly in the cervical spine (C4 to C6) than in any other area. Limitations imposed by the ribs, spinous processes, and joint capsules in the thoracic and lumbar spine make this area less flexible and less susceptible to flexion and extension injuries than the cervical spine.

A compression injury, causing the vertebral bones to shatter, squash, or even burst, occurs when there is spinal loading from a high-velocity blow to the top of the head or when landing forcefully on the feet (38-17A). This typically occurs at the cervical level (*e.g.*, diving injuries) or in the thoracolumbar area (*e.g.*, falling from a distance and landing on the feet). Compression injuries may occur when the vertebrae are weakened by conditions such as osteoporosis and cancer with bone metastasis. Axial rotation injuries can produce highly unstable injuries. Maximal axial rotation occurs in the cervical region, especially between C1 and C2 and at the lumbosacral joint (Fig. 38-17B). Coupling of vertebral motions is common in injury when two or more individual motions occur (*e.g.*, lateral bending and axial rotation).

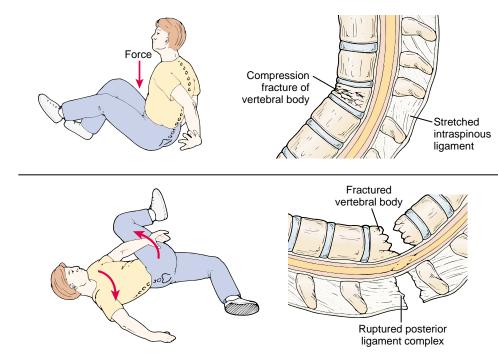
Types and Classification of Spinal Cord Injury

Spinal cord injury involves damage to the neural elements of the spinal cord. The damage may result from direct trauma to the cord from penetrating wounds or indirect injury resulting from vertebral fractures, fracture-dislocations, or subluxations of the spine. The spinal cord may be contused, not only at the site of injury, but also above and below the trauma site (Fig. 38-18). Traumatic injury may be complicated by blood flow to the cord, with resulting infarction.

Primary and Secondary Injuries

The pathophysiology of SCI can be divided into two types: primary and secondary.^{5,34} The *primary neurologic injury* occurs at the time of mechanical injury and is irreversible. It is charac-

Chapter 38: Alterations in Neuromuscular Function



■ **FIGURE 38-17** ■ (**A**) Compression vertebral fracture secondary to axial loading as occurs when a person falls from a height and lands on the buttocks. (B) Rotational injury, in which there is concurrent fracture and tearing of the posterior ligamentous complex, is caused by extreme lateral flexion or twisting of the head or neck.



■ FIGURE 38-18 ■ Cervical contusion. Hyperflexion injury caused forward angulation of the cervical cord, with fracture of the anterior lip of the underlying vertebral body. The cord is angulated over the superior-posterior ridge of the fixed underlying cervical body. (Rubin E., Farber J.L. [1999]. Pathology [3rd ed., p. 1465]. Philadelphia: Lippincott Williams & Wilkins)

terized by small hemorrhages in the gray matter of the cord, followed by edematous changes in the white matter that lead to necrosis of neural tissue. This type of pathology results from the forces of compression, stretch, and shear associated with fracture or compression of the spinal vertebrae, dislocation of vertebrae (e.g., flexion, extension, subluxation), and contusions caused by jarring of the cord in the spinal canal. Penetrating injuries produce lacerations and direct trauma to the cord and may occur with or without spinal column damage. Lacerations occur when there is cutting or tearing of the spinal cord, which injures nerve tissue and causes bleeding and edema.

Secondary injuries follow the primary injury and promote the spread of injury. Although there is considerable debate about the pathogenesis of secondary injuries, the tissue destruction that occurs ends in progressive neurologic damage. After SCI, several pathologic mechanisms come into play, including vascular damage, neuronal injury that leads to loss of reflexes below the level of injury, and release of vasoactive agents and cellular enzymes. Vascular pathology (i.e., vessel trauma and hemorrhage) can lead to ischemia, increased vascular permeability, and edema. Blood flow to the spinal cord may be further compromised by spinal shock that results from a loss of vasomotor tone and neural reflexes below the level of injury. The release of vasoactive substances (i.e., norepinephrine, serotonin, dopamine, and histamine) from the wound tissue causes vasospasm and impedes blood flow in the microcirculation, producing further necrosis of blood vessels and neurons. The release of proteolytic and lipolytic enzymes from injured cells causes delayed swelling, demyelination, and necrosis in the neural tissue in the spinal cord.

717

Classification

Alterations in body function that result from SCI depend on the level of injury and the amount of cord involvement. *Tetraplegia*, sometimes referred to as quadriplegia, is the impairment or loss of motor or sensory function (or both) after damage to neural structures in the cervical segments of the spinal cord.³⁵ It results in impairment of function in the arms, trunk, legs, and pelvic organs (see Fig. 38-7). *Paraplegia* refers to impairment or loss of motor or sensory function (or both) in the thoracic, lumbar, or sacral segments of the spinal cord from damage of neural elements in the spinal canal. With paraplegia, arm functioning is spared, but depending on the level of injury, functioning of the trunk, legs, and pelvic organs may be involved. Paraplegia includes conus medullaris and cauda equina injuries (to be discussed).

Further definitions of SCI describe the extent of neurologic damage as *complete* or *incomplete*. The prognosis for return of function is better in an incomplete injury because of preservation of axonal function.

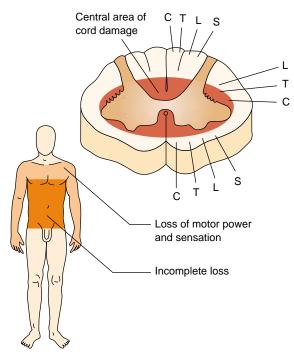
Incomplete Spinal Cord Injuries. Incomplete SCI implies there is some residual motor or sensory function below the level of injury.³⁶ Incomplete injuries may manifest in a variety of patterns but can be organized into certain patterns or "syndromes" that occur more frequently and reflect the predominant area of the cord that is involved. Types of incomplete lesions include the central cord syndrome, anterior cord syndrome, Brown-Séquard syndrome, and the conus medullaris syndrome.

A condition called *central cord syndrome* occurs when injury is predominantly in the central gray or white matter of the cord⁵ (Fig. 38-19). Because the corticospinal tract fibers are organized with those controlling the arms located more centrally and those controlling the legs located more laterally, some external axonal transmission may remain intact. Motor function of the upper extremities is affected, but the lower extremities may not be affected or may be affected to a lesser degree, with some sparing of sacral sensation. Bowel, bladder, and sexual functions usually are affected to various degrees and may parallel the degree of lower extremity involvement. This syndrome occurs almost exclusively in the cervical cord, rendering the lesion a UMN lesion with spastic paralysis. Central cord damage is more common in elderly persons with narrowing or stenotic changes in the spinal canal that are related to arthritis. Damage also may occur in persons with congenital stenosis.

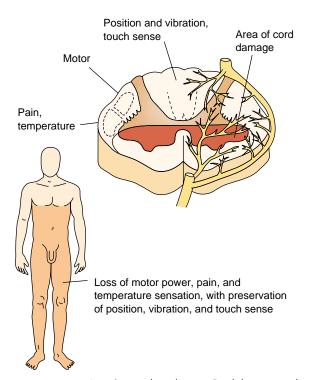
Anterior artery or anterior cord syndrome usually is caused by damage from infarction of the anterior spinal artery, resulting in damage to the anterior two thirds of the cord⁵ (Fig. 38-20). The deficits result in loss of motor function provided by the corticospinal tracts and loss of pain and temperature sensation from damage to the lateral spinothalamic tracts. The posterior one third of the cord is relatively unaffected, preserving the dorsal column axons that convey position, vibration, and touch sensation.

A condition called *Brown-Séquard syndrome* results from damage to a hemisection of the anterior and posterior cord⁵ (Fig. 38-21). The effect is a loss of voluntary motor function from the corticospinal tract, proprioception loss from the ipsilateral side of the body, and contralateral loss of pain and temperature sensation from the lateral spinothalamic tracts for all levels below the lesion.

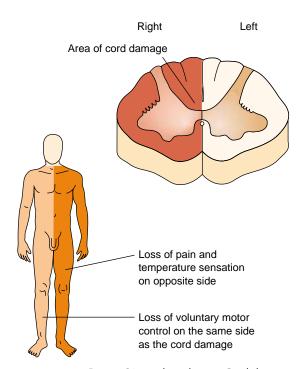
Conus medullaris syndrome involves damage to the conus medullaris or the sacral cord (*i.e.*, conus) and lumbar nerve



■ FIGURE 38-19 ■ Central cord syndrome. A cross-section of the cord shows central damage and the associated motor and sensory loss. (C, cervical; T, thoracic; L, lumbar; S, sacral). (Hickey J.V. [2003]. *The clinical practice of neurological and neurosurgical nursing* [5th ed., p. 420]. Philadelphia: Lippincott Williams & Wilkins)



■ FIGURE 38-20 ■ Anterior cord syndrome. Cord damage and associated motor and sensory loss are illustrated. (Hickey J.V. [2003]. *The clinical practice of neurological and neurosurgical nursing* [5th ed., p. 420]. Philadelphia: Lippincott Williams & Wilkins)



■ FIGURE 38-21 ■ Brown-Séquard syndrome. Cord damage and associated motor and sensory loss are illustrated. (Hickey J.V. [2003]. The clinical practice of neurological and neurosurgical nursing [5th ed., p. 421]. Philadelphia: Lippincott Williams & Wilkins)

roots in the neural canal. Functional deficits resulting from this type of injury usually result in flaccid bowel, bladder, and sexual function. Sacral segments occasionally show preserved reflexes if only the conus is affected. Motor function in the legs and feet may be impaired without significant sensory impairment. Damage to the lumbosacral nerve roots in the spinal canal usually results in LMN and sensory neuron damage known as cauda equina syndrome. Functional deficits present as various patterns of asymmetric flaccid paralysis, sensory impairment, and pain.

Complete Spinal Cord Injuries. Complete SCI implies there is an absence of motor and sensory function below the level of injury. Complete cord injuries can result from severance of the cord, disruption of nerve fibers although they remain intact, or interruption of blood supply to that segment, resulting in complete destruction of neural tissue and UMN or LMN paralysis. Approximately 3% of persons with signs of complete injuries on initial examination experience some recovery within 24 hours.36

Acute Spinal Cord Injury

Sudden complete transection of the spinal cord results in complete loss of motor, sensory, reflex, and autonomic function below the level of injury. This immediate response to spinal cord injury is often referred to as spinal cord shock. It is characterized by flaccid paralysis with loss of tendon reflexes below the level of injury, absence of somatic and visceral sensations below the level of injury, and loss of bowel and bladder function. Loss of systemic sympathetic vasomotor tone may result in vasodilation, increased venous capacity, and hypotension. These manifestations occur regardless of whether the level of the lesion eventually will produce spastic (UMN) or flaccid (LMN) paralysis. The basic mechanisms accounting for transient spinal shock are unknown. Spinal shock may last for hours, days, or weeks. Usually, if reflex function returns by the time the person reaches the hospital, the neuromuscular changes are reversible. This type of reversible spinal shock may occur in football-type injuries, in which jarring of the spinal cord produces a concussion-like syndrome with loss of movement and reflexes, followed by full recovery within days. In persons in whom the loss of reflexes persists, hypotension and bradycardia may become critical but manageable problems. In general, the higher the level of injury, the greater is the effect.

Management. The goal of management of acute SCI is to reduce the neurologic deficit and prevent any additional loss of neurologic function. The specific steps in resuscitation and initial evaluation can be carried out at the trauma site or in the emergency room, depending on the urgency of the situation.^{34,37} Most traumatic injuries to the spinal column render it unstable, mandating measures such as immobilization with collars and backboards and limiting the movement of persons at risk for or with known SCI. Every person with multiple trauma or head injury, including victims of traffic and sporting accidents, should be suspected of having sustained an acute SCI.34,37

The nature of the injury determines further methods of stabilization and treatment. In unstable injuries of the cervical spine, cervical traction improves or restores spinal alignment, decompresses neural structures, and facilitates recovery. Fractures and dislocations of the thoracic and lumbar vertebrae may be initially stabilized by restricting the person to bed rest and turning him or her in a log-rolling manner to keep the spine rigid. Gunshot or stab wounds of the spinal column may not produce structural instability and require immobilization. The goal of early surgical intervention for an unstable spine is to provide internal skeletal stabilization so that early mobilization and rehabilitation can occur.

One of the more important aspects of early SCI care is the prevention and treatment of spinal or systemic shock and the hypoxia associated with compromised respiration. Correcting hypotension or hypoxia is essential to maintaining circulation to the injured cord.³⁸ The use of high-dose methylprednisolone has been shown to improve the outcome from SCI when given shortly after injury. Methylprednisolone is a short-acting corticosteroid that has been used extensively in the treatment of inflammatory and allergic disorders. In acute SCI, it is thought to stabilize cell membranes, enhance impulse generation, improve blood flow, and inhibit free radical formation.

Disruption of Functional Abilities

Functional abilities after SCI are subject to various degrees of somatosensory and skeletal muscle function loss and altered reflex activity based on the level of cord injury and extent of cord damage (Table 38-2).

Somatosensory and Skeletal Muscle Function

Motor function in cervical injuries ranges from complete dependence to independence with or without assistive devices in activities of mobility and self-care. The functional levels of cervical injury are related to C5, C6, C7, or C8 innervation. At the

720 Unit Ten: Alterations in the Nervous System

Tunctional Abilities by Level of Cord injury					
lnjury Level	Segmental Sensorimotor Function	Dressing, Eating	Elimination	Mobility*	
C1	Little or no sensation or control of head and neck; no diaphragm con- trol; requires continuous ventilation	Dependent	Dependent	Limited. Voice or sip-n-puff con- trolled electric wheelchair	
C2 to C3	Head and neck sensation; some neck control. Independent of mechanical ventilation for short periods	Dependent	Dependent	Same as for C1	
C4	Good head and neck sensation and motor control; some shoulder eleva-	Dependent; may be able to eat with	Dependent	Limited to voice, mouth, head, chin, or shoulder-controlled electric wheelchair	
C5	tion; diaphragm movement Full head and neck control; shoulder strength; elbow flexion	adaptive sling Independent with assistance	Maximal assistance	Electric or modified manual wheel chair, needs transfer assistance	
C6	Fully innervated shoulder; wrist exten- sion or dorsiflexion	Independent or with minimal assistance	Independent or with minimal assistance	Independent in transfers and wheelchair	
C7 to C8	Full elbow extension; wrist plantar flexion; some finger control	Independent	Independent	Independent; manual wheelchair	
T1 to T5	Full hand and finger control; use of intercostal and thoracic muscles	Independent	Independent	Independent; manual wheelchair	
T6 to T10	Abdominal muscle control, partial to good balance with trunk muscles	Independent	Independent	Independent; manual wheelchair	
T11 to L5	Hip flexors, hip abductors (L1–3); knee extension (L2–4); knee flexion and ankle dorsiflexion (L4–5)	Independent	Independent	Short distance to full ambulation with assistance	
S1 to S5	Full leg, foot, and ankle control; inner- vation of perineal muscles for bowel, bladder, and sexual function (S2–4)	Independent	Normal to impaired bowel and blad- der function	Ambulate independently with or without assistance	

TABLE 38-2 Functional Abilities by Level of Cord Injury

*Assistance refers to adaptive equipment, setup, or physical assistance.

C5 level, deltoid and biceps function is spared, allowing full head, neck, and diaphragm control with good shoulder strength and full elbow flexion. At the C6 level, wrist dorsiflexion by the way of wrist extensors is functional, allowing tenodesis, which is the natural bending inward and flexion of the fingers when the wrist is extended and bent backward. Tenodesis is a key movement because it can be used to pick up objects when finger movement is absent. A functional C7 injury allows full elbow flexion and extension, wrist plantar flexion, and some finger control. At the C8 level, finger flexion is added.

Thoracic cord injuries (T1 to T12) allow full upper extremity control with limited to full control of intercostal and trunk muscles and balance. Injury at the T1 level allows full fine motor control of the fingers. Because of the lack of specific functional indicators at the thoracic levels, the level of injury usually is determined by sensory level testing.

Functional capacity in the L1 through L5 nerve innervations allows hip flexors, hip abductors (L1 to L3), movement of the knees (L2 to L5), and ankle dorsiflexion (L4 to L5). Sacral (S1 to S5) innervation allows for full leg, foot, and ankle control and innervation of perineal musculature for bowel, bladder, and sexual function.

Altered Reflex Activity. Spinal cord reflexes are fully integrated within the spinal cord and can function independent of input from higher centers. Altered spinal reflex activity after SCI is essentially determined by the level of injury and whether UMNs or LMNs are affected. With UMN injuries at T12 and above, the cord reflexes remain intact, while communication

pathways with higher centers have been interrupted. This results in spasticity of involved skeletal muscle groups and of smooth and skeletal muscles that control bowel, bladder, and sexual function. In LMN injuries at T12 or below, the reflex circuitry itself has been damaged at the level of the spinal cord or spinal nerve, resulting in a decrease or absence of reflex function. The LMN injuries cause flaccid paralysis of involved skeletal muscle groups and the smooth and skeletal muscles that control bowel, bladder, and sexual function. However, injuries near the T12 level may result in mixed UMN and LMN deficits (*e.g.*, spastic paralysis of the bowel and bladder with flaccid muscle tone).

After the period of spinal shock in a UMN injury, isolated spinal reflex activity and muscle tone that is not under the control of higher centers returns. This may result in hypertonia and spasticity of skeletal muscles below the level of injury. These spastic movements are involuntary, instead of voluntary, a distinction that needs to be explained to persons with SCI and their families. The antigravity muscles, the flexors of the arms and extensors of the legs, are predominantly affected. Spastic movements usually are heightened initially after injury, reaching a peak and then becoming stable in approximately 2 years.

The stimuli for reflex muscle spasm arise from somatic and visceral afferent pathways that enter the cord below the level of injury. The most common of these stimuli are muscle stretching, bladder infections or stones, fistulas, bowel distention or impaction, pressure areas or irritation of the skin, and infections. Because the stimuli that precipitate spasms vary from person to person, careful assessment needs to be done to identify the factors that precipitate spasm in each person. Passive range-of-motion exercises to stretch the spastic muscles help to prevent spasm induced by muscle stretching such as occurs with change in body position.

Spasticity in and of itself is not detrimental and may even facilitate maintenance of muscle tone to prevent muscle wasting, improve venous return, and aid in mobility. Spasms become detrimental when they impair safety or reduce the ability to make functional gains in mobility and activities of daily living. Spasms also may cause trauma to bones and tissues, leading to joint contractures and skin breakdown.

Respiratory Muscle Function. Ventilation requires movement of the expiratory and inspiratory muscles, all of which receive innervation from the spinal cord. The main muscle of ventilation, the diaphragm, is innervated by segments C3 to C5 through the phrenic nerves. The intercostal muscles, which function in elevating the rib cage and are needed for coughing and deep breathing, are innervated by spinal segments T1 through T7. The major muscles of expiration are the abdominal muscles, which receive their innervation from levels T6 to T12.

Although the ability to inhale and exhale may be preserved at various levels of SCI, functional deficits in ventilation are most apparent in the quality of the breathing cycle and the ability to oxygenate tissues, eliminate carbon dioxide, and mobilize secretions. Cord injuries involving C1 to C3 result in a lack of respiratory effort, and affected patients require assisted ventilation. Although a C3 to C5 injury allows partial or full diaphragmatic function, ventilation is diminished because of the loss of intercostal muscle function, resulting in shallow breaths and a weak cough. Below the C5 level, as less intercostal and abdominal musculature is affected, the ability to take a deep breath and cough is less impaired. Maintenance therapy consists of muscle training to strengthen existing muscles for endurance and mobilization of secretions. The ability to speak is compromised with assisted ventilation, whether continuous or intermittent. Thus, ensuring adequate communication of needs is also essential.

Autonomic Nervous System Function

In addition to its effects on skeletal muscle function, SCI interrupts ANS function below the site of injury. This includes sympathetic outflow from the thoracic and lumbar cord and parasympathetic outflow from the sacral cord. Because of their site of exit from the CNS, the cranial nerves, such as the vagus, are unaffected. Dependent upon the level of injury, the spinal reflexes that control ANS function are largely isolated from the rest of the CNS. Afferent sensory input that enters the spinal cord is unaffected, as is the efferent motor output from the cord. Lacking is the regulation and integration of reflex function by centers in the brain and brain stem. This results in a situation in which the autonomic reflexes below the level of injury are uncontrolled, while those above the level of injury function in a relatively controlled manner.

The sympathetic nervous system regulation of circulatory function and thermoregulation present some of the most severe problems in SCI. The higher the level of injury and the greater the surface area affected, the more profound are the effects on circulation and thermoregulation. Persons with injury at the T6 level or above experience problems in regulating vasomotor tone; those with injuries below the T6 level usually have sufficient sympathetic function to maintain adequate vasomotor function. The level of injury and its corresponding problems may vary among persons, and some dysfunctional effects may be seen at levels below T6. With lower lumbar and sacral injuries, sympathetic function remains essentially unaltered.

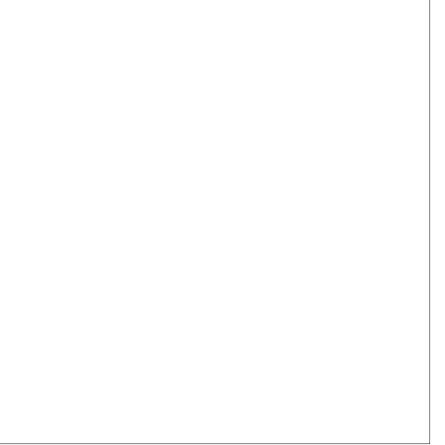
The Vasovagal Response. The vagus nerve (cranial nerve X) normally exerts a continuous inhibitory effect on heart rate. Vagal stimulation that causes a marked bradycardia by way of the vagus nerve is called the *vasovagal response.* Visceral afferent input to the vagal centers in the brain stem of persons with tetraplegia or high-level paraplegia can produce marked bradycardia when unchecked by a dysfunctional sympathetic nervous system. Severe bradycardia and even asystole can result when the vasovagal response is elicited by deep endotracheal suctioning or rapid position change. Preventive measures, such as hyperoxygenation before, during, and after suctioning, are advised. Rapid position changes should be avoided or anticipated, and anticholinergic drugs should be immediately available to counteract severe episodes of bradycardia.

Autonomic Dysreflexia. Autonomic dysreflexia, also known as *autonomic hyperreflexia*, represents an acute episode of exaggerated sympathetic reflex responses that occur in persons with injuries at T6 and above, in which CNS control of sympathetic responses is lost (see Fig. 38-22). It does not occur until spinal shock has resolved and autonomic reflexes return, most often within the first 6 months after injury. It is most unpredictable during the first year after injury but can occur throughout the person's lifetime.

Autonomic dysreflexia, is characterized by vasospasm, hypertension ranging from mild (20 mm Hg above baseline) to severe (as high as 240/120 mm Hg or higher), skin pallor, and goose flesh associated with the piloerector response. Because baroreceptor function and parasympathetic control of heart rate travels by way of the cranial nerves, these responses remain intact. Continued hypertension produces a baroreflex-mediated vagal slowing of the heart rate to bradycardic levels. There is an accompanying baroreflex-mediated vasodilatation with flushed skin and profuse sweating above the level of injury, headache ranging from dull to severe and pounding, nasal stuffiness, and feelings of anxiety. A person may experience one, several, or all of the symptoms with each episode.

The stimuli initiating the dysreflexic response include visceral distention, such as a full bladder or rectum; stimulation of pain receptors, such as occurs with pressure ulcers, ingrown toenails, dressing changes, and diagnostic or operative procedures; and visceral contractions, such as ejaculation, bladder spasms, or uterine contractions. In many cases, the dysreflexic response results from a full bladder.

Autonomic dysreflexia is a clinical emergency, and without prompt and adequate treatment, convulsions, loss of consciousness, and even death can occur. The major components of treatment include monitoring blood pressure while removing or correcting the initiating cause or stimulus. The person should be placed in an upright position, and all support hose or binders should be removed to promote venous pooling of blood and reduce venous return, thereby decreasing blood pressure. If the stimuli have been removed or the stimuli cannot



■ **FIGURE 38-22** ■ Mechanisms of autonomic hyperreflexia.

be identified and the upright position is established but the blood pressure remains elevated, drugs that block autonomic function are administered. Prevention of the type of stimuli that trigger the dysreflexic event is advocated.

Postural Hypotension. Postural, or orthostatic, hypotension usually occurs in persons with injuries at T4 to T6 and above and is related to the interruption of descending control of sympathetic outflow to blood vessels in the extremities and abdomen. Pooling of blood, along with gravitational forces, impairs venous return to the heart, and there is a subsequent decrease in cardiac output when the person is placed in an upright position. The signs of orthostatic hypotension include dizziness, pallor, excessive sweating above the level of the lesion, complaints of blurred vision, and possibly fainting. Postural hypotension usually is prevented by slow changes in position and measures to promote venous return.

Other Functions

Edema and Deep Vein Thrombosis. Edema and deep vein thrombosis are common problems in persons with SCI. The development of edema is related to decreased peripheral vascular resistance, areflexia or decreased tone in the paralyzed limbs, and immobility that causes increased venous pressure and abnormal pooling of blood in the abdomen, lower limbs, and upper extremities. Edema in the dependent body parts usually is relieved by positioning to minimize gravitational forces or by using compression devices (*e.g.*, support stockings, binders) that encourage venous return.

Deep vein thrombosis also is of concern because of the venous pooling and loss of skeletal muscle movement below the level of injury. Although it is seen more frequently in the postacute phase of SCI, it often has its origin during the events surrounding the initial injury. Prevention includes assessment of risk and measures to prevent venous pooling of blood, especially in the paralyzed limbs (*e.g.*, range-of-motion exercises and vascular compression devices).

Bladder, Bowel, and Sexual Function. Among the most devastating consequences of SCI is the loss of bowel and bladder function. Loss of bladder function results from disruption of neural pathways between the bladder and the reflex voiding center at the S2 to S4 level (*i.e.*, an LMN lesion) or between the reflex voiding center and higher brain centers for communication and coordinated sphincter control (*i.e.*, a UMN lesion). Persons with UMN lesions or spastic bladders lack awareness of bladder filling (*i.e.*, storage) and voluntary control of voiding (*i.e.*, evacuation). In LMN lesions or flaccid bladder dysfunction, lack of awareness of bladder filling and lack of bladder tone render the person unable to void voluntarily or involuntarily.

Bowel elimination is a coordinated function involving the enteric nervous system, the ANS, and the CNS. Persons with SCI above S2 to S4 develop spastic functioning of the defecation reflex and loss of voluntary control of the external anal sphincter. Damage to the cord at the S2 to S4 level causes flaccid functioning of the defecation reflex and loss of anal sphincter tone. Even though the enteric nervous system innervation of the bowel remains intact, without the defecation reflex, peristaltic movements are ineffective in evacuating stool.

Sexual function, as in bladder and bowel control, is mediated by the S2 to S4 segments of the spinal cord. The genital sexual response in SCI, which is manifested by an erection in men and vaginal lubrication in women, may be initiated by mental or touch stimuli, depending on the level of injury. The T11 to L2 cord segments have been identified as the mentalstimuli, or psychogenic, sexual response area, where autonomic nerve pathways in communication with the forebrain leave the cord and innervate the genitalia. The S2 to S4 cord segments have been identified as the sexual-touch reflex center. In T10 or higher injuries (UMN lesion), reflex sexual response to genital touch may occur freely. However, a sexual response to mental stimuli (T11 to L2) does not occur because of the spinal lesion blocking the communication pathway. In an injury at T12 or below (LMN lesion), the sexual reflex center may be damaged, and there may be no response to touch.

In men, the lack of erectile ability or inability to experience penile sensations or orgasm is not a reliable indicator of fertility, which should be evaluated by an expert. In women, fertility is parallel to menses; usually, it is delayed 3 months to 5 months after injury. There are hazards to pregnancy, labor, and birth control devices relative to SCI that require knowledgeable health care providers.

Skin Integrity. The entire surface of the skin is innervated by cranial or spinal nerves organized into dermatomes that show cutaneous distribution. The CNS and ANS also play a vital role in skin function. The sympathetic nervous system, through control of vasomotor and sweat gland activity, influences the health of the skin by providing adequate circulation, excretion of body fluids, and temperature regulation. The lack of sensory warning mechanisms and voluntary motor ability below the level of injury, coupled with circulatory changes, place the person with SCI at major risk for disruption of skin integrity. Significant factors associated with disruption of skin integrity are pressure, shearing forces, and localized trauma and irritation. Relieving pressure, allowing adequate circulation to the skin, and skin inspection are primary ways of maintaining skin integrity. Of all the complications after SCI, skin breakdown is the most preventable.

In summary, spinal cord injury is a disabling neurologic condition most commonly caused by motor vehicle accidents, falls, and sports injuries. Dysfunctions of the nervous system after SCI comprise various degrees of sensorimotor loss and altered reflex activity based on the level of injury and extent of cord damage. Depending on the level of injury, the physical problems of SCI include spinal shock; ventilation and communication problems; ANS dysfunction that predisposes to the vasovagal response, autonomic hyperreflexia, impaired body temperature regulation, and postural hypotension; impaired muscle pump and venous innervation leading to edema of dependent areas of the body and risk of deep vein thrombosis; altered sensorimotor integrity that contributes to uncontrolled muscle spasms, altered pain responses, and threat to skin integrity; alterations in bowel and bladder elimination; and impaired sexual function.

REVIEW QUESTIONS

Define a motor unit and characterize its mechanism of controlling muscle movement.

Describe muscle atrophy and differentiate between disuse and degenerative atrophy.

Describe the pathology associated with Duchenne muscular dystrophy.

Relate the clinical manifestations of myasthenia gravis to its cause.

Compare the cause and manifestations of peripheral mononeuropathies with peripheral polyneuropathies.

Describe the functional organization of the basal ganglia and communication pathways with the thalamus and cerebral cortex.

State the possible mechanisms responsible for the development of Parkinson's disease and characterize the manifestations and treatment of the disorder.

Relate the functions of the cerebellum to production of vestibulocerebellar ataxia, decomposition of movement, and cerebellar tremor.

Relate the pathologic UMN and LMN changes that occur in amyotrophic lateral sclerosis to the manifestations of the disease.

Explain the significance of demyelination and plaque formation in multiple sclerosis.

Describe the manifestations of multiple sclerosis.

Explain how loss of UMN function contributes to the muscle spasms that occur after recovery from spinal cord injury.

State the effects of spinal cord injury on ventilation and communication, the ANS, cardiovascular function, sensorimotor function, and bowel and bladder function.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- Kandel E.R., Schwartz J.H., Jessell T.M. (2000). Principles of neural science (4th ed., pp. 816–831, 833, 853–854). New York: McGraw-Hill.
- Guyton A., Hall J.E. (2000). *Textbook of* medical *physiology* (10th ed., pp. 634–637, 973). Philadelphia: W.B. Saunders.
- Penfield W., Rasmussen T. (1950). The cerebral cortex of man. New York: Macmillan.
- Bear M.E., Connors B.W., Paradiso M.A. (1999). Neuroscience: Exploring the brain (pp. 348–357). Baltimore: Williams & Wilkins.
- Hickey J.V. (1997). Neurological and neurosurgical nursing (4th ed., pp. 419–465, 469–480). Philadelphia: Lippincott-Raven.
- Cotran R.S., Kumar V., Collins T. (1999). Robbins pathologic basis of disease (6th ed., pp. 1275–1276, 1281–1282, 1289, 1326–1328). Philadelphia: W.B. Saunders.
- Rubin E., Farber J.L. (1999). Pathology (3rd ed., pp. 1496–1498, 1502–1505). Philadelphia: Lippincott Williams & Wilkins.
- Sarnat H.B. (2000). Muscular dystrophies. In Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.), *Nelson textbook of pediatrics* (16th ed., pp. 1873–1877). Philadelphia: W.B. Saunders.
- Katzung B.G. (2001). Basic and clinical pharmacology (8th ed., pp. 92– 102). New York: Lange Medical Books/McGraw-Hill.

724 Unit Ten: Alterations in the Nervous System

- Drachman D.B. (1994). Myasthenia gravis. New England Journal of Medicine 330, 1797–1810.
- Bedlack R.S., Sanders D.B. (2000). How to handle myasthenic crisis. Postgraduate Medicine 10, 211–222.
- Dawson D.M. (1993). Entrapment neuropathies of the upper extremities. New England Journal of Medicine 329, 2013–2018.
- 13. Ropper A.H. (1992). The Guillain-Barré syndrome. New England Journal of Medicine 326, 10–16.
- Rees J.H., Saudain S.E., Gregson N.A., et al. (1995). Campylobacter jejuni infection and Guillain-Barré syndrome. New England Journal of Medicine 333, 1374–1379.
- Langmuir A.D., Bregman D.J., Nathanson N., et al. (1984). An epidemic and clinical evaluation of Guillain-Barré syndrome reported in association of swine influenza vaccines. *American Journal of Epidemiology* 119, 841–879.
- 16. Acute Low Back Problems Guideline Panel. (1994). Acute low back problems in adults: Assessment and treatment. AHCPR publication no. 95-0642. Rockville, MD: Agency for Health Care Policy and Research, Public Health Service, U.S. Department of Health and Human Services.
- 17. Deyo R.A. (1998). Low-back pain. Scientific American 283, 49-53.
- Moore K.L., Dalley A.F. (1999). Clinically oriented anatomy (4th ed., pp. 452–453). Philadelphia: Lippincott Williams & Wilkins.
- Bratton R.L. (1999). Assessment and management of acute low back pain. American Family Physician 60, 2299–2308.
- 20. Ng D.C. (1996). Parkinson's disease: Diagnosis and treatment. Western Journal of Medicine 165, 234-240.
- Parkinson J. (1817). An essay on the shaking palsy. London: Sherwood, Nelley & Jones.
- Youdim M.B.H., Riederer P. (1997). Understanding Parkinson's disease. Scientific American 276, 52–59.
- Lücking C.B., Dürr A., Bonifati V., et al. (European Consortium on Genetic Susceptibility in Parkinson's Disease Genetic Study Group). (2000). Association between early-onset Parkinson's disease and mutations in the Parkin gene. *New England Journal of Medicine* 342, 1560–1567.
- 24. Colcher A., Simuni T. (1999). Clinical manifestations of Parkinson's disease. *Medical Clinics of North America* 83, 327–347.

- Young R. (1999). Update on Parkinson's disease. American Family Physician 59 (8), 2155–2167.
- Rowland L.P., Shneider N.A. (2001). Amyotrophic lateral sclerosis. New England Journal of Medicine 344, 1688–1700.
- Walling A.D. (1999). Amyotrophic lateral sclerosis: Lou Gehrig's disease. American Family Practitioner 59 (6), 1489–1496.
- 28. Noseworthy J.H., Lucchinett C., Rodrequez M., et al. (2000). Multiple sclerosis. *New England Journal of Medicine* 343, 938–952.
- Anderson D.W., Ellenberg J.H., Leventhal C.M., et al. (1992). Revised estimate of multiple sclerosis in the United States. *Annals of Neurology* 31, 333–336.
- Lublin F.D., Reingold S.C. (1996). Defining the clinical course of multiple sclerosis: Results of an international survey. *Neurology* 46, 907–911.
- Rudick R.A., Cohen J.A., Weinstock-Guttman B., et al. (1997). Management of multiple sclerosis. New England Journal of Medicine 337, 1604–1611.
- Brod S.A., Lindsey W., Wolinsky J.S. (1996). Multiple sclerosis: Clinical presentation, diagnosis and treatment. *American Family Physician* 54, 1301–1311.
- National Spinal Cord Injury Statistical Center. (2001). Spinal cord injury: Facts and figures at a glance. Birmingham: University of Alabama. [On-line]. Available: http://www.spinalcord.uab.edu.
- 34. Chiles B.W., Cooper P.R. (1996). Acute spinal cord injury. New England Journal of Medicine 334, 514-520.
- American Spinal Injury Association. (1992). Standards of neurological and functional classification of spinal cord injury. Chicago: American Spinal Cord Injury Association.
- Buckley D.A., Guanci M.K. (1999). Spinal cord trauma. Nursing Clinics of North America 34, 661–687.
- Fehling M.G., Louw D. (1996). Initial stabilization and medical management of acute spinal cord injury. *American Family Physician* 42, 155–162.
- Bracken M.B., Shepard M.J., Collins W.F., et al. (1997). Administration of methylprednisolone for 24 or 48 hours or tirilazad mesylate for 48 hours in the treatment of acute spinal cord injury: Results of the Third National Acute Spinal Cord Injury Study. *Journal of the American Medical Association* 277, 1597–1604.

39



Organization and Control of Somatosensory Function Sensory Systems

The Sensory Unit

Dermatomal Pattern of Dorsal Root Innervation Spinal Circuitry and Ascending Neural Pathways Central Processing of Somatosensory Information

Sensory Modalities

Tactile Sensation Thermal Sensation Position Sensation

Pain

Pain Theories

Pain Mechanisms and Pathways Pain Receptors and Mediators Spinal Cord Circuitry and Pathways **Brain Centers and Pain Perception Central Pathways for Pain Modulation** Pain Threshold and Tolerance **Types of Pain** Cutaneous and Deep Somatic Pain Visceral Pain **Referred Pain** Acute and Chronic Pain **Pain Management** Assessment Treatment Alterations in Pain Sensitivity and Special Types of Pain Alterations in Pain Sensitivity Special Types of Pain **Neuropathic Pain** Neuralgia **Complex Regional Pain Syndrome Phantom Limb Pain**

Headache Migraine Headache Migraine Headache in Children Cluster Headache Tension-type Headache Temporomandibular Joint Pain Pain in Children and Older Adults Pain in Children Pain in Older Adults

Sensory mechanisms provide individuals with a continuous stream of information about their bodies, the outside world, and the interactions between the two. The somatosensory component of the nervous system provides an awareness of body sensations such as touch, temperature, limb position, and pain.

ORGANIZATION AND CONTROL OF SOMATOSENSORY FUNCTION

The somatosensory system is designed to provide the central nervous system (CNS) with information about the body. Sensory neurons can be divided into three types that vary in distribution and the type of sensation detected: general somatic, special somatic, and general visceral afferent neurons (see Chapter 36). *General somatic afferent neurons* have branches with widespread distribution throughout the body and with many distinct types of receptors that result in sensations such as pain, touch, and temperature. *Special somatic afferent neurons* have receptors located primarily in muscles, tendons, and joints. These receptors sense position and movement of the body. *General visceral afferent neurons* have receptors on various visceral structures and sense fullness and discomfort.

Sensory Systems

Sensory systems can be conceptualized as a serial succession of neurons consisting of first-order, second-order, and third-order neurons. The *first-order neurons* contain the sensory receptors and transmit sensory information from the periphery to the CNS. The *second-order neurons* communicate with various reflex networks and sensory pathways in the spinal cord and contain the ascending pathways that travel to the thalamus. *Third-order neurons* relay information from the thalamus to the cerebral cortex (Fig. 39-1). Many interneurons process and modify the sensory information at the level of the second- and third-order neurons, and many more participate before coordinated and appropriate learned-movement responses occur.

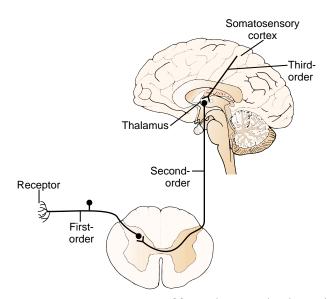
The Sensory Unit

The somatosensory experience arises from information provided by a variety of receptors distributed throughout the body. There are four major modalities of sensory experience: (1) discriminative touch, which is required to identify the size and shape of objects and their movement across the skin; (2) temperature sensation; (3) sense of movement of the limbs and joints of the body; and (4) nociception or pain sense.

Each of the somatosensory modalities is mediated by a distinct system of receptors and pathways to the brain. However, all somatosensory information from the limbs and trunk shares a common class of sensory neurons called *dorsal root ganglion neurons*. Somatosensory information from the face and cranial structures is transmitted by the trigeminal sensory neurons, which function in the same manner as the dorsal root ganglion neurons. The cell body of the dorsal root ganglion neuron, its receptor (which innervates a small area of periphery), and its central axon (which projects to the CNS) form a *sensory unit*. Individual dorsal root ganglion neurons respond selectively to specific types of stimuli because of their specialized peripheral terminals, or receptors.

Dermatomal Pattern of Dorsal Root Innervation

The somatosensory innervation of the body, including the head, retains a basic segmental organizational pattern that was established during embryonic development. The region of the body wall that is supplied by a single pair of dorsal root ganglia is called a *dermatome*. These dorsal root ganglion-



■ **FIGURE 39-1** ■ Arrangement of first-order, second-order, and third-order neurons of the somatosensory system.

KEY CONCEPTS

THE SOMATOSENSORY SYSTEM

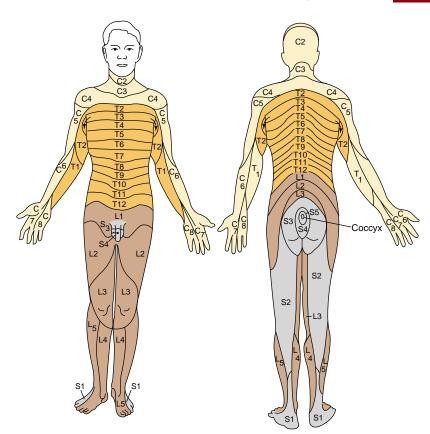
- The somatosensory system relays information to the CNS about four major body sensations: touch, temperature, pain, and body position. Stimulation of receptors on regions of the body wall is required to initiate the sensory response.
- The system is organized into dermatomes, with each segment supplied by a single dorsal root ganglion that sequentially relays the sensory information to the spinal cord, the thalamus, and the sensory cortex.
- Two pathways carry sensory information through the CNS. The discriminative pathway crosses in the medulla and relays touch and body position. The anterolateral pathway crosses in the spinal cord and relays temperature and pain sensation from the opposite side of the body.

innervated strips occur in a regular sequence moving upward from the second coccygeal segment through the cervical segments, reflecting the basic segmental organization of the body and the nervous system (Fig. 39-2). The cranial nerves that innervate the head send their axons to equivalent nuclei in the brain stem. Neighboring dermatomes overlap one another sufficiently so that a loss of one dorsal root or root ganglion results in reduced but not total loss of sensory innervation of a dermatome (Fig. 39-3). Dermatome maps are helpful in interpreting the level and extent of sensory deficits that are the result of segmental nerve and spinal cord damage.

Spinal Circuitry and Ascending Neural Pathways

On entry into the spinal cord, the central axons of the somatosensory neurons branch extensively and project to nuclei in the spinal gray matter. Some branches become involved in local spinal cord reflexes and directly initiate motor reflexes (*e.g.*, flexor-withdrawal reflex). Two parallel pathways, the rapid conducting *discriminative pathway* and the slower conducting *anterolateral pathway*, transmit information from the spinal cord to the thalamic level of sensation, each taking a different route through the CNS.

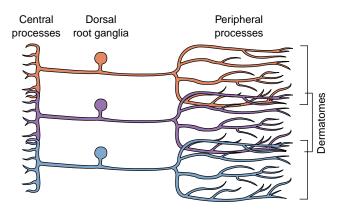
The Discriminative Pathway. The discriminative pathway, which crosses at the base of the medulla, is used for the rapid transmission of sensory information such as discriminative touch (Fig. 39-4). It contains branches of primary afferent axons that travel up the ipsilateral (*i.e.*, same side) dorsal columns of the spinal cord white matter and synapse with highly evolved somatosensory input association neurons in the medulla. The discriminative pathway uses only three neurons to transmit information from a sensory receptor to the somatosensory strip of parietal cerebral cortex of the opposite side of the brain: (1) the primary dorsal root ganglion neuron, which projects its central axon to the dorsal column nuclei; (2) the dorsal column neuron, which sends its axon through a rapid



■ FIGURE 39-2 ■ Cutaneous distribution of spinal nerves (dermatomes). (Barr, M. [1993]. *The human nervous system*. New York: Harper & Row)

conducting tract, called the *medial lemniscus*, that crosses at the base of the medulla and travels to the thalamus on the opposite side of the brain, where basic sensation begins; and (3) the thalamic neuron, which projects its axons through the somatosensory radiation to the primary sensory cortex. The medial lemniscus is joined by fibers from the sensory nucleus of the trigeminal nerve (cranial nerve V) that supplies the face. Sensory information arriving at the sensory cortex by this route can be discretely localized and discriminated in terms of intensity.

One of the distinct features of the discriminative pathway is that it relays precise information regarding spatial orientation. This is the only pathway taken by the sensations of muscle and joint movement, vibration, and delicate discriminative touch,

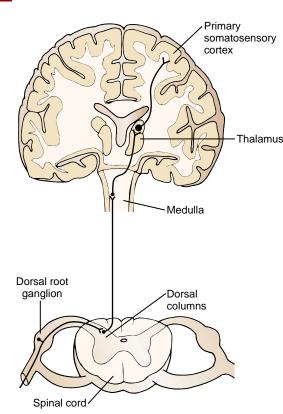


■ FIGURE 39-3 ■ The dermatomes formed by the peripheral processes of adjacent spinal nerves overlap on the body surface. The central processes of these fibers also overlap in their spinal distribution.

as is required to differentiate correctly the location of touch on the skin at two neighboring points (i.e., two-point discrimination). One of the important functions of the discriminative pathway is to integrate the input from multiple receptors. The sense of shape and size of an object in the absence of visualization, called stereognosis, is based on precise afferent information from muscle, tendon, and joint receptors. For example, a screwdriver is perceived as being different from a knife in terms of its texture (tactile sensibility) and shape based on the relative position of the fingers as they move over the object. This complex interpretive perception requires that the discriminative system must be functioning optimally and that higherorder parietal association cortex processing and prior learning must have occurred. If the discriminative somatosensory pathway is functional but the parietal association cortex has become discretely damaged, the person can correctly describe the object but does not recognize that it is a screwdriver. This deficit is called *astereognosis*.

The Anterolateral Pathway. The anterolateral pathways (anterior and lateral spinothalamic pathways), which crosses within the first few segments of entering the spinal cord, consists of bilateral multisynaptic slow-conducting tracts (Fig. 39-5). These pathways provide for transmission of sensory information such as pain, thermal sensations, crude touch, and pressure that does not require discrete localization of signal source or fine discrimination of intensity. The fibers of the anterolateral pathway originate in the dorsal horns at the level of the segmental nerve, where the dorsal root neurons enter the spinal cord. They cross in the anterior commissure of the cord, within a few segments of origin, to the opposite anterolateral pathway, where they ascend upward toward the brain. The spinothalamic tract

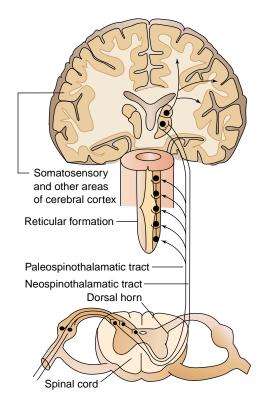
728 Unit Ten: Alterations in the Nervous System



■ FIGURE 39-4 ■ Discriminative pathway. This pathway is an ascending system for rapid transmission of sensations that relate joint movement (kinesthesis), body position (proprioception), vibration, and delicate touch. Primary afferents travel up the dorsal columns of the spinal cord white matter and synapse with somatosensory input association neurons in the medulla. Secondary neurons project through the brain stem to the thalamus and synapse with tertiary neurons, which relay the information to the primary somatosensory cortex on the opposite side of the brain.

fibers synapse with several nuclei in the thalamus, but en route they give off numerous branches that travel to the reticular activating system of the brain stem. These projections provide the basis for increased wakefulness or awareness after strong somatosensory stimulation and for the generalized startle reaction that occurs with sudden and intense stimuli. They also stimulate autonomic nervous system responses, such as an increase in blood pressure and heart rate, dilation of the pupils, and the pale, moist skin that results from constriction of the cutaneous blood vessels and activation of the sweat glands.

There are two subdivisions in the anterolateral pathway: the outer *neospinothalamic tract* and the inner *paleospinothalamic tract* (Fig. 39-5). The neospinothalamic tract, which carries bright pain, consists of a sequence of at least three neurons with long axons. It provides for relatively rapid transmission of sensory information to the thalamus. The paleospinothalamic tract, which is phylogenetically older than the neospinothalamic system, consists of bilateral, multisynaptic slowconducting tracts that transmit sensory signals that do not require discrete localization of signal source or discrimination of fine gradations in intensity. This slower-conducting pathway also projects into the intralaminar nuclei of the thalamus, which have close connections with the limbic cortical



■ FIGURE 39-5 ■ Neospinothalamic and paleospinothalamic subdivisions of the anterolateral sensory pathway. The neospinothalamic tract runs to the thalamic nuclei and has fibers that project to the somatosensory cortex. The paleospinothalamic tract sends collaterals to the reticular formation and other structures, from which further fibers project to the thalamus. These fibers influence the hypothalamus and the limbic system as well as the cerebral cortex.

systems. This circuitry provides touch with its affective or emotional aspects.

Central Processing of Somatosensory Information

Perception, or the final processing of somatosensory information, involves awareness of the stimuli, localization and discrimination of their characteristics, and interpretation of their meaning. As sensory information reaches the thalamus, it begins to enter the level of consciousness. In the thalamus, the sensory information is roughly localized and perceived as a crude sense. The full localization, discrimination of the intensity, and interpretation of the meaning of the stimuli require processing by the somatosensory cortex.

The somatosensory cortex is located in the parietal lobe, which lies behind the central sulcus and above the lateral sulcus (Fig. 39-6). The strip of parietal cortex that borders the central sulcus is called the *primary somatosensory cortex* because it receives primary sensory information by way of direct projections from the thalamus. A distorted map of the body and head surface, called the *sensory homunculus*, reflects the density of cortical neurons devoted to sensory input from afferents in corresponding peripheral areas. As depicted in Figure 39-7, most of the cortical surface is devoted to areas of the body such as the thumb, forefinger, lips, and tongue, where fine touch and pressure discrimination are essential for normal function.

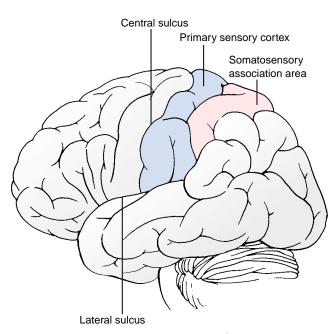


FIGURE 39-6 Primary somatosensory and association somatosensory cortex.

The somatosensory association cortex, which lies parallel to and just behind the primary somatosensory cortex, is required to transform the raw material of sensation into a meaningful experience. It is here that the stimulus pattern from the present sensory experience is integrated with past learning. For instance, a person's past learning plus present tactile sensation provide the perception of sitting on a soft chair, rather than on a hard bicycle seat.



Sensory Modalities

Somatosensory experience can be divided into *modalities*, a term used for qualitative, subjective distinctions between sensations such as touch, heat, and pain. Such experiences require the function of sensory receptors and forebrain structures in the thalamus and cerebral cortex. Sensory experience also involves quantitative sensory discrimination or the ability to distinguish between different levels of sensory stimulation.

The receptive endings of different afferent neurons are particularly sensitive to specific forms of physical and chemical energy. For instance, a receptive ending may be particularly sensitive to a small increase in local skin temperature. Other afferent sensory terminals are most sensitive to slight indentations of the skin, and their signals are subjectively interpreted as touch. Cool versus warm, sharp versus dull pain, and delicate touch versus deep pressure are all based on different populations of afferent neurons or on central integration of simultaneous input from several differently tuned afferents. For example, the sensation of itch results from a combination of high activity in pain- and touch-sensitive afferents, and the sensation of tickle requires a gently moving tactile stimulus over cool skin.

When information from different primary afferents reaches the forebrain, where subjective experience occurs, the qualitative differences between warmth and touch are called *sensory modalities*. Although the receptor-detected information is relayed to the thalamus and cortex over separate pathways, the experience of a modality, such as cold versus warm, is uniquely subjective.

Tactile Sensation

The tactile system, which relays sensory information regarding touch, pressure, and vibration, is considered the basic somatosensory system. Loss of temperature or pain sensitivity leaves the person with no awareness of deficiency. However, if the tactile system is lost, total anesthesia (*i.e.*, numbness) of the involved body part results.

Touch sensation results from stimulation of tactile receptors in the skin and in tissues immediately beneath the skin, pressure from deformation of deeper tissues, and vibration from rapidly repetitive sensory signals. There are at least six types of specialized tactile receptors in the skin and deeper structures: free nerve endings, Meissner's corpuscles, Merkel's disks, pacinian corpuscles, hair follicle end-organs, and Ruffini's endorgans^{1,2} (Fig. 39-8).

Free nerve endings are found in skin and many other tissues, including the cornea. They detect touch and pressure. *Meissner's corpuscles* are present in nonhairy parts of the skin. They are particularly abundant in the fingertips, lips, and other areas where the sense of touch is highly developed. *Meissner's corpuscles* are particularly sensitive to movement of very light objects over the surface of the skin and to low frequency vibration. *Merkel's disks* are found in nonhairy areas and in hairy parts of the skin. They are responsible for giving steady-state signals that allow for continuous determination of touch against the skin.

The *pacinian corpuscle* is located immediately beneath the skin and deep in the fascial tissues of the body and is important in detecting tissue vibration. The *hair follicle end-organs* detect movement on the surface of the body. *Ruffini's end-organs* are found in the skin and deeper structures, including the joint



capsules. These receptors are important for signaling continuous states of deformation, such as heavy and continuous touch and pressure.

The sensory information for tactile sensation enters the spinal cord through the dorsal roots of the spinal nerves. All tactile sensation that requires rapid transmission is transmitted through the discriminative pathway to the thalamus by way of the medial lemniscus. This includes touch sensation requiring a high degree of localization or fine gradations of intensity, vibratory sensation, and sensation that signals movement against the skin. In addition to the ascending discriminative pathway, tactile sensation also uses the more primitive and crude anterolateral pathway. The second-order dorsal horn neurons of this pathway have many branches or collaterals. After several synapses, axons are projected up both sides of the anterolateral aspect of the spinal cord to the thalamus. Few fibers travel all the way to the thalamus. Most synapse on reticular formation neurons that then send their axons on toward the thalamus, where a crude, poorly localized sensation from the opposite side of the body is received. From the thalamus, some projections travel to the somatosensory cortex.

Because of these multiple routes, total destruction of the pathways for tactile sensation seldom occurs. The only time this crude alternative system becomes essential is when the discriminative pathway is damaged. Then, despite projection of the anterolateral system information to the somatosensory cortex, only a poorly localized, high-threshold sense of touch remains. Such persons lose all sense of joint and muscle movement, body position, and two-point discrimination.

Thermal Sensation

Thermal sensation is discriminated by three types of receptors: cold receptors, warmth receptors, and pain receptors. The cold and warmth receptors are located immediately under the skin at discrete but separate points. In some areas, there are more cold receptors than warmth receptors. For example, the lips have 15 to 25 cold receptors per square centimeter, compared with 3 to 5 in the same-sized area of the finger.¹ There are also correspondingly fewer warmth receptors in these areas. The gradations of heat and cold result from selective stimulation of the different types of thermal receptors. The thermal receptors are very sensitive to differences between the temperature

of skin and temperature of objects that are touched. Warmth receptors respond proportionately to increases in skin temperature above resting values of 34°C and cool receptors to temperatures below 34°C.³

The thermal pain receptors are stimulated only by extremes of temperature such as "freezing cold" (temperatures below 5°C) and "burning hot" (temperatures above 45°C) sensations.³ With the exception of pain receptors, thermal receptors tend to adapt rapidly during the first few minutes and then more slowly during the next 30 minutes or so. However, these receptors do not appear to adapt completely, as evidenced by the experience of an intense sense of heat on entering a tub of hot water or the extreme degree of cold initially sensed when going outside on a cold day. On entering the dorsal horn, thermal signals are transmitted by neurons whose axons then cross to the opposite side of the cord and ascend in the multisynaptic, slow-conducting anterolateral system to the opposite side of the brain.

Conduction of thermal information through peripheral nerves is quite slow compared with the rapid tactile afferents that travel through the discriminative system. If a person places a foot in a tub of hot water, the tactile sensation occurs well in advance of the burning sensation. The foot has been removed from the hot water by the local withdrawal reflex well before the excessive heat is perceived by the forebrain. Local anesthetic agents block the small-diameter afferents that carry thermal sensory information before they block the large-diameter axons that carry touch information.

Position Sensation

Position sense refers to the sense of limb and body movement and position without using vision. It is mediated by input from proprioceptive receptors (muscle spindle receptors and Golgi tendon organs) found primarily in muscles, tendons, and joint capsules (see Chapter 38). There are two submodalities of proprioception: the stationary or static component (limb position sense) and the dynamic aspects of position sense (kinesthesia). Both of these depend on constant transmission of information to the CNS regarding the rate of change and degree of angulation of all joints. In addition, there are a number of mechanoreceptors in the joint capsules and ligaments. Many resemble Ruffini's end-organs, pacinian corpuscles, and Merkel's cells. Signals from these receptors are processed through the discriminative pathway. It appears that information from the joint receptors is combined with that from the muscle spindles and Golgi tendon organs, and probably from skin receptors that estimate joint angle. Removing one source of information can be compensated by use of other sources.

In summary, the somatosensory component of the nervous system provides an awareness of body sensations such as touch, temperature, position sense, and pain. There are three primary levels of neural integration in the somatosensory system: the sensory units containing the sensory receptors, the ascending pathways, and the central processing centers in the thalamus and cerebral cortex. A sensory unit consists of a single dorsal root ganglion neuron, its receptors, and its central axon that terminates in the dorsal horn of the spinal cord or medulla. The part of the body innervated by the somatosensory afferent neurons of one set of dorsal root ganglia is called a dermatome. Ascending pathways include the discriminative pathway, which crosses at the base of the medulla, and the anterolateral pathway, which crosses within the first few segments of entering the spinal cord. Perception, or the final processing of somatosensory information, involves centers in the thalamus and somatosensory cortex. In the thalamus, the sensory information is crudely localized and perceived. The full localization, discrimination of the intensity, and interpretation of the meaning of the stimuli require processing by the somatosensory cortex. A distorted map of the body and head surface, called the sensory homunculus, reflects the density of cortical neurons devoted to sensory input from afferents in corresponding peripheral areas.

The tactile system relays the sensations of touch, pressure, and vibration. It uses two anatomically separate pathways to relay touch information to the opposite side of the forebrain: the dorsal column discriminative pathway and the anterolateral pathway. Delicate touch, vibration, position, and movement sensations use the discriminative pathway to reach the thalamus, where third-order relay occurs to the primary somatosensory strip of parietal cortex. Crude tactile sensation is carried by the bilateral slow-conducting anterolateral pathway. Temperature sensations of warm-hot and cool-cold are the result of stimulation to thermal receptors of sensory units projecting to the thalamus and cortex through the anterolateral system on the opposite side of the body. Proprioception is the sense of limb and body movement and position without using vision. It is mediated by input muscle spindle receptors and Golgi tendon organs found in muscles, tendons, and joint capsules and by mechanoreceptors (e.g., Ruffini's endorgans, pacinian corpuscles, and Merkel's cells) in the joint capsules and ligaments.

PAIN

Pain is an "unpleasant sensory and emotional experience associated with potential tissue damage, or described in terms of such damage."⁴ It involves anatomic structures, physiologic behaviors, and psychological, social, cultural, and cognitive factors. Pain can be a prepotent or overwhelming experience, often disruptive of customary behavior, and when severe, it demands and directs all of a person's attention. Pain is the most common symptom that motives a person to seek professional help. It sends those who suffer to a health care facility more often and with more speed than any other symptoms. Its location, radiation, duration, and severity provide important clues as to its cause. Despite its unpleasantness, pain can serve a useful purpose because it warns of impending tissue injury, motivating the person to seek relief. For example, an inflamed appendix could progress in severity, rupture, and even cause death were it not for the warning afforded by the pain.

Pain Theories

Traditionally, two theories have been offered to explain the physiologic basis for the pain experience. The first, *specificity theory*, regards pain as a separate sensory modality evoked by the activity of specific receptors that transmit information to pain centers or regions in the forebrain where pain is experienced.⁵ The second theory includes a group of theories collectively referred to as *pattern theory*. It proposes that pain receptors share endings or pathways with other sensory modalities

KEY CONCEPTS

PAIN SENSATION

- Pain is both a protective and an unpleasant physical and emotionally disturbing sensation originating in pain receptors that respond to a number of stimuli that threaten tissue integrity.
- There are two pathways for pain transmission:
 - \Box The fast pathway for sharply discriminated pain that moves directly from the receptor to the spinal cord using myelinated Aδ fibers and from the spinal cord to the thalamus using the neospinothalamic tract
 - The slow pathway for continuously conducted pain that is transmitted to the spinal cord using unmyelinated C fibers and from the spinal cord to the thalamus using the more circuitous and slower-conducting paleospinothalamic tract.
- The central processing of pain information includes transmission to the somatosensory cortex, where pain information is perceived and interpreted; the limbic system, where the emotional components of pain are experienced; and to brain stem centers, where autonomic nervous system responses are recruited.
- Modulation of the pain experience occurs by way of the endogenous analgesic center in the midbrain, the pontine noradrenergic neurons, and the nucleus raphe magnus in the medulla, which sends inhibitory signals to dorsal horn neurons in the spinal cord.

732 Unit Ten: Alterations in the Nervous System

but that different patterns of activity (*i.e.*, spatial or temporal) of the same neurons can be used to signal painful and nonpainful stimuli.⁵ For example, light touch applied to the skin would produce the sensation of touch through low-frequency firing of the receptor; intense pressure would produce pain through high-frequency firing of the same receptor. Both theories focus on the neurophysiologic basis of pain, and both probably apply. Specific nociceptive afferents have been identified; however, almost all afferent stimuli, if driven at a very high frequency, can be experienced as painful.

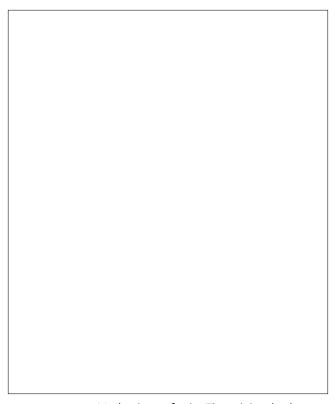
Gate control theory, a modification of specificity theory, was proposed by Melzack and Wall in 1965 to meet the challenges presented by the pattern theories.⁶ This theory postulated the presence of neural gating mechanisms at the segmental spinal cord level to account for interactions between pain and other sensory modalities. According to the gate control theory, the internuncial neurons involved in the gating mechanism are activated by large-diameter, faster-propagating fibers that carry tactile information. The simultaneous firing of the large-diameter touch fibers has the potential for blocking the transmission of impulses from the small-diameter myelinated and unmyelinated pain fibers.

More recently, Melzack has developed the neuromatrix theory to address further the brain's role in pain as well as the multiple dimensions and determinants of pain.⁷ This theory is particularly useful in understanding chronic pain and phantom limb pain, in which there is not a simple one-to-one relationship between tissue injury and pain experience. The neuromatrix theory proposes that the brain contains a widely distributed neural network, called the *body-self neuromatrix*, that contains somatosensory, limbic, and thalamocortical components. Genetic and sensory influences determine the synaptic architecture of an individual's neuromatrix that integrates multiple sources of input and evokes the sensory, affective, and cognitive dimensions of pain experience and behavior. These multiple input sources include somatosensory; other sensory impulses affecting interpretation of the situation; inputs from the brain addressing such things as attention, expectation, culture, and personality; intrinsic neural inhibitory modulation; and various components of stress-regulation systems.

Pain Mechanisms and Pathways

Pain usually is viewed in the context of tissue injury. The term *nociception*, which means "pain sense," comes from the Latin word *nocere* ("to injure"). Nociceptive stimuli are objectively defined as stimuli of such intensity that they cause or are close to causing tissue damage. Researchers often use the withdrawal reflex (*e.g.*, the reflexive withdrawal of a body part from a tissue-damaging stimulus) to determine when a stimulus is nociceptive. Stimuli used include pressure from a sharp object, strong electric current to the skin, or application of heat or cold of approximately 10°C above or below normal skin temperature. At low levels of intensity these noxious stimuli do activate nociceptors (pain receptors) but typically are perceived as painful only when the intensity reaches a level where tissue damage occurs or is imminent.

The mechanisms of pain are many and complex (Fig. 39-9). As with other forms of somatosensation, the pathways are composed of first-, second-, and third-order neurons. The first-order neurons and their receptive endings detect stimuli that threaten



■ FIGURE 39-9 ■ Mechanisms of pain. Tissue injury leads to release of inflammatory mediators with subsequent nociceptor stimulation. Pain impulses are then transmitted to dorsal horn of the spinal cord, where they make contact with second-order neurons that cross to the opposite side of the cord and ascend via the spinothalamic tract to the reticular activating system (RAS) and thalamus. The localization and meaning of pain occurs at the level of the somatosensory cortex.

the integrity of innervated tissues. Second-order neurons are located in the spinal cord and process nociceptive information. Third-order neurons project pain information to the brain. The thalamus and cortex integrate and modulate pain as well as the person's subjective reaction to the pain experience.

Pain Receptors and Mediators

Nociceptors, or pain receptors, are sensory receptors that are activated by noxious insults to peripheral tissues (Fig 39-9). Structurally, the receptive endings of the peripheral pain fibers are free nerve endings. These receptive endings, which are widely distributed in the skin, dental pulp, periosteum, meninges, and some internal organs, translate the noxious stimuli into action potentials that are transmitted by a dorsal root ganglion to the dorsal horn of the spinal cord. Nociceptive action potentials are transmitted through two types of afferent nerve fibers: myelinated Aδ fibers and unmyelinated C fibers. The larger Aδ fibers have considerably greater conduction velocities, transmitting impulses at a rate of 10 to 30 m/second. The C fibers are the smallest of all peripheral nerve fibers; they transmit impulses at the rate of 0.5 to 2.5 m/second. Pain conducted by A δ fibers traditionally is called *fast pain* and typically is elicited by mechanical or thermal stimuli. C-fiber pain often is described as slow-wave pain because it is slower in onset and longer in duration. It typically is incited by chemical stimuli or by persistent mechanical or thermal stimuli. The slow-wave potentials generated in C fibers are now believed to be responsible for central sensitization to chronic pain.

Stimulation of Nociceptors. Unlike other sensory receptors, nociceptors respond to several forms of stimulation, including mechanical, thermal, and chemical. Some receptors respond to a single type of stimuli (mechanical or thermal), and others, called *polymodal receptors*, respond to all three types of stimuli (mechanical, thermal, and chemical). Mechanical stimuli can arise from intense pressure applied to skin or from the violent contraction or extreme stretch of a muscle. Both extremes of heat and cold can stimulate nociceptors. Chemical stimuli arise from a number of sources, including tissue trauma, ischemia, and inflammation. A wide range of chemical mediators are released from injured and inflamed tissues, including hydrogen and potassium ions, prostaglandins, leukotrienes, histamine, bradykinin, acetylcholine, and serotonin. These chemical mediators produce their effects by directly stimulating nociceptors or sensitizing them to the effects of nociceptive stimuli; perpetuating the inflammatory responses that lead to the release of chemical agents that act as nociceptive stimuli; or inciting neurogenic reflexes that increase the response to nociceptive stimuli. For example, bradykinin, histamine, serotonin, and potassium activate and also sensitize nociceptors. Adenosine triphosphate, acetylcholine, and platelets act alone or in concert to sensitize nociceptors through other chemical agents such as prostaglandins. Aspirin and other nonsteroidal analgesic drugs are effective in controlling pain because they block the enzyme needed for prostaglandin synthesis.

Nociceptive stimulation that activates C fibers can cause a response known as *neurogenic inflammation* that produces vasodilation and an increased release of chemical mediators to which nociceptors respond. This mechanism is thought to be mediated by a dorsal root neuron reflex that produces retrograde transport and release of chemical mediators, which in turn causes increasing inflammation of peripheral tissues. This reflex can set up a vicious cycle, which has implications for persistent pain and hyperalgesia.⁸ Local anesthetics (*e.g.*, procaine [Novocain]) can prevent the spread of sensitization and secondary hyperalgesia caused by stimulation of cutaneous nociceptors by blocking the dorsal root neuron reflex.⁹

Mediators in the Spinal Cord. In the spinal cord, the transmission of impulses between the nociceptive neurons and the dorsal horn neurons is mediated by chemical neurotransmitters released from central nerve endings of the nociceptive neurons. Some of these neurotransmitters are amino acids (e.g., glutamate), others are amino acid derivatives (e.g., norepinephrine), and still others are low-molecular-weight peptides composed of two or more amino acids. The amino acid glutamate is a major excitatory neurotransmitter released from the central nerve endings of the nociceptive neurons. Substance P, a neuropeptide, also is released in the dorsal horn by C fibers in response to nociceptive stimulation. Substance P elicits slow excitatory potentials in dorsal horn neurons. Unlike glutamate, which confines its action to the immediate area of the synaptic terminal, some neuropeptides released in the dorsal horn can diffuse some distance because they are not inactivated by reuptake mechanisms. In persistent pain, this may help to explain the excitability and unlocalized nature of many painful

conditions. Neuropeptides such as substance P also appear to prolong and enhance the action of glutamate. If these neurotransmitters are released in large quantities or over extended periods, they can lead to secondary hyperalgesia, a condition in which the second-order neurons are overly sensitive to low levels of noxious stimulation. Understanding how chemical mediators function in nociception is an active area of research that has implications for the development of new treatments for pain.

Spinal Cord Circuitry and Pathways

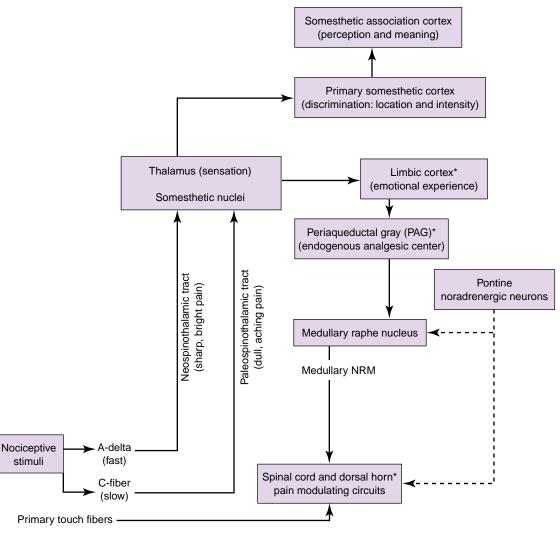
On entering the spinal cord through the dorsal roots, the pain fibers bifurcate and ascend or descend one or two segments before synapsing with association neurons in the dorsal horn. From the dorsal horn, the axons of association projection neurons cross through the anterior commissure to the opposite side and then ascend upward in the previously described neospinothalamic and paleospinothalamic pathways (Fig. 39-10).

The faster-conducting fibers in the neospinothalamic pathway (*i.e.*, lateral spinothalamic tract) are associated mainly with the transmission of sharp-fast pain information to the thalamus.¹ In the thalamus, synapses are made and the pathway continues to the contralateral parietal somatosensory area to provide the precise location of the pain. Typically, the pain is experienced as bright, sharp, or stabbing in nature. There is also a local cord-level withdrawal reflex that is designed to remove endangered tissue from a damaging stimulus.

The paleospinothalamic tracts are slower-conducting, multisynaptic pathways concerned with the diffuse, dull, aching, and unpleasant sensations that commonly are associated with chronic and visceral pain.¹ This information travels through the small, unmyelinated C fibers. Fibers of this system also project up the contralateral (i.e., opposite) anterolateral pathway to terminate in several thalamic regions, including those that project to the limbic system, where it is associated with the emotional aspects of pain. Fibers from the spinoreticular tract project bilaterally to the reticular formation of the brain stem. This component of the paleospinothalamic system facilitates avoidance reflexes at all levels. It also contributes to an increase in the electroencephalographic activity associated with alertness and indirectly influences hypothalamic functions associated with sudden alertness, such as increased heart rate and blood pressure. This may explain the tremendous arousal effects of certain pain stimuli.

Brain Centers and Pain Perception

The basic sensation of hurtfulness, or pain, occurs at the level of the thalamus. In the neospinothalamic system, interconnections between the lateral thalamus and the somatosensory cortex are necessary to add precision and discrimination to the pain sensation. Association areas of the parietal cortex are essential to the learned meaningfulness of the pain experience. For example, if a person is stung on the index finger by a bee and only the thalamus is functional, the person reports pain somewhere on the hand. With the primary sensory cortex functional, the person can localize the pain to the precise area on the index finger. With the association cortex functional, the person can interpret the buzzing and sight of the bee that preceded the pain as being related to the bee sting. The paleospinothalamic system projects diffusely from the intralaminar nuclei of the thalamus to large areas of the limbic cortex. These



* Location of opioid receptors

■ FIGURE 39-10 ■ Primary pain pathways. The transmission of incoming nociceptive impulses is modulated by dorsal horn circuitry that receives input from peripheral touch receptors and from descending pathways that involve the limbic cortical systems (orbital frontal cortex, amygdala, and hypothalamus), periaqueductal endogenous analgesic center in the midbrain, pontine noradrenergic neurons, and the nucleus raphe magnus (NRM) in the medulla. Dashed lines indicate inhibition or modulation.

connections probably are associated with the hurtfulness and the mood-altering and attention-narrowing effect of pain.

Central Pathways for Pain Modulation

A major advance in understanding pain was the discovery of neuroanatomic pathways that arise in the midbrain and brain stem, descend to the spinal cord, and modulate ascending pain impulses. One such pathway begins in an area of the midbrain called the *periaqueductal* gray (PAG) region. Through research it was found that focal stimulation of the midbrain PAG regions produced a state of analgesia. The resultant analgesia lasted for many hours and was sufficient to permit abdominal surgery, although levels of consciousness and reactions to auditory and visual stimuli remained unaffected. A few years later, opioid receptors were found to be highly concentrated in this and other regions of the CNS where electrical stimulation produced analgesia. Because of these findings, the PAG area of the midbrain often is referred to as the *endogenous analgesia center*.

The PAG area receives input from widespread areas of the CNS, including the cerebral cortex, hypothalamus, brain stem reticular formation, and spinal cord by way of the paleospinothalamic and neospinothalamic tracts. This region is intimately connected to the limbic system, which is associated with emotional experience. The neurons of the PAG area in the midbrain have axons that descend into an area called the *nucleus raphe magnus* (NRM) in the rostral medulla. The axons of these NRM neurons project to the dorsal horn of the spinal cord, where they terminate in the same layers as the entering primary pain fibers (Fig. 39-10). Stimulation of the NRM is thought to inhibit pain transmission by dorsal horn projection neurons.¹⁰ There also is evidence of noradrenergic neurons that can inhibit transmission of pain impulses at the level of the spinal cord. The discovery that norepinephrine can block pain transmission led to studies directed at the combined administration of opioids and clonidine, a central-acting α -adrenergic agonist for some types of pain relief.

Serotonin also has been identified as a neuromodulator in the NRM medullary nuclei that project to the spinal cord. It has been shown that tricyclic antidepressant compounds, such as amitriptyline, have analgesic properties independent of their antidepressant effects. These drugs, which enhance the effects of serotonin by blocking its presynaptic uptake, have been found to be effective in the management of certain types of chronic pain.¹¹

Endogenous Analgesic Mechanisms. There is evidence that the endogenous opioid peptides, morphine-like substances synthesized in many regions of the CNS including the spinal cord and PAG, modulate pain in the CNS. Three families of opioid peptides have been identified-the enkephalins, endorphins, and dynorphins. Although the endogenous opioid peptides appear to function as neurotransmitters, their full significance in pain control and other physiologic functions is not completely understood. Probably of greater importance in understanding mechanisms of pain control has been the characterization of receptors that bind the endogenous opioid peptides. The identification of these receptors has facilitated a more thorough understanding of the actions of available opioid drugs, such as morphine, and it also has facilitated ongoing research into the development of newer preparations that are more effective in relieving pain and have fewer side effects.

Pain Threshold and Tolerance

Pain threshold and tolerance affect an individual's response to a painful stimulus. Although the terms often are used interchangeably, pain threshold and pain tolerance have distinct meanings. Pain threshold is closely associated with tissue damage and the point at which a stimulus is perceived as painful. Pain tolerance relates more to the total pain experience; it is defined as the maximum intensity or duration of pain that a person is willing to endure before the person wants something done about the pain. Psychological, familial, cultural, and environmental factors significantly influence the amount of pain a person is willing to tolerate. The threshold to pain is fairly uniform from one person to another, whereas pain tolerance is extremely variable.12 Separation and identification of the role of each of these two aspects of pain continue to pose fundamental problems for the pain management team and for pain researchers.

Types of Pain

The most widely accepted classifications of pain are according to source or location, referral, and duration (acute or chronic). Classification based on associated medical diagnosis (*e.g.*, surgery, trauma, cancer, sickle cell disease, fibromyalgia) is useful in planning appropriate interventions.

Cutaneous and Deep Somatic Pain

Cutaneous pain arises from superficial structures, such as the skin and subcutaneous tissues. A paper cut on the finger is an example of easily localized superficial, or cutaneous, pain. It is

KEY CONCEPTS

TYPES OF PAIN

- Pain can be classified according to location, site of referral, and duration.
- Cutaneous pain is a sharp, burning pain that has its origin in the skin or subcutaneous tissues.
- Deep pain is a more diffuse and throbbing pain that originates in structures such as the muscles, bones, and tendons and radiates to the surrounding tissues.
- Visceral pain is a diffuse and poorly defined pain that results from stretching, distention, or ischemia of tissues in a body organ.
- Referred pain is pain that originates at a visceral site but is perceived as originating in part of the body wall that is innervated by neurons entering the same segment of the nervous system.
- Acute pain usually results from tissue damage and is characterized by autonomic nervous system responses.
- Chronic pain is persistent pain that is accompanied by loss of appetite, sleep disturbances, depression, and other debilitating responses.

a sharp, bright pain with a burning quality and may be abrupt or slow in onset. It can be localized accurately and may be distributed along the dermatomes. Because there is an overlap of nerve fiber distribution between the dermatomes, the boundaries of pain frequently are not as clear-cut as the dermatomal diagrams indicate.

Deep somatic pain originates in deep body structures, such as the periosteum, muscles, tendons, joints, and blood vessels. This pain is more diffuse than cutaneous pain. Various stimuli, such as strong pressure exerted on bone, ischemia to a muscle, and tissue damage, can produce deep somatic pain. This is the type of pain a person experiences from a sprained ankle. Radiation of pain from the original site of injury can occur. For example, damage to a nerve root can cause a person to experience pain radiating along its fiber distribution.

Visceral Pain

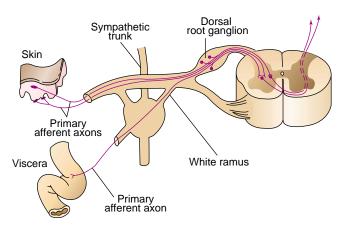
Visceral, or splanchnic, pain has its origin in the visceral organs. Common examples of visceral pain are renal colic, pain caused by cholecystitis, pain associated with acute appendicitis, and peptic ulcer pain. Although the viscera are diffusely and richly innervated, cutting or burning of viscera, as opposed to similar noxious stimuli applied to cutaneous or superficial structures, is unlikely to cause pain. Instead, strong abnormal contractions of the gastrointestinal system, distention, or ischemia affecting the walls of the viscera can induce severe visceral pain. Anyone who has had severe gastrointestinal distress or ureteral colic can readily attest to the misery involved. Visceral pain is transmitted by small unmyelinated pain fibers that travel with the axons of the autonomic nervous system and project to visceral input association neurons of the cord or brain stem. In addition to sending projections to the forebrain, these input association neurons also project through the paleospinal and spinoreticular pathways into visceral reflex circuits. Visceral pain typically is accompanied by autonomic nervous system responses such as nausea, vomiting, sweating, and pallor and, less commonly, is followed by shock.

The visceral pain pathways travel with the parasympathetic distribution for the upper and lower viscera and with the sympathetic distribution for the intervening viscera. Pain from the viscera may be localized only with difficulty. There are several explanations for this. First, innervation of visceral organs is poorly represented at the forebrain levels (*i.e.*, perception). A second possible explanation is that the brain does not easily learn to localize sensations that originate in organs that are only imprecisely visualized. For example, a cut on the third finger of the right hand can be readily seen, identified, and localized, whereas an inflamed internal organ can be localized only vaguely. A third explanation is that sensory information from thoracic and abdominal viscera can travel by two pathways to the CNS.

Referred Pain

Referred pain is pain that is perceived at a site different from its point of origin but innervated by the same spinal segment. It is hypothesized that visceral and somatic afferent neurons converge on the same dorsal horn projection neurons (Fig. 39-11). For this reason, it can be difficult for the brain to correctly identify the original source of pain. Pain that originates in the abdominal or thoracic viscera is diffuse and poorly localized and often perceived at a site far removed from the affected area. For example, the pain associated with myocardial infarction commonly is referred to the left arm, neck, and chest.

Referred pain may arise alone or concurrent with pain located at the origin of the noxious stimuli. This lack of correspondence between the location of the pain and the location



■ FIGURE 39-11 ■ Convergence of cutaneous and visceral inputs onto the same second-order projection neuron in the dorsal horn of the spinal cord. Although virtually all visceral inputs converge with cutaneous inputs, most cutaneous inputs do not converge with other sensory inputs. (From Conn P.M. [1995]. *Neuroscience in medicine*. Philadelphia: J.B. Lippincott)

of the painful stimuli can make diagnosis difficult. Although the term *referred* usually is applied to pain that originates in the viscera and is experienced as if originating from the body wall, it also may be applied to pain that arises from somatic structures. For example, pain referred to the chest wall could be caused by nociceptive stimulation of the peripheral portion of the diaphragm, which receives somatosensory innervation from the intercostal nerves. An understanding of pain referral is of great value in diagnosing illness. The typical pattern of pain referral can be derived from our understanding that the afferent neurons from visceral or deep somatic tissue enter the spinal cord at the same level as the afferent neurons from the cutaneous areas to which the pain is referred (Fig. 39-12).

The sites of referred pain are determined embryologically with the development of visceral and somatic structures that share the same site for entry of sensory information into the CNS and then move to more distant locations. For example, a person with peritonitis may report pain in the shoulder. Inter-

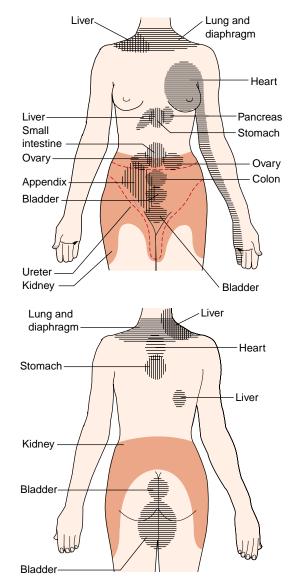


FIGURE 39-12 Areas of referred pain. (**Top**) Anterior view. (**Bottom**) Posterior view.

TADLE 20 1

nally, there is inflammation of the peritoneum that lines the central part of the diaphragm. In the embryo, the diaphragm originates in the neck, and its central portion is innervated by the phrenic nerve, which enters the cord at the level of the third to fifth segments (C3 to C5). As the fetus develops, the diaphragm descends to its adult position between the thoracic and abdominal cavities, while maintaining its embryonic pattern of innervation. Thus, fibers that enter the spinal cord at the C3 to C5 level carry information from both the neck area and the diaphragm, and the diaphragmatic pain is interpreted by the forebrain as originating in the shoulder or neck area.

Muscle spasm, or *guarding*, occurs when somatic structures are involved. Guarding is a protective reflex rigidity; its purpose is to protect the affected body parts (*e.g.*, an abscessed appendix or a sprained muscle). This protective guarding may cause blood vessel compression and give rise to the pain of muscle ischemia, causing local and referred pain.

Acute and Chronic Pain

It is common to classify pain according to its duration. Pain research of the past three decades has emphasized the importance of differentiating acute pain from chronic pain. The diagnosis and therapy for each is distinctive because they differ in cause, function, mechanisms, and psychological sequelae (Table 39-1).

Acute Pain. The classic definition of acute pain is pain that lasts less than 6 months. This somewhat arbitrary cutoff point reflects the notion that acute pain is the result of a tissuedamaging event, such as trauma or surgery, and usually is selflimited, ending when the injured tissues heal. The purpose of acute pain is to serve as a warning system. Besides alerting the person to the existence of actual or impending tissue damage, it prompts a search for professional help. The pain's location, radiation, intensity, and duration as well as those factors that aggravate or relieve it provide essential diagnostic clues.

Acute pain can lead to anxiety and secondary reflex musculoskeletal spasms, which in turn tend to worsen the pain.¹³ Interventions that alleviate the pain usually alleviate the anxiety and musculoskeletal spasms as well. Inadequately treated pain can provoke physiologic responses that alter circulation and tissue metabolism and produce physical manifestations, such as tachycardia, reflective of increased sympathetic activity. Inadequately treated acute pain tends to decrease mobility and respiratory movements such as deep breathing and coughing to the extent that it may complicate or delay recovery.

Chronic Pain. Chronic pain classically has been defined as pain lasting 6 months or longer. However, in practice one does not wait an arbitrary 6 months before deciding that the pain is chronic; rather, one considers the normal expected healing time for the underlying cause of the pain. The International Association for the Study of Pain defines chronic pain as that which persists beyond the expected normal time of healing.¹³ Chronic pain can be quite variable. It may be unrelenting and extremely severe, as in metastatic bone pain. It can be relatively continuous with or without periods of escalation, as with some forms of back pain. Some conditions with recurring episodes of acute pain are particularly problematic because they have characteristics of both acute and chronic pain. These include the pain associated with sickle cell crisis or migraine headaches.

and Chronic Pain					
Characteristic	Acute Pain	Chronic Pain			
Onset	Recent	Continuous or intermittent			
Duration	Short duration (<6 months)	6 months or more			

Autonomic Consistent with sym- Absence of	of auto-
responses pathetic fight-or- nomic r flight response* Increased heart rate Increased stroke volume Increased blood pressure Increased pupillary dilation Increased muscle tension Decreased gut motility Decreased salivary flow (dry mouth)	responses
component Associated depress Somatic p occupa Withdraw outside Decreased	sion pre- ition
Other types of Decreased response Decreased Appetite of	d libido

*Responses are approximately proportional to intensity of the stimulus.

Chronic pain is a leading cause of disability in the United States. Unlike acute pain, persistent chronic pain usually serves no useful function. In contrast, it imposes physiologic, psychological, familial, and economic stresses and may exhaust a person's resources.

Persons with chronic pain may not exhibit the somatic, autonomic, or affective behaviors often associated with acute pain. As painful conditions become prolonged and continuous, autonomic nervous system responses decrease. Decreased pain tolerance, which may result from the depletion of serotonin and endorphins, and depression are common in individuals with chronic pain. There is often loss of appetite, sleep disturbances, and depression.14 The link between depression and decreased pain tolerance may be explained by the similar manner in which both respond to changes in the biologic pathways of serotonergic and noradrenergic systems.¹⁴ Tricyclic antidepressants and other medications with serotonergic and noradrenergic effects have been shown to relieve a variety of chronic pain syndromes, lending credence to the theory that chronic pain and depression share a common biologic pathway.14

Pain Management

Assessment

Careful assessment of pain assists clinicians in diagnosing, managing, and relieving the patient's pain. Assessment includes such things as the nature, severity, location, and radiation of the pain. As with other disease states, it is preferable to eliminate the cause of the pain, rather than simply to treat the symptom. A careful history often provides information about the triggering factors (i.e., injury, infection, or disease) and the site of nociceptive stimuli (i.e., peripheral receptor or visceral organ). Although the observation of facial expression and posture may provide additional information, current AHCPR practice guidelines emphasize that "the single most reliable indicator of the existence and intensity of acute pain-and any resultant affective discomfort or distress—is the patient's self report."15,16 A comprehensive pain history should include pain onset; description, localization, radiation, intensity, quality, and pattern of the pain; anything that relieves or exacerbates it; and the individual's personal reaction to the pain.

Unlike many other bodily responses, such as temperature and blood pressure, the nature, severity, and distress of pain cannot be measured objectively. To overcome this problem, various methods have been developed for quantifying a person's pain. Most of these are based on the patient's report. They include numeric pain intensity, visual analog, and verbal descriptor scales.

Treatment

The therapeutic approaches to acute and chronic pain differ markedly. In acute pain, therapy is directed at providing pain relief by interrupting the nociceptive stimulus. Acute pain should be aggressively managed and pain medication provided before the pain becomes severe. This allows the person to be more comfortable and active and to assume a greater role in directing his or her own care. Because the pain is self-limited, in that it resolves as the injured tissues heal, long-term therapy usually is not needed.

Chronic pain management is much more complex and is based on multiple considerations. It usually requires early attempts to prevent pain and adequate therapy for acute bouts of pain. Specific treatment depends on the cause of the pain, the natural history of the underlying health problem, and the life expectancy of the individual. If the organic illness causing the pain cannot be cured, noncurative methods of pain control become the cornerstone of treatment. Treatment methods for chronic pain can include pharmacological and nonpharmacological interventions. Pharmacological treatment includes the use of non-narcotic and narcotic medications. Non-narcotic medications such as tricyclic antidepressants, antiseizure medications, and nonsteroidal anti-inflammatory drugs (NSAIDs) serve as useful adjuncts to opioids for the treatment of different types of chronic pain. Nonpharmacological interventions include neural blockade, electrical modalities (e.g., transcutaneous electrical nerve stimulation [TENS]), physical therapy, and cognitive behavioral interventions (e.g., imagery and relaxation strategies). Chronic pain is best handled by a multidisciplinary team that includes specialists in areas such as anesthesiology, nursing, physical therapy, social services, and surgery.

Cancer is a common cause of chronic pain. The goal of chronic cancer pain management should be pain alleviation and prevention.¹⁶ Preemptive therapy tends to reduce sensitization of pain pathways and provides for more effective pain control. Pharmacological and nonpharmacological interventions are the same as those used for other types of chronic pain. Depending on the form and stage of the cancer, other treatments such as palliative radiation, antineoplastic therapies, and palliative surgery may help to control the pain. The World Health Organization has created an analgesic ladder for cancer pain that assists clinicians in choosing the appropriate analgesic medications.¹⁷

Pharmacologic Treatment. An analgesic drug is a medication that acts on the nervous system to decrease or eliminate pain without inducing loss of consciousness. Analgesic drugs do not cure the underlying cause of the pain, but their appropriate use may prevent acute pain from progressing to chronic pain. Pain medications are commonly divided into three categories: non-narcotic analgesics, opioid or narcotic analgesics, and adjuvant analgesics.

Non-narcotic Analgesics. Common non-narcotic oral analgesic medications include aspirin, other NSAIDs, and acetaminophen. Aspirin, or acetylsalicylic acid, acts centrally and peripherally to block the transmission of pain impulses. It also has antipyretic and anti-inflammatory properties. Aspirin and the other NSAIDs inhibit several forms of prostaglandins through the inhibition of cyclooxygenase, an enzyme in the prostanoid pathway. Prostaglandins affect the sensation of pain by sensitizing nociceptors to chemical mediators such as bradykinin and histamine. Independent of prostaglandins, NSAIDs also decrease the sensitivity of blood vessels to bradykinin and histamine, affect lymphokine production by T lymphocytes, reverse vasodilation, and decrease the release of inflammatory mediators from granulocytes, mast cells, and basophils. Acetaminophen is an alternative to the NSAIDs. Although usually considered equivalent to aspirin as an analgesic and antipyretic agent, it lacks anti-inflammatory properties.

Opioid Analgesics. The term opioid or narcotic is used to refer to a group of medications, natural or synthetic, with morphine-like actions. The opioids (e.g., morphine, codeine, and many other semisynthetic congeners of morphine) exert their action through opioid receptors. There are three major categories of opioid receptors in the CNS, designated mu (μ , for "morphine"), delta (δ), and kappa (κ).¹⁸ Analgesia, as well as respiratory depression, miosis, reduced gastrointestinal motility (causing constipation), feelings of well-being or euphoria, and physical dependence result principally from morphine and morphine-like opioid analgesics that act at mu receptors. Part of the pain-relieving properties of exogenous opioids such as morphine involves the release of endogenous opioids.¹⁸

Opioids are used in the management of acute and chronic pain. When given for temporary relief of severe pain, such as that occurring after surgery, there is much evidence that opioids given routinely before the pain starts or becomes extreme are far more effective than those administered in a sporadic manner. Persons who are treated in this manner seem to require fewer doses and are able to resume regular activities sooner. Opioids also are used for persons with chronic pain such as that caused by cancer. Too often, because of undue concern about the possibility of addiction, many individuals with chronic pain receive inadequate pain relief. Addiction is not considered a problem in patients with cancer.¹⁹ Most pain experts agree that it is appropriate to provide the level of opioid necessary to relieve the severe intractable pain of persons whose life expectancy is limited. Morphine remains the most useful strong opioid, and the World Health Organization has recommended that oral morphine be part of the essential medication list and made available throughout the world as the medication of choice for cancer pain.²⁰

Adjuvant Analgesics. Adjuvant analgesics include medications such as tricyclic antidepressants, antiseizure medications, and neuroleptic anxiolytic agents. The fact that the pain suppression system has nonendorphin synapses raises the possibility that potent, centrally acting, nonopiate medications may be useful in relieving pain. Serotonin has been shown to play an important role in producing analgesia. The tricyclic antidepressant medications block the removal of serotonin from the synaptic cleft to produce pain relief in some persons. These medications are particularly useful in some chronic painful conditions, such as postherpetic neuralgia. Certain antiseizure medications, such as carbamazepine and phenytoin, have analgesic effects in some pain conditions. These medications, which suppress spontaneous neuronal firing, are particularly useful in the management of pain that occurs after nerve injury. Other agents, such as the corticosteroids, may be used to decrease inflammation and nociceptive stimuli responsible for pain.

Surgical Intervention. The effects of surgical interventions may be curative or palliative. Surgical interventions that remove the source of the pain (*e.g.*, inflamed appendix) are curative. In other instances, surgery is used for symptom management, rather than for cure. Surgery for severe, intractable pain of peripheral or central origin has met with some success. It can be used to remove the cause or block the transmission of intractable pain from phantom limb pain, severe neuralgia, inoperable cancer of certain types, and causalgia.

In summary, pain is an elusive and complex phenomenon; it is a symptom common to many illnesses. It is a highly individualized experience that is shaped by a person's culture and previous life experiences, and it is difficult to measure. Traditionally, there have been two principal theories of pain: the specificity and pattern theories. Scientifically, pain is viewed within the context of nociception. Nociceptors are receptive nerve endings that respond to noxious stimuli. Pain receptors respond to mechanical, thermal, and chemical stimuli. Nociceptive neurons transmit impulses to the dorsal horn neurons using chemical neurotransmitters. The neospinothalamic and the paleospinothalamic paths are used to transmit pain information to the brain. Several neuroanatomic pathways as well as endogenous opioid peptides modulate pain in the CNS.

Pain can be classified according to location, referral, and duration as well as associated medical diagnoses. Pain can arise from cutaneous, deep somatic, or visceral locations. Referred pain is pain perceived at a site different from its origin. Acute pain is self-limiting pain that ends when the injured tissue heals, whereas chronic pain is pain that lasts much longer than the anticipated healing time for the underlying cause of the pain. Pain threshold, pain tolerance, age, gender, and other factors affect an individual's reaction to pain.

Treatment modalities for pain include the use of nonpharmacologic and pharmacologic agents used singly or in combination. It is becoming apparent that even with chronic pain, the most effective approach is early treatment or even prevention. After pain is present, the greatest success in pain assessment and management is achieved with the use of an interdisciplinary approach.

ALTERATIONS IN PAIN SENSITIVITY AND SPECIAL TYPES OF PAIN

Alterations in Pain Sensitivity

Sensitivity and perception of pain varies among persons and in the same person under different conditions and in different parts of the body. Irritation, mild hypoxia, and mild compression of a peripheral nerve often result in hyperexcitability of the sensory nerve fibers or cell bodies. This is experienced as unpleasant hypersensitivity (*i.e.*, *hyperesthesia*) or increased painfulness (*i.e.*, *hyperalgesia*). Possible causes of increased sensitivity to noxious stimuli include a decrease in the threshold of nociceptors, an increase in pain produced by suprathreshold stimuli, and the windup phenomenon. Primary hyperalgesia occurs at the site of injury. Secondary hyperalgesia occurs in nearby uninjured tissue.

Hyperpathia is a syndrome in which the sensory threshold is raised, but when it is reached, continued stimulation, especially if repetitive, results in a prolonged and unpleasant experience. This pain can be explosive and radiates through a peripheral nerve distribution. It is associated with pathologic changes in peripheral nerves, such as localized ischemia. Spontaneous, unpleasant sensations called *paresthesias* occur with more severe irritation (*e.g.*, the pins-and-needles sensation that follows temporary compression of a peripheral nerve). The general term *dysesthesia* is given to distortions (usually unpleasant) of somesthetic sensation that typically accompany partial loss of sensory innervation.

More severe pathologic processes can result in reduced or lost tactile (*e.g., hypoesthesia, anesthesia*), temperature (*e.g., hypothermia, athermia*), and pain sensation (*i.e., hypalgesia*). Analgesia is the absence of pain on noxious stimulation or the relief of pain without loss of consciousness. The inability to sense pain may result in trauma, infection, and even loss of a body part or parts. Inherited insensitivity to pain may take the form of congenital indifference or congenital insensitivity to pain. In the former, transmission of nerve impulses appears normal, but the appreciation of painful stimuli at higher levels appears to be absent. In the latter, a peripheral nerve defect apparently exists such that transmission of painful nerve impulses does not result in perception of pain. Whatever the cause, persons who lack the ability to perceive pain are at constant risk of tissue damage because pain is not serving its protective function.

Allodynia (Greek allo, "other," and odynia, "painful") is the term used for the puzzling phenomenon of pain that follows a non-noxious stimulus to apparently normal skin. This term is

740 Unit Ten: Alterations in the Nervous System

intended to refer to instances in which otherwise normal tissues may be abnormally innervated or may be referral sites for other loci that give rise to pain with non-noxious stimuli. It may be that an area is hypersensitive because of inflammation, injury, or another cause, and a normally subthreshold stimulus is sufficient to trigger the sensation of pain. This response is thought to be chemically mediated, possibly the result of tissue damage in the surrounding area. Trigger points are highly localized points on the skin or mucous membrane that can produce immediate intense pain at that site or elsewhere when stimulated by light tactile stimulation. Myofascial trigger points are foci of exquisite tenderness found in many muscles and can be responsible for pain projected to sites remote from the points of tenderness. Trigger points are widely distributed in the back of the head and neck and in the lumbar and thoracic regions. These trigger points cause reproducible myofascial pain syndromes in specific muscles. These pain syndromes are the major source of pain in clients at chronic pain treatment centers.

Special Types of Pain

Neuropathic Pain

When peripheral nerves are affected by injury or disease, it can lead to unusual and sometimes intractable sensory disturbances. These include numbness, paresthesias, and pain. Depending on the cause, few or many axons could be damaged and the condition could be unilateral or bilateral. Causes of neuropathic pain can be categorized according to the extent of peripheral nerve involvement. Conditions that can lead to pain by causing damage to peripheral nerves in a single area include nerve entrapment, nerve compression from a tumor mass, and various neuralgias (e.g., trigeminal, postherpetic, and post-traumatic). Conditions that can lead to pain by causing damage to peripheral nerves in a wide area include diabetes mellitus, long-term alcohol use, hypothyroidism, renal insufficiency, and drug treatment with neurotoxic agents.²¹ Injury to a nerve also can lead to a multisymptom, multisystem syndrome called complex regional pain syndrome (previously known as causalgia or reflex sympathetic dystrophy). Nerve damage associated with amputation is believed to be a cause of phantom limb pain.

Neuropathic pain can vary with the extent and location of disease or injury. There may be allodynia or pain that is stabbing, jabbing, burning, or shooting. The pain may be persistent or intermittent. The diagnosis depends on the mode of onset, the distribution of abnormal sensations, the quality of the pain, and other relevant medical conditions (*e.g.*, diabetes, hypothyroidism, alcohol use, rash, or trauma). Injury to peripheral nerves sometimes results in pain that persists beyond the time required for the tissues to heal. Peripheral pathologic processes (*e.g.*, neural degeneration, neuroma formation, and generation of abnormal spontaneous neural discharges from the injured sensory neuron) and neural plasticity (*i.e.*, changes in CNS function) are the primary working hypotheses to explain persistent neuropathic pain.

Treatment methods include measures aimed at restoring or preventing further nerve damage (*e.g.*, surgery to resect a tumor causing nerve compression, improving glycemic control for diabetic patients with painful neuropathies), and interventions for the palliation of pain. Although many adjuvant analgesics are used for neuropathic pain, pain control often is difficult. The initial approach in seeking adequate pain control is to try these drugs in sequence and then in combination. If there has been a poor response to the adjuvant analgesics, opioids also can be used. Nonpharmacologic therapies such as electrical stimulation of the peripheral nerve or spinal cord can be used for radiculopathies and neuralgias. As a last resort, neurolysis or neurosurgical blockade sometimes is used.

Neuralgia

Neuralgia is characterized by severe, brief, often repetitive attacks of lightning-like or throbbing pain. It occurs along the distribution of a spinal or cranial nerve and usually is precipitated by stimulation of the cutaneous region supplied by that nerve.

Trigeminal Neuralgia. Trigeminal neuralgia, or *tic douloureux*, is one of the most common and severe neuralgias. It is manifested by facial tics or grimaces and characterized by stabbing, paroxysmal attacks of pain that usually are limited to the unilateral sensory distribution of one or more branches of the trigeminal nerve, most often the maxillary or mandibular divisions. Although intermittent, the pain often is excruciating and may be triggered by light touch. The treatment of trigeminal neuralgia may include carbamazepine, an antiseizure drug, or surgery to release the vessels or scar tissue, or destroy or block branches of cranial nerve V. Other interventions include avoidance of precipitating factors (*e.g.*, stimulation of trigger spots) and eye injury caused by irritation; provision for adequate nutrition; and avoidance of social isolation.

Postherpetic Neuralgia. Postherpetic pain is pain that persists as a complication of herpes zoster or shingles. It describes the presence of pain more than 1 month after the onset of the acute attack. Postherpetic neuralgia develops in from 10% to 70% of patients with shingles²²; the risk increases with age. The pain of postherpetic neuralgia occurs in the areas of innervation of the infected ganglia. During the acute attack of herpes zoster, the reactivated virus travels from the ganglia to the skin of the corresponding dermatomes, causing localized vesicular eruption and hyperpathia (i.e., abnormally exaggerated subjective response to pain). In the acute infection, proportionately more of the large nerve fibers are destroyed. Regenerated fibers appear to have smaller diameters. Because there is a relative loss of large fibers with age, elderly persons are particularly prone to suffering because of the shift in the proportion of large- to small-diameter nerve fibers. Older patients have pain, dysesthesia, and hyperesthesia after the acute phase; these are increased by minor stimuli.

Early treatment of shingles with high doses of systemic corticosteroids and an oral antiviral drug such as acyclovir or valacyclovir, a medication that inhibits herpesvirus DNA replication, may reduce the incidence of postherpetic neuralgia. Initially, postherpetic neuralgia can be treated with a topical anesthetic agent. A tricyclic antidepressant medication may be used for pain relief. Regional nerve blockade (*i.e.*, stellate ganglion, epidural, local infiltration, or peripheral nerve block) has been used with limited success.

Complex Regional Pain Syndrome

Recently, the International Association for the Study of Pain created the terms *complex regional pain syndrome I* and *complex regional pain syndrome II*. These terms refer, respectively, to re-

flex sympathetic dystrophy and causalgia. Trauma, frequently minor, to a nerve is the major cause. However, injury to soft tissue or a broken bone also can cause these pain syndromes. The hallmark is pain and mobility problems more severe than the injury warrants. Characteristically, the pain is severe and burning with or without deep aching. Usually, the pain can be elicited with the slightest movement or touch to the affected area; it increases with repetitive stimulation; and it lasts even after the stimulation has stopped. The pain can be exacerbated by emotional upsets or any increased peripheral sympathetic nerve stimulation. The variations of complex regional pain syndromes include sympathetic components. These are characterized by vascular and trophic (e.g., dystrophic or atrophic) changes to the skin, soft tissue, and bone, and can include rubor or pallor, sweating or dryness, edema (often sharply demarcated), skin atrophy, and with time, patchy osteoporosis.

According to the clinical practice guideline the cornerstone of treatment is promoting normal use of the affected part to the extent possible.²³ Initially, oral analgesics (including the adjuvant analgesics), TENS, and physical activity are used. Interruption of sympathetic innervation may considered. If treatment by sympathetic blockade provides relief from pain, sympathectomy may be considered. If not, electrical stimulation of the spinal cord or narcotics may be considered.

Phantom Limb Pain

Phantom limb pain, a type of neurologic pain, follows amputation of a limb or part of a limb. As many as 70% of those who under amputation experience phantom pain.²⁴ The pain often begins as sensations of tingling, heat and cold, or heaviness, followed by burning, cramping, or shooting pain. It may disappear spontaneously or persist for many years. One of the more troublesome aspects of phantom pain is that the person may experience painful sensations that were present before the amputation, such as that of a painful ulcer or bunion.

Several theories have been proposed as to the causes of phantom pain.²⁴ One theory is that the end of a regenerating nerve becomes trapped in the scar tissue of the amputation site. It is known that when a peripheral nerve is cut, the scar tissue that forms becomes a barrier to regenerating outgrowth of the axon. The growing axon often becomes trapped in the scar tissue, forming a tangled growth (*i.e.*, neuroma) of smalldiameter axons, including primary nociceptive afferents and sympathetic efferents. It has been proposed that these afferents show increased sensitivity to innocuous mechanical stimuli and to sympathetic activity and circulating catecholamines. A related theory moves the source of phantom limb pain to the spinal cord, suggesting that the pain is caused by the spontaneous firing of spinal cord neurons that have lost their normal sensory input from the body. In this case, a closed self-exciting neuronal loop in the posterior horn of the spinal cord is postulated to send impulses to the brain, resulting in pain. Even the slightest irritation to the amputated limb area can initiate this cycle. Other theories propose that the phantom limb pain may arise in the brain. In one hypothesis, the pain is caused by changes in the flow of signals through somatosensory areas of the brain. Treatment of phantom limb pain has been accomplished by the use of sympathetic blocks, TENS of the large myelinated afferents innervating the area, hypnosis, and relaxation training.

In summary, pain may occur with or without an adequate stimulus, or it may be absent in the presence of an adequate stimulus—either of which describes a pain disorder. There may be analgesia (absence of pain), hyperalgesia (increased sensitivity to pain), hypalgesia (a decreased sensitivity to painful stimuli), hyperpathia (an unpleasant and prolonged response to pain), hyperesthesia (an abnormal increase in sensitivity to sensation), hypoesthesia (an abnormal decrease in sensitivity to sensations), paresthesia (abnormal touch sensation such as tingling or "pins and needles" in the absence of external stimuli), or allodynia (pain produced by stimuli that do not normally cause pain).

Neuropathic pain may be caused by trauma or disease of neurons in a focal area or in a more global distribution (e.g., from endocrine disease or neurotoxic medications). Neuralgia is characterized by severe, brief, often repetitiously occurring attacks of lightning-like or throbbing pain that occurs along the distribution of a spinal or cranial nerve and usually is precipitated by stimulation of the cutaneous region supplied by that nerve. Trigeminal neuralgia, or tic douloureux, is one of the most common and severe neuralgias. It is manifested by facial tics or grimaces. Postherpetic neuralgia is a chronic pain that can occur after shingles, an infection of the dorsal root ganglia and corresponding areas of innervation by the herpes zoster virus. Complex regional pain syndrome is an extremely painful condition that may follow sudden and traumatic deformation of peripheral nerves. Phantom limb pain, a neurologic pain, can occur after amputation of a limb or part of a limb.

HEADACHE

Headache is a common health problem. Seventy-six percent of women and 57% of men report at least one headache a month.²⁵ Headache is caused by a number of conditions. Some headaches represent primary disorders, and others occur secondary to other disease conditions in which head pain is a symptom.

The most common types of primary or chronic headaches are migraine headache, tension-type headache, and cluster headache. In 1988, the International Headache Society published a proposed classification of headaches that lists diagnostic criteria for both primary headache syndromes and headaches that occur secondary to other medical conditions (see Chart 39-1 for a summary of the components of the system).²⁶

Although most causes of secondary headache are benign, some are indications of serious disorders, such as meningitis, brain tumor, or cerebral aneurysm. The sudden onset of a severe, intractable headache in an otherwise healthy person is more likely related to a serious intracranial disorder, such as subarachnoid hemorrhage or meningitis, than to a chronic headache disorder. Headaches that disturb sleep, exertional headaches, and headaches accompanied by neurologic symptoms such as drowsiness, visual or limb disturbances, or altered mental status also are indicative of underlying intracranial lesions or other pathology.

CHART 39-1 Classification and Diagnostic Criteria for Headache Disorders, Cranial Neuralgias, and Facial Pain

- 1. Migraine
 - 1.1 Migraine without aura
 - 1.2 Migraine with aura
 - 1.3 Ophthalmoplegic migraine
 - 1.4 Retinal migraine
 - 1.5 Childhood periodic syndromes that may be precursors to or associated with migraine
 - 1.6 Complications with migraine
 - 1.7 Migrainous disorder not fulfilling above criteria
- 2. Tension-type headache
 - 2.1 Episodic tension-type headache
 - 2.2 Chronic tension-type headache
 - 2.3 Headache of the tension type not fulfilling the above criteria
- 3. Cluster headache and chronic paroxysmal hemicrania
- 4. Miscellaneous headaches unassociated with structural lesion
- 5. Headache associated with head trauma
- 6. Headache associated with vascular disorders
- 7. Headache associated with nonvascular intracranial disorders
- 8. Headache associated with substances or their withdrawal
- 9. Headache associated with noncephalic infection
- 10. Headache associated with metabolic disorder
- 11. Headache or facial pain associated with disorder of cranium, neck, eyes, ears, nose, sinuses, teeth, mouth, or other facial or cranial structures
- 12. Cranial neuralgias, nerve trunk pain, and deafferentation pain
- 13. Headache not classifiable

(Adapted from Oleson J. [1988]. Classification and diagnostic criteria for headache disorders, cranial neuralgias, and facial pain. *Cephalgia* 8 [Suppl 7], 13–19)

The diagnosis and classification of headaches often is difficult. It requires a comprehensive history and physical examination to exclude secondary causes. The history should include factors that precipitate headache, such as foods and food additives, missed meals, and association with the menstrual period. A careful medication history is essential because many medications can provoke or aggravate headaches. Alcohol also can cause or aggravate headache. A headache diary in which the person records his or her headaches and concurrent or antecedent events may be helpful in identifying factors that contribute to headache onset. Appropriate laboratory and imaging studies of the brain may be done to rule out secondary headaches.

Migraine Headache

Migraine headaches affect more than 10 million persons in the United States. They occur more frequently in women than men and result in considerable time lost from work and other activities.²⁷ Migraine headaches tend to run in families and are

thought to be inherited as an autosomal dominant trait with incomplete penetrance.²⁵

There are two categories of migraine headache-migraine without aura, which accounts for approximately 85% of migraines, and migraine with aura, which accounts for most of the remaining migraines. Migraine without aura is a pulsatile, throbbing, unilateral headache that typically lasts 1 to 2 days and is aggravated by routine physical activity. The headache is accompanied by nausea and vomiting, which often is disabling, and a sensitivity to light and sound. Visual disturbances occur quite commonly and consist of visual hallucinations such as stars, sparks, and flashes of light. Migraine with aura has similar symptoms, but with the addition of visual or neurologic symptoms that precede the headache. The aura usually develops over a period of 5 to 20 minutes and lasts less than an hour. Although only a small percentage of persons with migraine experience an aura before an attack, many persons without aura have prodromal symptoms, such as fatigue and irritability, that precede the attack by hours or even days.

Subtypes of migraine include ophthalmoplegic migraine, hemiplegic migraine, aphasic migraine, and retinal migraine, in which transient visual and motor deficits occur. Ophthalmoplegic migraine is characterized by diplopia, caused by a transient paralysis of the muscles that control eye movement (usually the third cranial nerve), and localized pain around the eye. Migraine headache also can present as a mixed headache, including symptoms typically associated with tension-type headache or chronic daily headache. These are called *transformed migraine* and are difficult to classify.

The mechanisms of migraine attacks are poorly understood. There is increasing evidence to support a neurogenic basis for migraine. Supporting this neurogenic concept is the common presence of premonitory symptoms before the headache begins; the presence of focal neurologic disturbances, which cannot be explained in terms of cerebral blood flow; and the numerous accompanying symptoms, including autonomic and constitutional dysfunction.²⁵ The pathophysiologic process of migraine probably involves alterations in serotonin function and occurrence of inflammatory disturbances in the trigeminal vascular system.²⁵ Hormonal variations, particularly in estrogen levels, play a role in the pattern of migraine attacks. For many women, migraine headaches coincide with their menstrual periods. The greater predominance of migraine headaches in women is thought to be related to the aggravating effect of estrogen on the migraine mechanism.²⁵ Dietary substances, such as monosodium glutamate, aged cheese, and chocolate, also may precipitate migraine headaches. The actual triggers for migraine are the chemicals in the food, not allergens.

The treatment of migraine headaches includes preventative and abortive nonpharmacologic and pharmacologic treatment. In 1998, the U.S. Headache Consortium, a multidisciplinary panel, produced a set of evidence-based guidelines for the nonpharmacologic and pharmacologic management and prevention of migraine headaches in primary care settings.²⁷

Nonpharmacologic treatment includes the avoidance of migraine triggers, such as foods, that precipitate an attack. Many persons with migraines benefit from maintaining regular eating and sleeping habits. Measures to control stress, which also can precipitate an attack, also are important. During an attack, many persons find it helpful to retire to a quiet, darkened room until symptoms subside.

Pharmacologic treatment involves both abortive therapy for acute attacks and preventive therapy. A wide range of medications is used to treat the acute symptoms of migraine headache.²⁸ These include 5-HT₁ (serotonin) receptor agonists (*e.g.*, the triptans), ergotamine derivatives, analgesics (*e.g.*, acetylsalicylic acid, acetaminophen, and NSAIDs such as naproxen sodium), sedatives (*e.g.*, butalbital), and antiemetic medications. Frequent use of abortive headache medications may cause rebound headache or perpetuate chronic daily headaches.

Preventative pharmacologic treatment may be necessary if migrainous headaches are disabling or occur more than two or three times a month. In most cases, preventative treatment must be taken daily for months to years. The β -adrenergic blocking medications are usually the first choice for prophylactic treatment because of empiric support for their effectiveness, safety, efficacy, and favorable side effect profile. Several other medications that may be effective prophylactically for migraine headache are antidepressants, selective serotonin reuptake inhibitors, calcium channel blockers, antiseizure medications (*e.g.*, divalproex sodium, valproic acid), ergot derivatives (methysergide), and NSAIDs (naproxen sodium).²⁹ When a decision to discontinue preventive therapy is made, the medications should be gradually withdrawn.

There may be serious side effects with some of the antimigraine medications. Because of the risk of coronary vasospasm, the serotonin receptor agonists should not be given to persons with coronary artery disease. Ergotamine preparations can cause uterine contractions and should not be given to pregnant women. They also can cause vasospasm and should be used with caution in persons with peripheral vascular disease.

Migraine Headache in Children

Migraine headaches occur in children as well as adults.^{30,31} Before puberty, migraine headaches are equally distributed between the sexes. The essential diagnostic criterion for migraine in children is the presence of recurrent headaches separated by pain-free periods. Diagnosis is based on at least three of the following symptoms or associated findings: abdominal pain, nausea or vomiting, throbbing headache, unilateral location, associated aura (visual, sensory, motor), relief during sleep, and a positive family history.³¹ Symptoms vary widely among children, from those that interrupt activities and cause the child to seek relief in a dark environment, to those detectable only by direct questioning. A common feature of migraine in children is intense nausea and vomiting. The vomiting may be associated with abdominal pain and fever; thus, migraine may be confused with other conditions such as appendicitis. More than half of children with migraine undergo spontaneous prolonged remission after their 10th birthday. Because headaches in children can be a symptom of other, more serious disorders, including intracranial lesions, it is important that other causes of headache that require immediate treatment be excluded.

Cluster Headache

Cluster headaches are relatively uncommon headaches that affect men more often than women. These headaches tend to occur in clusters over weeks or months, followed by a long, headache-free remission period. Typically the symptoms in cluster headaches include severe, unrelenting, unilateral pain located, in order of decreasing frequency, in the orbital, retroorbital, temporal, supraorbital, and infraorbital region. The pain is of rapid onset and builds to a peak in approximately 10 to 15 minutes, lasting for 15 to 180 minutes. The pain behind the eye radiates to the ipsilateral trigeminal nerve (*e.g.*, temple, cheek, gum). The headache frequently is associated with one or more symptoms such as conjunctival redness, lacrimation, nasal congestion, rhinorrhea, forehead and facial sweating, miosis, ptosis, and eyelid edema. Because of their location and associated symptoms, cluster headaches are often mistaken for sinus infections or dental problems.²⁵

The underlying pathophysiologic mechanisms of cluster headaches are unknown. Hypotheses include the interplay of vascular, neurogenic, metabolic, and humoral factors. Although the trigeminovascular system appears to be involved in the pathogenesis of cluster headache, a theory to explain the symptoms, periodicity, and circadian regularity of cluster headaches does not exist. The regularity in the timing of cluster headache may be caused by dysfunction of the hypothalamic biologic clock mechanisms. Ipsilateral lacrimation, nasal stuffiness, and rhinorrhea are thought to result from parasympathetic overactivity. Pain and vasodilation are thought to result from activation of the trigeminovascular system.

Because of the relatively short duration and self-limited nature of cluster headache, oral preparations typically take too long to reach therapeutic levels. The most effective treatments are those that act quickly (*e.g.*, oxygen inhalation and subcutaneous sumatriptan). Intranasal lidocaine also may be effective.²⁵ Oxygen inhalation may be indicated for home use. Prophylactic medications for cluster headaches include ergotamine, verapamil, methysergide, lithium carbonate, corticosteroids, sodium valproate, and indomethacin.

Tension-type Headache

The most common type of headache is tension-type headache. Unlike migraine and cluster headaches, tension-type headache usually is not sufficiently severe that it interferes with daily activities. Tension-type headaches frequently are described as dull, aching, diffuse, nondescript headaches, occurring in a hatband distribution around the head, and not associated with nausea or vomiting or worsened by activity. They can occur infrequently or be episodic or chronic.

The exact mechanisms of tension-type headache are not known, and the hypotheses of causation are contradictory. One popular theory is that tension-type headache results from sustained tension of the muscles of the scalp and neck; however, some research has found no correlation between muscle contraction and tension-type headache. Many authorities now believe that tension-type headaches are forms of migraine headache.²⁵ It is thought that migraine headache may be transformed gradually into chronic tension-type headache. Tension-type headaches also may be caused by oromandibular dysfunction, psychogenic stress, anxiety, depression, and muscular stress. They also may result from overuse of analgesics or caffeine. Daily use of caffeine, whether in beverages or medications, can produce addiction, and a headache can develop in such persons who go without caffeine for several hours.²⁵

Tension-type headaches often are more responsive to nonpharmacologic techniques, such as biofeedback, massage, acupuncture, relaxation, imagery, and physical therapy, than are other types of headache. For persons with poor posture, a combination of range-of-motion exercises, relaxation, and posture improvement may be helpful.²⁶

The medications of choice for acute treatment of tensiontype headaches are analgesics, including acetylsalicylic acid, acetaminophen, and NSAIDs. Persons with infrequently occurring tension-type headaches usually self-medicate using over-the-counter analgesics to treat the acute pain and do not require prophylactic medication. These agents should be used cautiously because rebound headaches can develop when the medications are taken regularly.

Because the "dividing lines" between tension-type headache, migraine, and chronic daily headache often are vague, addition of medications as well as the entire range of migraine medications may be tried in refractory cases. Other medications used concomitantly with analgesics include sedatives (*e.g.*, butalbital), anxiolytics (*e.g.*, diazepam), and skeletal muscle relaxants (*e.g.*, orphenadrine). Prophylactic treatment can include antidepressants (*e.g.*, amitriptyline, doxepin).

Temporomandibular Joint Pain

A common cause of head pain is temporomandibular joint (TMJ) syndrome. It usually is caused by an imbalance in joint movement because of poor bite, bruxism (*i.e.*, teeth grinding), or joint problems such as inflammation, trauma, and degenerative changes.³² The pain almost always is referred and commonly presents as facial muscle pain, headache, neck ache, or earache. Referred pain is aggravated by jaw function. Headache associated with this syndrome is common in adults and children and can cause chronic pain problems.

Treatment of TMJ pain is aimed at correcting the problem, and in some cases this may be difficult. The initial therapy for TMJ should be directed toward relief of pain and improvement in function. Pain relief often can be achieved with use of the NSAIDs. Muscle relaxants may be used when muscle spasm is a problem. In some cases, the selected application of heat or cold, or both, may provide relief. Referral to a dentist who is associated with a team of therapists, such as a psychologist, physical therapist, or pain specialist, may be indicated.³²

Head pain is a common disorder that is caused by a number of conditions. Some headaches represent primary disorders and others occur secondary to another disease state in which head pain is a symptom. Primary headache disorders include migraine headache, tension-type headache, and cluster headache. Although most causes of secondary headache are benign, some are indications of serious disorders, such as meningitis, brain tumor, or cerebral aneurysm. TMJ syndrome is one of the major causes of headaches. It usually is caused by an imbalance in joint movement because of poor bite, teeth grinding, or joint problems such as inflammation, trauma, and degenerative changes.

PAIN IN CHILDREN AND OLDER ADULTS

Pain frequently is underrecognized and undertreated in both children^{33,34} and the elderly.³⁵ In addition to the common obstacles to adequate pain management, such as concern about

the effects of analgesia on respiratory status and the potential for addiction to opioids, there are additional deterrents to adequate pain management in children and the elderly. With regard to both children and the elderly, there are stereotypic beliefs that they feel less pain than do other patients.^{33–36} These beliefs may affect a clinician's opinion about the need for pain control. In very young children and confused elderly, there are several additional factors. These include the extreme difficulty of assessing the location and intensity of pain in individuals who are cognitively immature or cognitively impaired, and the argument that even if they feel pain, they do not remember it.

Pain in Children

Human responsiveness to painful stimuli begins in the neonatal period and continues through the life span. Although the specific and localized behavioral reactions are less marked in the younger neonate or the more cognitively impaired individual, protective or withdrawal reflexes in response to nociceptive stimuli are clearly demonstrated. Pain pathways, cortical and subcortical centers, and neurochemical responses associated with pain transmission are developed and functional by the last trimester of pregnancy. As infants mature and children grow, their responses to pain become more complex and reflective of their maturing cognitive and developmental processes.³⁶ Children do feel pain and have been shown reliably and accurately to report pain at as young as 3 years of age. They also remember pain, as evidenced in studies of children with cancer, whose distress during painful procedures increases over time without intervention, and in neonates in intensive care units, who demonstrate protective withdrawal responses to a heel stick after repeated episodes.

The assessment of pain in children is somewhat complicated, but research has led to the development of a variety of developmentally appropriate measurement tools. These include scales with faces of actual children or cartoon faces that can be used to elicit a pain report from young children. In older children and adolescents, numeric scales (*i.e.*, 1 to 10) and word graphic scales (*i.e.*, "none," "a little," "most I have ever experienced") can be used. Another strategy for assessing a child's pain is to use a body outline and ask the child to indicate where the hurt is located. Particular care must be taken is assessing children's reports of pain because their report may be influenced by a variety of factors, including age, anxiety and fear levels, and parental presence.

The management of children's pain basically falls into two categories: pharmacologic and nonpharmacologic. In terms of pharmacologic interventions, many of the analgesics used in adults can be used safely and effectively in children and adolescents. However, it is critical when using specific medications to determine that the medication has been approved for use with children and that it is dosed appropriately according to the child's weight. As with any person in pain, the type of analgesic used should be matched to the type and intensity of pain; and whether the patient is a child or adult, the management of chronic pain may require a multidisciplinary team. The overriding principle in all pediatric pain management is to treat each child's pain on an individual basis and to match the analgesic agent with the cause and the level of pain. A second principle involves maintaining the balance between the level of side effects and pain relief such that pain relief is obtained with

as little opioid and sedation as possible. One strategy toward this end is to time the administration of analgesia so that a steady blood level is achieved and, as much as possible, pain is prevented. This requires that the child receive analgesia on a regular dosing schedule, not "as needed."

Nonpharmacologic strategies can be very effective in reducing the overall amount of pain and amount of analgesia used. In addition, some nonpharmacologic strategies can reduce anxiety and increase the child's level of self-control during pain. In full-term infants, ingesting 2 mL of a sucrose solution has been found to relieve the pain from a heel stick.37 Children as young as 4 years of age can use TENS, and they can be taught to use simple distraction and relaxation and other techniques such as application of heat and cold.³⁸ Other nonpharmacologic techniques can be taught to the child to provide psychological preparation for a painful procedure or surgery. These include positive self-talk, imagery, play therapy, modeling, and rehearsal. The nonpharmacologic interventions must be developmentally appropriate and, if possible, the child and parent should be taught these techniques when the child is not in pain (e.g., before surgery or a painful procedure) so that it is easier to practice the technique.

Pain in Older Adults

Among adults, the prevalence of pain in the general population increases with age.³⁵ It is estimated that from 25% to 50% of community-dwelling elders³⁹ and 80% of individuals in longterm care facilities report experiencing pain.⁴⁰ Research is inconsistent about whether there are age-related changes in pain perception. Some apparent age-related differences in pain may be attributable to differences in willingness to report the pain, rather altered pain perception. The elderly may be reluctant to report pain so as not to be a burden or out of fear of the diagnoses, tests, medications, or costs that may result from an attempt to diagnose or treat their pain.

The assessment of pain in the elderly may be relatively simple in a well-informed, alert, cognitively intact individual with pain from a single source and no comorbidities. In contrast it may to extraordinarily difficult to assess pain in a frail individual with severe dementia and many concurrent health problems. When possible, patient report of pain is the gold standard, but outward signs of pain should be considered as well. Accurately diagnosing pain when the individual has many health problems or some decline in cognitive function can be particularly challenging. In recent years, there has been increased awareness of the need to address issues of pain in individuals with dementia. The Assessment for Discomfort in Dementia Protocol is one example of the efforts to improve assessment and pain management in these individuals. It includes behavioral criteria for assessing pain and recommended interventions for pain. Its use has been shown to improve pain management.35,41

Treatment of pain in the elderly can be complicated. The elderly may have physiologic changes that affect the pharmacokinetics of medications prescribed for pain management. These changes include decreased blood flow to organs, delayed gastric motility, reduced kidney function, and decreased albumin related to poor nutrition. Physiologic changes may affect the choice of medications or dosing. In addition, the elderly often have many coexisting health problems, leading to polypharmacy. When multiple medications are being taken, there is an increased risk of drug interactions and of noncompliance because of the complexity of the treatment regimen.

In summary, children experience and remember pain, and even fairly young children are able accurately and reliably to report their pain. Recognition of this has changed the clinical practice of health professionals involved in the assessment of children's pain. Pain management in children is improving as exaggerated fears and misconceptions concerning the risks of addiction and respiratory depression in children treated with opioids also are dispelled. Pharmacologic (including opioids) and nonpharmacologic pain management interventions have been shown to be effective in children. Nonpharmacologic techniques must be based on the developmental level of the child and should be taught to both children and parents.

Pain is a common symptom in the elderly. Assessment, diagnosis, and treatment of pain in the elderly can be challenging. The elderly may be reluctant or cognitively unable to report their pain. Diagnosis and treatment can be complicated by comorbidities and age-related changes in cognitive and physiologic function.

REVIEW QUESTIONS

Compare the tactile, thermal, and position sense modalities in terms of receptors, adequate stimuli, ascending pathways, and central integrative mechanisms.

- Describe the organization of the somatosensory system in terms of first-, second-, and third-order neurons.
- Compare the discriminative pathway with the anterolateral pathway, and explain the clinical usefulness of this distinction.
- Differentiate among the specificity, pattern, gate control, and neuromatrix theories of pain.
- State the difference between the Aδ- and C-fiber neurons in the transmission of pain information.
- Trace the transmission of pain signals with reference to the neospinothalamic, paleospinothalamic, and reticulospinal pathways, including the role of chemical mediators and factors that modulate pain transmission.

Differentiate acute pain from chronic pain in terms of mechanisms, manifestations, and treatment.

Describe the mechanisms of referred pain, and list the common sites of referral for cardiac and other types of visceral pain.

Describe the cause and characteristics and treatment of neuropathic pain, trigeminal neuralgia, postherpetic neuralgia, and complex regional pain syndrome.

Differentiate between the periodicity of occurrence and manifestations of migraine headache, cluster headache, tensiontype headache, and headache caused by temporomandibular joint syndrome.

State how the pain response may differ in children and older adults.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- Guyton A., Hall J.E. (2001). Textbook of medical physiology (10th ed., pp. 552–563). Philadelphia: W.B. Saunders.
- Rhoades R.A., Tanner G.A. (1996). *Medical physiology* (pp. 60–70). Boston: Little, Brown.
- Kandel E.R., Schwartz J.H., Jessell T.M. (2000). Principles of neural science (4th ed., pp. 430–450). New York: McGraw-Hill.
- 4. Lister B.J. (1996). Dilemmas in the treatment of chronic pain. *American Journal of Medicine* 101 (Suppl. 1A), 2S-4S.
- Bonica J.J. (1991). History of pain concepts and pain theory. Mount Sinai Journal of Medicine 58, 191–202.
- Melzack R., Wall P.D. (1965). Pain mechanisms: A new theory. Science 150, 971–979.
- Melzack R. (1999). From the gate to the neuromatrix. Pain 6 (Suppl.), S121–S126.
- Cross S.A. (1994). Pathophysiology of pain. Mayo Clinic Proceedings 69, 375–383.
- Beeson J., Chaouch A. (1987). Peripheral and spinal mechanisms of nociception. *Physiological Reviews* 67, 67–186.
- 10. Markenson J.A. (1996). Mechanisms of chronic pain. *American Journal* of *Medicine* 101 (Suppl. 1A), 6S–18S.
- Cooper J.R., Bloom F.E., Roth R.H. (1991). The biochemical basis of neuropharmacology (6th ed., p. 263). New York: Oxford University Press.
- Fields H.L., Heinricher M.M., Mason P. (1991). Neurotransmitters in nociceptive modulatory circuits. *Review of Neuroscience* 14, 219–245.
- Grichnick K., Ferrante F.M. (1991). The difference between acute and chronic pain. *Mount Sinai Journal of Medicine* 58, 217–220.
- 14. Ruoff G.E. (1996). Depression in the patient with chronic pain. *Journal of Family Practice* 43 (6 Suppl.), S25–S33.
- 15. Acute Pain Management Guideline Panel. (1992). Clinical practice guideline no. 1. Acute pain management: Operative or medical procedures and trauma. AHCPR Publication No. 92-0032. Rockville, MD: Agency for Health Care Policy and Research, Public Health Service, U.S. Department of Health and Human Services.
- Jacox A., Carr D.B., Payne R., et al. (1994). *Clinical practice guideline* no. 9. Management of cancer pain. AHCPR Publication No. 94-0592. Rockville, MD: Agency for Health Care Policy and Research, Public Health Service, U.S. Department of Health and Human Services.
- World Health Organization. (1990). Cancer pain relief and palliative care: Report of the WHO Expert Committee (Technical Report Series. 804). Geneva, Switzerland: Author.
- Way W.L., Field H.L., Schumacher M.A. (2001). Opioid analgesics and antagonists. In Katzung H. (Ed.), *Basic and clinical pharmacology* (8th ed., pp. 512–531). New York: Lange Medical Books/McGraw-Hill.
- 19. Melzack R. (1990). The tragedy of needless pain. *Scientific American* 262 (2), 2–8.
- 20. Swerdlow M., Stjerward J. (1982). Cancer pain relief: An urgent problem. *World Health Forum* 3, 325–330.

- Vaillancourt P.D., Langevin H.M. (1999). Painful peripheral neuropathies. *Medical Clinics of North America* 83, 627–643.
- Kost R.G., Straus S.E. (1996). Postherpetic neuralgia: Pathogenesis, treatment and prevention. *New England Journal of Medicine* 335, 32–42.
- Reflex Sympathetic Dystrophy Syndrome Association of America. (2000). Clinical practice guideline for treatment of reflex sympathetic dystrophy syndrome. [On-line]. Available: http://www.rsds.org/ cpgeng.htm.
- 24. Melzack R. (1992). Phantom limb. Scientific American 226, 120-126.
- Saper J.R. (1999). Headache disorders. Medical Clinics of North America 83, 663–670.
- Olesen J. (Chair). (1988). Headache Classification Committee of the International Headache Society: The classification and diagnostic criteria for headache disorders, cranial neuralgias, and facial pain. *Cephalalgia* 8 (Suppl. 7), 1–96.
- The U.S. Headache Consortium Participants. (1998). Evidencebased guidelines for migraine headaches in primary care setting: Pharmacological management of acute attacks. American Academy of Neurology. [On-line]. Available http://www.aan.com/public/ practiceguidelines/05.pdf.
- Moore K.L., Noble S.L. (1997). Drug treatment for migraine: Part I. Acute therapy and drug-rebound headache. *American Family Physician* 56 (8), 2039–2048.
- 29. Noble S., Moore K.L. (1997). Drug treatment of migraine: Part II. Preventative therapy. *American Family Physician* 56 (9), 2279–2286.
- Annequin D., Tourniare B., Massoui H. (2000). Migraine and headache in childhood and adolescence. *Pediatric Clinics of North America* 47, 617–631.
- Behrman R.E., Kliegman R.M., Jenson H.B. (2000). Nelson textbook of pediatrics (16th ed., pp. 1832–1834). Philadelphia: W.B. Saunders.
- 32. Okeson J.P. (1996). Temporomandibular disorders in the medical practice. *Journal of Family Practice* 43, 347–356.
- 33. Berde C.B., Sethna N.F. (2002). Analgesics for treatment of pain in children. *New England Journal of Medicine* 347(14), 1094–1103.
- Broome M., Richtsmeier A., Maikler V., et al. (1996). Pediatric pain practices: A survey of health professionals. *Journal of Pain and Symptom Management* 4, 315–319.
- Kovach C.R., Weissman, D.E., Griffie J., et al. (1999). Assessment and treatment of discomfort for people with late-stage dementia. *Journal* of Pain and Symptom Management 18, 412–419.
- 36. Anand K.J.S., Hickey P.R. (1987). Pain and its effects in the human neonate and fetus. *New England Journal of Medicine* 317 (21), 321.
- Haouari N., Wood C., Griffiths G., et al. (1995). The analgesic effect of sucrose in full term infants: A randomized controlled trial. *British Medical Journal* 310, 1498–1500.
- Vessey J., Carlson K., McGill J. (1995). Use of distraction with children during an acute pain experience. *Nursing Research* 43, 369–372.
- Brattberg G., Parker M.G., Thorslund M. (1996). A longitudinal study of pain: Reported from middle age to old age. *Clinical Journal of Pain* 13, 144–149.
- 40. Ferrell B.A., Ferrell B.R., Osterweil D. (1990). Pain in the nursing home. *Journal of the American Geriatric Society* 38, 409–414.
- Morrison R.S., Siu A.L. (2000). A comparison of pain and its treatment in advanced dementia and cognitively intact patients with hip fracture. *Journal of Pain and Symptom Management* 19, 240–248.

40

Alterations in Special Sensory Function

The Eye and Disorders of Vision The Conjunctiva Conjunctivitis The Cornea Disorders of the Cornea **Aqueous Humor and Intraocular Pressure** Glaucoma The Lens **Disorders of Refraction Disorders of Accommodation** Cataracts The Retina **Blood Supply and Vascular Lesions Retinopathies Retinal Detachment** Macular Degeneration **Neural Pathways Disorders of the Optic Pathways** Visual Field Defects The Extraocular Eye Muscles and **Disorders of Eye Movements Strabismus** Amblyopia The Ear and Disorders of Auditory Function **External Ear** Disorders of the External Ear The Middle Ear and Eustachian Tube Disorders of the Eustachian Tube **Otitis Media Otosclerosis** The Inner Ear and Auditory Pathways Disorders of the Inner Ear and Central Auditory Pathways

Hearing Loss Conductive Hearing Loss Sensorineural Hearing Loss Presbycusis **Detection and Treatment of Hearing Loss** The Vestibular System and Maintenance of Equilibrium **Nystagmus** Vertigo **Motion Sickness Disorders of Peripheral Vestibular Function** Ménière's Disease **Benign Paroxysmal Positional Vertigo Disorders of Central Vestibular Function Diagnosis and Treatment Tests of Vestibular Function** Treatment of Vestibular Disorders

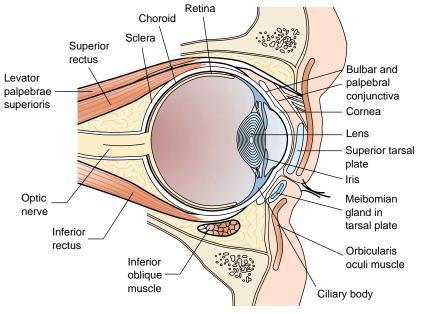
he special senses allow us to view and hear what is going on around us, to maintain our balance, and to communicate effectively with others. This chapter focuses on the eye and disorders of vision; the ear and disorders of hearing; and the vestibular system and disorders of equilibrium and balance.

THE EYE AND DISORDERS OF VISION

Almost 17.3 million persons in the United States have some degree of visual impairment; of these, 1.1 million are legally blind.¹ The prevalence of vision impairment increases with age. An estimated 26% of persons 75 years of age and older report visual impairment severe enough to interfere with recognizing a friend across the room or reading newspaper print even when wearing glasses.¹ At the other end of the age spectrum, an estimated 95,100 children younger than 18 years of age are severely visually impaired.¹

The optic globe, or eyeball, is a remarkable, mobile, nearly spherical structure contained in a pyramid-shaped cavity of the skull called the *orbit* (Fig. 40-1). The eyeball consists of

748 Unit Ten: Alterations in the Nervous System



■ **FIGURE 40-1** ■ The eye and its appendages, lateral view.

three layers: an outer supporting fibrous layer, the sclera; a vascular layer, the uveal tract; and a neural layer, the retina. Its interior is filled with transparent media, the aqueous and vitreous humors, which allow the penetration and transmission of light to photoreceptors in the retina. Exposed surfaces of the eyes are protected by the eyelids, which are mucous membrane-lined skin flaps that provide a means for shutting out most light. Tears bathe the anterior surface of the eye; they prevent friction between it and the lid, maintain hydration of the cornea, and protect the eye from irritation by foreign objects. The two eyes, with their associated extraocular muscles that permit directional rotation of the eyeball, provide different images of the same object. This results in binocular vision with depth perception.

The Conjunctiva

The conjunctiva is a thin transparent mucous membrane that lines the inner surface of both eyelids and covers the anterior surface of the optic globe to the limbus, or corneoscleral junction (Fig. 40-1). When the eyes are closed, the conjunctiva lines the closed conjunctival sac.

Conjunctivitis

Conjunctivitis, or inflammation of the conjunctiva (*i.e.*, red eye or pink eye), is one of the most common forms of eye disease. It may result from bacterial or viral infection, allergens, chemical agents, physical irritants, or radiant energy. Infections may extend from areas adjacent to the conjunctiva or may be bloodborne, such as in measles or chickenpox. Newborns can contract conjunctivitis during the birth process.

Dependent upon the cause, conjunctivitis can vary in severity from a mild hyperemia (redness) with tearing to severe conjunctivitis with purulent drainage. The conjunctiva is extremely sensitive to irritation and inflammation. Important symptoms

KEY CONCEPTS

VISION

- Vision is a special sensory function that incorporates the visual receptor functions of the eyeball, the optic nerve, and visual pathways that carry and distribute sensory information from the optic globe to the central nervous system, and the primary and visual association cortices that translate the sensory signals into visual images.
- The eyeball is a hollow spherical structure that functions in the reception of the light rays that provide the stimuli for vision. The refractive surface of the cornea and accommodative properties of the lens serve to focus the light signals from near and far objects on the photoreceptors in the retina.
- Visual information is carried to the brain by axons of the retinal cells that form the optic nerve. The two optic nerves fuse in the optic chiasm, where axons of the nasal retina of each eye cross to the contralateral side and travel with axons of the ipsilateral temporal retina to form the fibers of the optic radiations that travel to the visual cortex.
- Binocular vision depends on the coordination of three pairs of extraocular nerves that provide for the conjugate eye movements, with optical axes of the two eyes maintained parallel to one another.

of conjunctivitis are a foreign body sensation, a scratching or burning sensation, itching, and photophobia.

Bacterial Conjunctivitis. Bacterial conjunctivitis may present as a hyperacute, acute, or chronic infection. Common agents of bacterial conjunctivitis are *Streptococcus pneumoniae*, *Staphylococcus aureus*, and *Haemophilus influenzae*. The infection usually has an abrupt onset and is characterized by large amounts of yellow-green drainage. The eyelids are sticky, and there may be excoriation of the lid margins. Treatment may include local application of antibiotics. The disorder usually is self-limited, lasting approximately 10 to 14 days if untreated.²

Hyperacute conjunctivitis is a severe, sight-threatening ocular infection. The most common causes of hyperacute purulent conjunctivitis are *Neisseria gonorrhoeae* and *Neisseria meningitidis*, with *N. gonorrhoeae* being the most common.³ The symptoms, which typically are progressive, include conjunctival redness and edema (chemosis); lid swelling and tenderness; and swollen preauricular lymph nodes. Gonococcal ocular infections that are left untreated result in corneal ulceration with ultimate perforation and sometimes permanent loss of vision.³ Treatment includes systemic antibiotics supplemented with ocular antibiotics. Because of the increasing prevalence of penicillin-resistant *N. gonorrhoeae*, antibiotic choice should be determined by current information regarding antibiotic sensitivity.

Chronic bacterial conjunctivitis most commonly is caused by *Staphylococcus* species, although other bacteria may be involved. It often is associated with inflammation and bacterial colonization of eyelid margins. The symptoms of chronic bacterial conjunctivitis vary and can include itching, burning, foreign body sensation, and morning eyelash crusting. There may be flaky debris and erythema along the lid margins, as well as eyelash loss and eye redness. Some people with chronic bacterial conjunctivitis also have recurrent inflammation or infection of the meibomian glands (stye) in the lid margins.³ Treatment includes good eyelid hygiene and application of topical antibiotics.

Viral Conjunctivitis. One of the most common causes of viral conjunctivitis is adenovirus type 3. Conjunctivitis caused by this agent usually is associated with pharyngitis, fever, and malaise.^{4,5} It causes generalized hyperemia, excessive tearing, and minimal exudate. Children are affected more often than adults. Swimming pools contaminated because of inadequate chlorination are common sources of infection. Infections caused by adenoviruses types 4 and 7 often are associated with acute respiratory disease. These viruses are rapidly disseminated when large groups mingle with infected individuals (*e.g.*, military recruits). There is no specific treatment for this type of viral conjunctivitis; it usually lasts 7 to 14 days. Preventive measures include hygienic measures and avoiding shared use of eyedroppers, eye makeup, goggles, and towels.

Herpes simplex virus (HSV) conjunctivitis is characterized by unilateral infection, irritation, mucoid discharge, pain, and mild photophobia. Herpetic vesicles may develop on the eyelids and lid margins. Although the infection usually is caused by the HSV type 1, it also can be caused by the HSV type 2. It often is associated with herpes simplex virus keratitis, in which the cornea shows discrete epithelial lesions. Treatment involves the use of systemic or local antiviral agents. Local corticosteroid preparations increase the activity of the herpes virus, apparently by enhancing the destructive effect of collagenase on the collagen of the cornea. The use of these medications should be avoided in those suspected of having herpes simplex conjunctivitis or keratitis.

Chlamydial Conjunctivitis. Inclusion conjunctivitis usually is a benign suppurative conjunctivitis transmitted by the type of *Chlamydia trachomatis* (serotypes D through K) that causes venereal infections. It is spread by contaminated genital secretions and occurs in newborns of mothers with *C. trachomatis* infections of the birth canal. It also can be contracted through swimming in unchlorinated pools. The incubation period varies from 5 to 12 days, and the disease may last for several months if untreated. The infection usually is treated with appropriate oral antibiotics.

A more serious form of infection is caused by a different strain of *C. trachomatis* (serotypes A through C). This form of chlamydial infection affects the conjunctiva and causes ulceration and scarring of the cornea. It is the leading cause of preventable blindness in the world. Although the agent is widespread, it is seen mostly in developing countries, particularly those of Africa, Asia, and the Middle East.⁶ It is transmitted by direct human contact, contaminated objects (fomites), and flies.

Ophthalmia Neonatorum. Ophthalmia neonatorum is a form of conjunctivitis that occurs in infants younger than 1 month of age. It is usually contracted during or soon after vaginal delivery. There are many causes, including *N. gonorrhoeae, Pseudomonas*, and *C. trachomatis.*⁷ Epidemiologically, these infections reflect those sexually transmitted diseases most common in a particular area. Currently, *C. trachomatis* is the most common organism causing ophthalmia neonatorum. Drops of 0.5% erythromycin or 1% silver nitrate are applied immediately after birth to prevent gonococcal ophthalmia. Ophthalmia neonatorum is a potentially blinding condition, and it can cause serious and potentially systemic manifestations. It requires immediate diagnosis and treatment.

Allergic Conjunctivitis. Ocular allergy encompasses a spectrum of conjunctival conditions usually characterized by itching. The most common of these is seasonal allergic rhinoconjunctivitis or hay fever. Seasonal allergic conjunctivitis is an IgE-mediated hypersensitivity reaction precipitated by small airborne allergens such as pollens. It typically causes bilateral tearing, itching, and redness of the eyes.

The treatment of seasonal allergic rhinoconjunctivitis usually includes allergen avoidance, the use of cold compresses, oral antihistamines, and vasoconstrictor eye drops. Allergic conjunctivitis also has been successfully treated with topical mast cell stabilizers and topical nonsteroidal anti-inflammatory drugs.⁴

The Cornea

The cornea functions as a transparent membrane through which light passes as it moves toward the retina (Fig. 40-2). The cornea also contributes to the refraction (*i.e.*, bending) of light rays and focusing of vision. Three layers of tissue form the cornea: an extremely thin outer epithelial layer, which is continuous with the ocular conjunctiva; a middle layer called the *substantia propria* or

■ **FIGURE 40-2** ■ Transverse section of the eyeball.

stroma; and an inner endothelial layer (Bowman's membrane), which lies next to the aqueous humor of the anterior chamber. The substantia propria is composed of regularly arranged collagen bundles embedded in a mucopolysaccharide matrix. This organization of the collagen fibers makes the substantia propria transparent and is necessary for light transmission. Hydration within a limited range is necessary to maintain the spacing of the collagen fibers and transparency.

The cornea is avascular and obtains its nutrient and oxygen supply by diffusion from blood vessels of the adjacent sclera, from the aqueous humor at its deep surface, and from tears.

Disorders of the Cornea

The corneal epithelium is heavily innervated by sensory neurons. Epithelial injury causes discomfort that ranges from a foreign body sensation and burning of the eyes to severe pain. Because the cornea serves as the window of the eye and refracts light, vision is often blurred. Photophobia may occur as the result of painful contraction of the inflamed iris. Reflex tearing is common.

Trauma that causes abrasions of the cornea can be extremely painful, but if minor, the abrasions usually heal in a few days. The epithelium is an effective barrier to entrance of microorganisms into the cornea. It is capable of regeneration, and small defects heal without scarring. If the stroma is damaged, healing occurs more slowly, and the danger of infection is increased. Injuries to Bowman's membrane and the stromal layer heal with scar formation and permanent opacification. Opacities of the cornea impair the transmission of light. A minor scar can severely distort vision because it disturbs the refractive surface of the eye. **Corneal Edema.** The integrity of the epithelium and the endothelium is necessary to maintain hydration of the cornea within a limited range. Damage to either structure leads to edema and loss of transparency. Among the causes of corneal edema is the prolonged and uninterrupted wearing of hard contact lenses, which can deprive the epithelium of oxygen. The edema disappears spontaneously when the cornea comes in contact with the atmosphere. With corneal edema, the cornea appears dull, uneven, and hazy; visual acuity decreases; and iridescent vision (*i.e.*, rainbows around lights) occurs.

Keratitis. Keratitis refers to inflammation of the cornea. It can be caused by infections, hypersensitivity reactions, ischemia, defects in tearing, trauma, and interruption in sensory innervation, as occurs with local anesthesia. Scar tissue formation caused by keratitis is the leading cause of blindness and impaired vision throughout the world. Most of this vision loss is preventable if the condition is diagnosed early and appropriate treatment is instituted.

Keratitis can be divided into two types: nonulcerative and ulcerative. In nonulcerative or interstitial keratitis, all the layers of the epithelium are affected, but the epithelium remains intact. It is associated with a number of diseases, including syphilis, tuberculosis, and lupus erythematosus. It also may result from a viral infection entering through a small defect in the cornea. Treatment usually is symptomatic.

Ulcerative keratitis is an inflammatory process, in which parts of the epithelium, stroma, or both are destroyed. Causes of ulcerative keratitis include infectious agents such as those causing conjunctivitis (*e.g.*, *Staphylococcus*, *S. pneumoniae*, *Chlamydia*), exposure trauma, and use of extended-wear contact lens. Bacterial keratitis tends to be aggressive and demands immediate care. Exposure trauma may result from deformities of the lid, paralysis of the lid muscles, or severe exophthalmos. Mooren's ulcer is a chronic, painful, indolent ulcer that occurs in the absence of infection. It usually is seen in older persons and may affect both eyes. Although the cause is unknown, an autoimmune origin is suspected.

Herpes simplex virus keratitis is the most common cause of corneal ulceration and most common corneal cause of blindness in the United States.² Most cases are caused by HSV type 1 infections. An exception is neonatal keratitis, which is caused by HSV type 2 acquired during the birth process.8 HSV keratitis occurs as a primary or recurrent infection. Primary ocular herpes usually occurs in young children. It is manifested by vesicular blepharoconjunctivitis, occasionally with corneal involvement. It generally is self-limited, without causing corneal damage. After the initial primary infection, the virus may persist in a quiescent or latent state, remaining in the trigeminal ganglion and possibly in the cornea without causing signs of infection. Recurrent infection may be precipitated by various poorly understood, stress-related factors that reactivate the virus. The first symptoms are usually irritation, photophobia, and tearing. There is often a history of fever blisters, but the corneal lesion may be the only sign of recurrent HSV infection. Because corneal anesthesia occurs early in the course of the infection, symptoms may be minimal and the person may delay seeking medical attention. The most common lesion is a dendritic, or branching ulcer, involving the epithelial layer of the cornea. Epithelial ulcers usually respond to debridement and patching and usually heal without scarring. Topical antiviral agents may be used to increase the rate of healing. Corticosteroid drugs are contraindicated for use in the treatment of epithelial HSV lesions because they increase viral replication.

Stromal HSV keratitis produces increasingly severe corneal ulceration and scarring. It is often associated with an immunologic response to the virus, and topical antiviral agents alone are usually insufficient to control stromal keratitis. Thus, topical corticosteroids may be cautiously used in combination with appropriate antiviral agents. Corneal grafting may be indicated for persons with severe corneal scarring.

Corneal Transplantation. Advances in ophthalmologic surgery permit corneal transplantation using a cadaver cornea. Unlike kidney or heart transplantation procedures, which are associated with considerable risk of rejection of the transplanted organ, the use of cadaver corneas entails minimal danger of rejection because this tissue is not exposed to the vascular and therefore the immunologic defense system. Instead, the success of this type of transplantation operation depends on the prevention of scar tissue formation, which would limit the transparency of the transplanted cornea.

Aqueous Humor and Intraocular Pressure

The aqueous humor, which fills the space between the cornea and the lens, serves a nutritive function for the lens and posterior cornea (Fig. 40-2). The aqueous humor is produced by the ciliary epithelium in the posterior chamber and passes between the anterior surface of the lens and posterior surface of the iris, through the pupil, and into the anterior chamber. It leaves through the iridocorneal angle between the iris and the sclera. Here it filters through the trabecular meshwork and enters the canal of Schlemm for return to the venous circulation.

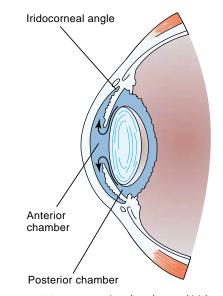
The hydrostatic pressure of the aqueous humor results from a balance of several factors, including the rate of aqueous secretion, resistance to flow through the narrow opening between the iris and the ciliary body at the entrance to the anterior chamber, and resistance to resorption at the trabeculated region of the sclera at the iridocorneal angle. Normally, the rate of aqueous production is equal to the rate of aqueous outflow, and the intraocular pressure is maintained within a normal range of 9 to 21 mm Hg.

Glaucoma

Glaucoma includes a group of conditions that produce an elevation in intraocular pressure. If left untreated, the pressure may increase sufficiently to cause ischemia and degeneration of the optic nerve, leading to progressive blindness. Glaucoma is a major contributor to blindness in the United States.^{1,9} The condition often is asymptomatic, and a significant loss of peripheral vision may occur before medical attention is sought, emphasizing the need for routine screening for early diagnosis and treatment.

Glaucoma commonly is classified as closed-angle (*i.e.*, narrow-angle) or open-angle (*i.e.*, wide-angle) glaucoma, depending on the location of the compromised aqueous humor circulation and resorption. Glaucoma may occur as a congenital or an acquired condition, and it may manifest as a primary or secondary disorder. Primary glaucoma occurs without evidence of pre-existing ocular or systemic disease. Secondary glaucoma can result from inflammatory processes that affect the eye, from tumors, or from the blood cells of traumaproduced hemorrhage that obstruct the outflow of aqueous humor.

Closed-Angle Glaucoma. In closed-angle glaucoma, the anterior chamber is narrow, and outflow becomes impaired when the iris thickens as the result of pupillary dilation (Fig. 40-3).



■ FIGURE 40-3 ■ Narrow anterior chamber and iridocorneal angle in closed-angle (narrow-angle) glaucoma.

As the iris thickens, it restricts the circulation between the base of the iris and the sclera, blocking the circulation between the posterior and anterior chambers and reducing or eliminating access to the angle where aqueous reabsorption occurs.⁹ Approximately 5% to 10% of all cases of glaucoma fall into this category. Closed-angle glaucoma usually occurs as the result of an inherited anatomic defect that causes a shallow anterior chamber. This defect is exaggerated by the anterior displacement of the peripheral iris that occurs in older persons because of the increase in lens size that occurs with aging.

The symptoms of closed-angle glaucoma are related to sudden, intermittent increases in intraocular pressure. These occur after prolonged periods in the dark, emotional upset, and other conditions that cause extensive and prolonged dilation of the pupil. Administration of pharmacologic agents such as atropine that cause pupillary dilation (mydriasis) also can precipitate an acute episode of increased intraocular pressure in persons with the potential for closed-angle glaucoma. Attacks of increased intraocular pressure are manifested by ocular pain and blurred or iridescent vision caused by corneal edema. The pupil may be enlarged and fixed. The symptoms often are spontaneously relieved by sleep and conditions that promote pupillary constriction. With repeated or prolonged attacks, the eye becomes reddened, and edema of the cornea may develop, giving the eye a hazy appearance. A unilateral, often excruciating, headache is common. Nausea and vomiting may occur, causing the headache to be confused with migraine.

Some persons with congenitally narrow anterior chambers never experience symptoms, and others experience symptoms only when they are elderly. Because of the dangers of vision loss, those with narrow anterior chambers should be warned about the significance of blurred vision, halos, and ocular pain. Sometimes, decreased visual acuity and an unreactive pupil may be the only clue to closed-angle glaucoma in the elderly.

The treatment of acute closed-angle glaucoma is primarily surgical. It involves creating an opening between the anterior and posterior chambers with laser or incisional iridectomy to allow aqueous humor to bypass the pupillary block. The anatomic abnormalities responsible for closed-angle glaucoma usually are bilateral, and prophylactic surgery often is performed on the other eye.

Primary Open-Angle Glaucoma. Primary open-angle glaucoma is the most common form of glaucoma. It tends to manifest after 35 years of age, with an incidence of 0.5% to 2% among persons 40 years of age and older.8 The condition is characterized by an abnormal increase in intraocular pressure that occurs in the absence of obstruction at the iridocorneal angle, thus the name open-angle glaucoma. Instead, it usually occurs because of an abnormality of the trabecular meshwork that controls the flow of aqueous humor into the canal of Schlemm. Risk factors for this disorder include an age of 40 years and older, family history of the disorder, diabetes mellitus, and myopia (nearsightedness). In some persons, the use of moderate amounts of topical corticosteroid medications can cause an increase in intraocular pressure. Sensitive persons also may sustain an increase in intraocular pressure with the use of systemic corticosteroid drugs.

Primary open-angle glaucoma usually is asymptomatic and chronic, causing progressive loss of visual field unless it is appropriately treated. The elevation in intraocular pressure in persons with open-angle glaucoma usually is treated pharmacologically or, in cases where pharmacologic treatment fails, by increasing aqueous outflow through a surgically created pathway.

Congenital or Infantile Glaucoma. Congenital glaucoma is caused by a disorder in which the anterior chamber retains its fetal configuration, with aberrant trabecular meshwork extending to the root of the iris, or is covered by a membrane. An X-linked recessive mode of inheritance is common, producing a high incidence among males.⁷ The earliest symptoms are excessive lacrimation and photophobia. Affected infants tend to be fussy, have poor eating habits, and rub their eyes frequently. Diffuse edema of the cornea usually occurs, giving the eye a grayish-white appearance. Chronic elevation of the intraocular pressure before the age of 3 years causes enlargement of the entire globe. Early surgical treatment is necessary to prevent blindness.

The Lens

The lens is a remarkable structure that functions to bring images into focus on the retina. The lens is an avascular, transparent, biconvex body, the posterior side of which is more convex than the anterior side. A thin highly elastic lens capsule, which is attached to the surrounding ciliary body by delicate suspensory radial ligaments called *zonules*, holds the lens in place (see Fig. 40-2). The suspensory ligaments and lens capsule normally are under tension, causing the lens to have a flattened shape for distant vision. Contraction of the muscle fibers of the ciliary body narrows the diameter of the ciliary body, relaxes the fibers of the suspensory ligaments, and allows the lens to relax to a more spherical or convex shape for near vision.

Disorders of Refraction

Refraction can be defined as a bending of light rays as they pass from one transparent medium (such as air) to a second transparent medium with a different density (such as a glass lens). When light rays pass through the center of a lens, their direction is not changed; however, other rays passing peripherally through a lens are bent (Fig. 40-4). The refractive power of a lens usually is described as the distance (in meters) from its surface to the point at which the rays come into focus (*i.e.*, focal length). Usually, this is reported as the reciprocal of this distance (*i.e.*, diopters). For example, a lens that brings an object into focus at 0.5 m has a refractive power of 2 diopters (1.0/0.5 = 2.0). With a fixed-power lens, the closer an object is to the lens, the further behind the lens is its focus point.

In the eye, the major refraction of light begins at the convex corneal surface. Further refraction occurs as light moves from the posterior corneal surface to the aqueous humor, from the aqueous humor to the anterior lens surface, and from the posterior lens surface to the vitreous humor. A perfectly shaped optic globe and cornea result in optimal visual acuity, producing a sharp image in focus at all points on the retinal surface in the posterior part, or fundus, of the eye (see Fig. 40-4). Unfortunately, individual differences in formation and growth of the eyeball and cornea frequently result in inappropriate focal image formation. If the anterior-posterior dimension of the eyeball is too short, the image is focused posterior to (behind)

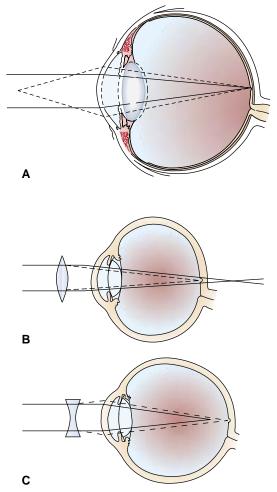


FIGURE 40-4 (A) Accommodation. The *solid lines* represent rays of light from a distant object, and the *dotted lines* represent rays from a near object. The lens is flatter for the former and more convex for the latter. In each case, the rays of light are brought to a focus on the retina. (B) Hyperopia corrected by a biconvex lens, shown by the *dotted lines*. (C) Myopia corrected by a biconcave lens, shown by the *dotted lines*.

the retina. This is called *hyperopia* or *farsightedness*. In such cases, the accommodative changes of the lens can bring distant images into focus, but near images become blurred. Appropriate biconvex lenses correct this type of defect. If the anterior-posterior dimension of the eyeball is too long, the focus point for an infinitely distant target is anterior to the retina. This condition is called *myopia* or *nearsightedness* (see Fig. 40-4). Persons with myopia can see close objects without problems because accommodative changes in their lens bring near objects into focus, but distant objects are blurred. Myopia can be corrected with an appropriate biconcave lens. Radial keratotomy, a form of refractive corneal surgery, can be performed to correct the defect. This surgical procedure involves the use of radial incisions to alter the corneal curvature.

Nonuniform curvature of the refractive medium (*e.g.*, horizontal vs. vertical plane) is called *astigmatism*. Astigmatism usually is the result of a defect in the cornea, but it can result from defects in the lens or the retina.

Disorders of Accommodation

Accommodation is the process whereby a clear image is maintained as gaze is shifted from a far to a near object. Accommodation requires convergence of the eyes, pupillary constriction, and thickening of the lens through contraction of the ciliary muscle.

In near vision, pupillary constriction (*i.e., miosis*) improves the clarity of the retinal image. This must be balanced against the resultant decrease in light intensity reaching the retina. During changes from near to far vision, pupillary dilation partially compensates for the reduced size of the retinal image by increasing the light entering the pupil. A third component of accommodation involves the reflex narrowing of the lid opening during near vision and widening during far vision.

The term *presbyopia* refers to changes in vision that occur because of aging. The lens consists of transparent fibers arranged in concentric layers, of which the external layers are the newest and softest. No loss of lens fibers occurs with aging; instead, additional fibers are added to the outermost portion of the lens. As the lens ages, it thickens, and its fibers become less elastic, so that the range of focus or accommodation is diminished to the point where reading glasses become necessary for near vision.

Cataracts

A cataract is a lens opacity that interferes with the transmission of light to the retina. Cataracts are the most common cause of age-related visual loss in the world; they are found in approximately 50% of those between 65 and 74 years of age and in 70% of those older than 75 years of age.¹

The cause of cataract development is thought to be multifactorial, with different factors being associated with different types of opacities. Several risk factors have been proposed, including the effects of aging, genetic influences, environmental and metabolic influences, drugs (*e.g.*, triparanol, chlorpromazine, corticosteroids), and injury.^{10,11} Metabolically induced cataracts are caused by disorders of carbohydrate metabolism (diabetes) or inborn errors of metabolism. Long-term exposure to sunlight (ultraviolet B radiation) and heavy smoking have been associated with increased risk of cataract formation.¹⁰ In some cases, cataracts occur as a developmental defect (*i.e.*, congenital cataracts) or secondary to trauma or diseases.

With normal aging, the nucleus and the cortex of the lens enlarge as new fibers are formed in the cortical zones of the lens. In the nucleus, the older fibers become more compressed and dehydrated, the lens proteins become more insoluble, and concentrations of calcium, sodium, potassium, and phosphate increase. During the early stages of cataract formation, a yellow pigment and vacuoles accumulate in the lens fibers. The unfolding of protein molecules, cross-linking of sulfhydryl groups, and conversion of soluble to insoluble proteins lead to the loss of lens transparency. The onset is usually gradual, and the only symptoms are increasingly blurred vision and visual distortion.

The manifestations of cataract depend on the extent of opacity and whether the defect is bilateral or unilateral. With the exception of traumatic or congenital cataract, most cataracts are bilateral. Age-related cataracts are characterized by increasingly blurred vision and visual distortion. Vision for far and near objects decreases. Dilation of the pupil in dim light improves vision. Central lens opacities may divide the visual axis and cause an optical defect in which two or more blurred images are seen. In addition to decreased visual acuity, cataracts tend to cause light entering the eye to be scattered, thereby producing glare or the abnormal presence of light in the visual field. On ophthalmoscopic examination, cataracts may appear as a gross opacity filling the pupillary aperture or as an opacity silhouetted against the red background of the fundus.

There is no effective medical treatment for cataract. Use of strong bifocals, magnification, appropriate lighting, and visual aids may be used as the cataract progresses. Surgery is the only treatment for correcting cataract-related vision loss. Surgery usually involves lens extraction and intraocular lens implantation. The cataract lens usually is removed using phacoemulsification techniques. Phacoemulsification involves ultrasonic fragmentation of the lens into fine pieces, which then are aspirated from the eye. Surgery commonly is performed on an outpatient basis and with the use of local anesthesia.

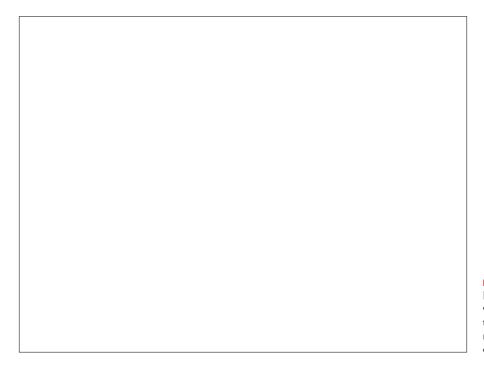
The Retina

The function of the retina is to receive visual images, partially analyze them, and transmit this modified information to the brain. Disorders of the retina and its function include ischemic conditions caused by disorders of the retinal blood supply; disorders of the retinal vessels such as retinopathies that cause hemorrhage and the development of opacities; separation of the pigment and sensory layers of the retina (*i.e.*, retinal detachment); and macular degeneration. Because the retina has no pain fibers, most diseases of the retina are painless and do not cause redness of the eye.

The retina is composed of two layers: an outer pigment (melanin-containing) epithelium and an inner neural layer (Fig. 40-2). The outer pigmented layer, a single-cell-thick lining abuts the choroid, and extends anteriorly to cover the ciliary body and the posterior side of the iris. Its pigmented epithelial cells, like those of the choroid, absorb light and prevent it from scattering. The pigment layer also stores large quantities of vitamin A, which is an important precursor of the photosensitive visual pigments.

The importance of melanin in the pigment layer is well illustrated by its absence in people with a condition called *albinism*. Albinism is a genetic deficiency of tyrosinase, the enzyme needed for the synthesis of melanin by the melanocytes. Affected persons have white hair, pink skin, and light blue eyes. In these persons, excessive light penetrates the unpigmented iris and choroid and is reflected in all directions, so that their photoreceptors are flooded with excess light, and visual acuity is markedly reduced. Excess stimulation of the photoreceptors at normal or high illumination levels is experienced as painful photophobia.

The light-sensitive inner neural retina covers the inner aspect of the eyeball. The neural retina is composed of three layers of neurons: a posterior layer of photoreceptors, a middle layer of bipolar cells, and an inner layer of ganglion cells that communicate with the photoreceptors. Light must pass through the transparent inner layers of the sensory retina before it reaches the photoreceptors. Local currents produced in response to light spread from the photoreceptors to the bipolar neurons and other interneurons and then to the innermost ganglionic cells, where action potentials are generated. The interneurons, which are composed of horizontal and amacrine cells, have cell bodies in the bipolar layer, and they play an important role in modulating retinal function. A superficial marginal layer contains the axons of the ganglion cells as they collect and leave the eye by way of the optic nerve (Fig. 40-5). The optic disk, where the optic nerve exits the eye, is the weak part of the eye because it is not reinforced by the sclera. The optic disk also forms the blind spot, because it is not reinforced by photoreceptors, and light focused on it cannot be seen. People do not notice the blind spot because of a sophisticated visual function



■ FIGURE 40-5 ■ Organization of the human retina. The visual pathway begins with photoreceptors (rods and cones) in the retina. The responses of the photoreceptors are transmitted by the bipolar cells to the ganglion cell layer of the retina.

called "filling in," which the brain uses to deal with missing visual input. Local retinal damage caused by small vascular lesions (*i.e.*, retinal stroke) and other localized pathologies can produce additional blind spots.

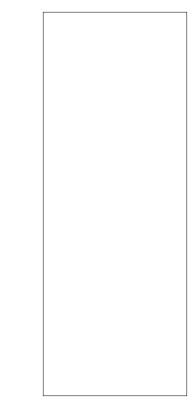
Two types of photoreceptors are present in the retina: rods, capable of black-white discrimination, and cones, capable of color discrimination. Rod-based vision is particularly sensitive to detecting light, especially moving light stimuli, at the expense of clear pattern discrimination. Rod vision is particularly adapted for night and low-level illumination. Dark adaptation is the process by which rod sensitivity increases to the optimum level. This requires approximately 4 hours in total or near-total darkness and involves only rods. Cone receptors, which are selectively sensitive to different wavelengths of light, provide the basis for color vision. Three types of cones, or cone-color systems, respond to the blue, green, and red portions of the visible electromagnetic spectrum. Cones do not have the dark adaptation of rods. Consequently, the dark-adapted eye is a rod receptor eye with only black-gray-white experience (scotopic or night vision). The light-adapted eye (photopic vision) adds the capacity for color discrimination.

Both rods and cones contain chemicals that decompose on exposure to light and, in the process, generate the currents that lead to the action potentials generated by the ganglionic cells. The light-sensitive chemical in the rods is called rhodopsin, and the light-sensitive chemicals in the cones are called cone or color pigments. Both types of photoreceptors are thin, elongated, mitochondria-filled cells with a single, highly modified cilium (see Fig. 40-6). The cilium has a short base, or inner segment, and a highly modified outer segment. The plasma membrane of the outer segment is tightly folded to form membranous disks (rods) or conical shapes (cones) containing visual pigment. Both rhodopsin and color pigment are incorporated into membranes of these disks in the form of transmembrane proteins. These disks are continuously synthesized at the base of the outer segment and shed at the distal end. The discarded membranes are phagocytized by the retinal pigment cells. If this phagocytosis is disrupted, as in retinitis pigmentosa, the sensory retina degenerates.

An area approximately 1.5 mm in diameter near the center of the retina, called the *macula lutea* (*i.e.*, "yellow spot"), is especially adapted for acute and detailed vision. This area is composed entirely of cones. In the central portion of the macula, the *fovea centralis*, the blood vessels and innermost layers are displaced to one side instead of resting on top of the cones (Fig. 40-2). This allows light to pass unimpeded to the cones without passing through several layers of the retina. Many of these cones are connected one-to-one with ganglion cells, an arrangement that favors high acuity.

Blood Supply and Vascular Lesions

The blood supply for the retina is derived from two sources: the choriocapillaris of the choroid and the branches of the central retinal artery. The cones and rods of the outer neural layer receive nutrients from the choriocapillaris, a fine layer of capillaries on the inner surface of the choroids against which the retina is pressed. Because the choriocapillary layer provides the only blood supply for the fovea centralis, detachment of this part of the sensory retina from the pigment epithelium causes irreparable visual loss. The central retinal artery, which is a branch of the ophthalmic, supplies the rest of the retina. A cor-



■ FIGURE 40-6 ■ Retinal rod, showing its component parts and the distribution of its organelles. Its outer segment contains the disks (rods). The connecting structure joined the outer and inner segments. The inner segment contains the mitochondria, the ribosomal endoplasmic reticulum, the free ribosomes, and the Golgi saccules. The synaptic body forms the site where the photo receptors synapse with subsequent nerve cells.

responding system of retinal veins unites to form the central vein of the retina.

Funduscopic examination of the eye with an ophthalmoscope provides an opportunity to examine the retinal blood vessels and other aspects of the retina (Fig. 40-7). Dilating the pupil pharmacologically enables more thorough examination of the retina.

Papilledema. The central retinal artery enters the eye through the *optic papilla* in the center of the optic nerve. The central vein of the retina exits the eye along the same path. The entrance and exit of the central artery and vein of the retina through the tough scleral tissue at the optic papilla can be compromised by any condition causing persistent increased intracranial pressure. The most common of these conditions are cerebral tumors, subdural hematomas, hydrocephalus, and malignant hypertension.

The thin-walled, low-pressure veins are the first to collapse, with the consequent backup and slowing of arterial blood flow. Under these conditions, capillary permeability increases, and leakage of fluid results in edema of the optic papilla, called *papilledema*. The interior surface of the papilla normally is cupshaped and can be evaluated through an ophthalmoscope.

KEY CONCEPTS

DISORDERS OF THE RETINAL BLOOD SUPPLY

- The blood supply for the retina is derived from the central retinal artery, which supplies blood flow for the entire inside of the retina, and from vessels in the choroid, which supply the rods and cones.
- Central retinal occlusion interrupts blood flow to the inner retina and results in unilateral blindness.
- The retinopathies, which are disorders of the retinal vessels, interrupt blood flow to the visual receptors, leading to visual impairment.
- Retinal detachment separates the visual receptors from the choroid, which provides their major blood supply.

With papilledema, sometimes called *choked disk*, the optic cup is distorted by protrusion into the interior of the eye (Fig. 40-8). Because this sign does not occur until the intracranial pressure is significantly elevated, compression damage to the optic nerve fibers may have begun. As a warning sign, papilledema occurs quite late. Unresolved papilledema results in the destruction of the optic nerve axons and blindness.

Retinopathies

The retinopathies involve the small blood vessels of the retina. They involve changes in blood vessel structure, development of microaneurysms, and formation of new fragile vessels (neovascularization). Breakdown of the blood-retinal barrier at the level of the capillary endothelium can develop, allowing fluid and plasma constituents to escape into the surrounding retina. Vessel weakness can lead to rupture of the vessel wall and hemorrhage. **Diabetic Retinopathy.** Diabetic retinopathy is the third leading cause of blindness for all ages in the United States. It ranks first as the cause of newly reported cases of blindness in persons between the ages of 20 and 74 years, and current estimates suggest it is responsible for 12,000 to 24,000 new cases of blindness in the United States each year.¹²

Diabetic retinopathy can be divided into two types: nonproliferative (i.e., background) and proliferative. Background or nonproliferative retinopathy is a progressive disorder of the small retinal vessels. It is characterized by thickening and hyperpermeability of the retinal capillary walls. The capillaries develop tiny outpouching called *microaneurysms*, while the retinal veins become dilated and tortuous. These vessels tend to leak plasma, resulting in localized edema that gives the retina a hazy appearance. Ruptured capillaries cause small intraretinal hemorrhages, and microinfarcts may cause cotton-wool exudates. A sensation of glare (because of the scattering of light) is a common complaint. The most common cause of decreased vision in persons with background retinopathy is macular edema.^{12,13} It represents fluid accumulation in the retina stemming from a breakdown in the bloodretina barrier.

Proliferative diabetic retinopathy represents a more severe retinal change than background retinopathy. It is characterized by formation of new fragile blood vessels (*i.e.*, neovascularization) at the disk and elsewhere in the retina. These vessels grow in front of the retina along the posterior surface of the vitreous or into the vitreous. These new blood vessels threaten vision in two ways. First, because they are abnormal, they tend to bleed easily, leaking blood into the vitreous cavity and decreasing visual acuity. Second, the blood vessels attach firmly to the retinal surface and posterior surface of the vitreous, such that normal movement of the vitreous may exert a pull on the retina, causing retinal detachment and progressive blindness.

Preventing diabetic retinopathy from developing or progressing is considered the best approach to preserving vision.¹⁴ Growing evidence suggests that careful control of blood sugar levels in persons with diabetes mellitus may retard the onset

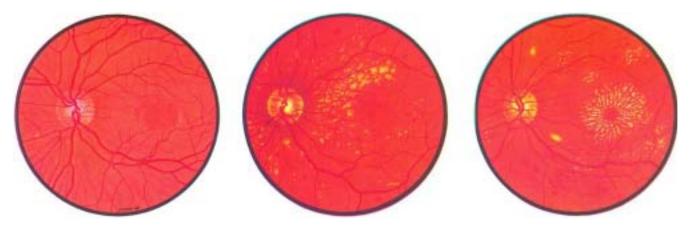
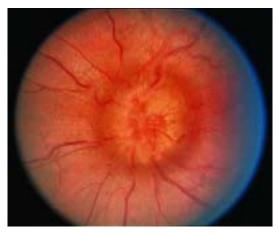


FIGURE 40-7 Fundus of the eye as seen in retinal examination with an ophthalmoscope: (**left**) normal fundus; (**middle**) diabetic retinopathy—combination of microaneurysms, deep hemorrhages, and hard exudates of background retinopathy; (**right**) hypertensive retinopathy with purulent exudates. Some exudates are scattered, while others radiate from the fovea to form a macular star. (Bates B. [1995]. *A guide to physical examination and history taking* (pp. 208, 210). Philadelphia: J.B. Lippincott Company)



■ FIGURE 40-8 ■ Chronic papilledema. The optic nerve head is congested and protrudes anteriorly toward the interior of the eye. It has blurred margins, and vessels within it are poorly seen. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 1558]. Philadelphia: Lippincott Williams & Wilkins)

and progression of retinopathy. Both hypertension and hyperlipidemia are thought to increase the risk of diabetic retinopathy in persons with diabetes.¹³

Because early proliferative diabetic retinopathy is likely to be asymptomatic, it must be identified early, before bleeding occurs and leads to fibrosis and retinal attachment. The American Diabetes Association, American College of Physicians, and American Academy of Ophthalmology have developed screening guidelines for diabetic retinopathy.^{13,15}

Photocoagulation using an argon laser provides the major direct treatment modality for diabetic retinopathy.¹³ Treatment strategies include laser photocoagulation applied directly to leaking microaneurysms and grid photocoagulation with a checkerboard pattern of laser burns applied to diffuse areas of leakage and thickening.¹² Vitrectomy has proved effective in removing vitreous hemorrhage and severing vitreoretinal membranes that develop.

Hypertensive Retinopathy. Long-standing systemic hypertension results in the compensatory thickening of arteriolar walls, which effectively reduces capillary perfusion pressure. Ordinarily, a retinal blood vessel is transparent and seen as a red line; in venules, the red cells resemble a string of boxcars. On ophthalmoscopy, arteries in persons with long-standing hypertension appear paler than veins because they have thicker walls. The thickened arterioles in chronic hypertension become opaque and have a copper-wiring appearance. Edema, microaneurysms, intraretinal hemorrhages, exudates, and cottonwool spots all are observed.⁶ Malignant hypertension involves swelling of the optic disk as a result of the local edema produced by escaped fluid. If the condition is permitted to progress long enough, serious visual deficits result.

Protective thickening of arteriolar walls cannot occur with sudden increases in blood pressure. Therefore, hemorrhage is likely to occur. Trauma to the optic globe or the head, sudden high blood pressure in eclampsia, and some types of renal disease characteristically are accompanied by edema of the retina and optic disk and by an increased likelihood of hemorrhage. Atherosclerosis of Retinal Vessels. In atherosclerosis, the lumen of the arterioles becomes narrowed. As a result, the retinal arteries become tortuous and narrowed. At sites where the arteries cross and compress veins, the red cell column of the vein appears distended. Exudate accumulates on arteriolar walls as "fluffy," white plaques (cotton wool patches). These patches

Chapter 40: Alterations in Special Sensory Function

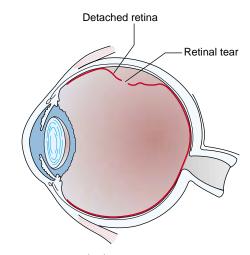
are damaged axons that on cross-section resemble "crystalloid bodies." Deep and superficial hemorrhages are common. Atheromatous plaques of the central artery are associated with increased danger of stasis, thrombi of the central veins, and occlusion.

Retinal Detachment

Retinal detachment involves the separation of the sensory retina from the pigment epithelium (Fig. 40-9). It occurs when traction on the inner sensory layer or a tear in this layer allows fluid, usually vitreous, to accumulate between the two layers. Retinal detachment that results from breaks in the sensory layer of the retina is called *rhegmatogenous detachment* (*rhegma* in Greek, meaning "rent" or "hole"). The vitreous normally is adherent to the retina at the optic disk, macula, and periphery of the retina. When the vitreous shrinks, it separates from the retina at the posterior pole of the eye (posterior vitreous detachment), but at the periphery, the vitreous pulls on the attached retina, which can lead to tearing of the retina. Vitreous fluid can enter the tear and contribute to further separation of the retina from its overlying pigment layer.

Persons with high grades of myopia may have abnormalities in the peripheral retina that predispose to sudden detachment. Intraocular surgery such as cataract extraction may produce traction on the peripheral retina that causes eventual detachment months or even years after surgery.¹² Detachment may result from exudates that separate the two retinal layers. Exudative detachment may be caused by intraocular inflammation, intraocular tumors, or certain systemic diseases. Inflammatory processes include posterior scleritis, uveitis, or parasitic invasion. Retinal detachment also can follow trauma immediately or at some later time.

Detachment of the neural retina from the retinal pigment layer separates the receptors from their major blood supply, the choroid. If retinal detachment continues for some time,



■ FIGURE 40-9 ■ Detached retina.

permanent destruction and blindness of that part of the retina occur. The primary symptom of retinal detachment is loss of vision. Sometimes, flashing lights or sparks, followed by small floaters or spots in the field of vision, occur as the retina pulls away from the posterior pole of the eye. There is no pain. As detachment progresses, the person perceives a dark curtain progressing across the visual field. Because the process begins in the periphery and spreads circumferentially and posteriorly, initial visual disturbances may involve only one quadrant of the visual field. Large peripheral detachments may occur without involvement of the macula, so that visual acuity remains unaffected. However, the tendency is for detachments to enlarge until the entire retina is detached.

Treatment is aimed at closing retinal tears and reattaching the retina. Rhegmatogenous detachment usually requires surgical treatment.

Macular Degeneration

Macular degeneration is characterized by destructive changes of the yellow-pigmented area surrounding the central fovea resulting from vascular disorders. Age-related macular degeneration is the leading cause of blindness among persons older than 75 years of age and of newly reported cases of blindness among those older than 65 years of age.¹ The cause of macular degeneration is unknown, although nutritional, hemodynamic, degenerative, and phototoxic factors are under investigation.

Macular degeneration is characterized by the loss of central vision, usually in both eyes. Age-related macular degeneration can be classified into early and late stages.¹ The early stage is associated with minimal visual impairment. There are pigmentary abnormalities and pale yellow spots that may occur individually or in groups throughout the macula, called *drusen*. Only eyes with large drusen are at risk for late-stage age-related macular degeneration. Persons with late-stage disease often find it difficult to see at long distances (*e.g.*, in driving), do close work (*e.g.*, reading), see faces clearly, or distinguish colors. However, the person may not be severely incapacitated because the peripheral retinal function usually remains intact.

There are two types of age-related macular degeneration: an atrophic nonexudative or "dry" form and an exudative or "wet" form.¹⁶ The atrophic form is characterized by a gradual, progressive bilateral vision loss from atrophy and degeneration of the rod and cone photoreceptors. It does not involve leakage of blood or serum; thus, it is called *dry age-related macular degen*eration. The exudative form is characterized by the formation of a choroidal neovascular membrane that separates the pigmented epithelium from the neuroretina. These new blood vessels have weaker walls than normal and are prone to leakage; thus, this condition is called wet age-related macular degeneration. The leakage of serous or hemorrhagic fluid into the subretinal space causes separation of the pigmented epithelium from the neurosensory retina. With time, the subretinal hemorrhages organize to form scar tissue. When this happens, retinal tissue death and loss of all visual function in the corresponding macular area occurs.

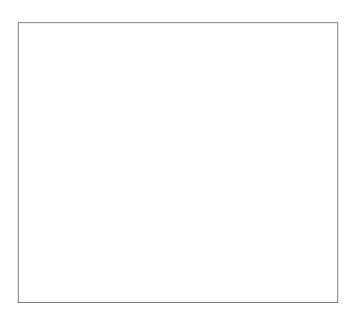
Although there is no treatment for the dry form of macular degeneration, argon laser photocoagulation may be useful in treating the neovascularization that occurs with the wet form.¹⁷ Another method that has been used to halt neovascularization is photodynamic therapy. It is a nonthermal process leading to localized production of reactive oxygen species that mediate

cellular, vascular, and immunologic injury and destruction of new blood vessels.

Neural Pathways. Full visual function requires the normally developed brain-related functions of photoreception and the pupillary reflex. These functions depend on the integrity of all visual pathways, including retinal circuitry and the pathway from the optic nerve to the visual cortex and other visual regions of the brain and brain stem.

Visual information is carried to the brain by axons of the retinal ganglion cells, which form the optic nerve. The two optic nerves meet and fuse in the optic chiasm, beyond which they are continued as the optic tracts (Fig. 40-10). In the optic chiasm, axons from the nasal retina of each eye cross to the opposite side and join with the axons of the temporal retina of the contralateral eye to form the optic tracts. One optic tract contains fibers from both eyes that transmit information from the same visual field.

Fibers of the optic tracts synapse in the lateral geniculate nucleus (LGN) of the thalamus. Axons from these neurons in the LGN form the optic radiations to the primary visual cortex in the occipital lobe. The pattern of information transmission established in the optic tract is retained in the optic radiations. For example, the axons from the right visual field, represented by the nasal retina of the right eye and the temporal retina of the left eye, are united at the chiasm. They continue through the left optic tract and left optic radiation to the left visual cortex, where visual experience is first perceived. The left primary visual cortex receives two representations of the right visual



■ FIGURE 40-10 ■ Diagram of optic pathways. The red lines indicate right visual field and the blue lines the left visual field. Note the crossing of fibers from the medial half of each retina at the optic chiasm. Lesion 1 (right optic nerve) produces unilateral blindness. Lesion 2 (optic chiasm) may involve only those fibers that originate in the nasal half of each retina and cross to the opposite side in the optic chiasm, visual loss involves the temporal half of each field (bitemporal hemianopia). Lesion 3 (right optic tract) interrupts fibers (and vision) originating on the same side of both eyes (homonymous) with loss of vision from half of each field (hemianopia).

field. Physical separation of information from the left and right visual fields is maintained in the visual cortex. Interaction between these disparate representations occurs and provides the basis for the sensation of depth in the near visual field.

The primary visual cortex (area 17) surrounds the calcarine fissure, which lies in the occipital lobe. It is at this level that visual sensation is first experienced (Fig. 40-11). Immediately surrounding area 17 are the visual association cortices (areas 18 and 19) and several other association cortices. These association cortices, with their thalamic nuclei, must be functional for added meaningfulness of visual perception. This higher-order aspect of the visual experience depends on previous learning.

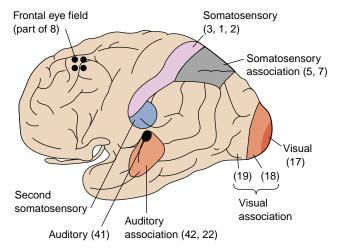
Disorders of the Optic Pathways

Among the disorders that can interrupt the visual pathway are trauma, tumors, and vascular lesions. Trauma and tumors can produce direct injury or impinge on the optic pathways. Vascular insufficiency in any one of the arterial systems of the retina or visual pathways can seriously affect vision. For example, normal visual function depends on adequacy of blood flow in the ophthalmic artery and its branches; the central artery of the retina; the anterior and middle cerebral arteries, which supply the intracranial optic nerve, chiasm, and optic tracts; and the posterior cerebral artery, which supplies the LGN, optic radiation, and visual cortex. The adequacy of posterior cerebral artery function depends on that of the vertebral and basilar arteries that supply the brain stem.

Visual Field Defects

The visual field refers to the area that is visible during fixation of vision in one direction. As with a camera, the simple lens system of the eye inverts the image of the external world on each retina (Fig. 40-10). In addition, the right and left sides of the visual field also are reversed. The right binocular visual field (the nasal half of the right eye and the temporal half of the left eye) is seen by the left retinal halves of each eye. Likewise, the left binocular field is seen by the right retinal halves of each eye.

Most of the visual field is *binocular*, or seen by both eyes. This binocular field is subdivided into central and peripheral portions. Central portions of the retina provide high visual acu-



■ FIGURE 40-11 ■ Lateral view of the cortex illustrating the location of the visual, visual association, auditory, and auditory association areas.

ity and correspond to the field focused on the central fovea. The peripheral and surrounding portion provides the capacity to detect objects, particularly moving objects. Beyond the visual field shared by both eyes, the left lateral periphery of the visual field is seen exclusively by the left nasal retina, and the right peripheral field by the right nasal retina.

Visual field defects result from damage to the visual pathways or the visual cortex (Fig. 40-10). The testing of visual fields of each eye and of the two eyes together are useful in localizing lesions affecting the system. Perimetry or visual field testing, in which the visual field of each eye is measured and plotted in an arc, is used to identify defects and determine the location of lesions.

Blindness in one eye is called anopia. If half of the visual field for one eye is lost, the defect is called *hemianopia*; loss of a quarter field is called quadrantanopia. Enlarging pituitary tumors can produce longitudinal damage through the optic chiasm with loss of the medial fibers of the optic nerve representing both nasal retinas and both temporal visual half-fields. Loss of the temporal or peripheral visual fields on both sides results in a narrow binocular field, commonly called tunnel vision. The loss of different half-fields in the two eyes is called a heteronymous loss, and the abnormality is called heteronymous hemianopia. Destruction of one or both lateral halves of the chiasm is common with multiple aneurysms of the circle of Willis (see Chapter 37). In this condition, the function of one or both temporal retinas is lost, and the nasal fields of one or both eves are lost. The loss of the temporal fields (nasal retina) of both eyes is called bitemporal heteronymous anopia. With both eyes open, the person with bilateral defects still has the full binocular visual field.

Loss of the optic tract, LGN, full optic radiation, or complete visual cortex on one side results in loss of the corresponding visual half-fields in each eye. Homonymous means "the same for both eyes. In left-side lesions, the right visual field is lost for each eve and is called *complete right homonymous hemianopia*. Partial injury to the left optic tract, LGN, or optic radiation can result in the loss of a quarter of the visual field in both eyes. This is called homonymous quadrantanopia, and depending on the lesion, it can involve the upper (superior) or lower (inferior) fields. The LGN, optic radiation, and visual cortex all receive their major blood supply from the posterior cerebral artery; unilateral occlusion of this artery results in complete loss of the opposite field (*i.e.*, homonymous hemianopia). Bilateral occlusion of these arteries results in total cortical blindness.

The Extraocular Eye Muscles and Disorders of Eye Movements

For complete function of the eyes, it is necessary that the two eves point toward the same fixation point and that the retinal and central nervous system (CNS) visual acuity mechanisms function. Despite slight variations in the view of the external world for each eye, it is important that these two images become fused, which is a forebrain function. Binocular fusion is controlled by ocular reflex mechanisms that adjust the orientation of each eye to produce a single image. If these reflexes fail, diplopia or double vision occurs.

Binocular vision depends on three pairs of extraocular muscles-the medial and lateral recti, the superior and inferior recti, and the superior and inferior obliques (Fig. 40-12). Each of the three sets of muscles in each eye is reciprocally innervated **FIGURE 40-12** Extraocular eye muscles of the right eye.

so that one muscle relaxes when the other contracts. Reciprocal contraction of the medial and lateral recti moves the eye from side to side (adduction and abduction); the superior and inferior recti move the eye up and down (elevation and depression). The oblique muscles rotate (intorsion and extorsion) the eye around its optic axis. A seventh muscle, the levator palpebrae superioris, elevates the upper lid.

The extraocular muscles are innervated by three cranial nerves. The trochlear nerve (CN IV) innervates the superior oblique, the abducens nerve (CN VI) innervates the lateral rectus, and the oculomotor nerve (CN III) innervates the remaining four muscles. Table 40-1 describes the function and innervation of the extraocular muscles.

Normal vision depends on the coordinated action of the entire visual system and a number of central control systems. It is through these mechanisms that an object is simultaneously imaged on the fovea of both eyes and perceived as a single image. Strabismus and amblyopia are two disorders that affect this highly integrated system. Although strabismus may develop in later life, it is seen most commonly in children.

Strabismus

Strabismus, or squint, refers to any abnormality of eye coordination or alignment that results in loss of binocular vision (Fig. 40-13). When images from the same spots in visual space

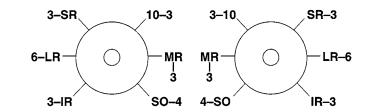
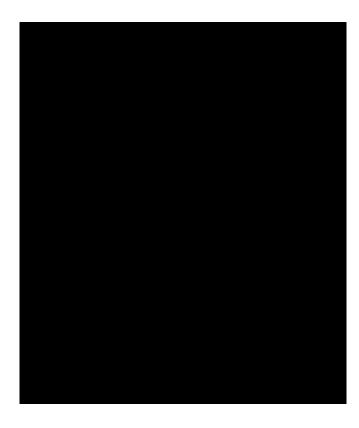


TABLE 40-1	Eye in Primary Position: Extrinsic Ocular Muscle Actions			
Muscle*	Innervation	Primary	Secondary	Tertiary
MR: medial rectus	III	Adduction		
LR: lateral rectus	VI	Abduction		
SR: superior rectus		Elevation	Intorsion	Adduction
IR: inferior rectus	III	Depression	Extorsion	Adduction
SO: superior oblig	ue IV	Intorsion	Depression	Abduction
IR: inferior oblique	e III	Extorsion	Elevation	Abduction

*In the schema of the functional roles of the six extraocular muscles, the major directional force applied by each muscle is indicated on the top. These muscles are arranged in functionally opposing pairs per eye and in parallel opposing pairs for conjugate movements of the two eyes. The numbers associated with each muscle indicate the cranial nerve innervation: 3, oculomotor (III) cranial nerve; 4, trochlear (IV) cranial nerve; 6, abducens (VI) cranial nerve.

Chapter 40: Alterations in Special Sensory Function

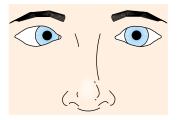


do not fall on corresponding points of the two retinas, diplopia, or double vision, occurs.

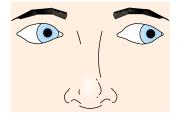
In standard terminology, the disorders of eve movement are described according to the direction of movement. Esotropia refers to medial deviation, exotropia refers to lateral deviation, hypertropia refers to upward deviation, hypotropia refers to downward deviation, and cyclotropia refers to torsional deviation. The term concomitance refers to equal deviation in all directions of gaze. A nonconcomitant strabismus is one that varies with the direction of gaze. Strabismus may be divided into paralytic (nonconcomitant) forms, in which there is weakness or paralysis of one or more of the extraocular muscles, and nonparalytic (concomitant) forms, in which there is no primary muscle impairment. Strabismus is called intermittent, or *periodic*, when there are periods in which the eyes are parallel. It is monocular when the same eye always deviates and the fellow eye fixates. Figure 40-14 illustrates abnormalities in eye movement associated with esotropia and exotropia.

Strabismus affects approximately 4% of children younger than 6 years of age.¹⁸ Because 30% to 50% of these children sustain permanent secondary loss of vision, or amblyopia, if the condition is left untreated, early diagnosis and treatment are essential.19

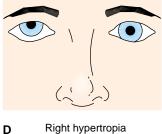
Paralytic Strabismus. Paralytic strabismus results from paresis (*i.e.*, weakness) or plegia (*i.e.*, paralysis) of one or more of the extraocular muscles. When the normal eye fixates, the affected



Primary position: right esotropia



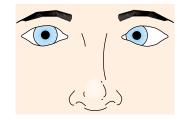
В Left gaze: no deviation



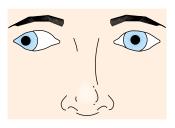
sociated with paralysis of the right lateral rectus muscle: (A) primary position (looking straight ahead) of the eyes; (**B**) left gaze with no deviation; and (**C**) right gaze with left esotropia; (D) primary position of the eyes with weakness of the right inferior rectus and right hypertropia; and (E) primary position of the eyes with weakness of the right medial rectus and right exotropia.

■ FIGURE 40-14 ■ Paralytic strabismus as-

Right hypertropia



Right gaze: left esotropia С



Right exotropia

Ε

eye is in the position of primary deviation. In the case of esotropia, there is weakness of one of the lateral rectus muscles, usually the result of weakness of the abducens nerve (CN VI). When the affected eye fixates, the unaffected eye is in a position of secondary deviation. The secondary deviation of the unaffected eye is greater than the primary deviation of the affected eye. This is because the affected eye requires an excess of innervational impulse to maintain fixation; the excess impulses also are distributed to the unaffected eye, causing overaction of its muscles.²

Paralytic strabismus is uncommon in children but accounts for nearly all cases of adult strabismus; it can be caused by a number of conditions. Paralytic strabismus is seen most commonly in adults who have had cerebral vascular accidents and also may occur as the first sign of a tumor or inflammatory condition involving the CNS. One type of muscular dystrophy exerts its effects on the extraocular muscles. Initially, eye movements in all directions are weak, with later progression to bilateral optic immobility. Weakness of eye movement and lid elevation often is the first evidence of myasthenia gravis. The pathway of the oculomotor (CN III), trochlear (CN IV), and abducens (CN VI) nerves through the cavernous sinus and the back of the orbit make them vulnerable to basal skull fracture and tumors of the cavernous sinus (e.g., cavernous sinus syndrome) or orbit (e.g., orbital syndrome).²⁰ In infants, paralytic strabismus can be caused by birth injuries affecting the extraocular muscles or the cranial nerves supplying these muscles. It also can result from congenital anomalies of the muscles. In general, paralytic strabismus in an adult with previously normal binocular vision causes diplopia. This does not occur in persons who have never developed binocular vision.

Nonparalytic Strabismus. In nonparalytic strabismus, there is no extraocular muscle weakness or paralysis, and the angle of deviation is always the same in all fields of gaze. With persistent deviation, secondary abnormalities may develop because of overaction or underaction of the muscles in some fields of gaze. Nonparalytic esotropia is the most common type of strabismus. The disorder may be accommodative, nonaccommodative, or a combination of the two. Accommodative strabismus is caused by disorders such as uncorrected hyperopia, in which the esotropia occurs with accommodation. The onset of this type of esotropia characteristically occurs between 18 months and 4 years of age because accommodation is not well developed until that time. The disorder most often is monocular but may be alternating. Approximately 50% of the cases of esotropia are accommodative. The causes of nonaccommodative strabismus are obscure. The disorder may be related to faulty muscle insertion, fascial abnormalities, or faulty innervation. There is evidence that idiopathic strabismus may have a genetic basis; siblings may have similar disorders.

Diagnosis and Treatment. All infants and children should be examined for visual alignment. Alignment of the visual axis occurs in the first 3 months of life. All infants should have consistent, synchronized eye movement by 5 to 6 months of age.²¹ Infants who have reached this age and whose eyes are not aligned at all times during waking hours should be examined by a qualified practitioner.

Treatment of strabismus is directed toward the development of normal visual acuity, correction of the deviation, and superimposition of the retinal images to provide binocular vision. Nonsurgical and surgical methods can be used. In children, early treatment is important; the ideal age to begin is 6 months. Nonsurgical treatment includes occlusive patching, pleoptics (*i.e.*, eye exercises), and prism glasses. Because prolonged occlusive patching leads to loss of useful vision in the covered eye, patching is alternated between the affected and unaffected eye. This improves the vision in the affected eye without sacrificing vision in the unaffected eye. Prism glasses compensate for an abnormal alignment of an optic globe. Occasionally, longacting miotics in low doses are used to cause pharmacologic accommodation in place of or in combination with corrective lenses.¹⁸ Surgical procedures may be used to strengthen or weaken a muscle by altering its length or attachment site.

Amblyopia

Amblyopia describes a condition of diminished vision (uncorrectable by lenses) in which no detectable organic lesion of the eye is present.²² This condition sometimes is referred to as *lazy eye*. Types of amblyopia include deprivation occlusion, strabismus, refractive, and organic amblyopia. It is caused by visual deprivation (*e.g.*, cataracts, severe ptosis) or abnormal binocular interactions (*e.g.*, strabismus, anisometropia) during visual immaturity. Normal development of the thalamic and cortical circuitry necessary for binocular visual perception requires simultaneous binocular use of each fovea during a critical period early in life (0 to 5 years of age).

In conditions causing abnormal binocular interactions, one image is suppressed to provide clearer vision. In esotropia, vision of the deviated eye is suppressed to prevent diplopia. A similar situation exists in anisometropia, in which the refractive indexes of the two eyes are different. Although the eyes are correctly aligned, they are unable to focus together, and the image of one eye is suppressed.

The reversibility of amblyopia depends on the maturity of the visual system at the time of onset and the duration of the abnormal experience. If esotropia is involved, some persons alternate eyes and do not experience diplopia. With late adolescent or adult onset, this habit pattern must be unlearned after correction.

Peripheral vision is less affected than central foveal vision in amblyopia. Suppression becomes more evident with high illumination and high contrast. It is as if the affected eye did not possess central vision and the person learns to fixate with the nonfoveal retina. If bilateral congenital blindness or near blindness (*e.g.*, from cataracts) occurs and remains uncorrected during infancy and early childhood, the person remains without pattern vision and has only overall field brightness and color discrimination. This is essentially bilateral amblyopia.

The treatment of children with the potential for development of amblyopia must be instituted well before the age of 6 years to avoid the suppression phenomenon. Severe refractive errors should be corrected. In strabismus, alternately blocking vision in one eye and then the other forces the child to use both eyes for form discrimination. The duration of occlusion of vision in the good eye must be short (2 to 5 hours per day) and closely monitored, or deprivation amblyopia can develop in the good eye as well.

In infants with congenital unilateral cataracts that are dense, central, and larger than 2 mm in diameter, treatment should be started before 2 months of age.² Although bilateral cataracts

Chapter 40: Alterations in Special Sensory Function 763

may require less urgent management, surgical treatment should be done as soon as possible to permit normal development of vision. The surgery for each eye is done separately, with as short an interval as possible between the surgery on the two eyes.

In summary, the optic globe, or eyeball, is a nearly spherical structure protected posteriorly by the bony structures of the orbit and anteriorly by the eyelids. A conjunctiva lines the inner surface of the eyelids and covers the optic globe to the junction of the cornea and sclera. Conjunctivitis, also called *red eye* or *pink eye*, may result from bacterial or viral infection, allergic reactions, or the injurious effects of chemical agents, physical agents, or radiant energy. Keratitis, or inflammation of the cornea, can be caused by infections, hypersensitivity reactions, ischemia, trauma, defects in tearing, or trauma. Trauma or disease that involves the stromal layer of the cornea heals with scar formation and permanent opacification. These opacities interfere with the transmission of light and may impair vision.

Interiorly, the eye is divided into a smaller, fluid-filled anterior cavity and a larger, vitreous-filled posterior segment. The anterior segment of the eye is divided into an anterior and posterior chamber, separated by the pupil and closely adjacent lens. Glaucoma, which is one of the leading causes of blindness in the United States, is characterized by an increase in intraocular pressure resulting from the overproduction or impeded outflow of aqueous humor from the anterior chamber of the eye. Closed-angle glaucoma is caused by a narrow anterior chamber and blockage of the outflow channels at the angle formed by the iris and the cornea. Open-angle glaucoma is caused by an imbalance between aqueous humor production and outflow.

Refraction refers to the ability to focus an object on the retina. In hyperopia, or farsightedness, the image falls behind the retina. In myopia, or nearsightedness, the image falls in front of the retina. Accommodation is the process by which a clear image is maintained as the gaze is shifted from afar to a near object. Presbyopia is a change in the lens that occurs because of aging such that the lens becomes thicker and less able to change shape and accommodate for near vision. A cataract is characterized by increased lens opacity. It can occur as the result of congenital influences, metabolic disturbances, infection, injury, and aging.

The retina covers the inner aspect of the posterior two thirds of the eyeball and is continuous with the optic nerve. It contains the photoreceptors for vision: the rods, for black and white discrimination, and the cones, for color vision. Disorders of retinal vessels can result from a number of local and systemic disorders, including diabetes mellitus and hypertension. They cause vision loss through changes that result in hemorrhage, the production of opacities, and the separation of the pigment epithelium and sensory retina. Retinal detachment involves separation of the sensory receptors from their blood supply; it causes blindness unless reattachment is accomplished promptly. Macular degeneration, which is a leading cause of blindness in the elderly, is characterized by loss of central vision resulting from destructive changes in the central fovea.

Visual information is carried to the brain by axons of the retinal ganglion cells that form the optic nerve. The two optic

nerves meet and cross at optic chiasm, with the axons from each nasal retina joining the uncrossed fibers of the temporal retina of the opposite eye in the optic tract. The fibers in the optic tract pass to the LGN in the thalamus and then to the primary visual cortex, which is located in the occipital lobe. Damage to the visual pathways leads to visual field defects that can be identified through visual field testing or perimetry.

Eve movement, which is controlled by the extraocular muscles, provides for alignment of the eyes and binocular vision. Strabismus refers to abnormalities in the coordination of eye movements with loss of binocular eye alignment. This inability to focus a visual image on corresponding parts of the two retinas results in diplopia. Paralytic strabismus is caused by weakness or paralysis of the extraocular muscles. Nonparalytic strabismus results from the inappropriate length or insertion of the extraocular muscles or from accommodation disorders. The neural pathways for vision develop during infancy. Amblyopia (i.e., lazy eye) is a condition of diminished vision that cannot be corrected by lenses and in which no detectable organic lesion in the eye can be observed. It results from inadequately developed CNS circuitry because of visual deprivation (e.g., cataracts) or abnormal binocular interactions (e.g., strabismus, anisometropia) during the period of visual immaturity.

THE EAR AND DISORDERS OF AUDITORY FUNCTION

The ears are paired organs consisting of an external and middle ear, which function in capturing, transmitting, and amplifying sound, and an inner ear that contains the receptive organs that are stimulated by sound waves (*i.e.*, hearing) or head position and movement (*i.e.*, vestibular function). Otitis media, or inflammation of the middle ear, is a common disorder of childhood. Hearing loss is one of the most common disabilities experienced by persons in the United States, particularly among the elderly.

External Ear

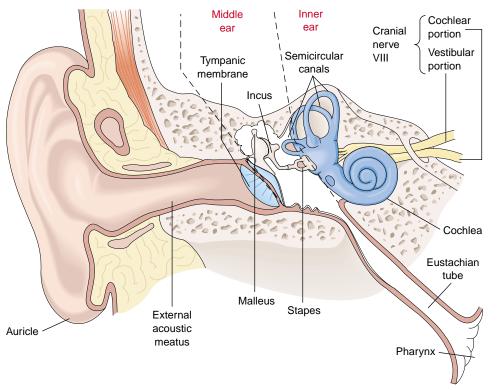
The external ear is a funnel-shaped structure that conducts sound waves to the tympanic membrane. It consists of the auricle, the external acoustic meatus, and the lateral surface of the tympanic membrane (see Fig. 40-15). A thin layer of skin containing fine hairs, sebaceous glands, and ceruminous glands lines the ear canal. Ceruminous glands secrete cerumen, or earwax, which has certain antimicrobial properties and is thought to serve a protective function.

Disorders of the External Ear

The function of the external ear is disturbed when sound transmission is obstructed by impacted cerumen, inflammation (*i.e.*, otitis externa), or drainage from the external ear (otorrhea).

Impacted Cerumen. Although the ear normally is self-cleaning, the cerumen can accumulate and narrow the canal. Impacted cerumen is a common cause of reversible hearing loss.²³ It usually produces no symptoms until the canal becomes completely

764 Unit Ten: Alterations in the Nervous System



■ **FIGURE 40-15** ■ External, middle, and internal subdivisions of the ear.

occluded, at which point the person experiences a feeling of fullness, loss of hearing, tinnitus (*i.e.*, ringing in the ears), or coughing because of vagal stimulation.

In most cases, cerumen can be removed by gentle irrigation using a bulb syringe and warm tap water. Warm water is used to avoid inducing a feeling of disequilibrium caused by the vestibular caloric response. Alternatively, health care professionals may remove cerumen using an otoscope and a wire loop or blunt cerumen curette.

Otitis Externa. Otitis externa is an inflammation of the external ear that can vary in severity from a mild eczematoid dermatitis to severe cellulitis. It can be caused by infectious agents, irritation (e.g., wearing earphones), or allergic reactions. Predisposing factors include moisture in the ear canal after swimming (*i.e.*, swimmer's ear) or bathing and trauma resulting from scratching or attempts to clean the ear. Most infections are caused by gram-negative bacteria (e.g., Pseudomonas, Proteus) or fungi that grow in the presence of excess moisture.²⁴ Otitis externa commonly occurs in the summer and is manifested by itching, redness, tenderness, and narrowing of the ear canal because of swelling. Inflammation of the auricle and external acoustic meatus makes movement of the ear painful. There may be watery or purulent drainage and intermittent hearing loss. Treatment usually includes the use of ear drops containing an appropriate antimicrobial or antifungal agent in combination with a corticosteroid to reduce inflammation.

The Middle Ear and Eustachian Tube

The middle ear is a small air-filled cavity located with the petrous (stony) portion of the middle ear. Its lateral wall is formed by the tympanic membrane and its medial wall by the

bone dividing the middle and inner ear (see Fig. 40-15). Posteriorly, the middle ear is connected with small air pockets in the temporal bone called *mastoid air spaces* or *cells*.

Three tiny bones, the auditory ossicles, are suspended from the roof of the middle ear cavity and connect the tympanic membrane with the oval window (see Fig. 40-15). They are connected by synovial joints and are covered with the epithelial lining of the cavity. The malleus ("hammer") has its handle firmly fixed to the upper portion of the tympanic membrane. The head of the malleus articulates with the incus ("anvil"), which articulates with the *stapes* ("stirrup"), which is inserted and sealed into the oval window by an annular ligament. Arrangement of the ear ossicles is such that their lever movements transmit vibrations from the tympanic membrane to the oval window and from there to the fluid in the inner ear. Two tissue-covered openings in the medial wall, the oval and the round windows, provide for the transmission of sound waves between the air-filled middle ear and the fluid-filled inner ear. It is the piston-like action of the stapes footplate that sets up compression waves in the inner ear fluid.

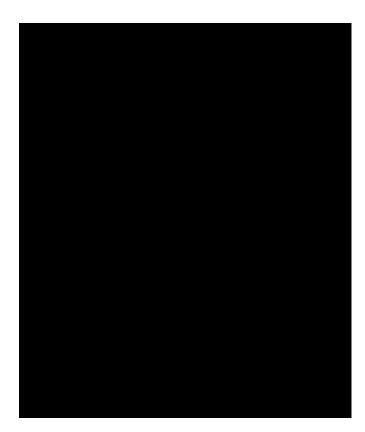
The middle ear is connected to the nasopharynx by the eustachian or auditory tube, which is located in a gap in the bone between the anterior and medial walls of the middle ear (Fig. 40-16). The eustachian tube is lined with a mucous membrane that is continuous with the pharynx and the mastoid air cells.²⁵ The nasopharyngeal entrance to the eustachian tube, which usually is closed, is opened by the action of the *tensor veli palatini muscles* (Fig. 40-17). Opening of the eustachian tube, which normally occurs with swallowing and yawning reflexes, provides the mechanism for equalizing the pressure of the middle ear with that of the atmosphere. This equalization ensures that the pressures on both sides of the tympanic membrane are the same, so that sound transmission is not reduced and rup-



ture does not result from sudden changes in external pressure, as occurs during plane travel.

Disorders of the Eustachian Tube

Abnormalities in eustachian tube function are important factors in the pathogenesis of middle ear infections. There are two important types of eustachian tube dysfunction: abnormal patency and obstruction. The *abnormally patent tube* does not close or does not close completely. In infants and children with an abnormally patent tube, air and secretions often are pumped into the eustachian tube during crying and nose blowing.



Obstruction can be functional or mechanical (see Fig. 40-17). *Functional obstruction* results from the persistent collapse of the eustachian tube due to a lack of tubal stiffness or poor function of the tensor veli palatini muscle that controls the opening of the eustachian tube. It is common in infants and young children because the amount and stiffness of the cartilage supporting the eustachian tube are less than in older children and adults. Changes in the craniofacial base also render the tensor

KEY CONCEPTS

DISORDERS OF THE MIDDLE EAR

- The middle ear is a small air-filled compartment in the temporal bone. It is separated from the outer ear by the tympanic membrane, contains tiny bony ossicles that aid in the amplification and transmission of sound to the inner ear, and is ventilated by the eustachian tube, which is connected to the nasopharynx.
- The eustachian tube, which is lined with a mucous membrane that is continuous with the nasopharynx, provides a passageway for pathogens to enter the middle ear.
- Otitis media (OM) refers to inflammation of the middle ear, usually associated with an acute infection (acute OM) or an accumulation of fluid (OME). It commonly is associated with disorders of eustachian tube function.
- Impaired conduction of sound waves and hearing loss occur when the tympanic membrane has been perforated; air in the middle ear has been replaced with fluid (OME); or the function of the bony ossicles has been impaired (otosclerosis).

muscle less efficient for opening the eustachian tube in this age group. In infants and children with craniofacial disorders, such as a cleft palate, or abnormalities in attachment of the tensor muscles may produce functional obstruction of the eustachian tube.

Mechanical obstruction results from internal obstruction or external compression of the eustachian tube. The most common cause of internal obstruction is swelling and secretions resulting from allergy and viral respiratory infections. With obstruction, air in the middle ear is absorbed, causing a negative pressure and the transudation of serous capillary fluid into the middle ear.

Otitis Media

Otitis media (OM) is an infection of the middle ear that is associated with a collection of fluid. Although OM may occur in any age group, it is the most common diagnosis made by health care providers who care for children. Almost all children have had at least one ear infection by 7 years of age, and one third have three or more episodes by 3 years of age.²⁶ Infants and young children are at highest risk, with the peak incidence between 6 and 13 months. The incidence is higher in boys, infants who are not breast-fed, those who use pacifiers beyond infancy, children in large day care settings, children exposed to tobacco smoke, those with siblings or parents with a significant history of OM, those with allergic rhinitis, and children with congenital or acquired immune deficiencies (e.g., acquired immunodeficiency syndrome).27 The incidence of OM also is higher among children with craniofacial anomalies (e.g., cleft palate, Down's syndrome) and among Canadian and Alaskan Eskimos, and Native Americans.²⁸ It is more common during the winter months, reflecting the seasonal patterns of upper respiratory tract infections.

There are two reasons for the increased risk of OM in infants and young children. First, the eustachian tube is shorter, more horizontal, and wider in this age group than in older children and adults. Second, the infection can spread more easily through the eustachian canal of infants who spend most of their day lying supine. Bottle-fed infants have a higher incidence of OM than breast-fed infants, probably because they are held in a more horizontal position during feeding, and swallowing while in the horizontal position facilitates the reflux of milk into the middle ear. Breast-feeding also provides for the transfer of protective maternal antibodies to the infant.

Otitis media may present as acute otitis media (AOM), recurrent OM, or OM with effusion (OME) or fluid in the middle ear. This effusion may be thin and watery (serous), thick and mucus-like (mucoid), or purulent (containing pus). The characteristics of the fluid vary depending on the type of OM.

Acute Otitis Media. Acute OM is characterized by the presence of fluid in the middle ear in combination with signs and symptoms of an acute or systemic infection.²⁹ AOM can fail to resolve despite antibiotic treatment (persistent OM), or it may resolve and then recur (recurrent OM). Most cases of AOM follow an upper respiratory tract infection that has been present for several days. AOM may be of either bacterial or viral origin. There may be more than one type of bacteria present in some children. S. pneumoniae causes the largest proportion (40% to 50%) of cases generated by a single organism, and it is the least likely to resolve without treatment.³⁰ Emergence of a multiple–drug-resistant strain of *S. pneumoniae* (DRSP) has led to increased numbers of treatment failures.

Acute OM is characterized by otalgia (earache), fever, and hearing loss. Children older than 3 years of age may have rhinorrhea or running nose, vomiting, and diarrhea. In contrast, younger children often have nonspecific signs and symptoms that manifest as ear tugging, irritability, nighttime awakening, and poor feeding. Ear pain usually increases as the effusion accumulates behind the tympanic membrane. Perforation of the tympanic membrane may occur acutely, allowing purulent material from the eustachian tube to drain into the external auditory canal. This may prevent spread of the infection into the temporal bone or intracranial cavity.

Diagnosis of AOM is made by associated signs and symptoms and otoscopic examination.³¹ The treatment of AOM includes the use of medications for fever and pain relief and the judicious use of antibiotic therapy in high-risk children. If the tympanic membrane is bulging and painful because of the accumulation of purulent drainage, a myringotomy (surgical incision in the eardrum) may be done to relieve the pressure, thus reducing pain and hearing loss. In addition, this procedure prevents the ragged opening that can follow spontaneous rupture of the tympanic membrane.

Residual middle ear effusions are part of the continuum of AOM and persist regardless of whether antibiotics have been used. The effusion usually clears spontaneously within 1 to 3 months and does not require further treatment unless it persists beyond this period.

Recurrent Otitis Media. Recurrent OM is defined as three new AOM episodes within 6 months or four episodes in 1 year that occur with almost every upper respiratory tract infection. Re-inforcement of environmental controls, such as avoidance of passive tobacco smoke, is important. Children with recurrent OM should be evaluated to rule out any anatomic variations (*e.g.*, enlarged adenoids) and immunologic disorders.

Recurrent OM may be managed with prophylactic antibiotic therapy. However, the emergence of bacterial resistance has raised concerns about the injudicious use of prophylactic antibiotics. Another approach to prevent recurrent OM is immunization with pneumococcal and influenza vaccines. Placement of tympanostomy tubes is another alternative, particularly for children who have experienced five or more OM episodes within a 12-month period.

Otitis Media With Effusion. Otitis media with effusion is a condition in which the tympanic membrane is intact and there is an accumulation of fluid in the middle ear without signs or symptoms of infection. The type of effusion often is described as serous, nonsuppurative, or secretory, but these terms may not be correct in all cases. In comparison to children with AOM, those with OME do not have a fever or other signs and symptoms of infection, although some may report a feeling of ear fullness.

Most cases of persistent middle ear effusion resolve spontaneously within 3 months. The management options for this duration include observation only, antibiotic therapy, or combination antibiotic and corticosteroid therapy. Because there is concern about hearing loss and its effect on learning and speech, a hearing evaluation may be indicated and usually is done after 6 weeks. If the effusion persists for 3 months or longer and is accompanied by hearing loss of 20 decibels (dB) or greater in children of normal development, tympanostomy tube placement may be indicated.²⁷

Complications of Otitis Media. The complications of OM include hearing loss and extratemporal complications, including those affecting the middle ear, mastoid, adjacent structures of the temporal bone, and intracranial structures.

Hearing loss, which is a common complication of OM, usually is conductive and temporary based on the duration of the effusion. Hearing loss that is associated with fluid collection usually resolves when the effusion clears. Permanent hearing loss may occur as the result of damage to the tympanic membrane or other middle ear structures. Cases of sensorineural hearing loss are rare. Persistent and episodic conductive hearing loss in children may impair their cognitive, linguistic, and emotional development. However, the degree and duration of hearing loss required to produce such effects are unknown.

Adhesive OM involves an abnormal healing reaction in an inflamed middle ear. It produces irreversible thickening of the mucous membranes and may cause impaired movement of the ossicles and possibly conductive hearing loss. Tympanosclerosis involves the formation of whitish plaques and nodular deposits on the submucosal surface of the tympanic membrane, with possible adherence of the ossicles and conductive hearing loss.

A *cholesteatoma* is a saclike mass containing silvery-white debris of keratin, which is shed by the squamous epithelial lining of the tympanic membrane. As the lining of the epithelium sheds and desquamates, the lesion expands and erodes the surrounding tissues. The lesion, which is associated with chronic middle ear infection, is insidiously progressive, and erosion may involve the temporal bone, causing intracranial complications. Treatment involves microsurgical techniques to remove the cholesteatomatous material.

The mastoid antrum and air cells constitute a portion of the temporal bone and may become inflamed as an extension of acute or chronic OM. The disorder causes necrosis of the mastoid process and destruction of the bony intercellular matrix, which are visible by radiologic examination. Mastoid tenderness and drainage of exudate through a perforated tympanic membrane can occur. Chronic mastoiditis can develop as the result of chronic middle ear infection. Mastoid or middle ear surgery, along with other medical treatment, may be indicated.

Intracranial complications are uncommon since the advent of antimicrobial therapy. Although rare, these complications can develop if the infection spreads through vascular channels, by direct extension, or through preformed pathways such as the round window. These complications are seen more often with chronic suppurative OM and mastoiditis. They include meningitis, focal encephalitis, brain abscess, lateral sinus thrombophlebitis or thrombosis, labyrinthitis, and facial nerve paralysis.

Otosclerosis

SHORT

Otosclerosis refers to the formation of new spongy bone around the stapes and oval window, which results in progressive deafness.³ In most cases, the condition is familial and follows an autosomal dominant pattern with variable penetrance. Otosclerosis may begin at any time in life but usually does not appear until after puberty, most frequently between the ages of 20 and 30 years. The disease process accelerates during pregnancy.

Otosclerosis begins with resorption of bone in one or more foci. During active bone resorption, the bone structure appears spongy and softer than normal (*i.e.*, osteospongiosis). The resorbed bone is replaced by an overgrowth of new, hard, sclerotic bone. The process is slowly progressive, involving more areas of the temporal bone, especially in front of and posterior to the stapes footplate. As it invades the footplate, the pathologic bone increasingly immobilizes the stapes, reducing the transmission of sound. The pressure of otosclerotic bone on inner ear structures or the vestibulocochlear nerve (CN VIII) may contribute to the development of tinnitus, sensorineural hearing loss, and vertigo (to be discussed later in this chapter).

The symptoms of otosclerosis involve an insidious hearing loss. Initially, the affected person is unable to hear a whisper or someone speaking at a distance. In the earliest stages, the bone conduction by which the person's own voice is heard remains relatively unaffected. At this point, the person's own voice sounds unusually loud, and the sound of chewing becomes intensified. Because of bone conduction, most of these persons can hear fairly well on the telephone, which provides an amplified signal. Many are able to hear better in a noisy environment, probably because the masking effect of background noise causes other persons to speak more loudly.

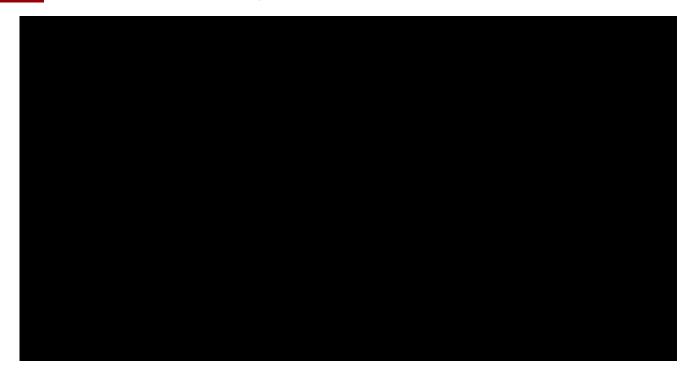
The treatment of otosclerosis can be medical or surgical. A carefully selected, well-fitting hearing aid may allow a person with conductive deafness to lead a normal life. Sodium fluoride has been used with some success in the medical treatment of osteospongiosis. Because much of the conductive hearing loss associated with otosclerosis is caused by stapedial fixation, surgical treatment involves stapedectomy with reconstruction using the patient's own stapes or a prosthetic device.

The Inner Ear and Auditory Pathways

The inner ear contains a labyrinth, or system of intercommunicating channels, and the receptors for hearing and position sense. An outer bony wall, or bony labyrinth, encloses a thin walled, membranous labyrinth, which floats in the bony labyrinth. Two separate fluids are found in the inner ear. The *periotic fluid* or *perilymph* separates the bony labyrinth from the membranous labyrinth, and the *otic fluid* or *endolymph* fills the membranous labyrinth. Periotic fluid composition is similar to that of the CSF, and a tubular perilymphatic duct connects the periotic fluid with the CSF in the arachnoid space of the posterior fossa. Otic fluid has a potassium content that is similar to that of intracellular fluid.

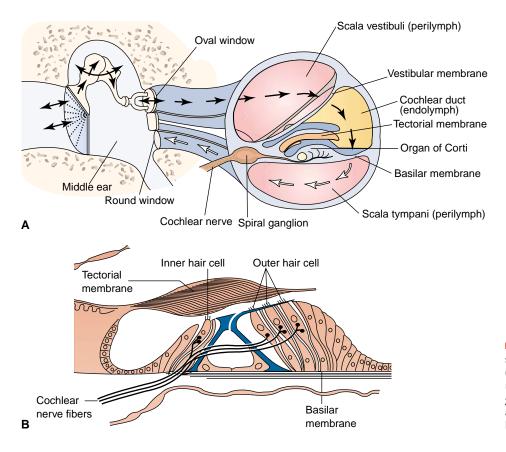
Localized dilatations of the membranous labyrinth develop into three specialized sensory regions: the ampulla of each semicircular canal, the maculae of the utricle and sacculus, and the cochlea (Fig. 40-18). The cochlea is enclosed in a bony tube shaped like a snail shell that winds around a central bone column called the *modiolus*. Running through its center is the triangular membranous cochlear duct, which houses the spiral organ of Corti, the receptor organ for hearing. The cochlear duct and the spiral lamina, a thin shelflike extension that spirals

768 Unit Ten: Alterations in the Nervous System



up the modiolus divides the cavity of the cochlea into three chambers: the scala vestibuli, the scala tympani, and the scala media (Fig. 40-19A). The organ of Corti, which rests atop the basilar membrane in the scala media, is composed of supporting cells and several long rows of cochlear hair cells: one row

of inner hair cells and three rows of outer hair cells (Fig. 40-19B). Afferent fibers from the cochlear nerve are coiled around the bases of the hair cells. Sound waves, delivered to the oval window by the stapes footplate, are transmitted to the periotic fluid in the scala vestibuli and scala tympani. Transduction of sound



■ FIGURE 40-19 ■ (A) Path taken by sound waves reaching the inner ear. (B) Spiral organ of Corti has been removed from the cochlear duct and greatly enlarged to show the inner and outer hair cells, the basilar membrane, and cochlear nerve fibers. stimuli occurs when the trapped cilia of the hair cells in the organ of Corti are bent by the sound-induced movement of the basilar membrane.

Afferent fibers from the organ of Corti have their cell bodies in the spiral ganglion in the central portion of the cochlea. Nerve fibers from the spiral ganglion (i.e., vestibulocochlear or auditory nerve [CN VIII]) travel to the cochlear nuclei in the caudal pons. Many secondary nerve fibers from the cochlear nuclei pass to the nuclei on the opposite side of the pons or rostrally toward the inferior colliculus of the midbrain. From the inferior colliculus, the auditory pathway passes to the primary auditory cortex (area 41) in the temporal lobe via relays in the medial geniculate nucleus of the thalamus (Fig. 40-11). The auditory association cortex (areas 42 and 22), which is necessary for the meaningfulness of sound, borders the primary cortex. Because some of the fibers from each ear cross, each auditory cortex receives impulses from both ears.

Disorders of the Inner Ear and Central Auditory Pathways

Disorders of the cochlear component of the inner ear and auditory pathways can lead to the presence of tinnitus or sensorineural hearing loss.

Tinnitus. Tinnitus (from the Latin *tinniere*, meaning "to ring") is the perception of abnormal ear or head noises not produced by an external stimulus.³³ Although it often is described as "ringing of the ears," it may also assume a hissing, roaring, buzzing, or humming sound. It has been estimated that 35 million people in the United States have the disorder. The condition affects males and females equally, is most prevalent between 40 and 70 years of age, and occasionally affects children.33

Tinnitus may be constant, intermittent, and unilateral or bilateral. Intermittent periods of mild, high-pitched tinnitus lasting for several minutes are common in normal-hearing persons. Impacted cerumen is a benign cause of tinnitus, which resolves after the ear wax is removed. Medications such as aspirin and stimulants such as nicotine and caffeine can cause transient tinnitus. Although tinnitus is a subjective experience, for clinical purposes it is subdivided into objective and subjective tinnitus. Objective tinnitus refers to those rare cases in which the sound is detected or potentially detectable by another observer. Typical causes of objective tinnitus include vascular abnormalities or neuromuscular disorders. For example, in some vascular disorders, sounds generated by turbulent blood flow (e.g., arterial bruits or venous hums) are conducted to the auditory system. Vascular disorders typically produce a pulsatile form of tinnitus. Subjective tinnitus refers to noise perception when there is no noise stimulation of the cochlea. The physiologic mechanism underlying subjective tinnitus is largely unknown. It seems likely that there are several mechanisms, including abnormal firing of auditory receptors, dysfunction of cochlear neurotransmitter function or ionic balance, and alterations in central processing of the signal.

Treatment measures for tinnitus are designed to treat the symptoms, rather than effect a cure. They include elimination of drugs or other substances such as caffeine, some cheeses, red wine, and foods containing monosodium glutamate that are suspected of causing tinnitus. The use of an externally produced sound (noise generators or tinnitus-masking devices) may be used to mask or inhibit the tinnitus.

Disorders of the Central Auditory Pathways. The auditory pathways in the brain involve communication between the two sides of the brain at many levels. As a result, strokes, tumors, abscesses, and other focal abnormalities seldom produce more than a mild reduction in auditory acuity on the side opposite the lesion. For intelligibility of auditory language, lateral dominance becomes important. On the dominant side, usually the left side, the more medial and dorsal portion of the auditory association cortex is of crucial importance. This area is called Wernicke's area, and damage to it is associated with auditory receptive aphasia (and agnosia of speech). Persons with damage to this area of the brain can speak intelligibly and read normally but are unable to understand the meaning of major aspects of audible speech.

Irritative foci that affect the auditory radiation or the primary auditory cortex can produce roaring or clicking sounds, which appear to come from the auditory environment of the opposite side (i.e., auditory hallucinations). Focal seizures that originate in or near the auditory cortex often are immediately preceded by the perception of ringing or other sounds preceded by a prodrome (*i.e.*, aura). Damage to the auditory association cortex, especially if bilateral, results in deficiencies of sound recognition and memory (*i.e.*, auditory agnosia). If the damage is in the dominant hemisphere, speech recognition can be affected (*i.e.*, sensory or receptive aphasia).

Hearing Loss

Nearly 30 million Americans have hearing loss. It affects persons of all age groups. One of every 1000 infants born in the United States is completely deaf, and more than 3 million children have hearing loss.³⁴ Thirty percent to 40% of people older than 75 years have hearing loss.

The level of hearing is measured in decibels, where 0 dB is the threshold for perception of sound at a given frequency in persons with normal hearing.³⁵ A 10-fold increase in sound pressure level from 0 dB is measured as 20 dB. Hearing loss is qualified as mild, moderate, severe, or profound. "Hard of hearing" is defined as hearing loss greater than 20 to 25 dB in adults and greater than 15 dB in children. Profound deafness is defined as hearing loss greater than 100 dB³⁶ or 70 dB in children.37

There are many causes of hearing loss or deafness. Most fit into the categories of conductive, sensorineural, or mixed

KEY CONCEPTS

HEARING LOSS

- Hearing loss represents impairment of the ability to detect and perceive sound.
- Conductive hearing loss is caused by disorders in which auditory stimuli are not transmitted through the structures of the outer and middle ears to the sensory receptors in the inner ear.
- Sensorineural hearing loss is caused by disorders that affect the inner ear, auditory nerve, or auditory pathways.

deficiencies that involve a combination of conductive and sensorineural function deficiencies of the same ear.³⁵ Chart 40-1 summarizes common causes of hearing loss. Hearing loss may be genetic or nongenetic, sudden or progressive, unilateral or bilateral, partial or complete, reversible or irreversible. Age and suddenness of onset provide important clues as to the cause of hearing loss.

Conductive Hearing Loss

Conductive hearing loss occurs when auditory stimuli are not adequately transmitted through the auditory canal, tympanic membrane, middle ear, or ossicle chain to the inner ear. Temporary hearing loss can occur as the result of impacted cerumen in the outer ear or fluid in the middle ear. Foreign bodies, including pieces of cotton and insects, may impair hearing. More permanent causes of hearing loss are thickening or damage of the tympanic membrane or involvement of the bony structures

CHART 40-1 Common Causes of Conductive and Sensorineural Hearing Loss

Conductive Hearing Loss

- External ear conditions
- Impacted ear wax or foreign body
- Otitis externa
- Middle ear conditions
- Trauma
- Otitis media (acute and with effusion)
- Otosclerosis
- Tumors

Sensorineural Hearing Loss

- Trauma
- Head injury
- Noise
- Central nervous system infections (e.g., meningitis)
- Degenerative conditions
- Presbycusis
- Vascular
- Atherosclerosis
- Sudden deafness
- Ototoxic drugs (*e.g.*, aminoglycosides, salicylates, loop diuretics)
- Tumors
 - Vestibular schwannoma (acoustic neuroma)
- Meningioma
- Metastatic tumors
- Idiopathic
 - Ménière's disease

Mixed Conductive and Sensorineural Hearing Loss

- Middle ear conditions
 - Barotrauma
 - Cholesteatoma
- Otosclerosis
- Temporal bone fractures

(ossicles and oval window) of the middle ear caused by otosclerosis or Paget's disease.

Sensorineural Hearing Loss

Sensorineural, or perceptive, hearing loss occurs with disorders that affect the inner ear, auditory nerve, or auditory pathways of the brain. With this type of deafness, sound waves are conducted to the inner ear, but abnormalities of the cochlear apparatus or auditory nerve decrease or distort the transfer of information to the brain. Tinnitus often accompanies cochlear nerve irritation. Abnormal function resulting from damage or malformation of the central auditory pathways and circuitry is included in this category.

Sensorineural hearing loss may have a genetic cause or may result from intrauterine infections, such as maternal rubella, or developmental malformations of the inner ear. Genetic hearing loss may result from mutation in a single gene (monogenetic) or from a combination of mutations in different genes and environmental factors (multifactorial).³⁶ It has been estimated that 50% of profound deafness in children has a monogenetic basis.^{36,38} The inheritance pattern for monogenetic hearing loss is autosomal recessive in approximately 75% of cases.³⁶ Hearing loss may begin before development of speech (prelingual) or after speech development (postlingual). Most prelingual forms are present at birth. Genetic forms of hearing loss also can be classified as being part of a syndrome in which other abnormalities are present, or as nonsyndromic, in which deafness is the only abnormality.

Sensorineural hearing loss also can result from trauma to the inner ear, tumors that encroach on the inner ear or sensory neurons, vascular disorders with hemorrhage, or thrombosis of vessels that supply the inner ear. Other causes of sensorineural deafness are infections and drugs. Sudden sensorineural hearing loss represents an abrupt loss of hearing that occurs instantaneously or on awakening. It most commonly is caused by viral infections, circulatory disorders, or rupture of the labyrinth membrane that can occur during tympanotomy.³⁹ Hypothyroidism is a potential cause of sensorineural hearing loss in older persons.

Environmentally induced deafness can occur through direct exposure to excessively intense sound, as in the workplace or at a concert. This is a particular problem in older adults who were working in noisy environments before the mid-1960s, when there were no laws mandating use of devices for protective hearing. Sustained or repeated exposure to noise pollution at sound intensities greater than 100 to 120 dB can cause corresponding mechanical damage to the organ of Corti. If damage is severe, permanent sensorineural deafness to the offending sound frequencies results. Wearing earplugs or ear protection is important under many industrial conditions and for musicians and music listeners exposed to high sound amplification.

A number of infections can cause sensorineural hearing loss. Deafness or some degree of hearing impairment is the most common serious complication of bacterial meningitis in infants and children, reportedly resulting in sensorineural hearing loss in 5% to 35% of persons who survive the infection.³⁸ The mechanism causing hearing impairment seems to be a suppurative labyrinthitis or neuritis resulting in the loss of hair cells and damage to the auditory nerve. Untreated suppurative

771

OM also can extend into the inner ear and cause sensorineural hearing loss through the same mechanisms.

Among the neoplasms that impair hearing are *acoustic neuromas*. Acoustic neuromas are benign Schwann cell tumors affecting CN VIII. These tumors usually are unilateral and cause hearing loss by compressing the cochlear nerve or interfering with blood supply to the nerve and cochlea. Other neoplasms that can affect hearing include meningiomas and metastatic brain tumors. The temporal bone is a common site of metastases. Breast cancer may metastasize to the middle ear and invade the cochlea.

Drugs that damage inner ear structures are labeled ototoxic. Vestibular symptoms of ototoxicity include light-headedness, giddiness, and dizziness; if toxicity is severe, cochlear symptoms consisting of tinnitus or hearing loss occur. Hearing loss is sensorineural and may be bilateral or unilateral, transient or permanent. Several classes of drugs have been identified as having ototoxic potential, including the aminoglycoside antibiotics and some other basic antibiotics, antimalarial drugs, some chemotherapeutic drugs, loop diuretics, and salicylates. The symptoms of drug-induced hearing loss may be transient, as often is the case with salicylates and diuretics, or they may be permanent. The risk of ototoxicity depends on the total dose of the drug and its concentration in the bloodstream. It is increased in persons with impaired kidney functioning and in those previously or currently treated with another potentially ototoxic drug.

Presbycusis

The term *presbycusis* is used to describe degenerative hearing loss that occurs with advancing age. Approximately 23% of persons between 65 and 75 years of age and 40% of the population older than 75 years are affected.⁴⁰ Men are affected earlier and experience a greater loss than women.

The degenerative changes that impair hearing may begin in the fifth decade of life and not be clinically apparent until later.⁴¹ Onset may be associated with chronic noise exposure or vascular disorders. The disorder involves loss of neuroepithelial (hair) cells, neurons, and the stria vascularis. High-frequency sounds are affected more than low-frequency sounds because high and low frequencies distort the base of the basilar membrane, but only low frequencies affect the distal (apical) region. Through the years, permanent mechanical damage to the organ of Corti is more likely to occur near the base of the cochlea, where the high sonic frequencies are discriminated. Loss of high-frequency discrimination is characterized by difficulty in understanding words in noisy environments, in hearing a speaker in an adjacent room, or hearing a speaker whose back is turned.

Although hearing loss is a common problem in the elderly, many older persons are not appropriately assessed for hearing loss. When assessing an older person's ability to hear, it is important to ask both the person and the family about awareness of hearing loss.

Detection and Treatment of Hearing Loss

Diagnosis of hearing loss is aided by careful history of associated otologic factors such as otalgia, otorrhea, tinnitus, and self-described hearing difficulties; physical examination to detect the presence of conditions such as otorrhea, impacted cerumen, or injury to the tympanic membrane; and hearing tests. Testing for hearing loss includes a number of methods, including a person's reported ability to hear an observer's voice, use of a tuning fork to test air and bone conduction, audioscopes, and auditory brain stem evoked responses (ABRs).

Tuning forks are used to differentiate conductive and sensorineural hearing loss. Audioscopes can be used to assess a person's ability to hear pure tones at 1000 to 2000 Hz (usual speech frequencies). The ABR is a noninvasive method that permits functional evaluation of certain defined parts of the central auditory pathways. Electroencephalographic (EEG) electrodes and high-gain amplifiers are required to produce a record of the electrical wave activity elicited during repeated acoustic stimulations of either or both ears. ABR recording involves subjecting the ear to loud clicks and using a computer to pick up nerve impulses as they are processed in the midbrain. Imaging studies such as computed tomography (CT) scans and magnetic resonance imaging (MRI) can be done to determine the site of a lesion and the extent of damage.³⁴

Because hearing impairment can have a major impact on the development of a child, early identification through screening programs is strongly advocated. The American Academy of Pediatricians endorses the goal of universal detection of hearing loss in infants before 3 months of age, with proper intervention no later than 6 months of age.⁴²

Treatment. Untreated hearing loss can have many consequences. In infants and children, hearing loss can greatly affect language development and hearing-associated learning. Social isolation and depressive disorders are common in hearingimpaired elderly. Hearing-impaired people may avoid social situations where background noise makes conversation difficult to hear. Safety issues, both in and out of the home, may become significant. Treatment of hearing loss ranges from simple removal of impacted cerumen in the external auditory canal to surgical procedures such as those used to reconstruct the tympanic membrane. For other people, particularly the frail elderly, hearing aids remain an option. Cochlear implants also are an option for those with profound hearing loss.

In summary, hearing is a specialized sense whose external stimulus is the vibration of sound waves. Anatomically, the auditory system consists of the outer ear, middle ear, and inner ear, the auditory pathways, and the auditory cortex. The middle ear is a tiny air-filled cavity in the temporal bone. The inner ear contains the receptors for hearing.

Disorders of the auditory system include infections of the external and middle ear, otosclerosis, and conduction and sensorineural deafness. Otitis externa is an inflammatory process of the external ear. The middle ear is a tiny, air-filled cavity located in the temporal bone. The eustachian tube, which connects the middle ear to the nasopharynx allows equalization of pressure between the middle ear and the atmosphere. Infections can travel from the nasopharynx to the middle ear along the eustachian tube, causing OM or inflammation of the middle ear. The eustachian tube is shorter and more horizontal in infants and young children, and infections of the middle ear are a common problem in these age groups. OM may present as AOM, recurrent OM, or OME. AOM usually follows an upper respiratory tract infection and is characterized by otalgia, fever, and hearing loss. The effusion that accompanies OM can persist for weeks or months, interfering with hearing and impairing speech development. Otosclerosis is a familial disorder of the otic capsule. It causes bone resorption followed by excessive replacement with sclerotic bone. The disorder eventually causes immobilization of the stapes and conduction deafness.

Deafness, or hearing loss, can develop as the result of a number of auditory disorders. It can be conductive, sensorineural, or mixed. Conduction deafness occurs when transmission of sound waves from the external to the inner ear is impaired. Sensorineural deafness can involve cochlear structures of the inner ear or the neural pathways that transmit auditory stimuli. Sensorineural hearing loss can result from genetic or congenital disorders, trauma, infections, vascular disorders, tumors, or ototoxic drugs. Treatment of hearing loss includes the use of hearing aids and, in some cases of profound deafness, implantation of a cochlear prosthesis.

THE VESTIBULAR SYSTEM AND MAINTENANCE OF EQUILIBRIUM

The vestibular receptive organs, which are located in the inner ear, and their CNS connections, contribute to the reflex activity needed for effective posture and movement and serve to maintain a stable visual field despite changes in head position. Because the vestibular apparatus is part of the inner ear and located in the head, it is head motion and acceleration that are sensed.

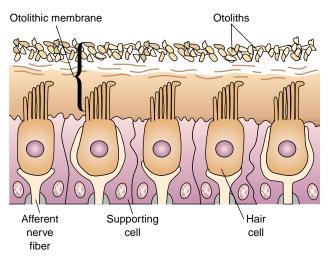
KEY CONCEPTS

DISORDERS OF THE VESTIBULAR SYSTEM

- The vestibular system, which is located in the inner ear and senses head motion and acceleration, contributes to the reflex activity needed for effective posture and movement, and it serves to maintain a stable visual field despite changes in head position.
- The vestibular system, which has extensive interconnections with neural pathways controlling vision, hearing, chemotactic receptor trigger zone, the cerebellum, and the autonomic nervous system, is characterized by vertigo, nystagmus, tinnitus, nausea and vomiting, and autonomic nervous system manifestations.
- Disorders of vestibular function can result from repeated stimulation of the vestibular system, such as during car, air, and boat travel (motion sickness); distention of the endolymphatic compartment of the inner ear (Ménière's disease); or dislodgment of otoliths that participate in the receptor function of the vestibular system (benign paroxysmal positional vertigo).

The peripheral apparatus of the vestibular system is contained in the bony labyrinth of the inner ear next to and continuous with the cochlea of the auditory system. It is divided into five prominent structures: three semicircular ducts, a utricle, and a saccule (Fig. 40-19). Receptors in these structures are differentiated into the angular acceleration-deceleration receptors of the semicircular ducts and the linear accelerationdeceleration and static gravitational receptors of the utricle and saccule. Small patches of hair cells are located in the floor of the utricle (utricular macula), in the sidewall of the saccule (saccular macula), at the base of each semicircular duct (cristae), and in the organ of Corti along the floor of the cochlear duct (Fig. 40-20). Each hair cell has several microvilli and one true cilium, called a kinocilium. At the apical end of each inner hair cell is a projecting bundle of rodlike structures called stereocilia. The stereocilia of the hair cells extend into a flattened gelatinous mass, the otolithic membrane, which is studded with tiny stones (calcium carbonate crystals) called otoliths. Although small, the density of the otoliths increases the membrane's weight and its resistance to change in motion. When the head is tilted, the gelatinous mass shifts its position because of the pull of the gravitational field, bending the stereocilia of the macular hair cells. Although each hair cell becomes less or more excitable, depending on the direction in which the cilia are bending, the hair cells are oriented in all directions, making these sense organs sensitive to static or changing head position in relation to the gravitational field.

The nerve fibers from the ganglionic cells that supply the vestibular apparatus become the superior and inferior vestibular nerves, which become part of the eighth cranial nerve. Impulses from the vestibular nerves initially pass to one of two destinations: the vestibular nuclear complex in the stem or the cerebellum. The vestibular nuclei, which form the main integrative center for balance, also receives input from visual and somatic receptors, particularly from proprioceptors in the neck muscles that report the angle or inclination of the head. The vestibular nuclei integrate this information and then send im-



■ FIGURE 40-20 ■ The relation of the otoliths to the sensory cells in the macula of the utricle and saccule. (Adapted from Selkurt F.D. [Ed.]. [1982]. *Basic physiology for the health sciences* [2nd ed.]. Boston: Little, Brown)

pulses to the brain stem centers that control the extrinsic eye movements (CN III, IV, and VI) and reflex movements of the neck, limb, and trunk muscles (via the vestibulospinal tracts). Reflex movements of the eyes and body allow for quick adjustment of body position to maintain or regain balance. Neurons of the vestibular nuclei also project to the thalamus, the temporal cortex, the somesthetic area of the parietal cortex, and the chemoreceptor trigger zone. The thalamic and cortical projections provide the basis for the subjective experiences of position in space and of rotation. Connections with the chemoreceptor trigger zone stimulate the vomiting center in the brain (see Chapter 26). This accounts for the nausea and vomiting that often are associated with vestibular disorders.

Nystagmus

The term nystagmus is used to describe the vestibulo-ocular reflexes that occur in response to ongoing head rotation. The vestibulo-ocular reflexes produce slow compensatory conjugate eve rotations that occur in the direction precisely opposite to ongoing head rotation and provide for continuous, ongoing reflex stabilization of the binocular fixation point. This reflex can be demonstrated by holding a pencil vertically in front of the eyes and moving it from side to side through a 10-degree arc at a rate of approximately five times per second. At this rate of motion, the pencil appears blurred, because a different and more complex reflex, smooth pursuit, cannot compensate quickly enough. However, if the pencil is maintained in a stable position and the head is moved back and forth at the same rate, the image of the pencil is clearly defined. The eye movements are the same in both cases. The reason that the pencil image remains clear in the second situation is because the vestibuloocular reflexes keep the image of the pencil on the retinal fovea. When compensatory vestibulo-ocular reflexes carry the conjugate eye rotations to their physical limit, a very rapid conjugate movement (i.e., saccade) moves the eyes in the direction of head rotation to a new fixation point, followed by a slow vestibulo-ocular reflex as the head continues to rotate past the new fixation point. This pattern of slow-fast-slow movements is called *nystagmus*.

Spontaneous nystagmus that occurs without head movement or visual stimuli is always pathologic. It seems to appear more readily and more severely with fatigue and to some extent can be influenced by psychological factors. Nystagmus caused by CNS pathology, in contrast to vestibular end-organ or vestibulocochlear nerve sources, seldom is accompanied by vertigo. If present, the vertigo is of mild intensity.

Vertigo

Disorders of vestibular function are characterized by a condition called *vertigo*, in which an illusion of motion occurs. It is either an exaggerated sense of motion when there is no motion or motion in response to a given bodily movement. Persons with vertigo frequently describe a sensation of spinning or tumbling, a "to-and-fro" motion, or falling forward or backward. Vertigo should be differentiated from light-headedness, faintness, unsteadiness, or syncope (loss of consciousness; Table 40-2).^{43,44} Vertigo or dizziness can result from peripheral or central vestibular disorders. Approximately 85% of persons with vertigo have a peripheral vestibular disorder, whereas only 15% have a central disorder. Vertigo caused by peripheral disorders tends to be severe in intensity and episodic or brief in duration. In contrast, vertigo attributable to central causes tends to be mild and constant and chronic in duration.

Diagnostic methods include the Romberg test (discussed later in this section), an evaluation of gait, and observations for the presence of nystagmus.⁴⁵ Laboratory investigations include audiologic evaluation, electronystagmography, CT scan or MRI, and auditory brain stem evoked responses (ABRs). These tests help to distinguish between central and peripheral causes of vertigo and to identify causes requiring specific treatment.

Motion Sickness

Motion sickness is a form of normal physiologic vertigo. It is caused by repeated rhythmic stimulation of the vestibular system, such as is encountered in car, air, or boat travel. Vertigo, malaise, nausea, and vomiting are the principal symptoms. Autonomic signs, including lowered blood pressure, tachycardia, and excessive sweating, may occur. Some persons experience a variant of motion sickness, reporting sensing the rocking motion of the boat after returning to ground. This usually resolves after the vestibular system becomes accustomed to the stationary influence of being back on land.

Motion sickness can usually be suppressed by supplying visual signals that more closely match the motion signals being supplied to the vestibular system. For example, looking out the window and watching the environment move when experiencing motion sickness associated with car travel provides the vestibular system with the visual sensation of motion. Among

TABLE 40-2	Differences in Pathology and Manifestations of Dizziness Associated
	Positional Vertigo, Presyncope, and Disequilibrium State

Type of Disorder	Pathology	Symptoms
Benign positional vertigo	Disorder of otoliths	Vertigo initiated by a change in head position, usually lasts less than a minute
Presyncope Disequilibrium	Orthostatic hypotension Sensory (<i>e.g.,</i> vision, proprioception) deficits	Light-headedness and feeling faint on assumption of standing position Dizziness and unsteadiness when walking, especially when turning; re- lieved by additional proprioceptive stimulation such as touching wall or table

the drugs used to suppress, reduce, or ameliorate the symptoms of motion sickness are the antihistamines (*e.g.*, meclizine, cyclizine, dimenhydrinate, and promethazine) and the anticholinergic drugs (*e.g.*, scopolamine, atropine).

Disorders of Peripheral Vestibular Function

Disorders of peripheral vestibular function occur when these signals from the vestibular organs in the inner ear are distorted, such as in benign paroxysmal positional vertigo, or unbalanced by unilateral involvement of one of the vestibular organs, such as in Ménière's disease. The inner ear is vulnerable to injury caused by fracture of the petrous portion of the temporal bones; by infection of nearby structures, including the middle ear and meninges; and by blood-borne toxins and infections. Alcohol can cause transient episodes of vertigo.

Ménière's Disease

Ménière's disease is a disorder of the inner ear caused by distention of the endolymphatic compartment of the inner ear, causing a triad of hearing loss, vertigo, and tinnitus.^{46,47} The primary lesion appears to be in the endolymphatic sac, which is thought to be responsible for endolymph filtration and excretion. A number of pathogenic mechanisms have been postulated, including an increased production of endolymph, decreased production of perilymph accompanied by a compensatory increase in volume of the endolymphatic sac, and decreased absorption of endolymph caused by malfunction of the endolymphatic sac or blockage of endolymphatic pathways.

The cause of Ménière's disease is unknown. A number of conditions, such as trauma, infection (*e.g.*, syphilis), and immunologic, endocrine (adrenal-pituitary insufficiency and hypothyroidism), and vascular disorders have been proposed as possible causes of Ménière's disease.⁴⁷ The most common form of the disease is an idiopathic form thought to be caused by a single viral injury to the fluid transport system of the inner ear.

Ménière's disease is characterized by fluctuating episodes of tinnitus, feelings of ear fullness, and violent rotary vertigo that often renders the person unable to sit or walk. There is a need to lie quietly with the head fixed in a comfortable position, avoiding all head movements that aggravate the vertigo. Symptoms referable to the autonomic nervous system, including pallor, sweating, nausea, and vomiting, usually are present. The more severe the attack, the more prominent are the autonomic manifestations. A fluctuating hearing loss occurs, with a return to normal after the episode subsides. Initially the symptoms tend to be unilateral, resulting in rotary nystagmus caused by an imbalance in vestibular control of eye movements. Because initial involvement usually is unilateral and because the sense of hearing is bilateral, many persons with the disorder are not aware of the full extent of their hearing loss. However, as the disease progresses, the hearing loss stops fluctuating and progressively worsens, with both ears tending to be affected so that the prime disability becomes one of deafness.⁴⁷ The episodes of vertigo diminish and then disappear, although the person may be unsteady, especially in the dark.

Methods used in the diagnosis of Ménière's disease include audiograms, vestibular testing by electronystagmography, and petrous pyramid radiographs. The administration of hyperosmolar substances, such as glycerin and urea, often produces acute temporary hearing improvement in persons with Ménière's disease and sometimes is used as a diagnostic measure of endolymphatic hydrops.

The management of Ménière's disease focuses on attempts to reduce the distention of the endolymphatic space and can be medical or surgical. Pharmacologic management consists of suppressant drugs (*e.g.*, prochlorperazine, promethazine, diazepam), which act centrally to decrease the activity of the vestibular system. Diuretics are used to reduce endolymph fluid volume. A low-sodium diet is recommended in addition to these medications. Corticosteroid medications (*e.g.*, prednisone) may be used to maintain satisfactory hearing and resolve dizziness. Gentamicin therapy has been used for ablation of the vestibular system.^{47,48} This treatment is mainly effective in controlling vertigo and does not alter the underlying pathology.

Surgical methods include the creation of an endolymphatic shunt, in which excess endolymph from the inner ear is diverted into the subarachnoid space or the mastoid (endolymphatic sac surgery), and vestibular nerve section.

Benign Paroxysmal Positional Vertigo

Benign paroxysmal positional vertigo (BPPV) is the most common cause of pathologic vertigo and usually develops after the fourth decade. It is characterized by brief periods of vertigo, usually lasting less than 1 minute, that are precipitated by a change in head position.⁴⁹ The most prominent symptom of BPPV is vertigo that occurs in bed when the person rolls into a lateral position. It also commonly occurs when the person is getting in and out of bed, bending over and straightening up, or extending the head to look up. It also can be triggered by amusement rides that feature turns and twists.

BPPV is thought to result from damage to the delicate sensory organs of the inner ear, the semicircular ducts, and otoliths. In persons with BPPV, the calcium carbonate particles (otoliths) from the utricle become dislodged and become freefloating debris in the endolymph (otic fluid) of the posterior semicircular duct, which is the most dependent part of the inner ear.⁴⁹ Movement of the free-floating debris causes this portion of the vestibular system to become more sensitive, such that any movement of the head in the plane parallel to the posterior duct may cause vertigo and nystagmus. There usually is a several-second delay between head movement and onset of vertigo, representing the time it takes to generate the exaggerated endolymph activity. Symptoms usually subside with continued movement, probably because the movement causes the debris to be redistributed throughout the endolymph system and away from the posterior duct.

Diagnosis is based on tests that involve the use of a change in head position to elicit vertigo and nystagmus. BPPV often is successfully treated with drug therapy to control vertigoinduced nausea. Nondrug therapies using habituation exercises and canalith repositioning are successful in many people (to be discussed).⁴⁹ Otolith repositioning involves a series of maneuvers in which the head is moved to different positions in an effort to reposition the free-floating debris in the endolymph of the semicircular canals.

Disorders of Central Vestibular Function

Abnormal nystagmus and vertigo can occur as a result of CNS lesions involving the cerebellum and lower brain stem. Central causes of vertigo include brain stem ischemia, tumors, and multiple sclerosis.⁴⁵ When brain stem ischemia is the cause of vertigo, it usually is associated with other brain stem signs, such as diplopia, ataxia, dysarthria, or facial weakness. Compression of the vestibular nuclei by cerebellar tumors invading the fourth ventricle results in progressively severe signs and symptoms. In addition to abnormal nystagmus and vertigo, vomiting and a broad-based and dystaxic gait become progressively more evident. The central demyelinating effects of multiple sclerosis can present with vertigo as often as 10% of the time, and as many as one third of persons with multiple sclerosis experience vertigo and nystagmus some time in the course of the disease.⁴⁵

Centrally derived nystagmus usually has equal excursion in both directions (*i.e.*, pendular). In contrast to peripherally generated nystagmus, CNS-derived nystagmus is relatively constant, rather than episodic; can occur in any direction, rather than being primarily in the horizontal or torsional (rotatory) dimensions; often changes direction through time; and cannot be suppressed by visual fixation. Repeated induction of nystagmus results in rapid diminution or "fatigue" of the reflex with peripheral abnormalities, but fatigue is not characteristic of central lesions. Abnormal nystagmus can make reading and other tasks that require precise eye positional control difficult.

Diagnosis and Treatment

Tests of Vestibular Function

Diagnosis of vestibular disorders is based on a description of the symptoms, a history of trauma or exposure to agents that are destructive to vestibular structures, and physical examination. Tests of eye movements (*i.e.*, nystagmus) and muscle control of balance and equilibrium often are used. The tests of vestibular function focus on the horizontal semicircular reflex because it is the easiest reflex to stimulate rotationally and calorically and to record using electronystagmography.

Electronystagmography. Electronystagmography (ENG) is a precise and objective diagnostic method of evaluating nystagmus eye movements. Electrodes are placed lateral to the outer canthus of each eye and above and below each eye. A ground electrode is placed on the forehead. With ENG, the velocity, frequency, and amplitude of spontaneous or induced nystagmus and the changes in these measurements brought by a loss of fixation, with the eyes open or closed, can be quantified. The advantages of ENG are that it is easily administered, is noninvasive, does not interfere with vision, and does not require head restraint.

Caloric Stimulation. Caloric testing involves elevating the head 30 degrees and irrigating each external auditory canal separately with 30 to 50 mL of ice water. The resulting changes in temperature, which are conducted through the petrous portion of the temporal bone, set up convection currents in the otic fluid that mimic the effects of angular acceleration. In an unconscious person with a functional brain stem and intact oculovestibular reflexes, the eyes exhibit a jerk nystagmus lasting 2 to 3 minutes, with the slow component toward the irrigated ear followed by rapid movement away from the ear. With impairment of brain stem function, the response becomes perverted and eventually disappears. An advantage of the caloric stimulation method is the ability to test the vestibular appara-

tus on one side at a time. The test is never done on a person who does not have an intact eardrum or those who have blood or fluid collected behind the eardrum.

Rotational Tests. Rotational testing involves rotation using a rotatable chair or motor-driven platform. Unlike caloric testing, rotational testing depends only on the inner ear and is unrelated to conditions of the external ear or temporal bone. A major disadvantage of the method is that both ears are tested simultaneously. Motor-driven chairs or platforms can be precisely controlled, and multiple graded stimuli can be delivered in a relatively short period. For rotational testing, the person is seated in a chair mounted on the motor-driven platform. Testing usually is performed in the dark without visual influence and with selected light stimuli. Eye movements are usually monitored using ENG.

Romberg Test. The Romberg test is used to demonstrate disorders of static vestibular function. The person being tested is requested to stand with feet together and arms extended forward so that the degree of sway and arm stability can be observed. The person then is asked to close his or her eyes. When visual clues are removed, postural stability is based on proprioceptive sensation from the joints, muscles, and tendons and from static vestibular reception. Deficiency in vestibular static input is indicated by greatly increased sway and a tendency for the arms to drift toward the side of deficiency. If vestibular input is severely deficient, the subject falls toward the deficient side.

Treatment of Vestibular Disorders

Pharmacologic Treatment. Depending on the cause, vertigo may be treated pharmacologically. There are two types of drugs used in the treatment of vertigo. First are the drugs used to suppress the illusion of motion. These include drugs such as antihistamines (*e.g.*, meclizine [Antivert], cyclizine [Marezine], dimenhydrinate [Dramamine], and promethazine [Phenergan]) and anticholinergic drugs (*e.g.*, scopolamine, atropine) that suppress the vestibular system. Although the antihistamines have long been used in treating vertigo, little is known about their mechanism of action. The second type includes drugs used to relieve the nausea and vomiting that commonly accompany the condition. Antidopaminergic drugs (*e.g.*, phenothiazines) and benzodiazepines commonly are used for this purpose.

Vestibular Rehabilitation. Vestibular rehabilitation, a relatively new treatment modality for peripheral vestibular disorders, has met with considerable success.^{50,51} It commonly is done by physical therapists and uses a home exercise program that incorporates habituation exercises, balance retraining exercises, and a general conditioning program.⁵¹ The habituation exercises take advantage of physiologic fatigue of the neurovegetative response to repetitive movement or positional stimulation and are done to decrease motion-provoked vertigo, light-headedness, and unsteadiness. The exercises are selected to provoke the vestibular symptoms. The person moves quickly into the position that causes symptoms, holds the position until the symptoms subside (*i.e.*, fatigue of the neurovegetative response), relaxes, and then repeats the exercise for a prescribed number of times. The exercises usually are repeated twice daily.

The habituation effect is characterized by decreased sensitivity and duration of symptoms. It may occur in as little as 2 weeks or take as long as 6 months.⁵¹

Balance-retraining exercises consist of activities directed toward improving individual components of balance that may be abnormal. General conditioning exercises, a vital part of the rehabilitation process, are individualized to the person's preferences and lifestyle. They should consist of motion-oriented activity that the person is interested in and should be done on a regular basis, usually four to five times per week.⁵¹

In summary, the vestibular system plays an essential role in the equilibrium sense, which is closely integrated with the visual and proprioceptive (position) senses. Receptors for the vestibular system, in the semicircular ducts of the inner ear, respond to changes in linear and angular acceleration of the head. The vestibular nerve fibers travel in CN VIII to the vestibular nuclei at the junction of the medulla and pons; some fibers pass through the nuclei to the cerebellum.

Disorders of peripheral vestibular function, which involve the inner ear sensory organs, include Ménière's disease and benign paroxysmal positional vertigo. Ménière's disease, which is caused by an overaccumulation of endolymph, is characterized by severe, disabling episodes of tinnitus, feelings of ear fullness, and violent rotary vertigo. Benign paroxysmal positional vertigo is thought to be caused by free-floating particles in the posterior semicircular canal. It presents as a sudden onset of dizziness or vertigo that is provoked by certain changes in head position. Among the methods used in treatment of the vertigo that accompanies vestibular disorders are habituation exercises (for BPPV) and antivertigo drugs.

REVIEW QUESTIONS

Describe the formation and outflow of aqueous humor from the eye and relate it to the development of glaucoma and to the pathogenesis of closed-angle and open-angle glaucoma.

Explain the difference between myopia and hyperopia.

Describe the changes in eye structure that occur with cataract.

Describe the pathogenesis of background and proliferative diabetic retinopathies and their mechanisms of visual impairment.

Discuss the cause of retinal detachment.

Explain the pathology and visual changes associated with macular degeneration.

Characterize what is meant by a visual field defect.

Explain the need for early diagnosis and treatment of eye movement disorders in children.

Relate the functions of the eustachian tube to the development of middle ear problems, including acute otitis media and otitis media with effusion.

Describe anatomic variations as well as risk factors that make infants and young children more prone to acute otitis media. Differentiate among conductive, sensorineural, and mixed hearing loss and cite the more common causes of each.

Define the term *presbycusis* and describe factors that contribute to its development.

Compare the manifestations and pathologic processes associated with benign positional vertigo and Ménière's disease.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- 1. Leonard R. (1999). Statistics on visual impairment: A resource manual. Lighthouse International. [On-line]. Available: http://www. lighthouse.org.
- Vaughan D.G., Ashbury T., Riordan-Eva P. (1999). *General ophthalmology* (15th ed., pp. 92–103, 119–140, 160, 188, 200–212, 216–233). Norwalk, CT: Appleton & Lange.
- Morrow G.L., Abbott R.L. (1998). Conjunctivitis. American Family Physician 57(4), 735–746.
- 4. Hara J.H. (1996). The red eye: Diagnosis and treatment. *American Family Physician* 54, 2423–2430.
- Riordan-Eva P., Vaughan D.G. (2001). Eye. In Tierney L.M., McPhee S.J., Papadakis M.A. (Eds.), *Current medical diagnosis and treatment* (40th ed., pp. 185–216). New York: Lange Medical Books/McGraw-Hill.
- Klintworth G.K. (1999). The eye. In Rubin E., Farber J.L. (Eds.), *Pathology* (3rd ed., pp. 1537–1563). Philadelphia: Lippincott Williams & Wilkins.
- Olitsky S.E., Nelson L. (2000). Disorders of the conjunctiva. In Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.), *Nelson textbook of pediatrics* (16th ed., pp. 1911–1913, 1918, 1925–1926). Philadelphia: W.B. Saunders.
- Evans N.M. (1995). Ophthalmology (2nd ed., pp. 43–44, 68–69, 113–116, 205). New York: Oxford University Press.
- Rosenberg L.F. (1995). Glaucoma: Early detection and therapy for prevention of vision loss. *American Family Physician* 52, 2289–2298.
- Agency for Health Care Policy and Research, Cataract Management Guideline Panel. (1993). *Clinical practice guideline 4. Cataract in adults: Management of functional impairment*. AHCPR no. 93-0543. Bethesda, MD: U.S. Department of Health and Human Services.
- Albert D.M., Dryja T.P. (1999). The eye. In Cotran R.S., Kumar V., Collins T. (Eds.), *Robbins pathologic basis of disease* (6th ed., pp. 1359– 1377). Philadelphia: W.B. Saunders.
- D'Amico D.J. (1994). Diseases of the retina. New England Journal of Medicine 331, 95–106.
- 13. Ferris F.L., Davis M.D., Aiello L.M. (1999). Treatment of diabetic retinopathy. *New England Journal of Medicine* 341, 667–678.
- 14. Klein R. (1994). Eye care delivery for people with diabetes. *Diabetes Care* 17, 614–615.
- Diabetes Control and Complications Trial Research Group. (1993). The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *New England Journal of Medicine* 329, 977–986.
- 16. Fine S.L., Berger J.W., MacGuire M., et al. (2000). Age-related macular degeneration. *New England Journal of Medicine* 342, 483–492.
- 17. Woods S. (1992). Macular degeneration. Nursing Clinics of North America 27, 755-761.
- 18. Mills M.D. (1999). The eye in childhood. *American Family Physician* 60, 907–918.
- Lavrich J.B., Nelson L.B. (1993). Diagnosis and management of strabismus disorders. *Pediatric Clinics of North America* 40, 737–751.

Chapter 40: Alterations in Special Sensory Function

- Kline L.B., Bajandas F.J. (1996). Neuro-ophthalmology review manual (Chapters 4–7). Thorofare, NJ: Slack.
- Broderick P. (1998). Pediatric vision screening for the family physician. American Family Practitioner 58 (3), 691–704.
- Rubein S.E., Nelson S.B. (1993). Amblyopia: Diagnosis and management. Pediatric Clinics of North America 40, 727–735.
- 23. Grossan M. (2000). Safe, effective techniques for cerumen removal. *Geriatrics* 55, 83–86.
- Jackler R.K., Kaplan M.J. (2001). Diseases of the ear. In Tierney L.M., McPhee S.J., Papadakis M.A. (Eds.), *Current medical diagnosis and treatment* (40th ed., pp. 217–231). New York: Lange Medical Books/ McGraw-Hill.
- 25. Bluestone C.D., Klein J. (1995). Otitis media in infants and children. Philadelphia: W.B. Saunders.
- Swanson J.A., Hoecker J.L. (1996). Otitis media in young children. Mayo Clinic Proceedings 71, 179–183.
- Kenna M. (2000). The ear. In Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.), *Nelson textbook of pediatrics* (16th ed., pp. 1951–1959). Philadelphia: W.B. Saunders.
- Pichichero, M.E. (2000). Acute otitis media: Part II. Treatment in an era of increasing antibiotic resistance. *American Family Physician* 61, 2410–2415.
- Berman S. (1995). Otitis media in children. New England Journal of Medicine 332, 1560–1565.
- 30. Dowell S.F., Butler J.C., Giebink G.S., et al. (1999). Acute otitis media: Management and surveillance in an era of pneumococcal resistance a report from the Drug Resistant *Streptococcus pneumoniae* Therapeutic Working Group. *Pediatric Infectious Disease Journal* 18 (1), 1–9.
- Pichichero M.E. (2000). Acute otitis media: Part I. Improving diagnostic accuracy. American Family Physician 61, 2051–2056.
- 32. Rubin E., Farber J.L. (1999). *Pathology* (3rd ed., pp. 1331–1332). Philadelphia: Lippincott Williams & Wilkins.
- Fortune D.S. (1999). Tinnitus: Current evaluation and management. Medical Clinics of North America 83, 153–162.
- 34. Weissman J.L. (1996). Hearing loss. Radiology 199, 593-611.
- Shohet J.A., Bent T. (1998). Hearing loss: The invisible disability. Postgraduate Medicine 104 (3), 81–83, 87–90.

- 36. Willems P.J. (2000). Genetic causes of hearing loss. *New England Journal of Medicine* 342, 1101–1109.
- Kenna M. (2000). Hearing loss. In Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.), *Nelson textbook of pediatrics* (16th ed., pp. 1940– 1947). Philadelphia: W.B. Saunders.
- Nadol J.G. (1993). Hearing loss. New England Journal of Medicine 329, 1092–1101.
- Yamasoba T., Kikuchi S., O'uchi T., et al. (1993). Sudden sensorineural hearing loss associated with slow blood flow of the vertebrobasilar system. *Annals of Otorhinolaryngology* 102, 873–877.
- Gates G.A. (Chairperson). (1989). Invitational Geriatric Otorhinolaryngology Workshop: Presbycusis. Otolaryngology—Head and Neck Surgery 100, 266–271.
- 41. Saeed S., Ramsden R. (1994). Hearing loss. Practitioner 238, 454-460.
- Task Force on Newborn and Infant Hearing of the American Academy of Pediatrics. (1999). Newborn and infant hearing loss: Detection and intervention. *Pediatrics* 103, 527–530.
- Derebery J.M. (1999). The diagnosis and treatment of dizziness. Medical Clinics of North America 83, 163–176.
- Ruckenstein M.J. (2001). The dizzy patient: How you can help. Consultant 41 (1), 29–33.
- Derebery M.J. (1999). The diagnosis and treatment of dizziness. Medical Clinics of North America 83, 163–176.
- Dickins J.R.E., Graham S.S. (1990). Ménière's disease: 1983–1989. American Journal of Otology 11, 51–65.
- Saeed S.R. (1998). Fortnightly review: Diagnosis and treatment of Ménière's disease. British Medical Journal 316, 368–372.
- Brooks C.B. (1996). The pharmacological treatment of Ménière's disease. Clinical Otolaryngology 21, 3–11.
- Furman J.M., Cass S.P. (1999). Benign paroxysmal positional vertigo. New England Journal of Medicine 341, 1590–1596.
- Horak F.B., Jones-Rycewicz C., Black F.W., et al. (1992). Effects of vestibular rehabilitation on dizziness and imbalance. *Otolaryngology— Head and Neck Surgery* 106, 175–180.
- Brandt T. (2000). Management of vestibular disorders. Journal of Neurology 247, 491–499.

Eleven

Alterations in the Skeletal and Integumentary Systems

CHAPTER

Structure and Function of the Skeletal System

Characteristics of Skeletal Tissue Cartilage Bone Types of Bone **Bone Cells** Periosteum and Endosteum Bone Growth and Remodeling Hormonal Control of Bone Formation and Metabolism Parathyroid Hormone Calcitonin Vitamin D **Skeletal Structures Classification of Bones Tendons and Ligaments** Joints and Articulations **Synarthroses Diarthroses Blood Supply and Innervation Bursae** Intra-articular Menisci

ithout the skeletal system, movement in the external environment would not be possible. The bones of the skeletal system serve as a framework for the attachment of muscles, tendons, and ligaments. The skeletal system protects and maintains soft tissues in their proper position, provides stability for the body, and maintains the body's shape. The bones act as a storage reservoir for calcium, and the central cavity of some bones contains the hematopoietic connective tissue in which blood cells are formed.

The skeletal system consists of the axial and appendicular skeleton. The axial skeleton, which is composed of the bones of the skull, thorax, and vertebral column, forms the axis of the body. The appendicular skeleton consists of the bones of the upper and lower extremities, including the shoulder and hip. For our purposes, the skeletal system is considered to include the bones and cartilage of the axial and appendicular skeleton, as well as the connective tissue structures (*i.e.*, ligaments and tendons) that connect the bones and join muscles to bone.

CHARACTERISTICS OF SKELETAL TISSUE

Two types of connective tissue are found in the skeletal system: cartilage and bone. Each of these connective tissue types consists of living cells, nonliving intercellular protein fibers, and an amorphous (shapeless) ground substance. The tissue cells are responsible for secreting and maintaining the intercellular substances in which they are housed. These substances provide the structural characteristics of the tissue. For example,

KEY CONCEPTS

THE SKELETAL SYSTEM

- The skeletal system consists of the bones of the skull, thorax, and vertebral column, which form the axial skeleton, and the bones of the upper and lower extremities, which form the appendicular skeleton.
- Two types of connective tissue are found in the skeletal system: (1) cartilage, a semirigid and slightly flexible structure that plays an essential role in prenatal and childhood development of the skeleton and as a surface for the articulating ends of skeletal joints; and (2) bones, which provide for the firm structure of the skeleton and serve as a reservoir for calcium and phosphate storage.
- Both bone and cartilage are composed of living cells and a nonliving intercellular matrix that is secreted by the living cells.
- Bone matrix is maintained by three types of cells: osteoblasts, which synthesize and secrete the constituents of bone; osteoclasts, which resorb surplus bone and are required for bone remodeling; and the osteocytes, which make up the osteoid tissue of bone.

the intercellular matrix of bone is impregnated with calcium salts, providing the hardness that is characteristic of this tissue.

Two main types of intercellular fibers are found in skeletal tissue: collagenous and elastic. Collagen is an inelastic and insoluble fibrous protein. Because of its molecular configuration, collagen has great tensile strength; the breaking point of collagenous fibers found in human tendons is reached with a force of several hundred kilograms per square centimeter. Fresh collagenous fibers generally appear white. The collagen fibers in tendons and ligaments give these structures their white color. Elastin is the major component of elastic fibers that allows them to stretch several times their length and rapidly return to their original shape when the tension is released. Ligaments and structures that must undergo repeated stretching contain a high proportion of elastic fibers.

Cartilage

Cartilage is a firm but flexible type of connective tissue consisting of cells and intercellular fibers embedded in an amorphous, gel-like material. It has a smooth and resilient surface and a weight-bearing capacity exceeded only by that of bone.

Cartilage is essential for growth before and after birth. It is able to undergo rapid growth while maintaining a considerable degree of stiffness. In the embryo, most of the axial and appendicular skeleton is formed first as a cartilage model and is replaced by bone. In postnatal life, cartilage continues to play an essential role in the growth of long bones and persists as articular cartilage in the adult. There are three types of cartilage: elastic cartilage, hyaline cartilage, and fibrocartilage. *Elastic cartilage* contains some elastin in its intercellular substance. It is found in areas, such as the ear, where some flexibility is important. Pure cartilage is called *hyaline cartilage* (from a Greek word meaning "glass") and is pearly white. It is the type of cartilage seen on the articulating ends of fresh soup bones found in the supermarket. *Fibrocartilage* has characteristics that are intermediate between dense connective tissue and hyaline cartilage. It is found in the intervertebral disks, in areas where tendons are connected to bone, and in the symphysis pubis.

Hyaline cartilage is the most abundant type of cartilage. It forms much of the cartilage of the fetal skeleton. In the adult, hyaline cartilage forms the costal cartilages that join the ribs to the sternum and vertebrae, many of the cartilages of the respiratory tract, the articular cartilages, and the epiphyseal plates.

Cartilage cells, which are called *chondrocytes*, are located in lacunae. These lacunae are surrounded by an uncalcified, gellike intercellular matrix of collagen fibers and ground substance. Cartilage is devoid of blood vessels and nerves. The free surfaces of most hyaline cartilage, with the exception of articular cartilage, are covered by a layer of fibrous connective tissue called the *perichondrium*.

It has been estimated that approximately 65% to 80% of the wet weight of cartilage is water held in its gel structure. Because cartilage has no blood vessels, this tissue fluid allows the diffusion of gases, nutrients, and wastes between the chondrocytes and blood vessels outside the cartilage. Diffusion cannot take place if the cartilage matrix becomes impregnated with calcium salts, and cartilage dies if it becomes calcified.

Bone

Bone is connective tissue in which the intercellular matrix has been impregnated with inorganic calcium salts so that it has great tensile and compressible strength but is light enough to be moved by coordinated muscle contractions. The intercellular matrix is composed of two types of substances—organic matter and inorganic salts. The organic matter, including bone cells, blood vessels, and nerves, constitutes approximately one third of the dry weight of bone; the inorganic salts make up the other two thirds.

The organic matter consists primarily of collagen fibers embedded in an amorphous ground substance. The inorganic matter consists of hydroxyapatite, an insoluble macrocrystalline structure of calcium phosphate salts, and small amounts of calcium carbonate and calcium fluoride. Bone may also take up lead and other heavy metals, thereby removing these toxic substances from the circulation. This can be viewed as a protective mechanism. The antibiotic tetracycline is readily bound to calcium deposited in newly formed bones and teeth. When tetracycline is given during pregnancy, it can be deposited in the teeth of the fetus, causing discoloration and deformity. Similar changes can occur if the drug is given for long periods to children younger than 6 years of age.

Types of Bone

There are two types of mature bones, cancellous and compact bone (Fig. 41-1). Both types are formed in layers and thus are called *lamellar bone*. Cancellous (spongy) bone is found in the

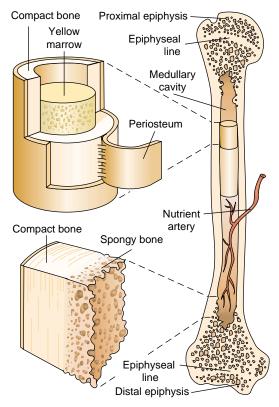


FIGURE 41-1 A long bone shown in longitudinal section.

interior of bones and is composed of *trabeculae*, or *spicules*, of bone that form a latticelike pattern. These latticelike structures are lined with osteogenic cells and filled with red or yellow bone marrow. Cancellous bone is relatively light, but its structure is such that it has considerable tensile strength and weight-bearing properties. Compact (cortical) bone, which forms the outer shell of a bone, has a densely packed calcified intercellular matrix that makes it more rigid than cancellous bone. The relative quantity of compact and cancellous bone varies in different types of bones throughout the body and in different parts of the same bone, depending on the need for strength and lightness. Compact bone is the major component of tubular bones. It is also found along the lines of stress on long bones and forms an outer protective shell on other bones.

Bone Cells

Four types of bone cells participate in the formation and maintenance of bone tissue: osteogenic cells, osteoblasts, osteocytes, and osteoclasts (Table 41-1).

Osteogenic Cells. The undifferentiated osteogenic cells are found in the periosteum, endosteum, and epiphyseal plate of growing bone. These cells differentiate into osteoblasts and are active during normal growth; they may also be activated in adult life during healing of fractures and other injuries. Osteogenic cells also participate in the continual replacement of worn-out bone tissue.

Osteoblasts. The osteoblasts, or bone-building cells, are responsible for the formation of the bone matrix. Bone formation

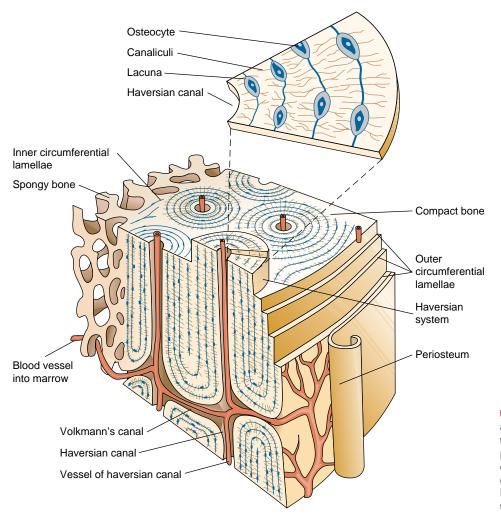
TABLE 41-1	Function of Bone Cells	
Type of Bone Cell	Function	
Osteogenic cells	Undifferentiated cells that differentiate into osteoblasts. They are found in the perios- teum, endosteum, and epiphyseal growth plate of growing bones.	
Osteoblasts	Bone-building cells that synthesize and se- crete the organic matrix of bone. Osteo- blasts also participate in the calcification of the organic matrix.	
Osteocytes	Mature bone cells that function in the main- tenance of bone matrix. Osteocytes also play an active role in releasing calcium into the blood.	
Osteoclasts	Bone cells responsible for the resorption of bone matrix and the release of calcium and phosphate from bone.	

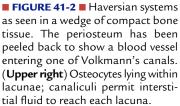
occurs in two stages: ossification and calcification. Ossification involves the formation of osteoid, or prebone. Calcification of bone involves the deposition of calcium salts in the osteoid tissue. The osteoblasts synthesize collagen and other proteins that make up osteoid tissue. They also participate in the calcification process of the osteoid tissue, probably by controlling the availability of calcium and phosphate. Osteoblasts secrete the enzyme *alkaline phosphatase*, which is thought to act locally in bone tissue to raise calcium and phosphate levels to the point at which precipitation occurs. The activity of the osteoblasts undoubtedly contributes to the increase in serum levels of alkaline phosphatase that follows bone injury and fractures.

Osteocytes. The osteocytes are mature bone cells that are actively involved in maintaining the bony matrix. Death of the osteocytes results in the resorption of this matrix. The osteocytes lie in a small lake filled with extracellular fluid, called a *lacuna*, and are surrounded by a calcified intercellular matrix. Extracellular fluid-filled passageways permeate the calcified matrix and connect with the lacunae of adjacent osteocytes. These passageways are called *canaliculi*. Because diffusion does not occur through the calcified matrix of bone, the canaliculi serve as communicating channels for the exchange of nutrients and metabolites between the osteocytes and the blood vessels on the surface of the bone layer.

The osteocytes, together with their intercellular matrix, are arranged in layers, or lamellae. In compact bone, 4 to 20 lamellae are arranged concentrically around a central haversian canal, which runs essentially parallel to the long axis of the bone. Each of these units is called a *haversian system*, or *osteon*. The haversian canals contain blood vessels that carry nutrients and wastes to and from the canaliculi (Fig. 41-2). The blood vessels from the periosteum enter the bone through tiny openings called *Volkmann's canals* and connect with the haversian systems. Cancellous bone is also composed of lamellae, but its trabeculae usually are not penetrated by blood vessels. Instead, the bone cells of cancellous bone are nourished by diffusion from the endosteal surface through canaliculi, which interconnect their lacunae and extend to the bone surface.

782 Unit Eleven: Alterations in the Skeletal and Integumentary Systems





Osteoclasts. Osteoclasts are "bone-chewing" cells that function in the resorption of bone, removing the mineral content and the organic matrix. They are large phagocytic cells of monocyte/ macrophage lineage. Although the mechanism of osteoclast formation and activation remains elusive, it is known that parathyroid hormone (PTH) increases the number and resorptive function of the osteoclasts. Calcitonin is thought to reduce the number and resorptive function of the osteoclasts. Estrogen also reduces the number and function of the osteoclasts; thus, the decrease in estrogen levels that occur at menopause results in increased reabsorption of bone. The mechanism whereby osteoclasts exert their resorptive effect on bone is unclear. These cells may secrete an acid that removes calcium from the bone matrix, releasing the collagenic fibers for digestion by osteoclasts or mononuclear cells. The osteoclastic cells, by virtue of their phagocytic lineage, also imbibe minute particles of bone matrix and crystals, eventually dissolving and releasing them into the blood.

Periosteum and Endosteum

Bones are covered, except at their articular ends, by a membrane called the *periosteum* (see Fig. 41-1). The periosteum has an outer fibrous layer and an inner layer that contains the osteogenic cells needed for bone growth and development. The peri-

osteum contains blood vessels and acts as an anchorage point for vessels as they enter and leave the bone. The endosteum is the membrane that lines the spaces of spongy bone, the marrow cavities, and the haversian canals of compact bone. It is composed mainly of osteogenic cells. These osteogenic cells contribute to the growth and remodeling of bone and are necessary for bone repair.

Bone Growth and Remodeling

The skeletal system develops from the mesoderm, the thin middle layer of embryonic tissue. Development of the vertebrae of the axial skeleton begins at approximately the fourth week in the embryo; during the ninth week, ossification begins with the appearance of ossification centers in the lower thoracic and upper lumbar vertebrae. The paddle-shaped limb buds of the lower extremities make their appearance late in the fourth week. The hand pads are developed by days 33 to 36, and the finger rays are evident on days 41 to 43 of embryonic development.¹

During the first two decades of life, the skeleton undergoes general overall growth. The long bones of the skeleton, which grow at a relatively rapid rate, are provided with a specialized structure called the *epiphyseal growth plate*. As long bones grow in length, the deeper layers of cartilage cells in the growth plate multiply and enlarge, pushing the articular cartilage farther away from the metaphysis and diaphysis of the bone.² As this happens, the mature and enlarged cartilage cells at the metaphyseal end of the plate become metabolically inactive and are replaced by bone cells (Fig. 41-1). This process allows bone growth to proceed without changing the shape of the bone or causing disruption of the articular cartilage. The cells in the growth plate stop dividing at puberty, at which time the epiphysis and metaphysis fuse.

Several factors can influence the growth of cells in the epiphyseal growth plate. Epiphyseal separation can occur in children as the result of trauma. The separation usually occurs in the zone of the mature enlarged cartilage cells, which is the weakest part of the growth plate. The blood vessels that nourish the epiphysis pass through the growth plate. These vessels are ruptured when the growth plate separates. This can cause cessation of growth and a shortened extremity.

The growth plate also is sensitive to nutritional and metabolic changes. Scurvy (*i.e.*, vitamin C deficiency) impairs the formation of the organic matrix of bone, causing slowing of growth at the epiphyseal plate and cessation of diaphyseal growth. In rickets (*i.e.*, vitamin D deficiency), calcification of the newly developed bone on the metaphyseal side of the growth plate is impaired. Thyroid and growth hormones are required for normal growth. Alterations in these and other hormones can affect growth (see Chapter 31).

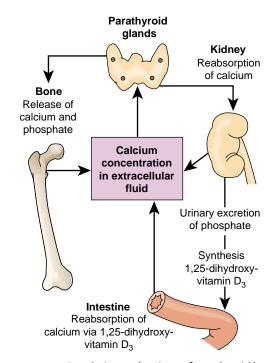
Growth in the diameter of bones occurs as new bone is added to the outer surface of existing bone along with an accompanying resorption of bone on the endosteal or inner surface. Such oppositional growth allows for widening of the marrow cavity while preventing the cortex from becoming too thick and heavy. In this way, the shape of the bone is maintained. As a bone grows in diameter, concentric rings are added to the bone surface, much as rings are added to a tree trunk; these rings form the lamellar structure of mature bone. Osteocytes, which develop from osteoblasts, become buried in the rings. Haversian channels form as periosteal vessels running along the long axis become surrounded by bone.

Hormonal Control of Bone Formation and Metabolism

The process of bone formation and mineral metabolism is complex. It involves the interplay among the actions of PTH, calcitonin, and vitamin D. Other hormones, such as cortisol, growth hormone, thyroid hormone, and the sex hormones, also influence bone formation directly or indirectly.

Parathyroid Hormone

PTH is one of the important regulators of calcium and phosphate levels in the blood. PTH prevents serum calcium levels from falling below and serum phosphate levels from rising above normal physiologic concentrations (see Chapter 31). The secretion of PTH is regulated by negative feedback levels of ionized calcium. PTH maintains serum calcium levels by initiation of calcium release from bone, by conservation of calcium by the kidney, by enhanced intestinal absorption of calcium through activation of vitamin D, and by reduction of serum phosphate levels (Fig. 41-3). PTH also increases the movement of calcium and phosphate from bone into the



■ FIGURE 41-3 ■ Regulation and actions of parathyroid hormone.

extracellular fluid. Calcium is immediately released from the canaliculi and bone cells; a more prolonged release of calcium and phosphate is mediated by increased osteoclast activity. In the kidney, PTH stimulates tubular reabsorption of calcium while reducing the reabsorption of phosphate. The latter effect ensures that increased release of phosphate from bone during mobilization of calcium does not produce an elevation in serum phosphate levels. This is important because an increase in calcium and phosphate levels could lead to crystallization in soft tissues. PTH increases intestinal absorption of calcium because of its ability to stimulate activation of vitamin D by the kidney.

Calcitonin

Whereas PTH increases blood calcium levels, the hormone calcitonin lowers blood calcium levels. Calcitonin, sometimes called *thyrocalcitonin*, is secreted by the parafollicular, or C, cells of the thyroid gland. Calcitonin inhibits the release of calcium from bone into the extracellular fluid. It is thought to act by causing calcium to become sequestered in bone cells and by inhibiting osteoclast activity. Calcitonin also reduces the renal tubular reabsorption of calcium and phosphate; the decrease in serum calcium level that follows administration of pharmacologic doses of calcitonin may be related to this action.

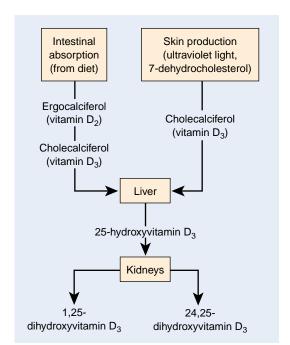
The major stimulus for calcitonin synthesis and release is an increase in serum calcium. The role of calcitonin in overall mineral homeostasis is uncertain. There are no clearly definable syndromes of calcitonin deficiency or excess, which suggests that calcitonin does not directly alter calcium metabolism. It has been suggested that the physiologic actions of calcitonin are related to the postprandial handling and processing of dietary calcium. This theory proposes that after meals, calcitonin maintains parathyroid secretion at a time when it normally would be reduced by calcium entering the blood from the digestive tract. Although excess or deficiency states associated with alterations in physiologic levels of calcitonin have not been observed, it has been shown that pharmacologic doses of the hormone reduce osteoclastic activity. Because of this action, calcitonin has proved effective in the treatment of Paget's disease (see Chapter 43). The hormone is also used to reduce serum calcium levels during hypercalcemic crises.

Vitamin D

Vitamin D and its metabolites are not vitamins but steroid hormones. There are two forms of vitamin D: vitamin D₂ (ergocalciferol) and vitamin D₃ (cholecalciferol). The two forms differ by the presence of a double bond, but they have identical biologic activity. The term *vitamin* D is used to indicate both forms.

Vitamin D has little or no activity until it has been metabolized to compounds that mediate its activity. Figure 41-4 depicts sources of vitamin D and pathways for activation. The first step of the activation process occurs in the liver, where vitamin D is hydroxylated to form the metabolite 25-hydroxyvitamin D₃ [25-(OH)D₃]. From the liver, 25-(OH)D₃ is transported to the kidneys, where it undergoes conversion to 1,25-dihydroxyvitamin D₃ [1,25-(OH)₂D₃] or 24,25-dihydroxyvitamin D₃ [24,25-(OH)₂D₃]. Other metabolites of vitamin D have been and still are being discovered.

There are two sources of vitamin D: intestinal absorption and skin production. Intestinal absorption occurs mainly in the jejunum and includes vitamin D_2 and vitamin D_3 . The most important dietary sources of vitamin D are fish, liver, and irradiated milk. Because vitamin D is fat soluble, its absorption is mediated by bile salts and occurs by means of the lymphatic vessels. In the skin, ultraviolet radiation from sunlight spontaneously converts 7-dehydrocholesterol provitamin D_3 to vitamin D_3 . A circulating vitamin D-binding protein provides a mechanism to remove vitamin D from the skin and make it available to the rest of the body.



■ FIGURE 41-4 ■ Sources and pathway for activation of vitamin D.

With adequate exposure to sunlight, the amount of vitamin D that can be produced by the skin is usually sufficient to meet physiologic requirements. The importance of sunlight exposure is evidenced by population studies that report lower vitamin D levels in countries, such as England, that have less sunlight than the United States. Elderly persons who are housebound or institutionalized frequently have low vitamin D levels. The deficiency often goes undetected until there are problems such as pseudofractures or electrolyte imbalances. Seasonal variations in vitamin D levels probably reflect changes in sunlight exposure.

The most potent of the vitamin D metabolites is 1,25- $(OH)_2D_3$. This metabolite increases intestinal absorption of calcium and promotes the actions of PTH on resorption by the osteoclasts is increased, and bone formation by the osteoblasts is decreased; there is also an increase in acid phosphatase and a decrease in alkaline phosphatase. Intestinal absorption and bone resorption increase the amount of calcium and phosphorus available to the mineralizing surface of the bone. The role of 24,25-(OH)_2D_3 is less clear. There is evidence that 24,25-(OH)_2D_3, in conjunction with 1,25-(OH)_2D_3, may be involved in normal bone mineralization.

The regulation of vitamin D activity is influenced by several hormones. PTH and prolactin stimulate $1,25-(OH)_2D_3$ production by the kidney. States of hyperparathyroidism are associated with increased levels of $1,25-(OH)_2D_3$, and hypoparathyroidism leads to lowered levels of this metabolite. Prolactin may have an ancillary role in regulating vitamin D metabolism during pregnancy and lactation. Calcitonin inhibits $1,25-(OH)_2D_3$ production by the kidney. In addition to hormonal influences, changes in the concentration of ions such as calcium, phosphate, hydrogen, and potassium exert an effect on $1,25-(OH)_2D_3$ and $24,25-(OH)_2D_3$ production. Under conditions of deprivation of phosphate and calcium, $1,25-(OH)_2D_3$ levels are increased, whereas hyperphosphatemia and hypercalcemia decrease the levels of metabolite.

In summary, skeletal tissue is composed of two types of connective tissue: cartilage and bone. These skeletal structures are composed of similar tissue types; each has living cells and nonliving intercellular fibers and ground substance that is secreted by the cells. Cartilage is a firm, flexible type of skeletal tissue that is essential for growth before and after birth. There are three types of cartilage: elastic, hyaline, and fibrocartilage. Hyaline cartilage, which is the most abundant type, forms the costal cartilages that join the ribs to the sternum and vertebrae, many of the cartilages of the respiratory tract, and the articular cartilages.

The characteristics of the various skeletal tissue types are determined by the intercellular matrix. In bone, this matrix is impregnated with calcium salts to provide hardness and strength. There are four types of bone cells: osteocytes, or mature bone cells; osteoblasts, or bone-building cells; osteoclasts, which function in bone resorption; and osteogenic cells, which differentiate into osteoblasts. Densely packed compact bone forms the outer shell of a bone, and latticelike cancellous bone forms the interior. The periosteum, the membrane that covers bones, contains blood vessels and acts as an anchorage point for vessels as they enter and leave the bone. The endosteum is the membrane that lines the spaces of spongy bone, the marrow cavities, and the haversian canals of compact bone.

The process of bone formation and mineral metabolism involves the interplay among the actions of PTH, calcitonin, and vitamin D. PTH acts to maintain serum levels of ionized calcium; it increases the release of calcium and phosphate from bone, the conservation of calcium and elimination of phosphate by the kidney, and the intestinal reabsorption of calcium through vitamin D. Calcitonin inhibits the release of calcium from bone and increases renal elimination of calcium and phosphate, thereby serving to lower serum calcium levels. Vitamin D functions as a hormone in regulating body calcium. It increases absorption of calcium from the intestine and promotes the actions of PTH on bone.

SKELETAL STRUCTURES

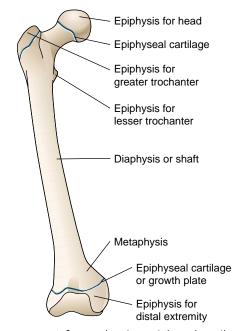
Classification of Bones

Bones are classified by shape as long, short, flat, and irregular. Long bones are found in the upper and lower extremities. Short bones are irregularly shaped bones located in the ankle and the wrist. Except for their surface, which is compact bone, these bones are spongy throughout. Flat bones are composed of a layer of spongy bone between two layers of compact bone. They are found in areas such as the skull and rib cage, where extensive protection of underlying structures is needed, or, as in the scapula, where a broad surface for muscle attachment must be provided. Irregular bones, because of their shapes, cannot be classified in any of the previous groups. This group includes bones such as the vertebrae and the bones of the jaw.

A typical long bone has a shaft, or *diaphysis*, and two ends, called *epiphyses*. Long bones usually are narrow in the midportion and broad at the ends so that the weight they bear can be distributed over a wider surface. The shaft of a long bone is formed mainly of compact bone roughly hollowed out to form a marrow-filled medullary canal. The ends of long bones are covered with articular cartilage that rests on a bony plate, the subchondral bone.

In growing bones, the part of the bone shaft that funnels out as it approaches the epiphysis is called the *metaphysis* (Fig. 41-5). It is composed of bony trabeculae that have cores of cartilage. In the child, the epiphysis is separated from the metaphysis by the cartilaginous growth plate. After puberty, the metaphysis and epiphysis merge, and the growth plate is obliterated.

Bone marrow occupies the medullary cavities of the long bones throughout the skeleton and the cavities of cancellous bone in the vertebrae, ribs, sternum, and flat bones of the pelvis. The cellular composition of the bone marrow varies with age and skeletal location. Red bone marrow contains developing red blood cells and is the site of blood cell formation. Yellow bone marrow is composed largely of adipose cells. At birth, nearly all of the marrow is red and hematopoietically active. As the need for red blood cell production decreases during postnatal growth, red marrow is gradually replaced with yellow



■ FIGURE 41-5 ■ A femur, showing epiphyseal cartilages for the head, metaphysis, trochanters, and distal end of the bone.

bone marrow in most of the bones. In the adult, red marrow persists in the vertebrae, ribs, sternum, and ilia.

Tendons and Ligaments

In the skeletal system, tendons and ligaments are dense connective tissue structures that connect muscles and bones. Tendons connect muscles to bone, and ligaments connect the movable bones of joints. Tendons can appear as cordlike structures or as flattened sheets, called *aponeuroses*, such as in the abdominal muscles.

The dense connective tissue found in tendons and ligaments has a limited blood supply and is composed largely of intercellular bundles of collagen fibers arranged in the same direction and plane. This type of connective tissue provides great tensile strength and can withstand tremendous pull in the direction of fiber alignment. At the sites where tendons or ligaments are inserted into cartilage or bone, a gradual transition from pure dense connective tissue to bone or cartilage occurs. In cartilage, this transitional tissue is called *fibrocartilage*.

Tendons that may rub against bone or other frictiongenerating surfaces are enclosed in double-layered sheaths. An outer connective tissue tube is attached to the structures surrounding the tendon, and an inner sheath encloses the tendon and is attached to it. The space between the inner and outer sheath is filled with a fluid similar to synovial fluid.

Joints and Articulations

Articulations, or joints, are areas where two or more bones meet. The term *arthro* is the prefix used to designate a joint. For example, *arthrology* is the study of joints, and *arthroplasty* is the repair of a joint. There are two classes of joints, based on movement and the presence of a joint cavity: synarthroses and diarthroses.

KEY CONCEPTS

SKELETAL JOINTS

- Joints, or articulations, are sites where two or more bones meet to hold the skeleton together and give it mobility.
- There are two types of joints: synarthroses, which are immovable joints, and diarthroses, which are freely movable joints.
- All limb joints are synovial diarthrodial joints, which are enclosed in a joint cavity containing synovial fluid.
- The articulating surfaces of synovial joints are covered with a layer of avascular cartilage that relies on oxygen and nutrients contained in the synovial fluid.
- Regeneration of articular cartilage of synovial joints is slow, and the healing of injuries often is slow and unsatisfactory.

Synarthroses

Synarthroses are joints that lack a joint cavity and move little or not at all. There are three types of synarthroses: synostoses, synchondroses, and syndesmoses. *Synostoses* are nonmovable joints in which the surfaces of the bones are joined by dense connective tissue or bone. The bones of the skull are joined by synostoses; they are joined by dense connective tissue in children and young adults and by bone in older persons. *Synchondroses* are joints in which bones are connected by hyaline cartilage and have limited motion. The ribs are attached to the sternum by this type of joint. *Syndesmoses* permit a certain amount of movement; they are separated by a fibrous disk and joined by interosseous ligaments. The symphysis pubis of the pelvis and the bodies of the vertebrae that are joined by intervertebral disks are examples of syndesmoses.

Diarthroses

Diarthrodial joints (*i.e.*, synovial joints) are freely movable joints. Most joints in the body are of this type. Although they are classified as freely movable, their movement ranges from almost none (*e.g.*, sacroiliac joint), to simple hinge movement (*e.g.*, interphalangeal joint), to movement in many planes (*e.g.*, shoulder or hip joint). The bony surfaces of these joints are covered with thin layers of articular cartilage, and the cartilaginous surfaces of these joints slide past each other during movement. As discussed in Chapter 43, diarthrodial joints are the joints most frequently affected by rheumatic disorders.

In a diarthrodial joint, the articulating ends of the bones are not connected directly but are indirectly linked by a strong fibrous capsule (*i.e.*, joint capsule) that surrounds the joint and is continuous with the periosteum (Fig. 41-6). This capsule supports the joint and helps to hold the bones in place. Additional support may be provided by ligaments that extend between the bones of the joint.

The joint capsule consists of two layers: an outer fibrous layer and an inner membrane, the synovium. The synovium

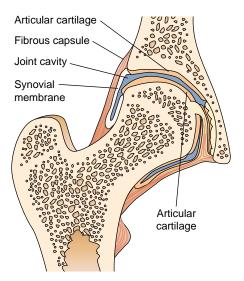


FIGURE 41-6 Diarthrodial joint, showing the articular cartilage, fibrous joint capsule, joint cavity, and synovial membrane.

surrounds the tendons that pass through the joints and the free margins of other intra-articular structures, such as ligaments and menisci. The synovium forms folds that surround the margins of articulations but do not cover the weightbearing articular cartilage. These folds permit stretching of the synovium so that movement can occur without tissue damage.

The synovium secretes a slippery fluid with the consistency of egg white called *synovial fluid*. This fluid acts as a lubricant and facilitates the movement of the articulating surfaces of the joint. Normal synovial fluid is clear or pale yellow, does not clot, and contains fewer than 100 cells/mm³. The cells are predominantly mononuclear cells derived from the synovium. The composition of the synovial fluid is altered in many inflammatory and pathologic joint disorders. Aspiration and examination of the synovial fluid play an important role in the diagnosis of joint diseases.

The articular cartilage is an example of hyaline cartilage and is unique in that its free surface is not covered with perichondrium. It has only a peripheral rim of perichondrium, and calcification of the portion of cartilage abutting the bone may limit or preclude diffusion from blood vessels supplying the subchondral bone. Articular cartilage is apparently nourished by the diffusion of substances contained in the synovial fluid bathing the cartilage. Regeneration of most cartilage is slow; it is accomplished primarily by growth that requires the activity of perichondrium cells. In articular cartilage, which has no perichondrium, superficial injuries heal slowly.

Blood Supply and Innervation

The blood supply to a joint arises from blood vessels that enter the subchondral bone at or near the attachment of the joint capsule and form an arterial circle around the joint. The synovial membrane has a rich blood supply, and constituents of plasma diffuse rapidly between these vessels and the joint cavity. Because many of the capillaries are near the surface of the synovium, blood may escape into the synovial fluid after relatively minor injuries. Healing and repair of the synovial membrane usually are rapid and complete. This is important because synovial tissue is injured in many surgical procedures that involve the joint.

The nerve supply to joints is provided by the same nerve trunks that supply the muscles that move the joints. These nerve trunks also supply the skin over the joints. As a rule, each joint of an extremity is innervated by all the peripheral nerves that cross the articulation; this accounts for the referral of pain from one joint to another. For example, hip pain may be perceived as pain in the knee.

The tendons and ligaments of the joint capsule are sensitive to position and movement, particularly stretching and twisting. These structures are supplied by the large sensory nerve fibers that form proprioceptor endings (see Chapter 38). The proprioceptors function reflexively to adjust the tension of the muscles that support the joint and are particularly important in maintaining muscular support for the joint. For example, when a weight is lifted, there is a proprioceptor-mediated reflex contraction and relaxation of appropriate muscle groups to support the joint and protect the joint capsule and other joint structures. Loss of proprioception and reflex control of muscular support leads to destructive changes in the joint.

The synovial membrane is innervated only by autonomic fibers that control blood flow. It is relatively free of pain fibers, as evidenced by the fact that surgical procedures on the joint are often done under local anesthesia. The joint capsule and the ligaments have pain receptors; these receptors are more easily stimulated by stretching and twisting than are other joint structures. Pain arising from the capsule tends to be diffuse and poorly localized.

Bursae

In some diarthrotic joints, the synovial membrane forms closed sacs that are not part of the joint. These sacs, called *bursae*, contain synovial fluid. Their purpose is to prevent friction on a tendon. Bursae occur in areas where pressure is exerted because of close approximation of joint structures (Fig. 41-7). Such conditions occur when tendons are deflected over bone or where skin must move freely over bony tissue. Bursae may

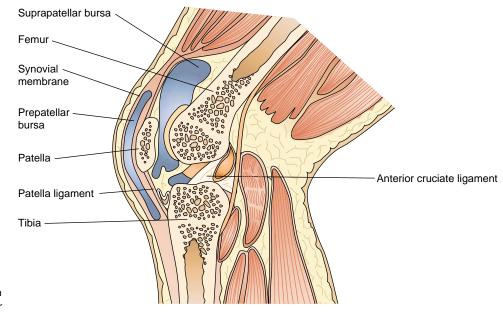
become injured or inflamed, causing discomfort, swelling, and limitation in movement of the involved area. A bunion is an inflamed bursa of the metatarsophalangeal joint of the great toe.

Intra-articular Menisci

Intra-articular menisci are fibrocartilage structures that develop from portions of the articular disk that occupied the space between articular cartilage surfaces during fetal development. Menisci may extend part way through the joint and have a free inner border, as at the lateral and medial articular surfaces of the knee, or they may extend through the joint, separating it into two separate cavities, as in the sternoclavicular joint. The menisci of the knee joint may be torn as the result of an injury (see Chapter 42).

In summary, bones are classified on the basis of their shape as long, short, flat, or irregular. Long bones are found in the upper and lower extremities; short bones in the ankle and wrist; flat bones in the skull and rib cage; and irregular bones in the vertebrae and jaw. Tendons and ligaments are dense connective skeletal tissue that connect muscles and bones. Tendons connect muscles to bones, and ligaments connect the movable bones of joints.

Articulations, or joints, are areas where two or more bones meet. Synarthroses are joints in which bones are joined together by fibrous tissue, cartilage, or bone; they lack a joint cavity and have little or no movement. Diarthrodial or synovial joints are freely movable. The surfaces of the articulating ends of bones in diarthrodial joints are covered with a thin layer of articular cartilage, and they are enclosed in a fibrous joint capsule. The joint capsule consists of two layers: an outer fibrous layer and an inner membrane, the synovium. The synovial fluid, which is secreted by the synovium into the joint capsule, acts as a lubricant and facilitates movement of the joint's articulating surfaces. Bursae, which are closed sacs containing synovial fluid, prevent friction in areas where tendons



■ FIGURE 41-7 ■ Sagittal section of knee joint, showing prepatellar and suprapatellar bursae. are deflected over bone or where skin must move freely over bony tissue.

Menisci are fibrocartilaginous structures that develop from portions of the articular disk that occupied the space between the articular cartilage during fetal development. The menisci may have a free inner border, or they may extend through the joint, separating it into two cavities. The menisci in the knee joint may be torn as a result of injury. Explain why pain is often experienced in all the joints of an extremity when only a single joint is affected by a disease process.

Describe the structure and function of a bursa.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

BIBLIOGRAPHY

- DeLuca H.F. (1988). The vitamin D story: A collaborative effort of basic science and clinical medicine. *FASEB Journal* 2, 236–242.
- Guyton A.C., Hall J.E. (2001). *Textbook of medical physiology* (10th ed., pp. 899–912). Philadelphia: W.B. Saunders.
- Junqueira L.C., Carneiro J., Kelly O. (1995). *Basic histology* (8th ed., pp. 124–151). Los Altos, CA: Lange Medical Publications.
- Moore K.L., Dalley A.F. (1998). *Clinically oriented anatomy* (4th ed.). Philadelphia: Lippincott Williams & Wilkins.
- Rhoades R.A., Tanner G.A. (1996). Medical physiology (pp. 725–735). Boston: Little, Brown.
- Rosenberg A. (1999). Bones, joints, and soft tissue tumors. In Cotran R.S., Kumar V., Collins T. (Eds.), *Robbins pathologic basis of disease* (6th ed., pp. 1215–1227). Philadelphia: W.B. Saunders.
- Schiller A.L., Teitelbaum S.L. (1999). Bones and joints. In Rubin E., Farber J.L. (Eds.), *Pathology* (3rd ed., pp. 1338–1347). Philadelphia: Lippincott Williams & Wilkins.

REVIEW QUESTIONS

Cite the common components of cartilage and bone.

Compare the properties of the intercellular collagen and elastic fibers of skeletal tissue.

Cite the characteristics and name at least one location of elastic cartilage, hyaline cartilage, and fibrocartilage.

State the function of parathyroid hormone, calcitonin, and vitamin D in terms of bone formation and metabolism.

State the location and function of the periosteum and the endosteum.

State the characteristics of tendons and ligaments.

State the difference between synarthrodial and diarthrodial joints.

42

Alterations in the Skeletal System: Trauma, Infection, and Developmental Disorders

Injury and Trauma of Musculoskeletal Structures

Soft Tissue Injuries Joint (Musculotendinous) Injuries Strains and Sprains Dislocations Loose Bodies Shoulder and Rotator Cuff Injuries **Knee Injuries Fractures** Classification **Manifestations Diagnosis and Treatment Bone Healing Complications of Fractures and** Other Musculoskeletal Injuries **Osteonecrosis Bone Infections Osteomyelitis Direct Contamination** Hematogenous Osteomyelitis **Chronic Osteomyelitis** Osteomyelitis With Vascular Insufficiency Tuberculosis of the Bone or Joint Neoplasms **Benign Neoplasms Malignant Bone Tumors** Osteosarcoma **Ewing's Sarcoma** Chondrosarcoma **Metastatic Bone Disease**

Disorders of Skeletal Growth and Development in Children Alterations During Normal Growth Periods Torsional Deformities Genu Varum and Genu Valgum Hereditary and Congenital Deformities Osteogenesis Imperfecta Congenital Clubfoot Developmental Dysplasia of the Hip Disorders of Later Development Developmental Disorders of the Hip Disorders of the Knee Scoliosis Types of Scoliosis

The musculoskeletal system includes the bones, joints, and muscles of the body together with associated structures such as ligaments and tendons. This system, which constitutes more than 70% of the body, is subject to a large number of disorders. These disorders affect persons in all age groups and walks of life and cause pain, disability, and deformity. The discussion in this chapter focuses on alterations and the effects of trauma, infections, ischemia, and neoplasms on structures of the musculoskeletal system. Disorders of skeletal growth and development in children are discussed at the end of the chapter.

INJURY AND TRAUMA OF MUSCULOSKELETAL STRUCTURES

A broad spectrum of musculoskeletal injuries results from numerous physical forces, including blunt tissue trauma, disruption of tendons and ligaments, and fractures of bony structures. Many of the forces that cause injury to the musculoskeletal system are typical for a particular environmental setting, activity, or age group. Trauma resulting from high-speed motor accidents is ranked as the number one killer of adults younger than 45 years of age.¹ Motorcycle accidents are especially common in young men, with fractures of the distal tibia, midshaft femur, and radius occurring most often. Trauma in children is usually the result of an accident. The most common causes of childhood injuries are falls, bicycle-related injuries, and sports injuries. Falls are the most common cause of injury in people 65 years of age and older. Impaired vision and hearing, dizziness, and unsteadiness of gait contribute to falls in older persons. These falls are often compounded by osteoporosis, or bone atrophy, which makes fractures more likely. Fractures of the hip and proximal humerus are particularly common in this age group.

Soft Tissue Injuries

Most skeletal injuries are accompanied by soft tissue injuries. These injuries include contusions, hematomas, and lacerations. They are discussed here because of their association with musculoskeletal injuries.

A *contusion* is an injury to soft tissue that results from direct trauma and is usually caused by striking a body part against a hard object. With a contusion, the skin overlying the injury remains intact. Initially, the area becomes ecchymotic (*i.e.*, black and blue) because of local hemorrhage; later, the discoloration gradually changes to brown and then to yellow as the blood is reabsorbed.

A large area of local hemorrhage is called a *hematoma*. Hematomas cause pain as blood accumulates and exerts pressure on nerve endings. The pain increases with movement or when pressure is applied to the area. The pain and swelling of a hematoma take longer to subside than that accompanying a contusion. A hematoma may become infected because of bacterial growth. Unlike a contusion, which does not drain, a hematoma may eventually split the skin because of increased pressures and produce drainage. The treatment for a contusion and a hematoma consists of elevating the affected part and applying cold for the first 24 hours to reduce the bleeding into the area. A hematoma may need to be aspirated.

A *laceration* is an injury in which the skin is torn or its continuity is disrupted. The seriousness of a laceration depends on the size and depth of the wound and on whether there is contamination from the object that caused the injury. Puncture wounds from nails or rusted material may result in the growth of toxic bacteria, leading to gas gangrene or tetanus. Lacerations are usually treated by wound closure, which is done after the area is sufficiently cleaned; the closed wound is covered with a sterile dressing. It is important to minimize contamination of the wound and to control bleeding. Contaminated wounds and open fractures are copiously irrigated and debrided, and the skin usually is left open to heal to prevent the development of an anaerobic infection or a sinus tract.

Joint (Musculotendinous) Injuries

Joints, or articulations, are sites where two or more bones meet. Joints (*i.e.*, diarthrodial) are supported by tough bundles of collagenous fibers called *ligaments* that attach to the joint capsule and bind the articular ends of bones together, and by *ten*- *dons* that join muscles to the periosteum of the articulating bones. Joint injuries involve mechanical overloading or forcible twisting or stretching.

Strains and Sprains

A *strain* is a stretching injury to a muscle or a musculotendinous unit caused by mechanical overloading. This type of injury may result from an unusual muscle contraction or an excessive forcible stretch. Although there usually is no external evidence of a specific injury, pain, stiffness, and swelling exist. The most common sites for muscle strains are the lower back and the cervical region of the spine. The elbow and the shoulder are also supported by musculotendinous units that are subject to strains. Foot strain is associated with the weight-bearing stresses of the feet; it may be caused by inadequate muscular and ligamentous support, overweight, or excessive exercise such as standing, walking, or running.

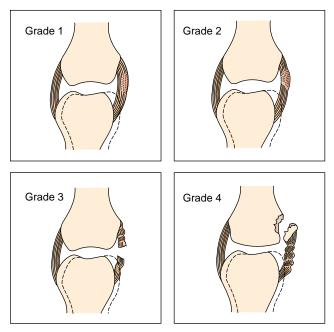
A sprain, which involves the ligamentous structures surrounding the joint, resembles a strain, but the pain and swelling subside more slowly. It usually is caused by abnormal or excessive movement of the joint. With a sprain, the ligaments may be incompletely torn or, as in a severe sprain, completely torn or ruptured (Fig. 42-1). The signs of sprain are pain, rapid swelling, heat, disability, discoloration, and limitation of function. Any joint may be sprained, but the ankle joint is most commonly involved, especially in higher-risk sports such as basketball. Most ankle sprains occur in the lateral ankle when the foot is turned inward under a person, forcing the ankle into inversion beyond the structural limits. Other common sites of sprain are the knee (the collateral ligament and anterior cruciate ligament) and elbow (the ulnar side). As with a strain, the soft tissue injury that occurs with a sprain is not evident on the radiograph. However, occasionally a chip of bone is evident when the entire ligament, including part of its bony attachment, has been ruptured or torn from the bone.

Healing of the dense connective tissues in tendons and ligaments is similar to that of other soft tissues. If properly treated,

KEY CONCEPTS

JOINT INJURIES

- Joints are the weakest part of the skeletal system and common sites for injury due to mechanical overloading or forcible twisting or stretching.
- Injury can include damage to the tendons, which connect muscle to bone; ligaments, which hold bones together; or the cartilage that covers the articular surface.
- Healing of the dense connective tissue involved in joint injuries requires time to restore the structures so that they are strong enough to withstand the forces imposed on the joint. Ligamentous injuries may require surgical intervention with approximation of many fibrous strands to facilitate healing.



■ FIGURE 42-1 ■ Degrees of sprain on the medial side of the right knee: grade 1, mild sprain of the medial collateral ligament; grade 2, moderate sprain with hematoma formation; grade 3, severe sprain with total disruption of the ligament; and grade 4, severe sprain with avulsion of the medial femoral condyle at the insertion of the medial collateral ligament. (Adapted from Spickler L.L. [1983]. Knee injuries of the athlete. *Orthopedic Nursing* 2 [5], 12–13)

injuries usually heal with the restoration of the original tensile strength. Repair is accomplished by fibroblasts from the inner tendon sheath or the loose connective tissue that surrounds the tendon. Capillaries infiltrate the injured area during the initial healing process and supply the fibroblasts with the materials they need to produce large amounts of collagen. Formation of the new collagen fibrils begins within 4 to 5 days and al-though tensile strength increases steadily thereafter, it is not sufficient to permit strong tendon pulls for about 7 weeks.² During the first 3 weeks, there is a danger that muscle contraction will pull the injured ends apart, causing the tendon to heal in the lengthened position. There is also a danger that adhesions will develop in areas where tendons pass through fibrous channels, such as in the distal palm of the hands, rendering the tendon useless.

The treatment of muscle strains and ligamentous sprains is similar in several ways. For an injured extremity, such as the ankle, elevation of the part followed by local application of cold may be sufficient.³ Compression, accomplished through the use of adhesive wraps or a removable splint, helps reduce swelling and provides support. A cast is applied for severe sprains, especially those severe enough to warrant surgical repair. Immobilization for a muscle strain is continued until the pain and swelling have subsided. In a sprain, the affected joint is immobilized for several weeks. Immobilization may be followed by graded active exercises. Early diagnosis, treatment, and rehabilitation are essential in preventing chronic ligamentous instability.

Dislocations

Dislocation of a joint is the loss of articulation of the bone ends in the joint capsule caused by displacement or separation of the bone end from its position in the joint. It usually follows a severe trauma that disrupts the holding ligaments. Dislocations are seen most often in the shoulder and acromioclavicular joints. A *subluxation* is a partial dislocation in which the bone ends in the joint are still in partial contact with each other.

Dislocations can be congenital, traumatic, or pathologic. Congenital dislocations occur in the hip and knee. Traumatic dislocations occur after falls, blows, or rotational injuries. For example, car accidents often cause dislocations of the hip and accompanying acetabular fractures because of the direction of impact. In the shoulder and patella, dislocations may become recurrent, especially in athletes. They recur with the same motion but require less and less force each time. Less common sites of dislocation, seen mainly in young adults, are the wrist and midtarsal region. They usually are the result of direct force, such as a fall on an outstretched hand. Pathologic dislocation in the hip is a late complication of infection, rheumatoid arthritis, paralysis, and neuromuscular diseases.

Diagnosis of a dislocation is based on history, physical examination, and radiologic findings. The symptoms are pain, deformity, and limited movement. The treatment depends on the site, mechanism of injury, and associated injuries such as fractures. Dislocations that do not reduce spontaneously usually require manipulation or surgical repair. Various surgical procedures also can be used to prevent redislocation of the patella, shoulder, or acromioclavicular joints. Immobilization is necessary for several weeks after reduction of a dislocation to allow healing of the joint structures. In dislocations affecting the knee, alternatives to surgery are isometric quadricepsstrengthening exercises and a temporary brace. Surgical procedures, such as joint replacement, may be necessary in certain pathologic dislocations.

Loose Bodies

Loose bodies are small pieces of bone or cartilage within a joint space. These can result from trauma to the joint or may occur when cartilage has worn away from the articular surface, causing a necrotic piece of bone to separate and become free floating. The symptoms are painful catching and locking of the joint. Loose bodies are commonly seen in the knee, elbow, hip, and ankle. The loose body repeatedly gets caught in the crevice of a joint, pinching the underlying healthy cartilage; unless the loose body is removed, it may cause osteoarthritis and restricted movement. The treatment consists of removal using operative arthroscopy.

Shoulder and Rotator Cuff Injuries

The glenohumeral (shoulder) joint is a ball-and-socket joint that permits a wide range of motion, a factor that makes the joint relatively unstable.⁴ The support and movement of the shoulder joint relies heavily on the support of four relatively small muscle-tendon groups collectively known as the rotator cuff (Fig. 42-2). The space between the acromion of the shoulder blade and superior part of the humeral head is called the impingement interval. This space is normally narrow and is maximally narrow when the arm if abducted or moved above the horizontal.



Rotator cuff impingement, tendonitis, and tears are common among athletes. The rotator cuff is commonly injured during repetitive movements that carry the arm above the shoulder, such as those used by baseball pitchers, tennis and racquet ball players, swimmers, and weight lifters.⁴ Among the relevant muscles, the supraspinous is the one most often affected. Rotator cuff tears may be partial or full thickness. Partial tears do not completely severe the tendon and respond well to nonsurgical treatment. Full thickness tears require surgery.

Recurrent inflammation or tendonitis of the rotator cuff inflammation, also known as *shoulder impingement syndrome*, is characterized by pain and swelling of the cuff tendons and the surrounding bursa. It occurs most often in people with loose joints, people with abnormal bony anatomy in the shoulder, and people who do repetitive lifting above the shoulder level. It is common in baseball pitchers. Lifting overhead or just moving the arm above the head may cause pinching or impingement of one of the tendons of the rotator cuff. Tendonitis may progress to a partial or complete rotator cuff tear caused by progressive weakening of the tendon fibers.

Calcific tendonitis is a self-limiting calcification of the rotator cuff.⁵ It involves the deposition of calcium crystals in the tendon of the rotator cuff. It is most common in women, ages 30 to 60 years, and workers in sedentary jobs. The disorder involves four stages: (1) precalcific phase with asymptomatic fibrocartilaginous transformation within the cuff, (2) deposition of calcium crystals, (3) the resorptive phase, and (4) the postcalcific healing and repair phase. The third phase is usually the most incapacitating because the movement of the calcium crystals in the subacromial bursa during the resorptive process may cause constant and severe pain and restriction of motion that typically lasts for several weeks.

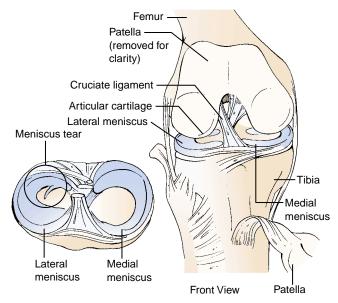
Many physical examination maneuvers are used to define shoulder pathology. The history and mechanism of injury are important. In addition to standard radiographs, an arthrogram, computed tomography (CT) scan, or magnetic resonance imaging (MRI) scan may be obtained. Arthroscopic examination under anesthesia is done for diagnostic purposes and operative arthroscopy to repair severe tears. Conservative treatment with anti-inflammatory agents, corticosteroid injections, and physical therapy often is done. A period of rest is followed by a customized exercise and rehabilitation program to improve strength, flexibility, and endurance.

Knee Injuries

The knee is a common site of injury, particularly sport-related injuries in which the knee is subjected to abnormal twisting and compression forces. These forces can result in injury to the menisci, rupture of the anterior cruciate ligament, patellar subluxation and dislocation, and chondromalacia. Knee injuries in young adulthood and both knee and hip injuries in middle age substantially increase the risk of osteoarthritis in the same joint later in life.⁶

Meniscus Injuries. The menisci are C-shaped plates of fibrocartilage that are superimposed between the condyles of the femur and tibia. The menisci play a major role in load bearing and shock absorption; they help to stabilize the knee; they assist in joint lubrication; and they serve as a source of nutrition for articular cartilage in the knee. There are two menisci in each knee, a lateral and medial meniscus (Fig. 42-3). The menisci are thicker at their external margins and taper to thin, unattached edges at their interior margin. They are firmly attached at their ends to the intercondylar area of the tibia and are supported by the coronary and transverse ligaments of the knee.

Any action of the knee that causes injury to the knee ligaments can also cause a meniscal tear.⁷ Meniscus injury commonly occurs as the result of a rotational injury from a sudden or sharp pivot or a direct blow to the knee, as in hockey, basketball, or football. The injured knee is edematous and painful,



■ FIGURE 42-3 ■ The knee showing the lateral and medial meniscus (with the patella removed for clarity). Insert (**lower left**) shows meniscus tear.

especially with hyperflexion and hyperextension. A loose fragment may cause knee instability and locking.

Diagnosis is made by examination and confirmed by methods such as arthroscopy and radiologic, CT scans, MRI, and radionuclide imaging.⁸ Initial treatment of meniscal injuries may be conservative. The knee may be placed in a removable knee immobilizer. Isometric quadriceps exercises may be prescribed. Activity usually is restricted until complete motion is recovered. Arthroscopic meniscectomy may be performed when there is recurrent or persistent locking, recurrent effusion, or disabling pain.

Rupture of the Anterior Cruciate Ligament. The cruciate ligaments (CL) secure the femur to the tibia in a crossed position (Fig. 42-3). The anterior CL (ACL) and posterior CL (PCL) control flexion and lateral rotation of the knee. The ACL is the weaker ligament and is often ruptured with a lateral blow to the knee. People with knees that are unstable as the result of ACL tears are at risk for early degenerative joint disease, as well as subsequent damage to other structures of the knee.⁹

About 85% of people who have an ACL tear are immediately disabled and not able to continue their activity. They often report the feeling of "giving way" and hearing or feeling a "pop" at the time of injury. Swelling is usually caused by hemarthrosis.

Diagnosis is based on the history, physical examination of the knee, and MRI. The selection of treatment methods, which include surgical repair, alterations in lifestyle, and bracing, depend on the person's activity level and expectation. Surgical repair is usually the treatment of choice for athletes who want to continue activities requiring twisting or rapid changes in direction.

Patellar Subluxation and Dislocations. Recurrent subluxation and dislocation of the patella (*i.e.*, knee cap) are common injuries in young adults. They account for approximately 10% of all athletic injuries and are more common in women. Sports such as skiing or tennis may cause stress on the patella. These sports involve external rotation of the foot and lower leg with knee flexion, a position that exerts rotational stresses on the knee. Congenital knee variations are also a predisposing factor.

There is often a sensation of the patella "popping out" when the dislocation occurs. Other complaints include the knee giving out, swelling, crepitus, stiffness, and loss of range of motion.

Treatment can be difficult, but nonsurgical methods are used first. They include immobilization with the knee extended, bracing, administration of anti-inflammatory agents, and isometric quadriceps-strengthening exercises. Surgical intervention often is necessary.

Chondromalacia. Chondromalacia, or softening of the articular cartilage, is seen most commonly on the undersurface of the patella and occurs most frequently in young adults. It can be the result of recurrent subluxation of the patella or overuse in strenuous athletic activities. Persons with this disorder typically report pain, particularly when climbing stairs or sitting with the knees bent. Occasionally, the person experiences weakness of the knee.

The treatment consists of rest, isometric exercises, and application of ice after exercise. Part of the patella may be surgically removed in severe cases. In less severe cases, the soft portion is shaved using a saw inserted through an arthroscope.

Fractures

A fracture, or discontinuity of the bone, is the most common type of bone lesion. Normal bone can withstand considerable compression and shearing forces and, to a lesser extent, tension forces. A fracture occurs when more stress is placed on the bone than it is able to absorb. Grouped according to cause, fractures can be divided into three major categories: (1) fractures caused by sudden injury, (2) fatigue or stress fractures, and (3) pathologic fractures. The most common fractures are those resulting from sudden injury. The force causing the fracture may be direct, such as a fall or blow, or indirect, such as a massive muscle contraction or trauma transmitted along the bone. For example, the head of the radius or clavicle can be fractured by the indirect forces that result from falling on an outstretched hand.

A *fatigue fracture* results from repeated wear on a bone. Pain associated with overuse injuries of the lower extremities, especially posterior medial tibial pain, is one of the most common symptoms that physically active persons, such as runners, experience. *Stress fractures* in the tibia may be confused with "shin splints," a nonspecific term for pain in the lower leg from overuse in walking and running, because they frequently do not appear on x-ray films until 2 weeks after the onset of symptoms.

A *pathologic fracture* occurs in bones that already are weakened by disease or tumors. Fractures of this type may occur spontaneously with little or no stress. The underlying disease state can be local, as with infections, cysts, or tumors, or it can be generalized, as in osteoporosis, Paget's disease, or disseminated tumors.

Classification

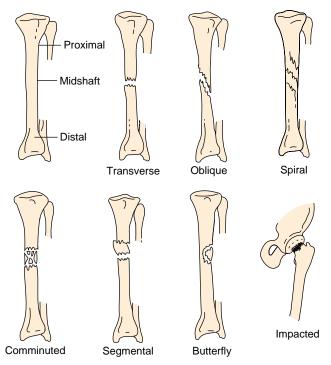
Fractures usually are classified according to location, type, and direction or pattern of the fracture line (Fig. 42-4). A fracture of the long bone is described in relation to its position in the bone—proximal, midshaft, and distal. Other descriptions are used when the fracture affects the head or neck of a bone,

KEY CONCEPTS

FRACTURE HEALING

- Fractures are caused by forces that disrupt the continuity of bone.
- Fracture healing depends on the extent of the injury, the ability to align the bone fragments, and immobilizing the fracture site so that healing can take place.
- Bone healing occurs by replacement of injured bone cells. It involves formation of a hematoma that provides the foundation for blood vessel and fibroblast infiltration, proliferation of bone repair cells (osteoblasts), callus formation, ossification of the callus, and remodeling of the fracture site.

Unit Eleven: Alterations in the Skeletal and Integumentary Systems



■ **FIGURE 42-4** ■ Classification of fractures. Fractures are classified according to location (proximal, midshaft, or distal), the direction of fracture line (transverse, oblique, or spiral), and type (comminuted, segmental, butterfly, or impacted).

involves a joint, or is near a prominence such as a condyle or malleolus.

The type of fracture is determined by its communication with the external environment, the degree of break in continuity of the bone, and the character of the fracture pieces.¹⁰ A fracture can be classified as open or closed. When the bone fragments have broken through the skin, the fracture is called an *open* or *compound fracture*. In a closed fracture, there is no communication with the outside skin.

The degree of a fracture is described in terms of a partial or complete break in the continuity of bone. A *greenstick fracture*, which is seen in children, is an example of a partial break in bone continuity and resembles that seen when a young sapling is broken. This kind of break occurs because children's bones, especially until approximately 10 years of age, are more resilient than the bones of adults.

The character of the fracture pieces may also be used to describe a fracture. A *comminuted fracture* has more than two pieces. A *compression fracture*, as occurs in the vertebral body, involves two bones that are crushed or squeezed together. A fracture is called *impacted* when the fracture fragments are wedged together. This type usually occurs in the humerus, often is less serious, and usually is treated without surgery.

The direction of the trauma or mechanism of injury produces a certain configuration or pattern of fracture. *Reduction* is the restoration of a fractured bone to its normal anatomic position. The pattern of a fracture indicates the nature of the trauma and provides information about the easiest method for reduction. *Transverse fractures* are caused by simple angulatory forces. A *spiral fracture* results from a twisting motion, or torque. A transverse fracture is not likely to become displaced or lose its position after it is reduced. On the other hand, spiral, oblique, and comminuted fractures often are unstable and may change position after reduction.

Manifestations

The signs and symptoms of a fracture include pain, tenderness at the site of bone disruption, swelling, loss of function, deformity of the affected part, and abnormal mobility. The deformity varies according to the type of force applied, the area of the bone involved, the type of fracture produced, and the strength and balance of the surrounding muscles.

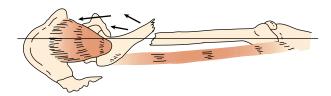
In long bones, three types of deformities-angulation, shortening, and rotation—are seen. Severely angulated fracture fragments may be felt at the fracture site and often push up against the soft tissue to cause a tenting effect on the skin. Bending forces and unequal muscle pulls cause angulation. Shortening of the extremity occurs as the bone fragments slide and override each other because of the pull of the muscles on the long axis of the extremity (Fig. 42-5). Rotational deformity occurs when the fracture fragments rotate out of their normal longitudinal axis; this can result from rotational strain produced by the fracture or unequal pull by the muscles that are attached to the fracture fragments. A crepitus or grating sound may be heard as the bone fragments rub against each other. In the case of an open fracture, there is bleeding from the wound where the bone protrudes. Blood loss from a pelvic fracture or multiple long bone fractures can cause hypovolemic shock in a trauma victim.

Shortly after the fracture has occurred, nerve function at the fracture site may be temporarily lost. The area may become numb, and the surrounding muscles may become flaccid. This condition has been called *local shock*. During this period, which may last for a few minutes to half an hour, fractured bones may be reduced with little or no pain. After this brief period, the pain sensation returns and, with it, muscle spasms and contractions of the surrounding muscles.

Diagnosis and Treatment

Diagnosis is the first step in the care of fractures and is based on history and physical manifestations. X-ray examination is used to confirm the diagnosis and direct the treatment. The ease of diagnosis varies with the location and severity of the fracture. In the trauma patient, the presence of other, more serious injuries may make diagnosis more difficult.

Treatment depends on the general condition of the person, the presence of associated injuries, the location of the fracture and its displacement, and whether the fracture is open or closed. A *splint* is a device for immobilizing the movable fragments of a fracture. When a fracture is suspected, the injured part always should be splinted before it is moved. This is essential for preventing further injury.



■ FIGURE 42-5 ■ Displacement and overriding of fracture fragments of a long bone (femur) caused by severe muscle spasm.

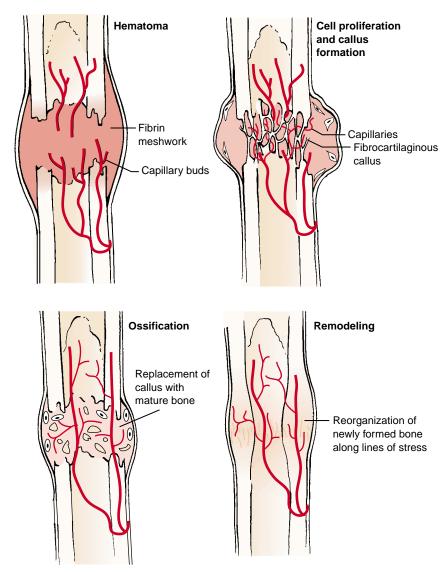
794

There are three objectives for treatment of fractures: (1) reduction of the fracture, (2) immobilization, and (3) preservation and restoration of the function of the injured part. Reduction of a fracture is directed toward replacing the bone fragments to as near normal anatomic position as possible. This can be accomplished by closed manipulation or surgical (open) reduction. Closed manipulation uses methods such as manual pressure and traction. Fractures are held in reduction by external or internal fixation devices. Surgical reduction involves the use of various types of hardware to accomplish internal fixation of the fracture fragments (Fig. 42-6). Immobilization prevents movement of the injured parts and is the single most important element in obtaining union of the fracture fragments. Immobilization can be accomplished through the use of external devices, such as splints, casts, external fixation devices, or traction, or by means of internal fixation devices inserted during surgical reduction of the fracture. Preservation and restoration of the function of muscles and joints are an ongoing process in the unaffected and affected extremities during the period of immobilization required for fracture healing. Exercises designed to preserve function, maintain muscle strength, and reduce joint stiffness should be started early.

Bone Healing

Bone healing occurs in a manner similar to soft tissue healing. However, it is a more complex process and takes longer. There are five stages involved in bone healing: (1) hematoma formation, (2) cellular proliferation, (3) callus formation, (4) ossification, and (5) remodeling (Fig. 42-6). The degree of response during each of these stages is in direct proportion to the extent of trauma.

Hematoma Formation. Hematoma formation occurs during the first 48 to 72 hours after fracture. It develops as blood from torn vessels in the bone fragments and surrounding soft tissue leaks between and around the fragments of the fractured bone. The hematoma is thought to be necessary for the initiation of the cellular events essential to fracture healing.¹¹⁻¹³ As the result of hematoma formation, clotting factors remain in the injured area to initiate the formation of a fibrin meshwork, which serves as a framework for the ingrowth of fibroblasts and new



■ FIGURE 42-6 ■ The stages of bone healing. The hematoma stage provides the fibrin meshwork and capillary buds needed for subsequent cellular invasion. Cellular proliferation and callus formation represent the stages during which osteoblasts enter the area and form the fibrocartilaginous callus that joins the bone fragments. The ossification stage involves the mineralization of the fibrocartilaginous callus; and the remodeling stage, the reorganization of mineralized bone along the lines of mechanical stress. capillary buds. Granulation tissue, the result of fibroblast and new capillary growth, gradually invades and replaces the clot. When a large hematoma develops, healing is delayed because macrophages, platelets, oxygen, and nutrients for callus formation are prevented from entering the area.

Cellular Proliferation. Three layers of bone structure are involved in the cellular proliferation that occurs during bone healing: the periosteum, or outer covering of the bone; the endosteum, or inner covering; and the medullary canal, which contains the bone marrow. During this process, the osteoblasts, or bone-forming cells, multiply and differentiate into a fibrocartilaginous callus. The fibrocartilaginous callus is softer and more flexible than callus. Cellular proliferation begins distal to the fracture site, where there is a greater supply of blood. After a few days, a fibrocartilage "collar" becomes evident around the fracture site. The collar edges on either side of the fracture eventually unite to form a bridge, which connects the bone fragments.

Callus Formation. During the early stage of callus formation, the fracture site becomes "sticky" as osteoblasts continue to move in and through the fibrin bridge to help keep it firm. Cartilage forms at the level of the fracture, where there is less circulation. In areas of the bone with muscle insertion, periosteal circulation is better, bringing in the nutrients necessary to bridge the callus. The bone calcifies as mineral salts are deposited. This stage usually occurs during the third to fourth week of fracture healing.

Ossification. Ossification involves the final laying down of bone. This is the stage at which the fracture has been bridged and the fracture fragments are firmly united. Mature bone replaces the callus, and the excess callus is gradually resorbed by the osteoclasts. The fracture site feels firm and immovable and appears united on the radiograph. At this point, it is safe to remove the cast.

Remodeling. Remodeling involves resorption of the excess bony callus that develops in the marrow space and encircles the external aspect of the fracture site. The remodeling process is directed by mechanical stress and direction of weight bearing.

Healing Time. Union of a fracture has occurred when the fracture is solid enough to withstand normal stresses and it is clinically and radiologically safe to remove the external fixation. Healing time depends on the site of the fracture, the condition of the fracture fragments, hematoma formation, and other local and host factors. In children, fractures usually heal within 4 to 6 weeks; in adolescents, they heal within 6 to 8 weeks; and in adults, they heal within 10 to 18 weeks. The increased rate of healing among children compared with adults may be related to the increased cellularity and vascularity of the child's periosteum.¹³ In general, fractures of long bones, displaced fractures, and fractures with less surface area heal more slowly. Function usually returns within 6 months after union is complete. However, return to complete function may take longer.

Impaired Bone Healing. Factors that influence bone healing are specific to the person, the type of injury sustained, and local factors that disrupt healing (Chart 42-1). Individual factors that

CHART 42-1 Factors Affecting Fracture Healing

- Nature of the injury or the severity of the trauma, including fracture displacement, edema, and arterial occlusion with crushing injuries
- Degree of bridge formation that develops during bone healing
- Amount of bone loss (*e.g.*, it may be too great for the healing to bridge the gap)
- Type of bone that is injured (*e.g.*, cancellous bone heals faster than cortical bone)
- Degree of immobilization that is achieved (*e.g.*, movement disrupts the fibrin bridge and cartilage forms instead of bone)
- Local infection, which retards or prevents healing
- Local malignancy, which must be treated before healing can proceed
- Bone necrosis, which prevents blood flow into the fracture site

may delay bone healing are the patient's age; current medications; debilitating diseases, such as diabetes and rheumatoid arthritis; local stress around the fracture site; circulatory problems and coagulation disorders; and poor nutrition.

Malunion is healing with deformity, angulation, or rotation that is visible on x-ray films.¹³ Early, aggressive treatment, especially of the hand, can prevent malunion and result in earlier alignment and return of function. Malunion is caused by inadequate reduction or alignment of the fracture.

Delayed union is the failure of a fracture to unite within the normal period (*e.g.*, 20 weeks for a fracture of the tibia or femur in an adult). Intra-articular fractures (those through a joint) may heal more slowly and may eventually produce arthritis. *Nonunion* is failure to produce union and cessation of the processes of bone repair. It is seen most often in the tibia, especially with open fractures or crushing injuries. It is characterized by mobility of the fracture site and pain on weight bearing. Muscle atrophy and loss of range of motion may occur. Non-union usually is established 6 to 12 months after the time of the fracture.¹³ The complications of fracture healing are summarized in Table 42-1.

Treatment methods for impaired bone healing encompass surgical interventions, including bone grafts, bracing, external fixation, or electrical stimulation of the bone ends. Electrical stimulation is thought to stimulate the osteoblasts to lay down a network of bone. Three types of commercial bone growth stimulators are available: a noninvasive model, which is placed outside the cast; a seminoninvasive model, in which pins are inserted around the fracture site; and a totally implantable type, in which a cathode coil is wound around the bone at the fracture site and operated by a battery pack implanted under the skin.

Complications of Fractures and Other Musculoskeletal Injuries

The complications of fractures and other orthopedic injuries are associated with loss of skeletal continuity, injury from bone fragments, pressure from swelling and hemorrhage (*e.g.*, frac-

TABLE 42-1	Complications of Fracture Healing		
Complication	Manifestations	Contributing Factors	
Delayed union	Failure of fracture to heal within predicted time as determined by x-ray	Large displaced fracture Inadequate immobilization Large hematoma Infection at fracture site Excessive loss of bone Inadequate circulation	
Malunion	Deformity at fracture site Deformity or angulation on x-ray	Inadequate reduction Malalignment of fracture at time of immobilization	
Nonunion	Failure of bone to heal before the process of bone repair stops Evidence on x-ray Motion at fracture site Pain on weight bearing	Inadequate reduction Mobility at fracture site Severe trauma Bone fragment separation Soft tissue between bone fragments Infection Extensive loss of bone Inadequate circulation Malignancy Bone necrosis Noncompliance with restrictions	

ture blisters, compartment syndrome), involvement of nerve fibers (*e.g.*, reflex sympathetic dystrophy and causalgia), or development of fat emboli.

Fracture Blisters. Fracture blisters are skin bullae and blisters representing areas of epidermal necrosis with separation of epidermis from the underlying dermis by edema fluid. They are seen with more severe, twisting types of injuries (e.g., motor vehicle accidents and falls from heights) but can also occur after excessive joint manipulation, dependent positioning, and heat application, or from peripheral vascular disease. They can be solitary, multiple, or massive, depending on the extent of injury. Most fracture blisters occur in the ankle, elbow, foot, knee, or areas where there is little soft tissue between the bone and the skin. The development of fracture blisters reportedly is reduced by early surgical intervention in persons requiring operative repair.¹⁴ This probably reflects the early operative release of the fracture hematoma, reapproximation of the disrupted soft tissues, ligation of bleeding vessels, and fixation of bleeding fracture surfaces. Prevention of fracture blisters is important because they pose an additional risk of infection.

Compartment Syndrome. Compartment syndrome is the result of increased pressure in a limited anatomic space that compromises circulation and threatens the viability and function of the nerves and muscles (see Chapter 15). It can be acute or chronic. Acute compartment syndrome can occur after a fracture or crushing injury, when excessive swelling around the site of injury results in increased pressure in a closed compartment. This increase in pressure occurs because fascia, which covers and separates muscles, is inelastic and unable to stretch and compensate for the extreme swelling. The most common sites are the four compartments of the lower leg (*i.e.*, deep posterior, superficial posterior, lateral, and anterior compartments) and the dorsal and volar compartments of the forearm.

The condition is characterized by pain that is out of proportion to the original injury or physical findings.^{14,15} Nerve compression may cause changes in sensation (*e.g.*, paresthesias such as burning or tingling or loss of sensation), diminished reflexes, and eventually the loss of motor function. Symptoms usually begin within a few hours but can be delayed as long as 64 hours.¹⁴ Compression of blood vessels may cause muscle ischemia and loss of function. In contrast to the diminished or absent pulses that occur when ischemia is caused by a tight bandage or cast, the arterial pulses often are normal in compartment syndrome. Pallor and loss of the pulse, when they occur, are late findings.

Treatment of compartment syndrome is directed at reducing the compression of blood vessels and nerves. Constrictive dressings and casts are loosened. Intracompartmental pressure can be measured by means of a catheter or needle inserted into the compartment. A fasciotomy, or surgical transection of the fascia that is restricting the muscle compartment, may be required. Delay in diagnosis and treatment of compartment syndrome can lead to irreversible nerve and muscle damage.

Reflex Sympathetic Dystrophy and Causalgia. Reflex sympathetic dystrophy, also known as the complex regional pain syndrome, is a complication of orthopedic injuries that causes pain out of proportion to the injury and autonomic nervous system dysfunction manifested by hyperhidrosis (increased sweating) and vasomotor instability (either flushed and warm or cold and pale).^{14,16} The disorder often produces long-term disability and chronic pain syndromes (see Chapter 39).

Pain, which is the prominent symptom of the disorder, is described as severe, aching, or burning. It usually increases in intensity with movement and with noxious and non-noxious stimuli. The cause of the pain is unclear but is thought to have a sympathetic nervous system component. Muscle wasting, thin and shiny skin, and abnormalities of the nails and bone can occur. Decreased muscle strength and disuse can lead to contractures and osteoporosis.

Treatment focuses on pain management and prevention of disability. Physical therapy interventions such as hot/cold baths and elevation of the limb are used to maximize range of motion and minimize pain. Medications include anti-inflammatory agents, vasodilators, and antidepressant medications. Sympathetic nerve blocks may be used.

Fat Embolism Syndrome. The fat embolism syndrome (FES) refers to a constellation of clinical manifestations resulting from the presence of fat droplets in the small blood vessels of the lung or other organs after a long-bone fracture or other major trauma.¹⁴ The fat emboli are thought to be released from the bone marrow or adipose tissue at the fracture site into the venous system through torn veins.

Clinically, the incidence of fat embolization is related to fractures of bones containing the most marrow (*i.e.*, long bones and the bones of the pelvis). An increase in intramedullary pressure in the femur is the most important pathogenic factor in the development of emboli. Although fat embolization occurs with fractures or operative fixation of fractures, FES occurs in only a small percentage of cases, supporting the hypothesis that factors other than mechanical forces may be necessary in the development of FES.

The main clinical features of FES are respiratory failure, cerebral dysfunction, and skin petechiae.^{14,17} Initial symptoms begin within a few hours to 3 to 4 days after injury and do not appear beyond 1 week after the injury. The first symptoms include a subtle change in behavior and signs of disorientation resulting from emboli in the cerebral circulation combined with respiratory depression. There may be complaints of substernal chest pain and dyspnea accompanied by tachycardia and a lowgrade fever. Diaphoresis, pallor, and cyanosis become evident as respiratory function deteriorates. A petechial rash that does not blanch with pressure often occurs 2 to 3 days after the injury. This rash usually is found on the anterior chest, axillae, neck, and shoulders. It also may appear on the soft palate and conjunctiva. The rash is thought to be related to embolization of the skin capillaries or thrombocytopenia.

Three degrees of severity are seen: subclinical, overt clinical, and fulminating. Although the subclinical and overt clinical forms of FES respond well to treatment, the fulminating form often is fatal. An important part of the treatment of fat emboli is early diagnosis. Arterial blood gases should be assayed immediately after recognition of clinical manifestations. Treatment is directed toward correcting hypoxemia and maintaining adequate fluid balance. Mechanical ventilation may be required. Corticosteroid drugs are administered to decrease the inflammatory response of lung tissues, decrease the edema, stabilize the lipid membranes to reduce lipolysis, and combat the bronchospasm. Corticosteroids are also given prophylactically to high-risk persons. The only preventive approach to FES is early stabilization of the fracture.

Osteonecrosis

Osteonecrosis, or death of a segment of bone, is a condition caused by the interruption of blood supply to the marrow, medullary bone, or cortex. It is a relatively common disorder and can occur in the medullary cavity of the metaphysis and the subchondral region of the epiphysis, especially in the proximal femur, distal femur, and proximal humerus. It is a common complicating disorder of Legg-Calvé-Perthes disease, sickle cell disease, steroid therapy, and hip surgery.^{18,19} The rates of osteonecrosis among persons treated with corticosteroids range from 5% to 25%. More than 10% of 500,000 joint replacements performed annually in the United States are for treatment of osteonecrosis.¹⁸

Although bone necrosis results from ischemia, the mechanisms producing the ischemia are varied and include mechanical vascular interruption such as occurs with a fracture; thrombosis and embolism (*e.g.*, sickle cell disease, nitrogen bubbles caused by inadequate decompression during deep sea diving); vessel injury (*e.g.*, vasculitis, radiation therapy); and increased intraosseous pressure with vascular compression (*e.g.*, steroidinduced osteonecrosis). In many cases, the cause of the necrosis is uncertain. Other than fracture, the most common causes of bone necrosis are idiopathic (*i.e.*, those of unknown cause) and prior steroid therapy. Chart 42-2 lists disorders associated with osteonecrosis.

Bone has a rich blood supply that varies from site to site. The flow in the medullary portion of bone originates in nutrient vessels from an interconnecting plexus that supplies the marrow, trabecular bone, and endosteal half of the cortex. The outer cortex receives its blood supply from periosteal, muscular, metaphyseal, and epiphyseal vessels that surround the bone. Some bony sites, such as the head of the femur, have only limited collateral circulation so that interruption of the flow, such as with a hip fracture, can cause necrosis and irreversible damage to a substantial portion of medullary and cortical bone.

The pathologic features of bone necrosis are the same, regardless of cause. The site of the lesion is related to the vessels involved. There is necrosis of cancellous bone and marrow. The cortex usually is not involved because of collateral blood flow. In subchondral infarcts (*i.e.*, ischemia below the cartilage), a triangular or wedge-shaped segment of tissue that has the subchondral bone plate as its base and the center of the epiphysis as its apex undergo necrosis. When medullary infarcts occur in

CHART 42-2 Causes of Osteonecrosis				
Mechanical disruption of blood vessels				
Fractures				
Legg-Calvé-Perthes disease				
Blount's disease				
Thrombosis and embolism				
Sickle cell disease				
Nitrogen bubbles in decompression sickness				
Vessel injury				
Vasculitis				
Connective tissue disease				
Systemic lupus erythematosus				
Rheumatoid arthritis				
Radiation therapy				
Gaucher's disease				
Increased intraosseous pressure				
Steroid-induced osteonecrosis				

fatty marrow, death of bone results in calcium release and necrosis of fat cells with the formation of free fatty acids. Released calcium forms an insoluble "soap" with free fatty acids. Because bone lacks mechanisms for resolving the infarct, the lesions remain for life.

One of the most common causes of osteonecrosis is that associated with administration of corticosteroids.^{18,19} Despite numerous studies, the mechanism of steroid-induced osteonecrosis remains unclear. The condition may develop after the administration of very high, short-term doses, during longterm treatment, or even from intra-articular injection. Although the risk increases with the dose and duration of treatment, it is difficult to predict who will be affected. The interval between corticosteroid administration and onset of symptoms rarely is less than 6 months and may be more than 3 years. There is no satisfactory method for preventing progression of the disease.

The symptoms associated with osteonecrosis are varied and depend on the extent of infarction. Typically, subchondral infarcts cause chronic pain that is initially associated with activity but that gradually becomes more progressive until it is experienced at rest. Subchondral infarcts often collapse and predispose the patient to severe secondary osteoarthritis.

Diagnosis of osteonecrosis is based on history, physical findings, radiographic findings, and the results of special imaging studies, including CT scans and technetium-99m bone scans. MRI is particularly effective in the diagnosis of osteonecrosis. Plain radiographs are used to define and classify the course of the disease, particularly of the hip.

Treatment of osteonecrosis depends on the underlying pathology. In some cases, only short-term immobilization, nonsteroidal anti-inflammatory drugs, exercises, and limitation in weight bearing are used. Osteonecrosis of the hip is particularly difficult to treat. In persons with early disease, limitation of weight bearing through the use of crutches may allow the condition to stabilize. Although several surgical approaches have been used, the most definitive treatment of advanced osteonecrosis of the knee or hip is total joint replacement.

In summary, many external physical agents can cause trauma to the musculoskeletal system. Particular factors, such as environments, activity, or age, can place a person at greater risk for injury. Some soft tissue injuries such as contusions, hematomas, and lacerations are relatively minor and easily treated. Muscle strains and ligamentous sprains are caused by mechanical overload on the connective tissue. They heal more slowly than the minor soft tissue injuries and require some degree of immobilization. Healing of soft tissue begins within 4 to 5 days of the injury and is primarily the function of fibroblasts, which produce collagen. Joint dislocation is caused by trauma to the supporting structures. Repeated trauma to the joint can cause articular softening (i.e., chondromalacia) or the separation of small pieces of bone or cartilage, called loose bodies, in the joint. The knee and shoulder are common sites for injuries in athletes. The rotator cuff is a common site for shoulder injuries. Knee injuries include injury to the menisci, anterior cruciate ligament tears, patellar subluxation and dislocation, and chondromalacia.

Fractures occur when more stress is placed on a bone than the bone can absorb. The nature of the stress determines the type of fracture and the character of the resulting bone fragments. Healing of fractures is a complex process that takes place in five stages: hematoma formation, cellular proliferation, callus formation, ossification, and remodeling. For satisfactory healing to take place, the affected bone has to be reduced and immobilized. This is accomplished with external fixation devices (e.g., splints, casts, or traction) or surgically implanted internal fixation devices. The complications of fractures are associated with loss of skeletal continuity (malunion or nonunion), pressure from swelling and hemorrhage (fracture blisters and compartment syndrome), involvement of nerve fibers (reflex sympathetic dystrophy and causalgia), or development of fat emboli.

Osteonecrosis, or death of a segment of bone, is a condition caused by the interruption of blood supply to the marrow, medullary bone, or cortex. Sites with poor collateral circulation, such as the femoral head, are most seriously affected. It is a common complicating disorder of Legg-Calvé-Perthes disease, sickle cell disease, steroid therapy, and hip surgery. Symptoms include pain that varies in severity, depending on the extent of infarction. Total joint replacement is the most frequently used treatment for advanced osteonecrosis.

BONE INFECTIONS

Bone infections are difficult to treat and eradicate. Their effects can be devastating; they can cause pain, disability, and deformity. Chronic bone infections may drain for years because of a sinus tract. This occurs when a passageway develops from an abscess or cavity in the bone to an opening through the skin.

Osteomyelitis

Osteomyelitis represents an acute or chronic pyogenic infection of the bone. The term osteo refers to bone, and myelo refers to the marrow cavity, both of which are involved in this disease. Osteomyelitis can be caused by direct extension or contamination of an open fracture or wound (contiguous invasion); seeding through the bloodstream (hematogenous

KEY CONCEPTS

BONE INFECTIONS

- Bone infections may be caused by a wide variety of microorganisms introduced during injury, during operative procedures, or from the bloodstream.
- Once localized in bone, the microorganisms proliferate, produce cell death, and spread within the bone shaft, inciting a chronic inflammatory response with further destruction of bone.
- Bone infections are difficult to treat and eradicate. Measures to prevent infection include careful cleaning and debridement of skeletal injuries and strict operating room protocols.

spread); or from skin infections in persons with vascular insufficiency. In most cases, *Staphylococcus aureus* is the infecting organism.^{20,21} *S. aureus* has two characteristics that favor its ability to produce osteomyelitis: (1) it is able to produce a collagen-binding adhesion molecule that allows it to adhere to the connective tissue elements of bone; and (2) it has the ability to be internalized and survive in osteoblasts, making the microorganism more resistant to antibiotic therapy.²⁰ The term *acute osteomyelitis* is used to describe a newly recognized bone infection. *Chronic osteomyelitis* refers to recurrence of a previously treated or untreated infection.

The pathogenesis of osteomyelitis includes the presence of the infecting agent, inflammation and the protective efforts of inflammatory cells, and bone destruction. The infective microorganisms present in osteomyelitis incite an inflammatory process with recruitment of phagocytic cells. In an attempt to contain the invading microorganisms, the phagocytes generate toxic oxygen radicals and release proteolytic enzymes that destroy surrounding tissues.²⁰ The purulent drainage that ensues spreads into blood vessels in the bone, increasing the intraosseous pressure and impairing blood flow. The loss of blood flow leads to ischemic necrosis of bone. The blood supply to the bone may become obstructed by septic thrombi, in which case the ischemic bone becomes necrotic. As the process continues, the necrotic bone separates from the viable surrounding bone to form devascularized fragments, called sequestra (Fig. 42-7). One of the characteristics of chronic osteomyelitis is the presence of necrotic bone and the absence of living osteocytes.

Direct Contamination

The most common cause of osteomyelitis is the direct contamination of bone from an open wound. It may be the result of an open fracture, a gunshot wound, or a puncture wound. Inadequate irrigation or debridement, introduction of foreign material into the wound, and extensive tissue injury increase the bone's susceptibility to infection. If the infection is not sufficiently treated, the acute infection may become chronic. Osteomyelitis may also occur as a complication of surgery, such as in the sternum after open heart surgery or in extremities after bone allograft or total joint replacement.

Osteomyelitis after trauma or bone surgery usually is associated with persistent or recurrent fevers, increased pain at the operative or trauma site, and poor incisional healing, which often is accompanied by continued wound drainage and wound separation. Prosthetic joint infections present with joint pain, fever, and cutaneous drainage.

Diagnosis requires both confirming the infection and identifying the offending microorganism. The diagnosis of skeletal infection entails use of various imaging strategies, including conventional radiology, nuclear imaging studies, CT scans, and MRI. Bone biopsy may be used to identify the causative microorganisms. Treatment includes the use of antibiotics and selective use of surgical interventions.

N=+ S=+ → S

■ FIGURE 42-7 ■ Hematogenous osteomyelitis of the fibula of 3 months' duration. The entire shaft has been deprived of its blood supply and has become a sequestrum (S) surrounded by new immature bone, involucrum (Iv). Pathologic fractures are present in the lower tibia and fibula. (Wilson F.C. [1980]. *The musculoskeletal system*. [2nd ed., p. 150]. Philadelphia: J.B. Lippincott)

Hematogenous Osteomyelitis

Hematogenous osteomyelitis occurs as the result of localization of a blood-borne infection in the bone. It is seen most commonly in children younger than 10 years of age²¹ but is seen occasionally in the elderly.²⁰

In children, it commonly begins in the metaphyseal region of long bones and usually is preceded by staphylococcal or streptococcal infections of the skin, sinuses, teeth, or middle ear. Thrombosis occurring as the result of local trauma may predispose to localization of the infection consequent to bacteremia.²¹

In the adult, hematogenous osteomyelitis usually affects the axial skeleton and the irregular bones in the wrist and ankle. It is most common in debilitated patients and in those with a history of chronic skin infections, chronic urinary tract infections, intravenous drug use, and in those who are immunologically suppressed. Intravenous drug users are at risk for infections with *Streptococcus* and *Pseudomonas*.

The signs and symptoms of acute hematogenous osteomyelitis are those of bacteremia accompanied by symptoms referable to the site of the bone lesion. Bacteremia is characterized by chills, fever, and malaise. There often is pain on movement of the affected extremity, loss of movement, and local tenderness followed by redness and swelling. X-ray studies may appear normal initially, but they show evidence of periosteal elevation and increased osteoclastic activity after an abscess has formed.

The treatment of hematogenous osteomyelitis begins with identification of the causative organism through blood cultures, aspiration cultures, and Gram's stain.²² Antibiotics are given first intravenously and then orally. Debridement and surgical drainage also may be necessary.

Chronic Osteomyelitis

SHORT

Chronic osteomyelitis has long been recognized as a disease. However, the incidence has decreased in the past century because of improvements in surgical techniques and antibiotic therapy. Chronic osteomyelitis includes all inflammatory processes of bone, excluding those in rheumatic diseases, that are caused by microorganisms. It may be the result of delayed or inadequate treatment of acute hematogenous osteomyelitis or osteomyelitis caused by direct contamination of bone. Acute osteomyelitis is considered to have become chronic when the infection persists beyond 6 to 8 weeks or when the acute process has been adequately treated and is expected to resolve but does not. Chronic osteomyelitis can persist for years; it may appear spontaneously, after a minor trauma, or when resistance is lowered.

The hallmark feature of chronic osteomyelitis is the presence of infected dead bone, a sequestrum, that has separated from the living bone. A sheath of new bone, called the involucrum, forms around the dead bone. Radiologic techniques such as x-ray films, bone scans, and sinograms are used to identify the infected site. Chronic osteomyelitis or infection around a total joint prosthesis can be difficult to diagnose because the classic signs of infection are not apparent and the blood leukocyte count may not be elevated. A subclinical infection may exist for years. Bone scans are used in conjunction with bone biopsy for a definitive diagnosis.²³

Treatment includes the use of antibiotics. Wound cultures are used to identify the microorganism and its sensitivity to

antibiotic therapy. Initial antibiotic therapy is followed by surgery to remove foreign bodies (*e.g.*, metal plates, screws) or sequestra and by long-term antibiotic therapy. Immobilization of the affected part usually is necessary, with restriction of weight bearing on a lower extremity.

Osteomyelitis With Vascular Insufficiency

In persons with vascular insufficiency, osteomyelitis may develop from a skin lesion. It is seen most commonly associated with chronic or ischemic foot ulcers in persons with longstanding diabetes or other chronic vascular disorders. It is characterized by local cellulitis with inflammation and necrosis. Treatment depends on the oxygen tension of the involved tissues. Debridement and antibiotic therapy may benefit persons who have good oxygen tension in the infected site. Amputation is indicated when oxygen tension is inadequate.

Tuberculosis of the Bone or Joint

Tuberculosis can spread from one part of the body, such as the lungs or the lymph nodes, to the bones and joints. When this happens, it is called *extrapulmonary* or *miliary tuberculosis*. It is caused by *Mycobacterium tuberculosis*. The disease is localized and progressively destructive but not as contagious as primary pulmonary tuberculosis. In approximately 50% of cases, it affects the vertebrae, but it also frequently is seen in the hip and knee.²⁴ Tuberculosis also can affect the joints and soft tissues. The disease is characterized by bone destruction and abscess formation. Local symptoms include pain, immobility, and muscle atrophy; joint swelling, mild fever, and leukocytosis also may occur. Diagnosis is confirmed by a positive culture. The most important part of the treatment is antituberculosis drug therapy. Conservative treatment is usually as effective as surgery, especially for earlier and milder cases.

Because of improved methods to prevent and treat tuberculosis, its incidence had diminished in recent decades. However, tuberculosis has re-emerged as a health problem, affecting one third of the global population and 10 million people in the United States.²⁵ Tuberculosis has increased because of the spread of disease in communal settings (*e.g.*, jails, shelters, nursing homes), the human immunodeficiency virus epidemic, and the influx of immigrants who have come to the United States from countries where the disease is endemic.²⁶ Unfortunately, the diagnosis of tuberculosis in the bones and joints still may be missed, especially when the musculoskeletal infection is the sole presenting sign. CT scans and MRI can be used as aids for early diagnosis.

In summary, bone infections occur because of the direct or indirect invasion of the skeletal circulation by microorganisms, most commonly *S. aureus*. Osteomyelitis, or infection of the bone and marrow, can be an acute or chronic disease. Acute osteomyelitis is seen most often as a result of the direct contamination of bone by a foreign object. Chronic osteomyelitis is a long-term process that can recur spontaneously at any time throughout a person's life. Tuberculosis of the bone, which is characterized by bone destruction and abscess formation, is caused by spread of the infection from the lungs or lymph nodes.

NEOPLASMS

Neoplasms in the skeletal system usually are referred to as *bone tumors*. Like other types of neoplasms, bone tumors may be benign or malignant (see Table 42-2). Primary bone tumors may arise from any of the skeletal components, including osseous bone tissue, cartilage, and bone marrow. Primary malignant tumors of the bone are uncommon, constituting approximately 1% of all adult cancers and 15% of pediatric malignancies.²⁷ However, metastatic disease of the bone is relatively common.

There are three major symptoms of bone tumors: pain, presence of a mass, and impairment of function (Chart 42-3). Pain is a feature common to almost all malignant tumors but may or may not occur with benign tumors. For example, a benign bone cyst usually is asymptomatic until a fracture occurs. Pain that persists at night and is not relieved by rest suggests malignancy. A mass or hard lump may be the first sign of a bone tumor. A malignant tumor is suspected when a painful mass exists that is enlarging or eroding the cortex of the bone. The ease of discovery of a mass depends on the location of the tumor; a small lump arising on the surface of the tibia is easy to detect, whereas a tumor that is deep in the medial portion of the thigh may grow to a considerable size before it is noticed. Benign and malignant tumors may cause the bone to erode to the point where it cannot withstand the strain of ordinary use. In such cases, even a small amount of bone stress or trauma precipitates a pathologic fracture. A tumor may produce pressure on a peripheral nerve, causing decreased sensation, numbness, a limp, or limitation of movement.

Benign Neoplasms

Benign bone tumors usually are limited to the confines of the bone, have well-demarcated edges, and are surrounded by a thin rim of sclerotic bone. The four most common types of be-

TABLE 42-2Classification of PrimaryBone Neoplasms				
Tissue Type	Benign Neoplasm	Malignant Neoplasm		
Bone	Osteoid osteoma Benign osteo- blastoma	Osteosarcoma Parosteal osteogenic sarcoma		
Cartilage	Osteochondroma Chondroma Chrondroblastoma Chondromyxoid fibroma	Chondrosarcoma		
Lipid	Lipoma	Liposarcoma		
Fibrous and fibroosseous tissue	Fibrous dysplasia	Fibrosarcoma Malignant fibrous histiocytoma		
Miscellaneous Bone marrow	Giant cell tumor	Malignant giant cell Ewing's sarcoma Multiple myeloma Reticulum cell sarcoma		

CHART 42-3 Symptoms of Bone Cancer

- Bone pain in an adult or child that comes on slowly but lasts for as long as a week, is constant or intermittent, and may be worse at night.
- Unexplained swelling or lump on the bones of the arms, legs, thighs, or other parts of the body that is firm and slightly tender and may be felt through the skin. It may interfere with normal movement and can cause the bone to break.

These symptoms are not sure signs of cancer. They also may be caused by other, less serious problems. Only a physician can tell for sure.

(Adapted from U.S. Department of Health and Human Services [1993]. What you need to know about cancers of the bone. NIH publication no. 93-1517. Bethesda, MD: U.S. Government Printing Office)

nign bone tumors are osteoma, chondroma, osteochondroma, and giant cell tumor.

An *osteoma* is a small bony tumor found on the surface of a long bone, flat bone, or the skull. It usually is composed of hard, compact (ivory osteoma), or spongy (cancellous) bone. It may be excised or left alone.

A *chondroma* is a tumor composed of hyaline cartilage.¹⁰ It may arise on the surface of the bone (*i.e.*, ecchondroma) or within the medullary cavity (*i.e.*, endochondroma). These tumors may become large and are especially common in the hands and feet. A chondroma may persist for many years and then take on the attributes of a malignant chondrosarcoma. A chondroma usually is not treated unless it becomes unsightly or uncomfortable.

An *osteochondroma* is the most common form of benign tumor in the skeletal system, representing 50% of all benign bone tumors and approximately 15% of all primary skeletal lesions.²⁷ It grows only during periods of skeletal growth, originating in the epiphyseal cartilage plate and growing out of the bone like a mushroom. An osteochondroma is composed of cartilage and bone and usually occurs singly but may affect several bones in a condition called *multiple exostoses*. Malignant changes are rare, and excision of the tumor is done only when necessary.

A giant cell tumor, or osteoclastoma, is an aggressive tumor of multinucleated cells that often behaves like a malignant tumor, metastasizing through the bloodstream and recurring locally after excision. It occurs most often in young adults, predominantly females, and is found most commonly in the knee, wrist, or shoulder. The tumor begins in the metaphyseal region, grows into the epiphysis, and may extend into the joint surface. Pathologic fractures are common because the tumor destroys the bone substance. Clinically, pain may occur at the tumor site, with gradually increasing swelling. X-ray films show destruction of the bone with expansion of the cortex.

The treatment of giant cell tumors depends on their location. If the affected bone can be eliminated without loss of function, such as the clavicle or fibula, the entire bone or part of it may be removed. When the tumor is near a major joint, such as the knee or shoulder, a local excision is done. Irradiation may be used to prevent recurrence of the tumor.

Malignant Bone Tumors

In contrast to benign tumors, malignant tumors tend to be ill defined, lack sharp borders, and extend beyond the confines of the bone, showing that it has destroyed the cortex. Specific types of bone tumors affect different age groups. They are virtually unknown in infancy, rare in children younger than 10 years of age, and peak during the teenage years. Adolescents have the highest incidence, with a rate of 3 cases per 100,000.²⁷ The two major forms of bone cancer in children and young adults are osteosarcoma and Ewing's sarcoma. It is unusual for either condition to be seen after 20 years of age and older.¹⁰

The methods used in the diagnosis of malignant bone tumors include radiographic studies, CT scans, MRI, and bone biopsy. Radiographs give the most general diagnostic information, such as malignant versus benign and primary versus metastatic status. The radiograph demonstrates the region of bone involvement, extent of destruction, and amount of reactive bone formed. Radioisotope scans are used to estimate the local intramedullary extent of the tumor and screen for other skeletal areas of involvement. CT scans further aid diagnosis and anatomic localization and can identify small pulmonary metastases not seen by conventional radiographs. MRI is the most accurate method of evaluating the intramedullary extent of bone tumor and can demarcate the soft structures in relation to neurovascular structures without the use of contrast media. It is best used in conjunction with a CT scan.²⁷ A bone biopsy is used to determine the type of tumor that is present.

Osteosarcoma

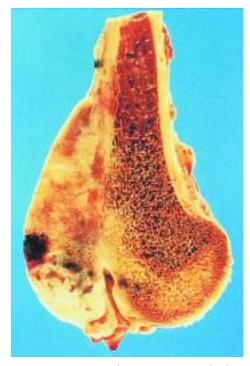
Osteosarcoma represents 60% of all bone tumors occurring in children and adolescents. Osteosarcoma has a bimodal distribution, with 75% occurring in persons younger than 20 years of age. A second peak occurs in the elderly with predisposing factors such as Paget's disease, bone infarcts, or prior irradiation.¹⁰ The male-to-female ratio increases to approximately 1.6 to 1 during late adolescence and adulthood. It is seen most commonly during periods of maximal growth. In younger persons, the primary tumor most often is located at the anatomic sites associated with maximum growth velocity—the distal femur, proximal tibia, and proximal humerus. Persons affected with osteosarcoma usually are tall and are found to have a high plasma level of somatomedin. Bone tumors in the elderly are more common in the humerus, pelvis, and proximal femur.

Osteosarcoma is a malignant tumor of mesenchymal cells, characterized by the direct formation of osteoid or immature bone by malignant osteoblasts. These cells synthesize thin, wispy, and purposeless fragments of bone. Osteogenic sarcomas are aggressive tumors that grow rapidly; they often are eccentrically placed in the bone and move from the metaphysis of the bone out to the periosteum, with subsequent spread to adjacent soft tissues.

The causes of osteosarcoma are unknown. The correlation of age and location of most of the tumors with the period of maximum growth suggests some relation to increased osteoblastic activity. Paget's disease, which is linked to osteosarcoma in adults, also is associated with increased osteoblastic activity. Irradiation from an internal source, such as the radioactive pharmaceutical technetium used in bone scans, or an external source, such as x-ray films, also has been associated with osteosarcoma.

The primary clinical feature of osteosarcoma is localized pain and swelling in the affected bone, usually of sudden onset. Patients and their families often associate the symptoms with recent trauma.28 The skin overlying the tumor may be warm, shiny, and stretched, with prominent superficial veins. The range of motion of the adjacent joint may be restricted. Osteosarcoma usually begins as a firm white or reddish mass and later becomes softer with a viscous interior (Fig. 42-8). The tumor infrequently metastasizes to the lymph nodes because the cells are unable to grow in the node. Nodal metastases usually occur only in the late course of disseminated disease. Most often, the tumor cells exit the primary tumor through the venous end of the capillary, and early metastasis to the lung is common. Lung metastases, even if massive, usually are relatively asymptomatic. The prognosis for a patient with osteosarcoma depends on the aggressiveness of the disease, radiologic features, presence or absence of pathologic fracture, size of the tumor, rapidity of tumor growth, and gender of the person.

The treatment for sarcomas is surgery in combination with multiagent chemotherapy used both before and after surgery.^{28,29} Osteosarcomas are relatively resistant to radiation therapy. In the past, treatment usually entailed amputation above the level of the tumor. Limb salvage surgical procedures, using a metal prosthesis or cadaver allografts, are becoming a standard alternative. Studies have shown that limb salvage surgery has no adverse effects on the long-term survival



■ FIGURE 42-8 ■ Juxtacortical osteosarcoma. The lower femur contains a malignant tumor arising from the periosteal surface of the bone and sparing the medullary cavity. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 1386]. Philadelphia: Lippincott Williams & Wilkins)

of persons with osteosarcoma. Chemotherapy using various drug combinations is the most effective treatment for metastatic osteosarcoma.

Ewing's Sarcoma

Ewing's sarcoma is the third most common type of primary bone tumor, and it is highly malignant. It commonly occurs in males younger than 25 years of age, with the incidence highest among teenagers.²⁸ Ewing's tumor arises from immature bone marrow cells and causes bone destruction from within. It usually occurs in the shaft of long bones or any portion of the pelvis and often metastasizes to bone marrow.

Manifestations of Ewing's tumor include pain, limitation of movement, and tenderness over the involved bone or soft tissue. It often is accompanied by systemic manifestations such as fever or weight loss, which may serve to confuse the diagnosis. There also may be a delay in diagnosis when the pain and swelling associated with the tumor are attributed to a sports injury. Pathologic fractures are common because of bone destruction.

Treatment methods incorporate a combination of multiagent chemotherapy, surgery, and radiation therapy. There is controversy about the potential superiority of surgical resection compared with radiation therapy.²⁸ Radiation therapy is associated with risk of radiation-induced second malignancies, especially osteosarcoma.

Chondrosarcoma

Chondrosarcoma, a malignant tumor of cartilage that can develop in the medullary cavity or peripherally, is the second most common form of malignant bone tumor. It occurs primarily in middle or later life and slightly more often in males. The tumor arises from points of muscle attachment to bone, particularly the knee, shoulder, hip, and pelvis. Chondrosarcomas can arise from underlying benign lesions such as osteo-chondroma, chondroblastoma, or fibrous dysplasia.¹⁰

Chondrosarcomas are slow growing, metastasize late, and often are painless. They can remain hidden in an area such as the pelvis for a long time. This type of tumor, like many primary malignancies, tends to destroy bone and extend into the soft tissues beyond the confines of the bone of origin. Chondrosarcomas mainly affect the bones of the trunk, pelvis, or proximal femur and rarely develop in the distal portion of a bone. Irregular flecks and ringlets of calcification often are prominent radiographic findings. Early diagnosis is important because chondrosarcoma responds well to early radical surgical excision. It usually is resistant to radiation therapy and available chemotherapeutic agents. Not infrequently, these tumors transform into a highly malignant tumor, mesenchymal chondrosarcoma, which requires a more aggressive treatment, including combination chemotherapy.

Metastatic Bone Disease

Skeletal metastases are the most common malignancy of osseous tissue. Approximately half of all people with cancer have bone metastasis at some point in their disease.³⁰ Metastatic lesions are seen most often in the spine, femur, pelvis, ribs, sternum, proximal humerus, and skull, and are less common in anatomic sites that are further removed from the trunk of the body. Tumors that frequently spread to the skeletal system are those of the breast, lung, prostate, kidney, and thyroid, although any cancer can ultimately involve the skeleton. More than 85% of bone metastases result from primary lesions in the breast, lung, or prostate.³¹ The incidence of metastatic bone disease is highest in persons older than 40 years of age.

The major symptom of bone metastasis is pain with evidence of an impending pathologic fracture. It usually develops gradually, over weeks, and is more severe at night. Pain is caused by stretching of the periosteum of the involved bone or by nerve entrapment, as in the nerve roots of the spinal cord by the vertebral body. Pathologic fractures occur in approximately 10% to 15% of persons with metastatic bone disease. The affected bone appears to be eaten away on x-ray images and, in severe cases, crumbles on impact, much like dried toast. Many pathologic fractures occur in the femur, humerus, and vertebrae. In the femur, fractures occur because the proximal aspect of the bone is under great mechanical stress.³¹

X-ray examinations are used along with CT or bone scans to detect, diagnose, and localize metastatic bone lesions. Approximately one third of persons with skeletal metastases have positive bone scans without radiologic findings. This is because 50% of the trabecular bone must be destroyed before a lesion is visible on plain radiographs.³² Arteriography using radiopaque contrast media may be helpful in outlining the tumor margins. A bone biopsy usually is done when there is a question regarding the diagnosis or treatment. A closedneedle biopsy with CT localization is particularly useful with spine lesions. Serum levels of alkaline phosphatase and calcium often are elevated in persons with metastatic bone disease.

The primary goals in treatment of metastatic bone disease are to prevent pathologic fractures and promote survival with maximum functioning, allowing the person to maintain as much mobility and pain control as possible. Treatment methods include chemotherapy, irradiation, and surgical stabilization. Radiation therapy is primarily used as a palliative treatment to alleviate pain and prevent pathologic fractures. After a pathologic fracture has occurred, bracing, intramedullary nailing of the femur, and spine stabilization may be done. Because adequate fixation often is difficult in diseased bone, cement (*i.e.*, methylmethacrylate) often is used with internal fixation devices to stabilize the bone.

In summary, bone tumors, like any other type of neoplasm, may be benign or malignant. Benign bone tumors grow slowly and usually do not destroy the surrounding tissues. Malignant tumors can be primary or metastatic. Primary bone tumors are rare, grow rapidly, metastasize to the lungs and other parts of the body through the bloodstream, and are associated with a high mortality rate. Metastatic bone tumors usually are multiple, originating primarily from cancers of the breast, lung, and prostate. The incidence of metastatic bone disease is increasing probably because improved treatment methods enable persons with cancer to live longer. Advances in chemotherapy, radiation therapy, and surgical procedures have substantially increased the survival and cure rates for many types of bone cancers. A primary goal in metastatic bone disease is the prevention of pathologic fractures.

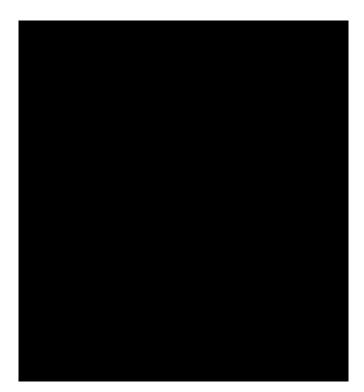
During childhood, skeletal structures grow in length and diameter and sustain a large increase in bone mass. Alterations in musculoskeletal structure and function may develop as a result of normal growth and developmental processes or as a result of impairment of skeletal development caused by hereditary or congenital influences.

Alterations During Normal Growth Periods

Infants and children undergo changes in muscle tone and joint motion during growth and development. Toeing-in, toeingout, bowlegs, and knock-knees occur frequently in infancy and childhood.³³ These changes usually cause few problems and are corrected during normal growth processes. The normal folded position of the fetus in utero causes physiologic flexion contractures of the hips and a froglike appearance of the lower extremities (Fig. 42-9). The hips are externally rotated, and the patellae point outward, whereas the feet appear to point forward because of the internal pulling force of the tibiae. During the first year of life, the lower extremities begin to straighten out in preparation for walking. Internal and external rotation become equal, and the hips extend. Flexion contractures of the shoulders, elbows, and knees also are commonly seen in newborns, but they should disappear by 4 to 6 months of age.34

Torsional Deformities

All infants and toddlers have lax ligaments that become tighter with age and assumption of the weight-bearing posture. The hypermobility that accompanies joint laxity coupled with the torsional, or twisting, forces exerted on the limbs during



KEY CONCEPTS

DEVELOPMENTAL SKELETAL DISORDERS

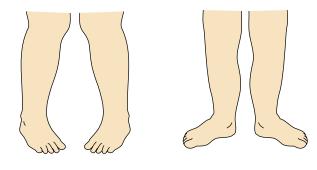
- Many disorders of early infancy are caused by intrauterine positions and resolve as the child grows.
- Infants and toddlers have lax ligaments that predispose to skeletal disorders caused by twisting or torsional forces.
- Bone growth in infants and children occurs at the epiphysis. Separation of the epiphyseal growth plate ruptures the blood vessels that nourish the epiphysis, causing cessation of growth and shortened extremity length.

growth are responsible for a number of variants seen in young children. Torsional forces caused by intrauterine positions or sleeping and sitting patterns twist the growing bones and can produce the deformities as a child grows and develops.

In infants, the femur normally is rotated to an anteverted position with the femoral head and neck rotated anteriorly with respect to the femoral condyles. Femoral anteversion (*i.e.*, medial rotation) decreases from an average 40 degrees at birth to approximately 15 degrees at maturity. The normal tibia is externally rotated approximately 5 degrees at birth and 15 degrees at maturity. Torsional abnormalities frequently demonstrate a familial tendency.

Toeing-in and Toeing-out. The foot progression angle describes the angle between the axis of the foot and the line of progression. It is determined by watching the child walking and running. Figure 42-10 illustrates the position of the foot in toeing-in and toeing-out.

Toeing-in (*i.e.*, metatarsus adductus) is the most common congenital foot deformity, affecting boys and girls equally. It can be caused by torsion in the foot, lower leg, or entire leg. Toeing-in caused by adduction of the forefoot (*i.e.*, congenital metatarsus adductus) usually is the result of the fetal position maintained in utero. It may occur in one foot or both feet. A supple deformity can be passively manipulated into a straight position and requires no treatment. Treatment consisting of



Toeing-in Toeing-out ■ FIGURE 42-10 ■ Position of feet in toeing-in and toeing-out. serial long leg casting or a brace that pushes the metatarsals (not the hindfoot) into abduction usually is required in a fixed deformity (*i.e.*, one in which the forefoot cannot be passively manipulated into a straight position).

Toeing-out is a common problem in children and is caused by external femoral torsion. This occurs when the femur can be externally rotated to approximately 90 degrees but internally rotated only to a neutral position or slightly beyond. Because the femoral torsion persists when a child habitually sleeps in the prone position, an external tibial torsion also may develop. If external tibial torsion is present, the feet point lateral to the midline of the medial plane. External tibial torsion rarely causes toeing-out; it only intensifies the condition. Toeing-out usually corrects itself as the child becomes proficient in walking. Occasionally, a night splint is used. Toeingin and toeing-out are less noticeable when the child is running or barefoot. Overcorrection of a supple foot deformity can cause flatfoot deformity, but a rigid deformity that is untreated can cause pain and improper fitting of footwear.

Tibial Torsion. Tibial torsion is determined by measuring the thigh-foot angle, which is done with the ankle and knee positioned at 90 degrees (Fig. 42-11A). In this position, the foot



normally rotates outward. *Internal tibial torsion* (*i.e.*, bowing of the tibia) is a rotation of the tibia that makes the feet appear to turn inward. It is the most common cause of toeing-in in children younger than 2 years of age. It is present at birth and may fail to correct itself if children sleep on their knees with the feet turned in or sit on in-turned feet. It is thought to be caused by genetic factors and intrauterine compression, such as an unstretched uterus during a first pregnancy or intrauterine crowding with twins or multiple fetuses. Tibial torsion improves naturally with growth, but this may take years.³⁵ The Denis Browne splint, a bar to which shoes are attached, may be used to put the feet into mild external rotation while the child is sleeping.

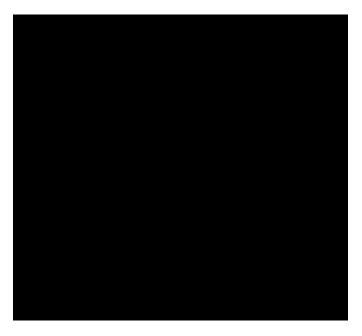
External tibial torsion, a much less common disorder, is associated with calcaneovalgus foot and is caused by a normal variation of intrauterine positioning or a neuromuscular disorder. It is characterized by an abnormally positive thigh-foot angle of 30 to 50 degrees. The condition corrects itself naturally, and treatment is observational. Significant improvement begins during the first year with the onset of ambulation and usually is complete by 2 to 3 years of age.³⁴ The normal adult exhibits 20 degrees of tibial torsion.

Femoral Torsion. Femoral torsion refers to abnormal variations in hip rotation. Hip rotation is measured at the pelvic level with the child in the prone position and the knees flexed at a 90-degree angle. In this position, the hip is in a neutral position. Rotating the lower leg outward produces internal or medial femoral rotation; rotating it inward produces external or lateral rotation (see Fig. 42-11B). During measurement of hip rotation, the legs are allowed to fall to full internal rotation by gravity alone; lateral rotation is measured by allowing the legs to fall inward and cross. Hip rotation in flexion and extension also can be measured with computed tomography (CT). By 1 year of age, there is normally approximately 45 degrees of internal and 45 degrees of external rotation.³⁴

Internal femoral torsion, also called femoral anteversion, is a normal variant commonly seen during the first 6 years of life, especially in 3- and 4-year-old girls.³³ Characteristically, there is 80 to 90 degrees of internal rotation of the hip in the prone position.³⁴ The condition is thought to be related to increased laxity of the anterior capsule of the hip such that it does not provide the stable pressure needed to correct the anteversion that is present at birth. Children are most comfortable sitting in the "W" position with their hips between their knees (Fig. 42-12). It is believed that this position allows the lower leg to act as a lever, producing torsional changes in the femur. When the child stands, the knees turn in and the feet appear to point straight ahead; when the child walks, knees and toes point in. Children with this problem are encouraged to sit cross-legged or in the so-called *tailor position*. If left untreated, the tibiae compensate by becoming externally rotated so that by 8 to 12 years of age the knees may turn in but the feet no longer do. This can result in patellofemoral malalignment with patellar subluxation or dislocation and pain. A derotational osteotomy may be done in severe cases or if there is functional disability.

Genu Varum and Genu Valgum

Genu varum or *bowlegs* is an outward bowing of the knees greater than 1 inch when the medial malleoli of the ankles are touching (Fig. 42-13). Most infants and toddlers have some



bowing of their legs up to age 2 years. If there is a large separation between the knees (>15 degrees) after 2 years of age, the child may require bracing. The child also should be evaluated for diseases such as rickets or tibia vara (*i.e.*, Blount's disease to be discussed).



■ FIGURE 42-13 ■ Normal genu varum (bowlegs) in a toddler (*left*) and genu valgum (knock-knees) in a toddler (*right*), which is often seen in children between 2 and 6 years of age. (From Weinstein S.L., Buckwalter J.A. [1994]. *Turek's orthopaedics* [5th ed.]. Philadelphia: J.B. Lippincott)

Genu valgum or knock-knees is a deformity in which there is decreased space between the knees (see Fig. 42-13). The medial malleoli in the ankles cannot be brought in contact with each other when the knees are touching. It is seen most frequently in children between the ages of 3 and 5 years and should resolve by 5 to 8 years of age.³⁴ The condition usually is the result of lax medial collateral ligaments of the knee and may be exacerbated by sitting in the "M" position. Genu valgum can be ignored to age 7 years, unless it is more than 15 degrees, unilateral, or associated with short stature. It usually resolves spontaneously and rarely requires treatment. If genu varum or genu valgum persists and is uncorrected, osteoarthritis may develop in adulthood as a result of abnormal intra-articular stress. Genu varum can cause gait awkwardness and increased risk of sprains and fractures. Uncorrected genu valgum may cause subluxation and recurrent dislocation of the patella, with a predisposition to chondromalacia and joint pain and fatigue.

Hereditary and Congenital Deformities

Congenital deformities are abnormalities that are present at birth. They range in severity from mild limb deformities, which are relatively common, to major limb malformations, which are relatively rare. There may be a simple webbing of the fingers or toes (syndactyly), the presence of an extra digit (*i.e.*, polydactyly), or the absence of a bone such as the phalanx, rib, or clavicle. Joint contractures and dislocations produce more severe deformity, as does the absence of entire bones, joints, or limbs. An epidemic of limb deformities occurred from 1957 to 1962 as a result of maternal ingestion of thalidomide. This drug was withdrawn from the market in 1961.

Congenital deformities are caused by many factors, some unknown. These factors include genetic influences, external agents that injure the fetus (*e.g.*, radiation, alcohol, drugs, viruses), and intrauterine environmental factors. Many of the organic bone matrix components have been identified only recently, and their interactions found to be more complex than originally thought. Diseases associated with abnormalities in bone matrix include those with deficient collagen synthesis and decreased bone mass.

Osteogenesis Imperfecta

Osteogenesis imperfecta is a hereditary disease characterized by defective synthesis of type I collagen.³⁶ It is one of the most common hereditary bone diseases, with an occurrence rate of approximately 1 case in 10,000 births.¹² Although it usually is transmitted as an autosomal dominant trait, a distinct form of the disorder with multiple lethal defects is thought to be inherited as an autosomal recessive trait.¹⁰ In some cases the defect is caused by a spontaneous mutation.

The clinical manifestations of osteogenesis imperfecta include a spectrum of disorders marked by extreme skeletal fragility. Four major subtypes have been identified.¹⁰ Type II is at one end of the spectrum and is uniformly fatal in utero or during early perinatal life. Less severe disease occurs when the disorder is inherited as a dominant trait. The skeletal system is not so weakened, and fractures often do not appear until the child becomes active and starts to walk, or even later in childhood. These fractures heal rapidly, although with a poor-quality callus. Other problems associated with defective connective tissue synthesis include short stature, thin skin, blue or gray sclera, abnormal tooth development, hypotonic muscles, loose-jointedness, scoliosis, and a tendency toward hernia formation. Hearing loss caused by otosclerosis of the middle and inner ear is common in affected adults.

There is no known medical treatment for correction of the defective collagen synthesis that is characteristic of osteogenesis imperfecta. Instead, treatment modalities focus on preventing and treating fractures.

Congenital Clubfoot

Clubfoot, or talipes, is a congenital deformity of the foot that can affect one or both feet. Like congenital dislocation of the hip, its occurrence follows a multifactorial inheritance pattern. The condition has an incidence of 1 case per 1000 live births and occurs twice as often in males as in females.³⁷ Clubfoot is associated with chromosomal abnormalities and may be associated with other congenital syndromes that are transmitted by mendelian inheritance patterns (see Chapter 4). However, it is most commonly idiopathic and found in healthy infants in whom no genetic or chromosomal abnormality or other extrinsic cause can be found.

Although the exact cause of clubfoot is unknown, three theories are generally accepted: an anomalous development occurs during the first trimester of pregnancy, the leg fails to rotate inward and move from the equinovarus position at approximately the third month, or the soft tissues in the foot do not mature and lengthen. Maternal smoking is associated with occurrence of clubfoot, and the risk increases enormously when combined with a family history.³⁸

In forefoot adduction, which accounts for approximately 95% of idiopathic cases, the foot is plantar flexed and inverted. This is the so-called *equinovarus type* in which the foot resembles a horse's hoof (Fig. 42-14). The other 5% of cases are of the calcaneovalgus type, or reverse clubfoot, in which the foot is dorsiflexed and everted. The reverse clubfoot can occur as an isolated condition or in association with multiple congenital

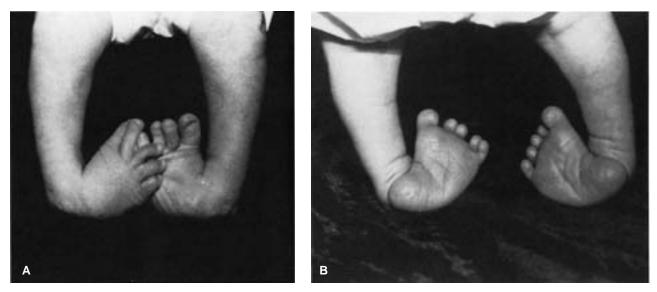
defects. At birth, the feet of many infants assume one of these two positions, but they can be passively overcorrected or brought back into the opposite position. If the foot cannot be overcorrected, some type of correction may be necessary.

Clubfoot varies in severity from a mild deformity to one in which the foot is completely inverted. Treatment is begun as soon as the diagnosis is made and usually is effective within a short period. Serial manipulations and casting are used to gently correct each component in the forefoot varus, the hindfoot varus, and the equinus. The treatment is continued until the foot is in a normal position with full correction evident clinically and on radiographic studies. Surgery may be required for severe deformities or when nonoperative treatment methods are unsuccessful.

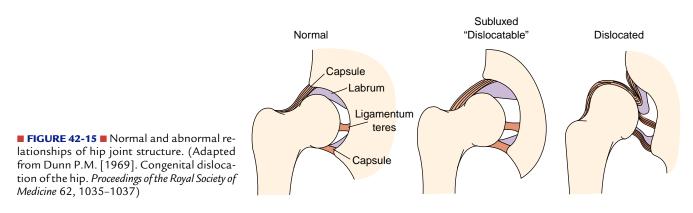
Developmental Dysplasia of the Hip

Developmental dysplasia of the hip, formerly known as *congenital dislocation of the hip*, is an abnormality in hip development that leads to a wide spectrum of hip problems in infants and children, including hips that are unstable, malformed, subluxated, or dislocated.¹⁰⁻¹² In less severe cases, the hip joint may be unstable, with excessive laxity of the joint capsule, or subluxated, so that the joint surfaces are separated and there is a partial dislocation (Fig. 42-15). With dislocated hips, the head of the femur is located outside of the acetabulum.

The results of newborn screening programs have shown that 1 of 100 infants have some evidence of hip instability; however, dislocation of the hip is seen in 1.5 of every 1000 live births.¹² The left hip is involved three times more frequently than the right hip because of the left occipital intrauterine positioning of most infants.³⁹ In white infants, developmental dysplasia of the hips occurs most frequently in first-born children and is six times more common in female than in male infants.⁴⁰ The cause of developmental dysplasia of the hip is multifactorial, with physiologic, mechanical, and postural factors playing a role. A positive family history and generalized



■ FIGURE 42-14 ■ Severe clubfoot deformity. (A) The heel is in severe varus, and the forefoot is adducted and inverted. (B) The cavus deformity results from the slightly pronated position of the forefoot in relation to the hindfoot. (From Weinstein S.L., Buckwalter J.A. [1994]. *Turek's orthopaedics* [5th ed.]. Philadelphia: J.B. Lippincott)



laxity of the ligaments are related. The increased frequency in girls is thought to result from their susceptibility to maternal estrogens and other hormones associated with pelvic relaxation. Dislocation also may result from environmental factors such as fetal position, a tight uterus that prevents fetal movement, and breech delivery.

Early diagnosis of a developmental dysplasia of the hip is important because treatment is easiest and most effective if begun during the first 6 months of life. Repeated dislocation causes damage to the femoral head and the acetabulum. Clinical examinations to detect dislocation of the hip should be done at birth and every several months during the first year of life. In infants, signs of dislocation include asymmetry of the hip or gluteal folds, shortening of the thigh so that one knee (on the affected side) is higher than the other, and limited abduction of the affected hip (Fig. 42-16). The asymmetry of gluteal folds is not definitive but indicates the need for further evaluation. In the older child, instability of the hip may produce a delay in walking or eventually cause a characteristic waddling gait. Diagnosis is confirmed by radiography. Ultrasound is used to diagnose the conditions in newborns and infants from birth to 4 months of age.40

The treatment of a developmental dysplasia should be individualized and depends on whether the hip is subluxated or dislocated. Mild instability often resolves without treatment. The best results are obtained if the treatment is begun before



■ FIGURE 42-16 ■ An 18-month-old girl has congenital dysplasia of the left hip. (From Weinstein S.L., Buckwalter J.A. [1994]. *Turek's orthopaedics* [5th ed.]. Philadelphia: J.B. Lippincott)

changes in the hip structure (*e.g.*, 2 to 3 months) prevent it from being reduced by gentle manipulation or abduction devices. Infants with dislocated hips caused by anatomic changes and toddlers who may lack development of the acetabular socket require more aggressive treatment, such as open reduction and joint reconstruction. Treatment at any age includes reduction of the dislocation and immobilization of the legs in an abducted position. With children younger than 3 years, skin traction is used when reduction cannot be easily obtained. This treatment is followed by several months of immobilization in a hip spica cast, plaster splints, or an abduction splint. Older children or adults with an unreduced dislocatable hip may require hip surgery because of damage to the articulating surface of the joint.

Disorders of Later Development

Developmental disorders of bones and joints are those that are not present at birth but make their appearance during later periods of development. Some may have a genetic component, some present during periods of rapid growth, and others may develop as the result of trauma and stress imposed by activities associated with various periods of development.

Developmental Disorders of the Hip

Legg-Calvé-Perthes Disease. Legg-Calvé-Perthes disease, or coxa plana, is an osteonecrotic disease of the proximal femoral (capital) epiphysis, which is the growth center for the head of the femur. It occurs in 1 of 1200 children, affecting primarily those between ages 2 and 13 years, with a peak incidence between 4 and 9 years.⁴¹ It occurs primarily in boys and is much more common in whites than African Americans. Although no definite genetic pattern has been established, it occasionally affects more than one family member.

The cause of Legg-Calvé-Perthes disease is unknown. The disorder usually is insidious in onset and occurs in otherwise healthy children. However, it may be associated with acute trauma. Affected children usually have a shorter stature. Undernutrition has been suggested as a causative factor. When girls are affected, they usually have a poorer prognosis than boys because they are skeletally more mature and have a shorter period for growth and remodeling than do boys of the same age. Although both legs can be affected, in 85% of cases, only one leg is involved.³⁵

The primary pathologic feature of Legg-Calvé-Perthes disease is an avascular necrosis of the bone and marrow involving the epiphyseal growth center in the femoral head. The disorder may be confined to part of the epiphysis, or it may involve the entire epiphysis. In severe cases, there is a disturbance in the growth pattern that leads to a broad, short femoral neck. The necrosis is followed by slow absorption of the dead bone during a period of 2 to 3 years. Although the necrotic trabeculae eventually are replaced by healthy new bone, the epiphysis rarely regains its normal shape.

Legg-Calvé-Perthes disease has an insidious onset with a prolonged course. The main symptoms are pain in the groin, thigh, or knee and difficulty in walking. The child may have a painless limp with limited abduction and internal rotation and a flexion contracture of the affected hip. The age of onset is important because young children have a greater capability for remodeling of the femoral head and acetabulum, so less flattening of the femoral head occurs. Early diagnosis is important and is based on correlating physical symptoms with radiographic findings that are related to the stage of the disease.

The goal of treatment is to reduce deformity and preserve the integrity of the femoral head. Conservative and surgical interventions are used in the treatment of Legg-Calvé-Perthes disease. Treatment involves abduction casts or braces to keep the legs separated in abduction with mild internal rotation. Surgery may be done to contain the femoral head in the acetabulum. The best surgical results are obtained when surgery is done early, before the epiphysis becomes necrotic.

Slipped Capital Femoral Epiphysis. Normally, the proximal femoral epiphysis unites with the neck of the femur between 16 and 19 years of age. Before this time (10 to 14 years of age in girls and 10 to 16 years in boys), the femoral head may slip from its normal position directly at the head of the femur and become displaced medially and posteriorly.⁴¹ The head is held in the acetabulum by the ligamentum teres, and the neck of the femur is pulled upward and outward. This produces an anterolateral and superior adduction with extension deformity. The condition occurs with an estimated frequency of between 1 in 100,000 to 1 in 800,000.³⁵ It is the most common disorder of the hip in adolescents.

The cause of slipped capital femoral epiphysis is obscure, but it may be related to the child's susceptibility to stress on the femoral neck as a result of genetics or abnormal structure. Boys are affected twice as often as girls, and in approximately one half of cases, the condition is bilateral. Affected children often are overweight with poorly developed secondary sex characteristics or, in some instances, are extremely tall and thin. In many cases, there is a history of rapid skeletal growth preceding displacement of the epiphysis. The condition also may be affected by nutritional deficiencies or endocrine disorders such a hypothyroidism, hypopituitarism, and hypogonadism. Rapid growth after administration of growth hormone has been associated with displacement of the epiphysis.

Children with the condition often report referred knee pain accompanied by difficulty in walking, fatigue, and stiffness. The diagnosis is confirmed by radiographic studies in which the degree of slipping is determined and graded according to severity. Early treatment is imperative to prevent lifelong crippling. Avoidance of weight bearing on the femur and bed rest are essential parts of the treatment. Traction or gentle manipulation under anesthesia is used to reduce the slip. Surgical insertion of pins to keep the femoral neck and head of the femur aligned is a common method of treatment for children with moderate or severe slips. Crutches are used for several months after surgical correction to prevent full weight bearing until the growth plate is sealed by the bony union.

Children with the disorder must be followed up closely until the epiphyseal plate closes. Long-term prognosis depends on the amount of displacement that occurs. Complications include avascular necrosis, leg shortening, malunion, and problems with the internal fixation. Degenerative arthritis may develop, requiring joint replacement later in life.

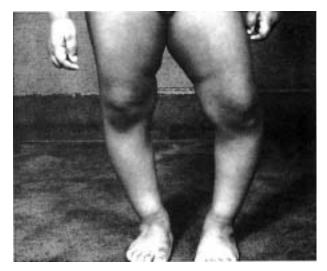
Disorders of the Knee

Blount's Disease. Idiopathic tibia vara, or Blount's disease, is a developmental deformity of the medial half of proximal tibial epiphysis that results in a progressive varus angulation below the knee (Fig. 42-17). It is the most common cause of pathologic genu varum and seen most often in black children, females, obese children, and early walkers.⁴² Onset can occur early in infancy or later, during adolescence. Adolescent Blount's disease occurs in the second decade of life, is seen in persons who are above the 95th percentile in height and weight, and is usually unilateral.⁴²

Long leg braces are used for treatment in early-onset disease. If progression occurs, or onset is late, surgery is done to correct the angulation and prevent further progression.

Osgood-Schlatter Disease. Osgood-Schlatter disease involves microfractures in the area where the patellar tendon inserts into the tibial tubercle, which is an extension of the proximal tibial epiphysis.³⁴ This area is particularly vulnerable to injury caused by sudden or continued strain from the patellar tendon during periods of growth, particularly in athletic individuals. It occurs most frequently in boys between the ages of 11 and 15 years and in girls between 8 and 13 years.

The disorder is characterized by pain in the front of the knee that is associated with inflammation and thickening of the patellar tendon. The pain usually is associated with specific activities, such as kneeling, running, bicycle riding, or stair climbing. There is swelling, tenderness, and increased prominence of the tibial tubercle. The symptoms usually are self-limiting.



■ FIGURE 42-17 ■ Rotational deformity of the proximal tibia, especially when unilateral, suggests tibia vera (Blount's disease). (From Weinstein S.L., Buckwalter J.A. [1994]. *Turek's orthopaedics* [5th ed.]. Philadelphia: J.B. Lippincott)

They may recur during growth periods but usually resolve after closure of the tibial growth plate. In some cases, limitations on activity, tibial bands or braces to immobilize the knee, antiinflammatory agents, and application of cold are necessary to relieve the pain. The objective of treatment is to release tension on the quadriceps to permit revascularization and reossification of the tibial tubercle. Complete resolution of symptoms through healing (physical close) of the tibia tubercle usually requires 12 to 24 months.³⁴ Occasionally, minor symptoms or an increased prominence of the tibial tubercle may continue into adulthood. In some cases, a high-riding patella can cause dislocation with chondromalacia of the patella and result in degenerative arthritis.

Scoliosis

Scoliosis is a lateral deviation of the spinal column that may or may not include rotation or deformity of the vertebrae. It has been estimated that more than 500,000 adults in the United States have scoliosis.⁴³ It is most commonly seen during adolescence and is eight times more common among girls than boys. Scoliosis can develop as the result of another disease condition, or it can occur without known cause. Idiopathic scoliosis accounts for 75% to 80% of cases of the disorder. The other 20% to 25% of cases result from more than 50 different causes, including poliomyelitis, congenital hemivertebrae, neurofibromatosis, and cerebral palsy. Although minor curves are relatively common (affecting approximately 2% of the population), it has been estimated that less than 0.1% of U.S. schoolchildren have severe idiopathic scoliosis.

Types of Scoliosis

Scoliosis is classified as postural or structural. With postural scoliosis, there is a small curve that corrects with bending. It can be corrected with passive and active exercises. Structural scoliosis does not correct with bending. It is a fixed deformity classified according to the cause: congenital, neuromuscular, and idiopathic.

Congenital Scoliosis. Congenital scoliosis is caused by disturbances in vertebral development during the sixth to eighth week of embryologic development. There are structural anomalies in the vertebrae that can cause a severe curvature. The child may have other anomalies and neurologic complications if the spine is involved. Early diagnosis and treatment of progressive curves are essential for children with congenital scoliosis.

Neuromuscular Scoliosis. Neuromuscular scoliosis develops from neuropathic or myopathic diseases. Neuropathic scoliosis is seen with cerebral palsy, myelodysplasia, and poliomyelitis. There is often a long, "C"-shaped curve from the cervical to the sacral region. In children with cerebral palsy, severe deformity may make treatment difficult. Myopathic neuromuscular scoliosis develops with Duchenne's muscular dystrophy and usually is not severe.

Idiopathic Scoliosis. Idiopathic scoliosis is a structural spinal curvature for which no cause has been established. It seems likely that genetics is involved, and mother-daughter pairings are common. Growth and mechanical factors also seem to play a role.

Idiopathic scoliosis can be divided into three groups on the basis of age at onset: infantile (birth to 3 years), juvenile (4 to

10 years), and adolescent (11 years and older).⁴⁴ Adolescent scoliosis is the most common type of idiopathic scoliosis. It accounts for approximately 80% of cases, and is seen most commonly in girls. An increase in joint laxity, which causes excessive joint motion and is found commonly in girls, has been associated with development of idiopathic scoliosis. Delayed puberty and menarche are other risk factors for the development of scoliosis.⁴⁵ Although the curve may be present in any area of the spine, the most common curve is a right thoracic curve, which produces a rib prominence on the convex side and hypokyphosis from rotation of the vertebral column around its long axis as the spine begins to curve.

Scoliosis usually is first noticed because of the deformity it causes. A high shoulder, prominent hip, or projecting scapula may be noticed by a parent or in a school screening program. Idiopathic scoliosis usually is a painless process, although pain may be present in severe cases, usually in the lumbar region. The pain may be caused by pressure on the ribs or on the crest of the ilium. There may be shortness of breath as a result of diminished chest expansion and gastrointestinal disturbances from crowding of the abdominal organs. Adults with less severe deformity may experience mild backache. If scoliosis is left untreated, the curve may progress to an extent that compromises cardiopulmonary function and creates a risk for neurologic complications.

Early diagnosis of scoliosis can be important in the prevention of severe spinal deformity. The cardinal signs of scoliosis are uneven shoulders or iliac crest, prominent scapula on the convex side of the curve, malalignment of spinous processes, asymmetry of the flanks, asymmetry of the thoracic cage, and rib hump or paraspinal muscle prominence when bending forward (Fig. 42-18). A complete physical examination is necessary for



children with scoliosis because the defect may be indicative of other, underlying pathology. Diagnosis of scoliosis is confirmed by radiographs, CT scans, MRI, or myelography.

The treatment of scoliosis depends on the severity of the deformity and the likelihood of progression. A brace may be used to control the progression of the curvature during growth and can provide some correction. Surgical intervention with instrumentation and spinal fusion is done in severe cases. Unlike bracing, which is intended to halt progression of the curvature, surgical intervention is used to decrease the curve.

In summary, skeletal disorders can result from congenital or hereditary influences or from factors that occur during normal periods of skeletal growth and development. Newborn infants undergo normal changes in muscle tone and joint motion, causing torsional conditions of the femur or tibia. Many of these conditions are corrected as skeletal growth and development take place.

Osteogenesis imperfecta is a rare autosomal hereditary disorder characterized by defective synthesis of connective tissue, including bone matrix. It results in poorly developed bones that fracture easily. Developmental dysplasia of the hip includes a range of structural abnormalities. Dislocated hips are always treated to prevent changes in the anatomic structure. Other childhood skeletal disorders, such as the osteochondroses, slipped capital femoral epiphysis, and scoliosis, are not corrected by the growth process. These disorders are progressive, can cause permanent disability, and require treatment. Disorders such as congenital dislocation of the hip and congenital clubfoot are present at birth. Both of these disorders are best treated during infancy.

Scoliosis represents a lateral deviation of the spinal column that may or may not include rotation or deformity of the vertebrae. It can occur as the result of congenital deformities of the vertebrae, neuromuscular diseases that produce weakness of the muscles that support the spinal column, or curvature for which no cause has been established (idiopathic scoliosis). Idiopathic scoliosis, which is the most common form, occurs more frequently in girls than boys. Treatment depends on the severity of the deformity and the likelihood of progression. If left untreated, large curves may progress to the extent that they compromise cardiopulmonary function and create a risk of neurologic complications.

REVIEW QUESTIONS

- Compare muscle strains and ligamentous sprains.
- Describe the healing process of soft tissue injuries.
- Differentiate open fractures from closed ones.

Describe the fracture healing process and relate individual and local factors to the healing process in bone.

Explain the importance of immobilization for fracture healing.

Describe the pathogenesis of osteomyelitis and differentiate among osteomyelitis caused by spread from a contaminated wound, hematogenous osteomyelitis, and osteomyelitis caused by vascular insufficiency in terms of etiologies, manifestations, and treatment.

Define osteonecrosis and cite major causes of osteonecrosis.

Differentiate between the properties of benign and malignant bone tumors.

Describe common torsional deformities that occur in infants and small children, proposed mechanisms of development, diagnostic methods, and treatment.

Define genu varum and genu valgum.

List the problems that occur because of defective tissue synthesis in osteogenesis imperfecta.

Characterize the abnormalities associated with developmental dysplasia of the hip and methods of diagnosis.

Define the term osteochondroses and describe the pathology and symptomatology of Legg-Calvé-Perthes disease and Osgood-Schlatter disease.

Describe the pathology associated with a slipped capital femoral epiphysis and explain why early treatment is important.

Discuss the diagnosis and treatment of idiopathic scoliosis.

connection-J

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- 1. Peters K.D., Kromchamek K., Murphy S.L. (1998). *Deaths: Final data for 1996. National Vital Statistics Report* (Vol. 47). Hyattsville, MD: National Center for Health Statistics.
- Liu S.H., Yang R., Road A., et al. (1995). Collagen in tendon, ligament, and bone healing. *Clinical Orthopedics and Related Research* 318, 265–278.
- Wolfe M.W., Uhl T.L., McCluskey L.C. (2001). Management of ankle strains. American Family Physician 63, 93–104.
- 4. Fongeimie A.E., Buss D.B., Rolnick S.J. (1998). Management of shoulder impingement syndrome and rotator cuff tears. *American Family Physician* Feb
- 5. Speed C.A., Hazleman B.L. (YEAR). Calcific tendonitis of the shoulder. *New England Journal of Medicine* 340, 1582–1584.
- 6. American Academy of Orthopedic Surgeons. (1995). *Play it safe: A guide to safety for young athletes.* Des Plaines, IL: Author.
- Muellner T., Nikolic A., Vecsei V. (1999). Recommendations for the diagnosis of traumatic meniscal injuries in athletes. *Sports Medicine* 27, 337–345.
- Maitra R.S., Miller M.D., Johnson D.L. (1999). Meniscal reconstruction. Part I: Indications, techniques, and graft considerations. *American Journal of Orthopedics* 28 (4), 213–218.
- 9. Ballas M.T., Tytko J., Mannarino F. (1998). Commonly missed orthopedic problems. *American Family Physician* Jan 15
- Rosenberg A. (1998). Bones, joints, and soft tissue tumors. In Cotran R.S., Kumar V., Collins T. (Eds.), *Robbins pathologic basis of disease* (6th ed., pp. 1215–1269). Philadelphia: W.B. Saunders.
- Einhorn T.A. (1998). The cell and molecular biology of fracture healing. Clinical Orthopaedics and Related Research 355 (Suppl.), S7–S21.
- Schiller A.L., Teitelbaum S.L. (1999). Bones and joints. In Rubin E., Farber J.L. (Eds.), *Pathology* (pp. 1359–1363). Philadelphia: Lippincott Williams & Wilkins.

- Hayda R.A., Brighton C.T., Esterhai J.L. (1998). Pathophysiology of delayed healing. *Clinical Orthopaedics and Related Research* 355 (Suppl.), S31–S36.
- Hoover T.J., Siefert J.A. (2000). Soft tissue complications of orthopedic emergencies. *Emergency Medicine Clinics of North America* 18, 115–139.
- Swain R., Ross D. (1999). Lower extremity compartment syndrome. Postgraduate Medicine 105, 159–168.
- Hellman D.B., Stone J.H. (2002). Arthritis and musculoskeletal disorders. In Tierney L.M., McPhee S.J., Papadakis M.A. (Eds.), *Current medical diagnosis and treatment* (41st ed., pp. 848–849). New York: Lange Medical Books/McGraw-Hill.
- Richards R.R. (1997). Fat emboli syndrome. Canadian Journal of Surgery 40, 334–339.
- Simkin P.S., Gardner G.C. (1994). Osteonecrosis: Pathogenesis and practicalities. *Hospital Practice* 29 (3), 73–84.
- Mont M.A., Jones J.C., Einhorn T.A., et al. (1998). Osteonecrosis of the femoral head. *Clinical Orthopaedics and Related Research* 355 (Suppl.), S314–S335.
- Lew D.P., Waldvogel F.A. (1997). Osteomyelitis. New England Journal of Medicine 336, 999–1007.
- Narashimhan N., Marks M. (1996). Osteomyelitis and septic arthritis. In Behrman R.F., Kliegman R.M., Arvin A.M. (Eds.), *Nelson textbook* of pediatrics (15th ed., pp. 724–728). Philadelphia: W.B. Saunders.
- Mader J.T., Landon G.C., Calhoun J. (1993). Antimicrobial treatment of osteomyelitis. *Clinical Orthopaedics and Related Research* 195, 87–95.
- Haas D.W., McAndrew M.P. (1996). Bacterial osteomyelitis in adults: Evolving considerations in diagnosis and treatment. *American Journal* of Medicine 101, 550–561.
- Childs S.G. (1996). Osteoarticular Mycobacterium tuberculosis. Orthopedic Nursing 15 (3), 28–33.
- Silber J.S., Whitfield S.B., Anbari K., et al. (2000). Insidious destruction of the hip by *Mycobacterium tuberculosis* and why early diagnosis is critical. *Journal of Arthroplasty* 15, 392–397.
- Alland D., Kalkut G., Moss A., et al. (1994). Transmission of tuberculosis in New York City: An analysis of DNA fingerprinting and epidemiologic methods. *New England Journal of Medicine* 330, 1710–1716.
- Rosen G., Forscher C.A., Mankin H.J., et al. (2000). Neoplasms of the bone and soft tissue. In Bast R.C., Kufe D.W., Pollock R.E., et al. (Eds.), *Cancer medicine* (pp. 1870–1902). Hamilton, Ontario, Canada: B.C. Decker.
- Arndt C.S. (2000). Neoplasms of bone. In Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.), *Nelson textbook of pediatrics* (16th ed., pp. 1558–1561). Philadelphia: W.B. Saunders.

- Arndt C.A.S., Crist W.M. (1999). Common musculoskeletal tumors of childhood and adolescence. *New England Journal of Medicine* 431, 342–352.
- American Cancer Society. (1999). Bone cancer. [On-line]. Available: http://www.cancer.org.
- O'Keefe R.J., Schwartz E.M., Boyce B.F. (2000). Bone metastasis: An update on bone resorption and therapeutic strategies. *Current Opinion* in Orthopedics 11, 353–359.
- Rubens R.D., Coleman R.E. (1995). Bone metastases. In Abeloff M.D., Armitage J.O., Lichter A.S., et al. (Eds.), *Clinical oncology* (pp. 643–666). New York: Churchill Livingstone.
- Bruce R.W. (1996). Torsional and angular deformities. Pediatric Clinics of North America 43, 867–881.
- Thompson G.H., Scoles P.V. (2000). Bone and joint disorders. In Behrman R.E., Kliegman R.M., Jenson H.B. (Eds.), *Nelson textbook of pediatrics* (16th ed., pp. 2055–2112). Philadelphia: W.B. Saunders.
- Johnson K.B., Oski F.S. (1997). Oski's essential pediatrics (pp. 149–169). Philadelphia: Lippincott-Raven.
- National Institutes of Health. (1999). Osteoporosis and related bone diseases. Fast facts on osteogenesis imperfecta. [On-line]. Available: http://www.oif.org/tier2/fastfact.htm.
- Weinstein S.L. (1994). The pediatric foot. In Weinstein S.L., Buckwalter J.A. (Eds.), *Turek's orthopaedics: Principles and their application* (5th ed., pp. 615–653). Philadelphia: J.B. Lippincott.
- Honein M., Paulozzi L.J., Moore C.A. (2000). Family history, maternal smoking, and clubfoot: An indication of a gene–environment interaction. *American Journal of Epidemiology* 152, 658–665.
- Novacheck T.F. (1996). Developmental dysplasia of the hip. *Pediatrics* 43, 829–848.
- 40. American Academy of Pediatrics. (2000). Clinical practice guideline: Early detection of developmental dysplasia of the hip. *Pediatrics* 105, 896–905.
- 41. Koops S., Quanbeck D. (1996). Three common causes of childhood hip pain. *Pediatric Clinics of North America* 43, 1056–1065.
- 42. Schoppee K. (1995). Blount disease. Orthopedic Nursing 14 (5), 31-34.
- U.S. Preventative Services Task Force. (1993). Screening for adolescent idiopathic scoliosis. *Journal of the American Medical Association* 269, 2664–2672.
- Weinstein S. (1994). The thoracolumbar spine. In Weinstein S.L., Buckwalter J.A. (Eds.), *Turek's orthopaedics: Principles and their application* (5th ed., pp. 447–483). Philadelphia: J.B. Lippincott.
- Omey M.L., Micheli L.J., Gerbino P.G. (2000). Idiopathic scoliosis and spondylolysis in the female athlete: Tips for treatment. *Clinical Orthopaedics and Related Research* 372, 74–84.

CHAPTER

Alterations in the Skeletal System: Metabolic and Rheumatoid Disorders

Metabolic Bone Disease Osteopenia Osteoporosis Pathogenesis Clinical Manifestations Diagnosis and Treatment Osteomalacia and Rickets Osteomalacia **Rickets Paget's Disease Rheumatic Disorders** Systemic Autoimmune Rheumatic Diseases **Rheumatoid Arthritis** Systemic Lupus Erythematosus Systemic Sclerosis Seronegative Spondyloarthropathies **Ankylosing Spondylitis Reactive Arthritis Psoriatic Arthritis Osteoarthritis Syndrome Pathogenesis Clinical Manifestations Diagnosis and Treatment Crystal-Induced Arthropathies** Gout **Rheumatic Diseases in Children and the Elderly Rheumatic Diseases in Children Juvenile Rheumatoid Arthritis** Juvenile Spondyloarthropathies **Rheumatic Diseases in the Elderly** Polymyalgia Rheumatica Pseudogout

he skeletal system provides the basic framework that supports the body, protects its organs, and provides for movement. For example, the bones of the lower extremities act as a pillar when we stand, and the ribs provide a cage that supports and protects our heart and lungs. Joints hold the bones of our skeleton together, making movement possible. This chapter focuses on two types of skeletal disorders: metabolic bone disorders, which disrupt bone integrity, and joint disorders, which disrupt mobility.

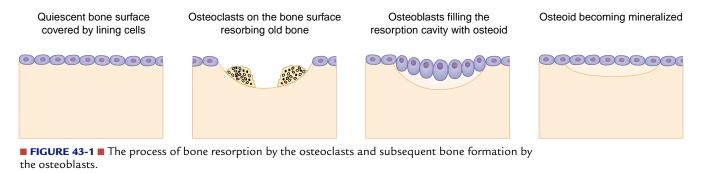
METABOLIC BONE DISEASE

Bone integrity depends on a process of bone resorption and formation, or bone remodeling, which is continuous throughout life. In the adult, approximately 25% of cancellous or spongy bone is replaced each year, compared with 3% of compact bone.¹ In the adult skeleton, bone remodeling proceeds in cycles that involve resorption of old bone by osteoclasts and subsequent formation of new bone by osteoblasts (Fig. 43-1). The sequence of bone resorption and bone formation is activated by many stimuli, including the actions of parathyroid hormone (PTH) and calcitonin. It begins with osteoclastic resorption of existing bone, during which the organic (protein matrix) and the inorganic (mineral) components are removed. The sequence proceeds to the formation of new bone by osteoblasts. In the adult, the length of one sequence (i.e., bone resorption and formation) is approximately 4 months. Ideally, the replaced bone should equal the absorbed bone. If it does not, there is a net loss of bone. In the elderly, for example, bone resorption and formation no longer are perfectly coupled, and bone mass is lost.

There are three major factors that influence bone remodeling: (1) mechanical stress, (2) extracellular calcium and phosphate levels, and (3) hormones, local growth factors, and cytokines. Mechanical stress stimulates osteoblastic activity and formation of the organic matrix. It is important in preventing bone atrophy and in healing fractures. Bone serves as a storage site for extracellular calcium and phosphate ions. Consequently, alterations in the extracellular levels of these

Dr Murtadha Alshareifi e-Library

Chapter 43: Alterations in the Skeletal System: Metabolic and Rheumatoid Disorders



ions affect their deposition in bone (see Chapter 6). Blood levels of calcium and phosphate are regulated by PTH and calcitonin (see Chapter 41). PTH promotes bone resorption, and calcitonin inhibits bone resorption.

Osteoclasts and osteoblasts have their origin in the bone marrow and the periosteum (see Chapter 41).¹ The osteoblasts, which are bone-building cells, originate from osteoprogenitor cells in the periosteum and the stromal or supporting cells within the marrow cavity. The osteoclasts, which are boneresorptive cells, are a member of the hematopoietic monocyte/ macrophage family. However, the development of osteoclasts from their monocyte/macrophage precursors cannot take place unless stromal-osteoblastic cells are present. The differentiation and function of osteoclasts and osteoblasts are regulated by chemical messengers, including colony-stimulating factors (CSF) and other cytokines (see Chapter 8). Interleukin-6, which is produced in response to systemic hormones such as PTH and vitamin D, stimulates the early stages of osteoclast development. Interleukin-6 is thought to be involved in the abnormal bone resorption associated with Paget's disease.

Recent evidence suggests that an interaction between a chemical messenger called the *RANK ligand*, which is produced by the stromal/osteoblastic cells, and RANK receptors on the macrophage/osteoclastic precursor cell is essential for the differentiation and proliferation of osteoclasts (Fig. 43-2). There also is evidence of a blocking molecule called *osteoprotegrin* (OPG) that can prevent the RANK ligand from binding to the RANK receptor, thus inhibiting the formation of osteoclasts. It is now believed that dysregulation of the RANK ligand/receptor pathway plays a prominent role in the pathogenesis of bone diseases such as osteoprorsis.²

815

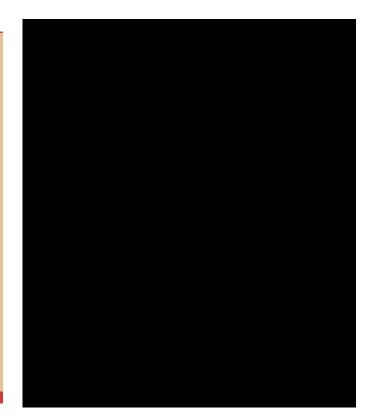
Osteopenia

Osteopenia is a condition that is common to all metabolic bone diseases. It is characterized by a reduction in bone mass greater than expected for age, race, or sex, and it occurs because of a decrease in bone formation, inadequate bone mineralization, or excessive bone deossification. *Osteopenia* is not

KEY CONCEPTS

METABOLIC BONE DISORDERS

- Metabolic bone disorders have their origin in the bone remodeling process that involves an orderly sequence of osteoclastic bone reabsorption, the formation of new bone by the osteoblasts, and mineralization of the newly formed osteoid tissue.
- Osteoporosis represents an increased loss of total bone mass due to an imbalance between bone absorption and bone formation, most often related to the aging process and decreased estrogen levels in postmenopausal women.
- Osteomalacia and rickets represent a softening of bone due to inadequate mineralization of the bone matrix caused by a deficiency of calcium or phosphate.
- Paget's disease is a disorder involving excessive bone destruction and repair, resulting in structural deformities of long bones, spine, pelvis, and cranium.



a diagnosis but a term used to describe an apparent lack of bone seen on x-ray studies. The major causes of osteopenia are osteoporosis, osteomalacia, malignancies such as multiple myeloma, and endocrine disorders such as hyperparathyroidism and hyperthyroidism.

Osteoporosis

Osteoporosis is skeletal disorder characterized by the loss of bone mass and deterioration of the architecture of cancellous bone with a subsequent increase in bone fragility and susceptibility to fractures.³ Osteoporosis can be classified as primary or secondary. Primary osteoporosis occurs in postmenopausal women and in elderly persons of both sexes. Secondary osteoporosis is associated with a definite cause, including a variety of endocrine disorders and genetic abnormalities.

Primary osteoporosis is extremely common in the United States. Ten million Americans already have osteoarthritis, and another 34 million have low bone mass placing them at risk for the disease. Osteoporosis is responsible for more than 1.5 million fractures annually, including 300,000 hip fractures and approximately 700,000 vertebral fractures.³

Pathogenesis

Regardless of cause, osteoporosis reflects enhanced bone resorption relative to bone formation. Although both of these factors play a role in most cases of osteoporosis, their relative contribution to bone loss may vary dependent upon age, sex, nutritional status, and genetic predisposition.

Under normal conditions, bone mass increases steadily during childhood, reaching a peak in the young adult years. The peak bone mass is an important determinant of the subsequent risk of osteoporosis. It is determined in part by genetic factors, gonadal (estrogen) levels, exercise, calcium intake and absorption, and environmental factors. Genetic factors are linked, in largest part, to the maximal amount of bone in a given person, referred to as peak bone mass. Bone mass positively correlates with the amount of skin pigmentation; whites have the least bone mass, and African Americans have the most. Mexican-American women have bone mass intermediate between non-Hispanic white women and African American women. Although osteoporosis is uncommon among African-American women, many cases are seen among postmenopausal women with brown and yellow skin.⁴ Exercise may prevent osteoporosis by increasing peak bone mineral density during periods of growth. Poor nutrition or an age-related decrease in intestinal absorption of calcium because of deficient activation of vitamin D may contribute to the development of osteoporosis, particularly in the elderly.

Hormonal factors play a significant role in the development of osteoporosis, particularly in postmenopausal women. Postmenopausal osteoporosis, which is caused by an estrogen deficiency, is manifested by a loss of cancellous bone and a predisposition to fractures of the vertebrae and distal radius. The loss of bone mass is greatest during early menopause, when estrogen levels are withdrawing. Several factors appear to influence the increased loss of bone mass associated with an estrogen deficiency. Decreased estrogen levels are associated with an increase in cytokines (*e.g.*, interleukins-1, interleukin-6, and tumor necrosis factor [TNF]) that stimulate the production of osteoclast precursors. Recent studies indicate that estrogen deficiency also influences osteoclast differentiation via the RANK receptor pathways.² Estrogen stimulates the production of OPG and thus inhibits the formation of osteoclasts, and it also blunts the responsiveness of osteoclast precursors to the RANK ligand. With menopause and its accompanying estrogen deficiency, this inhibition of osteoclast production is lost.² Evidence also suggests that estrogen deficiency, as well as normal aging, may lead to decreased osteoblastic activity and new bone formation. Thus, the bone loss associated with estrogen deficiency may be caused by a combination of increased bone resorption and decreased bone formation. Testosterone deficiency may contribute to bone loss in men with senile osteoporosis, although the effect is not of the same magnitude as that caused by estrogen deficiency.

Age-related changes in bone density occur in all individuals and contribute to the development of osteoporosis in both sexes. After maximal bone mass is attained at about 30 years of age, the rate of bone loss for both sexes is approximately 0.7% per year, and it increases to approximately 1% per year or more in menopausal women.² The age-related loss of bone reflects decreased osteoblast activity as well as an increase in osteoclastic activity. The greatest losses occur in areas containing abundant cancellous bone, such as the spine and femoral neck. Thus, these are common sites for fractures in persons with osteoporosis.

Secondary osteoporosis is associated with many conditions, including endocrine disorders, malabsorption disorders, malignancies, alcoholism, and certain medications. Persons with endocrine disorders such as hyperthyroidism, hyperparathyroidism, Cushing's syndrome, or diabetes mellitus are at high risk for the development of osteoporosis. Hyperthyroidism causes an acceleration of bone turnover. Some malignancies (e.g., multiple myeloma) secrete osteoclast-activating factor, causing significant bone loss. Alcohol is a direct inhibitor of osteoblasts and may also inhibit calcium absorption. Glucocorticoid use is the common form of drug-related osteoporosis, and its long-term use in treatment of disorders such as rheumatoid arthritis and chronic obstructive lung disease is associated with a high rate of fractures.³ The prolonged use of medications that increase calcium excretion, such as aluminum-containing antacids, corticosteroids, and anticonvulsants, also is associated with bone loss.⁵ Persons with human immunodeficiency virus (HIV) infection or acquired immunodeficiency syndrome (AIDS) who are being treated with antiretroviral therapy may have a lower bone density and signs of osteoporosis and osteopenia.6

Several groups of children and adolescents are at increased risk of decreased bone mass, including premature and lowbirth-weight infants who have lower than expected bone mass in the early weeks of life; children who require treatment with corticosteroid drugs (*e.g.*, those with childhood inflammatory diseases and transplant recipients); children with cystic fibrosis; and those with hypogonadal states (*e.g.*, anorexia nervosa and the female athlete triad).² Children with cystic fibrosis often have impaired gastrointestinal function that reduces the absorption of calcium and other nutrients, and many also require the frequent use of corticosteroid drugs.

Premature osteoporosis is being seen increasingly in female athletes because of an increased prevalence of eating disorders and amenorrhea.⁷ The *female athlete triad* refers to a pattern of disordered eating that leads to amenorrhea and eventually osteoporosis. Poor nutrition, combined with intense training, can lead to an energy deficit that causes a lack of estrogen production by the ovary and secondary amenorrhea.⁸ The lack of estrogen combined with the lack of calcium and vitamin D from dietary deficiencies results in a loss of bone density and increased risk of fractures.⁸ There is a concern that athletes with low bone mineral density will be at increased risk for fractures during their competitive years. It is unclear if osteoporosis induced by amenorrhea is reversible. It most frequently affects women engaged in endurance sports, such as running and swimming; in activities where appearance is important, such as figure skating, diving, and gymnastics; or sports with weight categories, such as horse racing, martial arts, and rowing.⁹

Clinical Manifestations

Osteoporotic changes occur in the diaphysis and metaphysis of bone. The diameter of the bone enlarges with age, causing the outer supporting cortex to become thinner. In severe osteoporosis, the bones begin to resemble the fragile structure of a fine porcelain vase. There is loss of trabeculae from cancellous bone and thinning of the cortex to such an extent that minimal stress causes fractures (Fig. 43-3).

The first clinical manifestations of osteoporosis are pain accompanied by skeletal fractures—a vertebral compression fracture or fractures of the hip, pelvis, humerus, or any other bone. Fractures usually represent an end stage of the disease. Fracture occurs with a force less than typically is needed. Women who present with fractures are much more likely to sustain another fracture than are women of the same age without osteoporosis. Wedging and collapse of vertebrae causes a loss of height in the vertebral column and kyphosis, a condition commonly referred to as *dowager's hump*. Usually, there is no generalized



■ FIGURE 43-3 ■ Osteoporosis. A section of the vertebral column, in which the bone marrow has been washed out, demonstrates a loss of bone tissue and a compression fracture of a vertebral body (*top*). (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 1367]. Philadelphia: Lippincott Williams & Wilkins)

bone tenderness. When pain occurs, it is related to fractures. Systemic symptoms such as weakness and weight loss suggest that the osteoporosis may be caused by underlying disease.

Diagnosis and Treatment

An important advance in diagnostic methods used for the identification of osteoporosis has been the use of bone mass density (BMD) studies. Several different techniques have been developed to measure BMD in multiple skeletal sites, including the vertebral column, hip, and spine. Excessive loss of height indicates some probability of low bone mass.¹⁰ Measurement of serial heights in older adults is another simple way to screen for osteoporosis. A further advance in the diagnosis of osteoporosis is the refinement of risk factors, permitting better analysis of risk pertaining to particular persons.

Prevention and early detection of osteoporosis are essential to the prevention of the associated deformities and fractures. It is important to identify persons in high-risk groups so treatment can begin early. Postmenopausal women of small stature or lean body mass, those with sedentary lifestyles, those with poor calcium intake, and those with diseases that demineralize bone are at greatest risk. Other risk factors include an age of 80 years or greater, maternal history of hip fracture, consumption of caffeine-containing beverages, previous hyperthyroidism, and current anticonvulsant therapy. Excessive intake of diet soda that is high in phosphate also can deplete calcium stores. Other risk factors found to be associated with osteoporosis are a diet high in protein, cigarette smoking, and alcohol ingestion. Risk factors for osteoporosis are listed in Chart 43-1.

Regular exercise and adequate calcium intake are important factors in preventing osteoporosis. Weight-bearing exercises

CHART 43-1 Risk Factors Associated With Osteoporosis

Primary

Advanced Age Female White (fair, thin skin) Small bone structure Postmenopausal Family History Contributing factors Sedentary lifestyle Calcium deficiency (long-term)

Secondary

Cushing's disease Diabetes Hyperparathyroidism Hyperthyroidism Malignancy Malabsorption disorders Chronic alcoholism Medications Anticonvulsants Aluminum-containing antacids Corticosteroids Heparin such as walking, jogging, rowing, and weight lifting are important in the maintenance of bone mass. Studies have indicated that premenopausal women need more than 1000 mg and postmenopausal women need 1500 mg of calcium daily.¹¹ This means that adults should drink three to four glasses of milk daily or substitute other foods that are high in calcium. Because most older American women do not consume a sufficient quantity of dairy products to meet their calcium needs, calcium supplementation is recommended. Deficient activation of vitamin D may be an important factor in the impaired intestinal absorption of calcium in the elderly. A daily intake of 400 to 800 IU of vitamin D is recommended because vitamin D optimizes calcium absorption and inhibits parathyroid secretion, which stimulates calcium resorption from bone.¹⁰

Active treatment of osteoporosis uses four types of agents: gonadal hormones (estrogen), calcitonin, and bisphosphonates. The development of selective estrogen receptor modulators (SERMs) has been an important advance in osteoporosis research. These agents maximize the beneficial effects of estrogen on bone, while minimizing the effects on breast and endometrial tissues.² Calcitonin can be used to decrease osteoclastic activity and has some effect on bone pain. The bisphosphonates are analogs of endogenous inorganic pyrophosphate that the body cannot break down. In bone, they bind to hydroxyapatite and prevent bone resorption through the inhibition of osteoclast activity.

Osteomalacia and Rickets

In contrast to osteoporosis, which causes a loss of total bone mass and results in brittle bones, osteomalacia and rickets produce a softening of the bones and do not involve the loss of bone matrix.

Osteomalacia

Osteomalacia is a generalized bone condition in which inadequate mineralization of bone results from a calcium or phosphate deficiency, or both. It is sometimes referred to as the adult form of rickets.

There are two main causes of osteomalacia: (1) insufficient calcium absorption from the intestine because of a lack of dietary calcium or deficiency or resistance to the action of vitamin D and (2) phosphate deficiency caused by increased renal losses or decreased intestinal absorption. Vitamin D deficiency is caused most commonly by reduced vitamin D absorption as a result of biliary tract or intestinal diseases that impair fat and fat-soluble vitamin absorption. Lack of vitamin D in the diet is rare in the United States because many foods are fortified with the vitamin. Anticonvulsant medications, such as phenobarbital and phenytoin, induce hepatic hydroxylases that accelerate breakdown of the active forms of vitamin D.

A form of osteomalacia called *renal rickets* occurs in persons with chronic renal failure. It is caused by the inability of the kidney to activate vitamin D and excrete phosphate and is accompanied by hyperparathyroidism, increased bone turnover, and increased bone resorption (see Chapter 24). Another form of osteomalacia results from renal tubular defects that cause excessive phosphate losses. This form of osteomalacia is commonly referred to as *vitamin D-resistant rickets* and often is a familial disorder.¹² It is inherited as an X-linked dominant gene passed by mothers to one half of their children and by fathers to their daughters only. This form of osteomalacia affects boys more severely than girls. Long-standing primary hyperparathyroidism causes increased calcium resorption from bone and hypophosphatemia, which can lead to rickets in children and osteomalacia in adults.

The incidence of osteomalacia is high among the elderly because of diets deficient in calcium and vitamin D and often is compounded by the intestinal malabsorption problems that accompany aging. Osteomalacia often is seen in cultures in which the diet is deficient in vitamin D, such as in northern China, Japan, and northern India. Women in these areas have a higher incidence of the disorder than do men because of the combined effects of pregnancy, lactation, and more indoor confinement. Osteomalacia occasionally is seen in strict vegetarians, persons who have had a gastrectomy, and those on long-term anticonvulsant, tranquilizer, sedative, muscle relaxant, or diuretic drugs. There also is a greater incidence of osteomalacia in the colder regions of the world, particularly during the winter months, probably because of lessened exposure to sunlight.

The clinical manifestations of osteomalacia are bone pain, tenderness, and fractures as the disease progresses. In severe cases, muscle weakness often is an early sign. The cause of muscle weakness is unclear. The combined effects of gravity, muscle weakness, and bone softening contribute to the development of deformities. There may be a dorsal kyphosis in the spine, rib deformities, a heart-shaped pelvis, and marked bowing of the tibiae and femurs. Osteomalacia predisposes a person to pathologic fractures in the weakened areas, especially in the distal radius and proximal femur. In contrast to osteoporosis, it is not a significant cause of hip fractures. There may be delayed healing and poor retention of internal fixation devices. Osteomalacia usually is accompanied by a compensatory or secondary hyperparathyroidism stimulated by low serum calcium levels.

Diagnostic methods include x-ray studies and laboratory tests such as serum calcium, phosphate, PTH, and vitamin D levels. Bone density studies help to confirm the diagnosis. A bone biopsy may be done to confirm the diagnosis of osteomalacia.

The treatment of osteomalacia is directed at the underlying cause. If the problem is nutritional, restoring adequate amounts of calcium and vitamin D to the diet may be sufficient. Vitamin D is specific for adult osteomalacia and vitamin Dresistant rickets, but large doses usually are needed to overcome the resistance to its calcium-absorption action and to prevent renal loss of phosphate. If osteomalacia is caused by malabsorption, the treatment is directed toward correcting the primary disease. For example, adequate replacement of pancreatic enzymes is of paramount importance in pancreatic insufficiency. In renal tubular disorders, the treatment is directed at the altered renal physiology.

Rickets

Rickets is a disorder of vitamin D deficiency, inadequate calcium absorption, and impaired mineralization of bone in children. Children with rickets manifest inadequate mineralization not only of bone, but also of the cartilaginous matrix of the epiphyseal growth plate. Rickets occurs primarily in underdeveloped areas of the world and among immigrants to developed countries. The causes are inadequate exposure to sunlight (e.g., children are often kept clothed and indoors) and prolonged breast-feeding without vitamin D supplementation.¹³ Although the vitamin D content of human milk is low, the combination of breast milk and sunlight exposure usually provides sufficient vitamin D. Another cause of rickets is the use of commercial alternative milks (e.g., soy or rice beverages) that are not fortified with vitamin D.14 A dietary deficiency in calcium and phosphorous may also contribute to the development of rickets. A newly discovered genetic mutation also can cause vitamin D deficiency rickets, a condition that does not respond to simple vitamin supplementation. The mutation results in the absence of a critical enzyme in vitamin D metabolism.15

The pathology of rickets is the same as that of osteomalacia seen in adults. Because rickets affects children during periods of active growth, the structural changes seen in the bone are somewhat different. Bones become deformed; ossification at epiphyseal plates is delayed and disordered, resulting in widening of the epiphyseal cartilage plate. Any new bone that does grow is unmineralized.

The symptoms of rickets usually are noticed between 6 months and 3 years of age. The child usually has stunted growth, with a height sometimes far below the normal range. Weight often is not affected so that the children, many of whom present with a protruding abdomen (i.e., rachitic potbelly), have been described as presenting a Buddha-like appearance when sitting. Early symptoms are lethargy and muscle weakness, which may be accompanied by convulsions or tetany related to hypocalcemia. Irritability is common. In severe cases, children lose their skin pigment, acquire flabby subcutaneous tissue, and have poorly developed musculature. The ends of long bones and ribs are enlarged. The thorax may be abnormally shaped, with prominent rib cartilage (*i.e.*, rachitic rosary). The legs exhibit bowlegged or knock-kneed deformities. The skull is enlarged and soft, and closure of the fontanels is delayed. Teeth are slow to develop, and the child may have difficulty standing.

Rickets is treated with a balanced diet sufficient in calcium, phosphorus, and vitamin D. Exposure to sunlight also is important, especially for premature infants and those receiving artificial milk feedings. Supplemental vitamin D in excess of normal requirements is given for several months. Maintenance of good posture, positioning, and bracing in older children are used to prevent deformities. After the disease is controlled, deformities may have to be surgically corrected as the child grows.

Paget's Disease

Paget's disease (*i.e.*, osteitis deformans) is not a true metabolic disease. It is a progressive skeletal disorder that involves excessive bone destruction and repair and is characterized by increasing structural changes of the long bones, spine, pelvis, and cranium. The disease usually begins during midadulthood and becomes progressively more common thereafter.²

The cause of Paget's disease is unknown. It may be caused by a virus capable of inciting osteoclastic activity.^{2,12} It has been suggested that the virus may induce secretion of interferon-6, which is a potent stimulator of osteoclastic recruitment and resorptive activity.² The disease usually begins insidiously and progresses slowly over many years. An initial osteolytic phase is followed by an osteoblastic sclerotic phase. During the initial osteolytic phase, abnormal osteoclasts proliferate. Bone resorption occurs so rapidly that new bone formation cannot keep up, and the bone is replaced by fibrous tissue. The two processes of destruction and rebuilding occur simultaneously. The bones increase in size and thickness because of accelerated bone resorption followed by abnormal regeneration. Irregular bone formation results in sclerotic and osteoblastic lesions. The result is a thick layer of coarse, thick bundles of trabecular bone with a rough and pitted outer surface that has the appearance of pumice (Fig. 43-4).

The clinical manifestations of Paget's disease depend on the specific area involved. Approximately 20% of persons with the disorder are totally asymptomatic, and the disease is discovered incidentally.¹⁶ Involvement of the skull causes headaches, intermittent tinnitus, vertigo, and eventual hearing loss. In the spine, collapse of the anterior vertebrae causes kyphosis of the thoracic spine. The femur and tibia become bowed (Fig. 43-4). Softening of the femoral neck can cause coxa vara (*i.e.*, reduced angle of the femoral neck). Coxa vara, in combination with softening of the sacral and iliac bones, causes a waddling gait. When the lesion affects only one bone, it may cause only mild pain and stiffness. Progressive deossification weakens and distorts the bone structure. The deossification process begins along the inner cortical surfaces and continues until the substance of the bone disappears. Pathologic fractures may occur, especially in the bones subjected to the greatest stress (e.g., upper femur, lower spine, pelvic bones). These fractures often heal poorly, with excessive and poorly distributed callus.

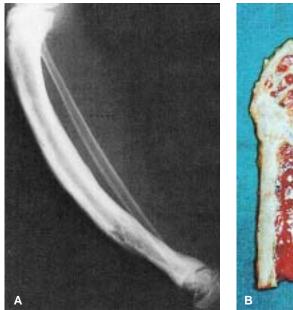
Other manifestations of Paget's disease include nerve palsy syndromes from lesions in the upper extremities, mental deterioration, and cardiovascular disease. Cardiovascular disease is the most serious complication and is listed as the most common cause of death in those with advanced generalized Paget's disease. It is caused by vasodilation of the vessels in the skin and subcutaneous tissues overlying the affected bones. When one third to one half of the skeleton is affected, the increased blood flow may lead to high-output cardiac failure. Ventilatory capacity may be limited by rib and spine involvement.

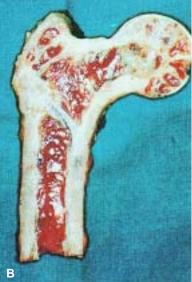
Osteogenic sarcomas occur in 5% to 10% of persons with severe disease.² One fifth of all osteogenic sarcomas in persons 50 years of age or older originate in people with Paget's disease.¹⁷ The bones most often affected, in order of frequency, are the femur, pelvis, humerus, and tibia.

Diagnosis of Paget's disease is based on characteristic bone deformities and x-ray changes. Elevated levels of serum alkaline phosphatase and urinary hydroxyproline support the diagnosis, and continued surveillance of these levels may be used to monitor the effectiveness of treatment. Technetium pyrophosphate bone scans are used to detect the rapid bone turnover indicative of active disease and to monitor the response to treatment. The scan cannot identify bone activity resulting from malignant lesions. Bone biopsy may be done to differentiate the lesion from osteomyelitis or a primary or metastatic bone tumor.

The treatment of Paget's disease is based on the degree of pain and the extent of the disease. Pain can be reduced with nonsteroidal or other anti-inflammatory agents. Suppressive agents such as the hormone calcitonin, mithramycin, and

820 Unit Eleven: Alterations in the Skeletal and Integumentary Systems





■ FIGURE 43-4 ■ Paget disease. (A) Radiograph of the leg shows marked involvement of the tibia by Paget disease with thickening and disorganization of the cortex. Note the normal appearance of the fibula. (B) The proximal end of a femur affected by Paget disease shows replacement of the normal cancellous architecture by coarse, thick bundles of trabecular bone. The cortical bone is irregularly thickened and exhibits a coarse, granular appearance instead of the normally smooth cortical bone. Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 1377]. Philadelphia: Lippincott Williams & Wilkins)

bisphosphonates are used to manage pain and prevent further spread of the disease and neurologic defects. Calcitonin, administered by nasal spray, inhibits osteoclast-mediated bone resorption. Bisphosphonates are the treatment of choice for Paget's disease. They act by binding directly to bone minerals, inhibiting bone loss by rapidly decreasing bone resorption, followed by a secondary slower decrease in the rate of bone formation.

In summary, in addition to its structural function, the skeleton is a homeostatic organ. Metabolic bone diseases such as osteoporosis, osteomalacia, rickets, and Paget's disease are the result of a disruption in the equilibrium of bone formation and resorption. Osteoporosis, which is the most common of the metabolic bone diseases, occurs when the rate of bone resorption is greater than that of bone formation. It is seen frequently in postmenopausal women and is the major cause of fractures in persons older than 45 years of age. Osteomalacia and rickets are caused by inadequate mineralization of bone matrix, primarily because of a deficiency of vitamin D. Paget's disease results from excessive osteoclastic activity and is characterized by the formation of poor-quality bone.

RHEUMATIC DISORDERS

Rheumatic disorders are characterized by inflammation, pain, and stiffness in the musculoskeletal system. *Arthritis* is a descriptive term applied to more than 100 rheumatic diseases, ranging from localized, self-limiting conditions to those that are systemic, autoimmune processes.^{18,19} Arthritis affects persons in all age groups and is the leading cause of disability in the United States.

The common use of the term *arthritis* oversimplifies the nature of the varied disease processes, the difficulty in differentiating one form of arthritis from another, and the complexity of treatment of these usually chronic conditions. These diverse rheumatic conditions share inflammation of the joint as a prominent or accompanying symptom. In the systemic rheumatic diseases, such as rheumatoid arthritis, the inflammation is primary, resulting from an immune response, probably autoimmune in origin. In rheumatic conditions limited to a single or few diarthrodial joints, such as osteoarthritis, the inflammation is secondary, resulting from the degenerative process and joint irregularities.

Systemic Autoimmune Rheumatic Diseases

Systemic autoimmune rheumatic diseases are a group of chronic disorders characterized by diffuse inflammatory vascular lesions and degenerative changes in connective tissue that share clinical features and may affect many of the same organs. They

KEY CONCEPTS

ARTHRITIS

- Arthritis represents a diverse group of rheumatic conditions that share inflammation of the joint as a prominent or accompanying symptom.
- In the systemic rheumatic diseases, the inflammation is primary, resulting from an immune response, probably autoimmune in origin.
- In rheumatic conditions, such as osteoarthritis, which are limited to a single or few diarthrodial joints, the inflammation is secondary, resulting from the degenerative process and joint irregularities.

include rheumatoid arthritis, systemic lupus erythematosus, and systemic sclerosis, all of which share an autoimmune systemic pathogenesis.

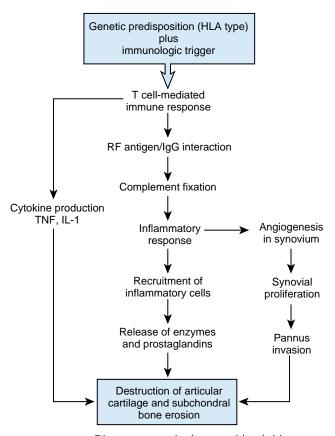
Rheumatoid Arthritis

Rheumatoid arthritis (RA) is a systemic inflammatory disease that affects 0.3% to 1.5% of the population, with women affected two to three times more frequently than men.¹⁸ Although the disease occurs in all age groups, its prevalence increases with age. The peak incidence among women is between the ages of 40 and 60 years, with the onset at 30 to 50 years of age.

Although the cause of RA remains uncertain, evidence points to a genetic predisposition and the development of joint inflammation that is immunologically mediated. It has been suggested that the disease is initiated in a genetically predisposed individual by the activation of a T cell-mediated response to an immunologic trigger, such as a microbial agent. The importance of genetic factors in the pathogenesis of RA is supported by the increased frequency of the disease among first-degree relatives and monozygotic twins. There is also a strong association of human leukocyte antigen (HLA) DR4 and/or HLA-DRB1 with RA.¹⁸ Thus, certain HLA-DR molecules may predispose to RA by their capacity to bind arthrogenic antigens, which in turn activate helper T cells and initiate the disease.

Pathogenesis. The pathogenesis of RA can be viewed as an aberrant immune response that leads to synovial inflammation and destruction of the joint architecture. Approximately 85% of those with the disease have a substance called the *rheumatoid factor* (RF), which is an autoantibody that reacts with a fragment of immunoglobulin G (IgG).^{12,20} The RF and IgG form immune complexes that fix complement; attract neutrophils, macrophages, and lymphocytes; and incite the inflammatory response (Fig. 43-5). The presence of high titers of RF is frequently associated with severe and unremitting disease, many systemic complications, and a serious prognosis.¹² However, it is important to note that about 15% persons with RA do not have the RF, and the factor is occasionally found in other disease states (and even healthy people).

Characteristic of RA is the development of an extensive network of new blood vessels in the synovial membrane that contributes to the advancement of the rheumatoid synovitis. This destructive vascular granulation tissue, which is called *pannus*, extends from the synovium to involve the "bare area," a region of unprotected bone at the junction between cartilage and subchondral bone. Pannus is a feature of RA that differentiates it from other forms of inflammatory arthritis²¹ (Fig. 43-6). The inflammatory cells found in the pannus have a destructive effect on the adjacent cartilage and bone. Eventually, pannus develops between the joint margins, leading to reduced joint motion and the possibility of eventual ankylosis. With progression of the disease, joint inflammation and the resulting structural changes can lead to joint instability, muscle atrophy from disuse, stretching of the ligaments, and involvement of the tendons and muscles. The effect of the pathologic changes on joint structure and function is related to the degree of disease activity, which can change at any time. Unfortunately, the destructive changes are irreversible.



■ FIGURE 43-5 ■ Disease process in rheumatoid arthritis.

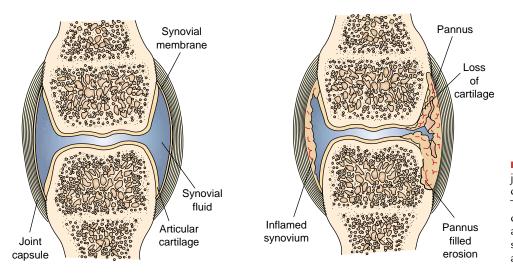
Clinical Manifestations. Rheumatoid arthritis often is associated with systemic as well as joint manifestations. It usually has an insidious onset marked by systemic manifestations such as fatigue, anorexia, weight loss, and low-grade fever when the disease is active. The erythrocyte sedimentation rate (ESR), which commonly is elevated during inflammatory processes, has been found to correlate with the amount of disease activity.²² Anemia associated with a low serum iron level or low iron-binding capacity is common.¹⁸ This anemia usually is resistant to iron therapy.

The disease, which is characterized by exacerbations and remissions, may involve only a few joints for brief durations, or it may be relentlessly progressive and debilitating. Approximately 3% of those with the disease have a progressive, unremitting form that does not respond to aggressive therapy.¹⁸

Joint Manifestations. Joint involvement usually is symmetric and polyarticular. Any diarthrodial joint can be involved. The person may report joint pain and stiffness that lasts 30 minutes and frequently for several hours. The limitation of joint motion that occurs early in the disease usually is caused by pain; later, it is caused by fibrosis. The most frequently affected joints initially are the fingers, hands, wrists, knees, and feet. Later, other diarthrodial joints may become involved. Spinal involvement usually is limited to the cervical region.

In the hands, there usually is bilateral and symmetric involvement of the proximal interphalangeal (PIP) and metacarpophalangeal (MCP) joints in the early stages of RA; the distal interphalangeal (DIP) joints rarely are affected. The fingers

822 Unit Eleven: Alterations in the Skeletal and Integumentary Systems



■ FIGURE 43-6 ■ (Left) Normal joint structures. (Right) Joint changes in rheumatoid arthritis. The left side denotes early changes occurring within the synovium, and the right side shows progressive disease that leads to erosion and the formation of pannus.

often take on a spindle-shaped appearance because of inflammation of the PIP joints (Fig. 43-7). Progressive joint destruction may lead to subluxation (i.e., dislocation of the joint resulting in misalignment of the bone ends) and instability of the joint and in limitation of movement. Swelling and thickening of the synovium can result in stretching of the joint capsule and ligaments. When this occurs, muscle and tendon imbalances develop, and mechanical forces applied to the joints through daily activities produce joint deformities. In the MCP joints, the extensor tendons can slip to the ulnar side of the metacarpal head, causing ulnar deviation of the finger (Fig. 43-8). Subluxation of the MCP joints may develop when this deformity is present. Hyperextension of the PIP joint and partial flexion of the DIP joint is called a swan neck deformity. After this condition becomes fixed, severe loss of function occurs because the person can no longer make a fist. Flexion of the PIP joint with hyperextension of the DIP joint is called a *boutonnière deformity*.

The knee is one of the most commonly affected joints and is responsible for much of the disability associated with the disease.¹⁸ Active synovitis may be apparent as visible swelling that obliterates the normal contour over the medial and lateral aspects of the patella. The *bulge sign*, which involves milking fluid from the lateral to the medial side of the patella, may be used to determine the presence of excess fluid when it is not visible. Joint contractures, instability, and genu valgus (knock-knee) deformity are other possible manifestations. Severe quadriceps atrophy can contribute to the disability. *Baker's cyst* may occur in the popliteal area behind the knee. This is caused by enlargement of the bursa and usually does not cause symptoms unless the cyst ruptures, in which case symptoms mimicking thrombophlebitis appear.

Because the lower extremity joints are weight-bearing structures, involvement of the foot and ankle causes greater dysfunction and pain than does involvement of the upper extrem-



ities. Disease activity can limit flexion and extension of the ankle, which can create difficulty in walking. Involvement of the metatarsophalangeal joints of the feet can cause subluxation, hallux valgus, and hammer toe deformities.

Neck discomfort is common. In rare cases, long-standing disease can lead to neurologic complications such as occipital headaches, muscle weakness, and numbness and tingling in the upper extremities. More severe but less common neurologic complications are dislocation of the first cervical vertebra and subluxation of the odontoid process of the second vertebra into the foramen magnum, which can lead to paralysis and is potentially fatal.

Extra-articular Manifestations. Although characteristically a joint disease, RA can affect a number of other tissues. Extra-articular manifestations probably occur with a fair degree of frequency but usually are mild enough to cause few problems. They are most likely to occur in persons with the RF.

Rheumatoid nodules are granulomatous lesions that develop around small blood vessels. The nodules may be tender or nontender, movable or immovable, and small or large. Typically, they are found over pressure points such as the extensor surfaces of the ulna. The nodules may remain unless surgically removed, or they may resolve spontaneously.

Vasculitis, involving the small and medium-size arterioles, is an uncommon manifestation of RA in persons with a long history of active arthritis and high titers of RF (see Chapter 15). Manifestations include ischemic areas in the nail fold and digital pulp that appear as brown spots. Ulcerations may occur in the lower extremities, particularly around the malleolar areas. In some cases, neuropathy may be the only symptom of vasculitis. The visceral organs, such as the heart, lungs, and gastrointestinal tract, also may be affected.

Other extra-articular manifestations include eye lesions such as episcleritis and scleritis, hematologic abnormalities, pulmonary disease, cardiac complications, infection, and Felty's syndrome (*i.e.*, leukopenia with or without splenomegaly).

Diagnosis and Treatment. The diagnosis of RA is based on findings of the history, physical examination, and laboratory tests. The criteria for RA developed by the American Rheumatism Association are useful in establishing the diagnosis²³ (Chart 43-2). At least four of the criteria must be present to make a diagnosis of RA. Although these criteria were developed for classification purposes and for use in epidemiological studies, they can be used as guidelines for diagnosing the illness in individual patients.

In the early stages, the disease often is difficult to diagnose. On physical examination, the affected joints show signs of inflammation, swelling, tenderness, and possibly warmth and reduced motion. The joints have a soft, spongy feeling because of the synovial thickening and inflammation. Body movements may be guarded to prevent pain. Changes in joint structure usually are not visible early in the disease.

The RF test results are not diagnostic for RA, but they can be of value in differentiating RA from other forms of arthritis. Radiologic findings also are not diagnostic in RA because joint erosions often are not seen on radiographic images in the early stages of the disorder. Synovial fluid analysis can be helpful in the diagnostic process. The synovial fluid has a cloudy appear-



ance, the white blood cell count is elevated as a result of inflammation, and the complement components are decreased.

The treatment plan for a person with RA includes education about the disease and its treatment, rest, therapeutic exercises, and medications. Because of the chronicity of the disease and the need for continuous, long-term adherence to the prescribed treatment modalities, it is important that the treatment be integrated into the person's lifestyle.

The goals of pharmacologic therapy for RA are to reduce pain, decrease inflammation, maintain or restore joint function, and prevent bone and cartilage destruction. Nonsteroidal anti-inflammatory drugs (NSAIDs) usually are the first choice in the treatment of RA. The NSAIDs, including the salicylates (e.g., aspirin), and the newer COX-2 inhibitors, provide antiinflammatory and analgesic effects. Second-line drug therapy is initiated early in the disease if joint symptoms persist despite use of NSAIDs. Disease-modifying antirheumatic drugs (DMARDs) include gold salts, hydroxychloroquine, sulfasalazine, methotrexate, and azathioprine. Methotrexate, a potent immunosuppressive drug, has become the drug of choice because of its potency. It is also relatively fast acting (i.e., improvement is seen in 1 month) compared with the slower-acting DMARDs, which can take 3 to 4 months to work. Corticosteroid drugs may be used to reduce discomfort.²⁴ To avoid long-term side effects, they are used only in specific situations for short-term therapy at a low dose level. This medication does not modify the disease and is unable to prevent joint destruction. Intra-articular corticosteroid injections can provide rapid relief of acute or subacute inflammatory synovitis in a few joints.

Newer antirheumatic drugs include leflunomide, etanercept, and infliximab. Leflunomide is a pyrimidine synthesis inhibitor that blocks the expansion of T cells.²⁵ Infliximab (Remicade) and etanercept (Enbrel) are biologic responsemodifying agents that block tumor necrosis factor- α (TNF- α), one of the key proinflammatory cytokines in RA.

Surgery also may be a part of the treatment of RA. Synovectomy may be indicated to reduce pain and joint damage when synovitis does not respond to medical treatment. Total joint replacements (*i.e.*, arthroplasty) may be performed to reduce pain and increase motion. Arthrodesis (*i.e.*, joint fusion) is indicated only in extreme cases when there is so much soft tissue damage and scarring or infection that a replacement is impossible.

Systemic Lupus Erythematosus

Systemic lupus erythematosus (SLE) is a chronic autoimmune disease that can affect virtually any organ system, including the skin, joints, kidneys, serosal membranes, and heart. SLE is a major rheumatic disease, and more than 90% of persons with the disease have polyarthralgias. SLE is a fairly common disease with a prevalence of approximately 1 case per 2000 persons in certain populations.²⁶ The peak incidence occurs between ages 15 to 40 years. There is a female-to-male ratio of 9 to 1, with the ratio becoming closer to 30 to 1 during the childbearing years. SLE is more common in African Americans, Hispanics, and Asians than whites, and the incidence in some families is higher than in others.¹⁸

The cause of SLE is unknown. The presence of a wide array of autoantibodies suggests a breakdown in the normal surveillance function of the immune system (see Chapter 10). Antibodies have been identified against a host of nuclear and cytoplasmic components of the cell that are neither tissue nor organ specific. Antinuclear antibodies (ANAs) are directed against several nuclear antigens, including DNA, histones, nonhistone proteins bound to RNA, and nucleolar antigens.²⁶ Another group of antibodies is directed against cell surface antigens of blood elements.

The development of autoantibodies is thought to result from a combination of factors, including genetic, hormonal, immunologic, and environmental factors. Genetic predisposition is evidenced by the occurrence of familial cases of SLE, especially among identical twins. In North American white populations there is a positive association between SLE and class II HLA genes, particularly at the HLA-DQ locus (see Chapter 9).²⁷ Studies also suggest that an imbalance in sex hormone levels may play a role in the development of the disease, especially because the disease is so prevalent among women. Androgens appear to protect against and estrogens seem to favor the development of SLE.¹⁸ Possible environmental triggers include ultraviolet (UV) light, certain drugs, and possibly infectious agents. UV light, specifically UVB associated with exposure to the sun or unshielded fluorescent bulbs, may trigger exacerbations. Photosensitivity occurs in approximately one third of patients with SLE.

Pathogenesis. The pathologic process probably begins with exaggerated production of autoantibodies. The autoantibodies combine with corresponding antigens to form immune complexes. These immune complexes are deposited in vascular and tissue surfaces, triggering an inflammatory response and ultimately causing local tissue injury. Some autoantibodies that have been identified in SLE include anti-deoxyribonucleic acid (DNA). Other antibodies may be produced against various cells, including red blood cell surface antigens, platelets, and coagulation factors. Autoantibodies against red blood cells can lead to anemia and those against platelets to thrombocytopenia. Certain drugs may provoke a lupus-like disorder in sus-

ceptible persons, particularly in the elderly. The most common of these drugs are hydralazine and procainamide. Other drugs, such as quinidine, methyldopa, isoniazid, and phenytoin, also have been known to produce this syndrome. The disease usually recedes when use of the drug is discontinued.

Clinical Course. Systemic lupus erythematosus can manifest in a variety of ways. The disease has been called the *great imitator* because it has the capacity for affecting many different body systems, including the musculoskeletal system, the skin, the cardiovascular system, the lungs, the kidneys, the central nervous system (CNS), and the red blood cells and platelets. The onset may be acute or insidious, and the course of the disease is characterized by exacerbations and remissions.²⁷

Arthralgias and arthritis are among the most commonly occurring early symptoms of SLE; approximately 90% of all persons with the disease report joint pain at some point during the course of their disease.²⁷ The polyarthritis of SLE initially can be confused with other forms of arthritis, especially RA, because of the symmetric arthropathy. Ligaments, tendons, and the joint capsule may be involved, causing varied deformities in approximately 30% of persons with the disease. Flexion contractures, hyperextension of the interphalangeal joint, and subluxation of the carpometacarpal joint contribute to the deformity and subsequent loss of function in the hands. Other musculoskeletal manifestations include tenosynovitis, rupture of the intrapatellar and Achilles tendons, and avascular necrosis, frequently of the femoral head.

Skin manifestations can vary greatly and may be classified as acute, subacute, or chronic. The acute skin lesions include the classic malar or "butterfly" rash on the nose and cheeks (Fig. 43-9). This rash is seen in SLE but may be associated with



other skin lesions, such as hives or livedo reticularis (*i.e.*, reticular cyanotic discoloration of the skin, often precipitated by cold) and fingertip lesions, such as periungual erythema, nail fold infarcts, and splinter hemorrhages. Hair loss is common. Mucous membrane lesions tend to occur during periods of exacerbation. Sun sensitivity may occur in SLE even after mild sun exposure. Discoid SLE (*i.e.*, chronic cutaneous lupus) involves plaquelike lesions on the head, scalp, and neck. These lesions first appear as red, swollen patches of skin, and later there can be scarring, depigmentation, and plugging of hair follicles. Ninety percent of patients with discoid lupus have disease that involves only the skin.

Renal involvement occurs in approximately 50% of persons with SLE. Several forms of glomerulonephritis may occur, including mesangial, focal proliferative, diffuse proliferative, and membranous (see Chapter 23). Interstitial nephritis also may occur. Nephrotic syndrome causes proteinuria with resultant edema in the legs, abdomen, and around the eyes.

The heart and lungs frequently are sites of complications in people with SLE.²⁷ Pulmonary involvement in SLE occurs in 40% to 50% of patients and is manifested primarily by pleural effusions or pleuritis. Less frequently occurring pulmonary problems include acute pneumonitis, pulmonary hemorrhage, chronic interstitial lung disease, and pulmonary embolism. Pericarditis, often accompanied by pleural effusions, is the most common of the cardiac manifestations of SLE. Myocarditis affects as many as 25% of those with SLE. Hypertension may be associated with lupus nephritis and long-term corticosteroid use. Ischemic heart disease can occur in older patients with longer-duration SLE. Hematologic disorders may manifest as hemolytic anemia, leukopenia, lymphopenia, or thrombocytopenia.

The CNS is involved in 30% to 75% of persons with SLE. The pathologic basis for the CNS symptoms is not entirely clear. It has been ascribed to an acute vasculitis that impedes blood flow, causing strokes or hemorrhage; an immune response involving antineuronal antibodies that attack nerve cells; or production of antiphospholipid antibodies that damage blood vessels and cause blood clots in the brain. Seizures can occur and are more common when renal failure is present. Psychotic symptoms, including depression and unnatural euphoria, as well as decreased cognitive functioning, confusion, and altered levels of consciousness, may develop.

Subacute cutaneous lupus erythematosus (SCLE) is a less severe form of lupus. The skin lesions in this condition may resemble psoriasis. These lesions are found in sun-exposed areas, such as the face, chest, upper back, and arms. Patients with SCLE may have mild systemic problems, which usually are limited to joint and muscle pains.

Diagnosis and Treatment. The diagnosis of SLE is based on a complete history, physical examination, and analysis of blood work. No single test can diagnose SLE in all persons. The most common laboratory test performed is the immunofluorescence test for ANA. Ninety-five percent of persons with untreated SLE have high ANA levels. Although the ANA test is not specific for lupus, it establishes that the differential diagnosis includes autoimmunity. The anti-DNA antibody test is more specific for the diagnosis of SLE.²⁷ Other serum tests may reveal moderate to severe anemia, thrombocytopenia, and leukocytosis or leukopenia. Additional immunologic tests may be done to give

support to the diagnosis or to differentiate SLE from other connective tissue diseases.

Treatment of SLE focuses on managing the acute and chronic symptoms of the disease. The goals of treatment include preventing progressive loss of organ function, reducing the possibility of exacerbations, minimizing disability from the disease process, and preventing complications from medication therapy.²⁸ Treatment with medications may be as simple as a drug to reduce inflammation, such as an NSAID. NSAIDs can control fever, arthritis, and mild pleuritis. An antimalarial drug (*e.g.*, hydroxychloroquine) may be the next medication considered to treat cutaneous and musculoskeletal manifestations of SLE. Corticosteroids are used to treat more significant symptoms of SLE, such as renal and CNS disorders. Immunosuppressive drugs are used in cases of severe disease.

Systemic Sclerosis

Systemic sclerosis, sometimes called *scleroderma*, is an autoimmune disease of connective tissue characterized by excessive collagen deposition in the skin and internal organs, such as the lungs, gastrointestinal tract, heart, and kidneys. In this disorder, the skin is thickened through fibrosis, with an accompanying fixation of subdermal structures, including the sheaths or fascia covering tendons and muscles.²⁹ Systemic sclerosis affects women four times as frequently as men, with a peak incidence in the 35-year to 50-year age group.¹² The cause of this rare disorder is poorly understood. There is evidence of both humoral and cellular immune system abnormalities.

Scleroderma presents as two distinct clinical entities: the diffuse or generalized form of the disease and the limited or CREST variant. In the CREST syndrome, hardening of the skin (scleroderma) is limited to the hands and face; whereas, the skin changes in diffuse scleroderma also involve the trunk and proximal extremities. Almost all persons with scleroderma develop polyarthritis and Raynaud's phenomenon, a vascular disorder characterized by reversible vasospasm of the arteries supplying the fingers (see Chapter 15).

Diffuse scleroderma is characterized by severe and progressive disease of the skin and the early onset of organ involvement. The typical person has a "stone facies" caused by tightening of the facial skin with restricted motion of the mouth. Involvement of the esophagus leads to hypomotility and difficulty swallowing. Malabsorption may develop if the submucosal and muscular atrophy affect the intestine. Pulmonary involvement leads to dyspnea and eventually respiratory failure. Vascular involvement of the kidneys is responsible for malignant hypertension and progressive renal insufficiency. Cardiac problems include pericarditis, heart block, and myocardial fibrosis.

The CREST syndrome is manifest by calcinosis (*i.e.*, calcium deposits in the subcutaneous tissue that erupt through the skin), Raynaud's phenomenon, esophageal dysmotility, sclerodactyly (localized scleroderma of the fingers), and telangiectasia.³⁰

Treatment of systemic sclerosis is largely symptomatic and supportive. The 9-year survival rate is about 40%.³⁰ Studies have indicated that if heart, lung, or kidney involvement is to become severe, it tends to do so early in disease and is a predictor of shortened survival. Patients who survive the first few years without experiencing severe organ involvement are less likely to experience life-threatening involvement later in their illness.³⁰

Seronegative Spondyloarthropathies

The *spondyloarthropathies* are an interrelated group of multisystem inflammatory disorders that primarily affect the axial skeleton, particularly the spine. Because there is an absence of the RF, these disorders often are referred to as *seronegative spondyloarthropathies*. In contrast to RA, the inflammation begins at sites where tendons and ligament insert into bone, rather than in the synovium. Sacroiliitis is the pathologic hallmark. Persons with the disorder may also have inflammation and involvement of the peripheral joints, in which case the signs and symptoms overlap with other inflammatory types of arthritis. The seronegative spondyloarthropathies include ankylosing spondylitis, reactive arthritis, and psoriatic arthritis.

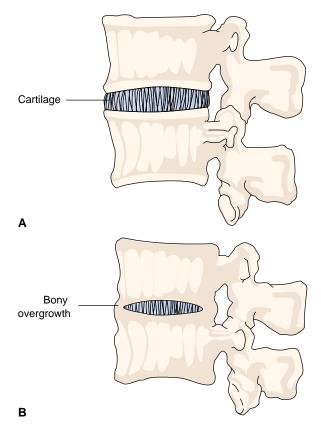
The pathogenesis of the spondyloarthropathies is unclear. However, there is a striking association with the HLA-B27 antigen. The HLA-B27 antigen is also found in the normal population; thus, it is neither necessary nor sufficient for the development of any of the diseases.

Ankylosing Spondylitis

Ankylosing spondylitis is a chronic, systemic inflammatory disease of the joints of the axial skeleton, manifested by pain and progressive stiffening of the spine. The disorder begins in the sacroiliac joints bilaterally and then smaller joints of the posterior elements of the spine. The result is the ultimate destruction of these joints with ankylosis or posterior fusion of the spine. The vertebrae take on a squared appearance, and bone bridges fuse one vertebral body to the next across the intervertebral discs (Fig. 43-10). Progressive spinal changes usually follow an ascending pattern up the spine. Occasionally, large synovial joints (i.e., hips, knees, and shoulders) may be involved. The small peripheral joints usually are not affected. The disease affects approximately 2% to 8% of the HLA-B27positive white population.¹⁸ Clinical manifestations usually begin in late adolescence or early adulthood and are slightly more common in men than in women. The disease spectrum ranges from an asymptomatic sacroiliitis to a progressive disease that can affect many body systems.

The pathogenesis of ankylosing spondylitis remains unclear. The presence of mononuclear cells in acutely involved tissue suggests an immune response in which genetic and environmental factors play a role. The HLA-B27 antigen remains one of the best-known examples of an association between a disease and a hereditary marker. Approximately 90% of those with ankylosing spondylitis have the HLA-B27 antigen, and nearly 100% of those who also have uveitis or aortitis have the marker. Although the mechanism by which HLA-B27 influences the development of ankylosing spondylitis remains to be discovered, it has been proposed that the HLA-B27 antigen and some environmental trigger are structurally related in a manner that induces an autoimmune response.

Clinical Manifestations. The person with ankylosing spondylitis typically reports low back pain, which may be persistent or intermittent. The pain, which becomes worse when resting, particularly when lying in bed, initially may be blamed on muscle strain or spasm from physical activity. Lumbosacral pain also may be present, with discomfort in the buttocks and hip areas. Sometimes, pain can radiate to the thigh in a manner similar to that of sciatic pain. Prolonged stiffness is present in the morning and after periods of rest. Mild physical activity



■ FIGURE 43-10 ■ The bony overgrowth (**B**) of the vertebra characteristic of ankylosing spondylitis is evident when compared with normal vertebra (**A**).

or a hot shower helps reduce pain and stiffness. Sleep patterns frequently are interrupted because of these manifestations. Walking or exercise may be needed to provide the comfort needed to return to sleep. Muscle spasm also may contribute to discomfort.

Loss of motion in the spinal column is characteristic of the disease. The severity and duration of disease activity influence the degree of mobility. Loss of lumbar lordosis occurs as the disease progresses, and this is followed by kyphosis of the thoracic spine and extension of the neck. A kyphotic spine makes it difficult for the patient to look ahead and to maintain balance while walking. The image is one of a person bent over looking at the floor and unable to straighten. X-ray films show a rigid, bamboo-like spine. The heart and lungs are constricted in the chest cavity. Abnormal weight bearing can lead to degeneration and destruction of the peripheral joints, most commonly the hips, shoulders and knees.

The most common extraskeletal involvement is acute anterior uveitis (inflammation of uvea of the eye), which occurs in 25% to 30% of patients sometime in the course of their disease.¹⁸ Systemic features of weight loss, fever, and fatigue may be apparent. Sometimes, the fatigue is a greater problem than pain or stiffness. Osteoporosis can occur, especially in the spine, which contributes to the risk of spinal fracture. Fusion of the costovertebral joints can lead to reduced lung volume.

Diagnosis and Treatment. The diagnosis of ankylosing spondylitis is based on history, physical examination, and x-ray examination. Laboratory findings frequently include an elevated ESR. A mild normocytic normochromic anemia may be present. Because HLA-B27 is found in 8% of the normal population, it is not a specific diagnostic test for the disease. Radiologic evaluations help differentiate sacroiliitis from other diseases.

Treatment is directed at controlling pain and maintaining mobility by suppressing inflammation. Therapeutic exercises are important to assist in maintaining motion in peripheral joints and in the spine. Muscle-strengthening exercises for extensor muscle groups also are prescribed. Immobilizing joints is not recommended. NSAIDs are used to reduce inflammation, which helps to control pain and reduce muscle spasm. Most peripheral joint pain and limitations of motion occur in the hip. Total hip replacement surgery may be used to reduce pain and increase mobility.

Reactive Arthritis

Reactive arthritis refers to a form of peripheral arthritis, often accompanied by one or more extra-articular manifestations, that occurs shortly after certain infections of genitourinary or gastrointestinal systems.^{18,31,32} The majority of affected persons are young men who have inherited the HLA-B27 marker. Two forms of reactive arthritis are Reiter's syndrome and enteropathic arthritis.

Reiter's Syndrome. Reiter's syndrome is considered to be a clinical manifestation of reactive arthritis that may be accompanied by extra-articular symptoms, such as uveitis, bowel inflammation, and carditis. Reiter's syndrome develops in a genetically susceptible host after an infection by bacteria, *Chlamydia trachomatis* in the genitourinary tract, or *Salmonella, Shigella, Yersinia*, or *Campylobacter* in the gastrointestinal tract.

The arthritis is usually asymmetric and frequently involves the large weight-bearing joints (*i.e.*, the knee and ankle); sacroiliitis or ankylosing spondylitis, especially after frequent recurrences. Systemic symptoms include fever and weight loss, both of which are common at the onset of the disease. Several mucocutaneous manifestations, including balanitis, oral lesions, and skin rash are common. The skin rash, which resembles pustular psoriasis, appears most commonly on the soles of feet and palms of hands but may affect any cutaneous area. The toenails and fingernails may become thickened and opaque and may crumble, resembling the changes seen in fungal infections of the nails. Aortitis, with aortic regurgitation occurs in 1% to 2% of persons, typically after longstanding active arthritis. Although most signs of the disease disappear within days or weeks, the arthritis may persist for several months or even years. Recurrences involving any of the combination of clinical manifestations are common and sometimes followed by permanent sequelae, especially in the joints.

The treatment is largely symptomatic. NSAIDs are used to treat the arthritic symptoms. Antibiotics may be of some use in treating the triggering infection.

Enteropathic Arthritis. Arthritis that is associated with an inflammatory bowel disease usually is considered an enteropathic arthritis because the intestinal disease is directly involved in the pathogenesis. Most cases of enteropathic arthritis are classified among the spondyloarthropathies. These include cases in which the arthritis is associated with inflammatory bowel disease (*i.e.*, ulcerative colitis and Crohn's disease), the reactive arthritides triggered by enterogenic bacteria, some of

the undifferentiated spondyloarthropathies, Whipple's disease, and reactions after intestinal bypass surgery.¹⁸ There is no direct relation between the activity of the bowel disease and the degree of arthritis activity.

Psoriatic Arthritis

Psoriatic arthritis occurs in approximately 5% to 7% of people with psoriasis. It is heterogenous disorder with features typical of the spondyloarthropathies in some persons, features of RA in others, and features of both diseases coexisting in yet others.¹⁸

The etiology of psoriasis and psoriatic arthritis is unknown. Genetic, environmental, and immunologic factors appear to influence susceptibility and expression of disease. Environmental factors including infectious agents and physical trauma may play a role in the pathogenesis of the disorder. T cellmediated immune responses also seem to play an important role in the skin and joint manifestations of the disease. Psoriasis improves after treatment with immunosuppressant agents such as cyclosporine.

Psoriatic arthritis may present with a variety of forms, including monoarthritis, asymmetric oligoarthritis, or symmetric polyarthritis. Usually the arthritis is asymmetric with "sausage" appearance of the fingers and toes. Sacroiliac joint involvement is common, and ankylosis of the sacroiliac joints may occur.

Although the arthritis can antedate detectable skin rash, the definitive diagnosis of psoriatic arthritis cannot be made without evidence of skin or nail changes typical of psoriasis. This heterogeneous clinical presentation suggests more than one disease is associated with psoriasis, or various clinical responses to a common cause. At least 20% of those with psoriatic arthritis have an elevated serum level of uric acid. The abnormally elevated serum uric acid level is caused by the rapid skin turnover of psoriasis, the breakdown of nucleic acid, and the metabolism to uric acid. This finding may lead to a misdiagnosis of gout. Psoriatic arthritis tends to be slowly progressive but is associated with a more favorable prognosis than RA.

Basic management is similar to that for the treatment of RA. Suppression of the skin disease may be important in helping control the arthritis. Often, affected joints are surprisingly functional and only minimally symptomatic.

Osteoarthritis Syndrome

Osteoarthritis (OA), formerly called *degenerative joint disease*, is the most prevalent form of arthritis. It is second only to cardiovascular disease as the cause of chronic disability in adults.³³ The term *osteoarthritis* encompasses a heterogeneous collection of syndromes and is more of a disease process than a specific entity. It can occur as a primary idiopathic or a secondary disorder, although this distinction is not always clear. Idiopathic or primary variants of OA occur as localized or generalized (*i.e.*, more than three joints) syndromes. Secondary OA has a known underlying cause, such as congenital or acquired defects of joint structures, trauma, metabolic disorders, or inflammatory diseases (see Chart 43-3).

One third of all adults in the United States have radiographic evidence of osteoarthritis of the hand, foot, knee, or hip. Sex and age interact to influence the time of onset and, with race, the pattern of joint involvement. Men are affected more commonly at a younger age than women, but the rate of women affected exceeds that of men by middle age. Hand OA is more likely to affect white women, whereas knee OA is more

CHART 43-3 Causes of Osteoarthritis	Mechanical injury
CHART 43-3 Causes of Osteoarthritis Postinflammatory disorders Rheumatoid arthritis Septic joint Post-traumatic disorders Acute fracture Ligament or meniscal injury Cumulative occupational or recreational trauma Anatomic or bony disorders Hip dysplasia Avascular necrosis Paget's disease Slipped capital femoral epiphysis Legg–Calvé–Perthes disease Metabolic disorders Calcium crystal deposition Hemochromatosis Acromegaly Wilson's disease Ochronosis Neuropathic arthritis Charcot joint Hereditary disorders of collagen	Mechanical injury Chondrocyte response Release of cytokines (e.g., TNF, IL-1) Production, release of protease enzymes Development of surface cracks Destruction of joint structures Osteophyte formation
Idiopathic or primary variants	FIGURE 43-11 Disease process in osteoarthritis.

common in black women. Obesity is a risk factor for OA of the knee in women and a contributory biomechanical factor in the pathogenesis of the disease. Excess fat may have a direct metabolic effect on cartilage beyond the effects of excess joint stress. Heredity influences the occurrence of hand OA in the DIP joint. Bone mass may also influence the risk of developing OA. In theory, thinner subchondral bone mass may provide a greater shock-absorbing function than denser bone, allowing less direct trauma to the cartilage.

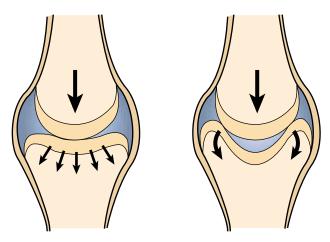
Osteoarthritis is a disorder of the articular cartilage and subchondral bone (*i.e.*, bony plate that supports the articular cartilage) of diarthrodial joints. The joint changes associated with OA are progressive loss of articular cartilage and synovitis resulting from the inflammation caused by the attempts of the bone to remold itself, creating osteophytes or spurs (Fig. 43-11). These changes are accompanied by joint pain, stiffness, limitation of motion, and possibly by joint instability and deformity. Although there may be periods of mild inflammation, it is not the severe, destructive type seen in the inflammatory forms of rheumatic diseases such as RA.

Pathogenesis

The pathogenesis of OA resides in the homeostatic mechanisms that maintain the articular cartilage. Cartilage is a specialized type of connective tissue. As with other types of tissue, it consists of cells (*i.e.*, chondrocytes) nested in an extracellular matrix. In articular cartilage, the extracellular matrix is composed of water, ground substance, collagen, and proteoglycans. The ground substance constitutes a highly hydrated, semisolid gel. Collagen molecules consist of polypeptide chains that form long fibrous strands. The primary function of the collagen fibers is to provide a rigid scaffold to support the chondrocytes and ground substance of cartilage. The proteoglycans, which are large macromolecules made up of disaccharides and amino acids, afford elasticity and stiffness, permitting articular cartilage to resist compression. They also provide a film of interstitial fluid that contributes to the lubrication of the joint. Under high loads such as weight bearing, fluid squeezes out of the cartilage with compression of the opposing surfaces of the joint. The greater the load is, the better the lubrication. With depletion of proteoglycans from the cartilage matrix in OA, the mechanisms that normally operate under high loads to produce a pressurized lubricating film may be impaired.

Articular cartilage plays two essential mechanical roles in joint physiology. First, the articular cartilage serves as a remarkably smooth weight-bearing surface. In combination with synovial fluid, the articular cartilage provides extremely low friction during movement of the joint. Second, the cartilage transmits the load down to the bone, dissipating the mechanical stress. The subchondral bone protects the overlying articular cartilage, providing it with a pliable bed and absorbing the energy of the force (Fig. 43-12).

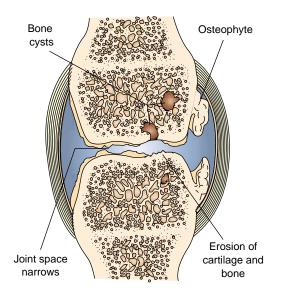
Popularly known as *wear and tear* arthritis, the changes that occur to the articular cartilage in OA are complex. The first recognizable change in OA is edema of the extracellular matrix, principally the intermediate layer. The cartilage loses its smooth aspect, and surface cracks occur allowing synovial fluid to enter and widen the crack. As the crack deepens, the vertical clefts form in the subchondral bone cartilage. Clusters of chondrocytes appear around these clefts and at the surface. Fissures cause fragments of cartilage to become dislodged and enter the articular cavity, creating osteocartilaginous loose bodies and



■ FIGURE 43-12 ■ (Left) A joint normally undergoes deformation of the articular cartilage and the subchondral bone when carrying a load. This maximizes the contact area and spreads the force of the load. (**Right**) If the joint does not deform with a load, the stresses are concentrated and the joint breaks down. (Redrawn from Brandt K.D. & Radin E. [1987]. The physiology of articular stress: Osteoarthroses. *Hospital Practice* [January 15], 111)

uncovering areas of subchondral bone. Sclerosis, or formation of new bone and cysts, usually occurs in the juxta-articular bone (*i.e.*, bone near the joint). New bone that forms at the joint margins is called an *osteophyte*, or spur.

The articular cartilage injury that occurs in OA is thought to result from the release of cytokines such as interleukin-1 and TNF (Fig. 43-13).¹⁸ These chemical messengers stimulate the production and release of proteases (enzymes) that are destructive to joint structures. The resulting damage predisposes the chondrocytes to more injury. The earliest changes in OA are



■ FIGURE 43-13 ■ Joint changes in osteoarthritis. The left side denotes early changes and joint space narrowing with cartilage breakdown. The right side shows more severe disease progression with lost cartilage and osteophyte formation.

the loss of proteoglycans from the surface of the articular cartilage, followed by death of the chondrocytes. Inadequate repair mechanisms and imbalances between the proteases and their inhibitors may contribute further to disease progression.

Immobilization also can produce degenerative changes in articular cartilage. Cartilage degeneration caused by immobility may result from loss of the pumping action of lubrication that occurs with joint movement. These changes are more marked and appear earlier in areas of contact but occur also in areas not subject to mechanical compression. Although cartilage atrophy is rapidly reversible with activity after a period of immobilization, impact exercise during the period of remobilization can prevent reversal of the atrophy. Thus, slow and gradual remobilization may be an important aspect in preventing cartilage injury. Clinically, it has implications for instructions concerning the recommended level of physical activity after removal of a cast.

Clinical Manifestations

The manifestations of OA may occur suddenly or insidiously. Initially, pain may be described as aching and may be somewhat difficult to localize. It worsens with use or activity and is usually relieved by rest. In later stages of disease activity, pain may be experienced during rest and for several hours after the use of the involved joints. Crepitus and grinding may be evident when the joint is moved. As the disease advances, even minimal activity may cause pain because of the limited range of motion resulting from intra-articular and periarticular structural damage.

The most frequently affected joints are the hips, knees, lumbar and cervical vertebrae, proximal and distal joints of the hand, the first carpometacarpal joint, and the first metatarsophalangeal joints of the feet. A single joint or several may be affected. Although a single weight-bearing joint may be involved initially, other joints often become affected because of the additional stress placed on them while trying to protect the original joint. It is not unusual for a person having a knee replacement to discover soon after the surgery is done that the second knee also needs to be replaced. Other clinical features are limitations of joint motion and joint instability. Joint enlargement usually results from new bone formation; the joint feels hard, in contrast to the soft, spongy feeling characteristic of the joint in RA. Sometimes, mild synovitis or increased synovial fluid can cause joint enlargement.

Diagnosis and Treatment

The diagnosis of OA usually is determined by history and physical examination, x-ray studies, and laboratory findings that exclude other diseases. Although OA often is contrasted with RA for diagnostic purposes, the differences are not always readily apparent. Other rheumatic diseases may be superimposed on OA.

Characteristic radiologic changes initially include medial joint space narrowing, followed by subchondral bony sclerosis, formation of spikes on the tibial eminence, and osteophytes. The results of laboratory studies usually are normal because the disorder is not a systemic disease. The ESR may be slightly elevated in generalized OA or erosive inflammatory variations of the disease. If inflammation is present, there may be a slight increase in the blood cell count. The synovial fluid usually is normal. Because there is no cure, the treatment of OA is symptomatic and includes physical rehabilitative, pharmacologic, and surgical measures. Physical measures are aimed at improving the supporting structures of the joint and strengthening opposing muscle groups involved in cushioning weight-bearing forces. These include a balance of rest and exercise, use of splints to protect and rest the joint, use of heat and cold to relieve pain and muscle spasm, and adjusting the activities of daily living. The involved joint should not be further abused, and steps should be taken to protect and rest it. This includes weight reduction (when weight-bearing surfaces are involved) and the use of a cane or walker if the hips and knees are involved. Muscle-strengthening exercises may help protect the joint and decrease pain.³⁴

Pharmacologic treatment is aimed at reducing inflammation or providing analgesia. The most common medications used in the treatment of OA are the NSAIDs, many of which are available without a prescription. For many persons, acetaminophen in doses as high as 4000 mg/day may be as effective and less toxic than NSAIDs. Corticosteroid injections may be used for relieving symptoms, especially for those who have an effusion of the joint. Injections usually are limited to a total of four and not more than three within 1 year because their use is thought to accelerate joint destruction.³⁴

Surgery is considered when the person is having severe pain and joint function is severely reduced. Procedures include arthroscopic lavage and debridement, bunion resections, osteotomies to change alignment of the knee and hip joints, and decompression of the spinal roots in osteoarthritic vertebral stenosis. Total hip replacements have provided effective relief of symptoms and improved range of motion for many persons, as have total knee replacements, although the latter procedure has produced less consistent results.

Crystal-Induced Arthropathies

Crystal deposition in joints produces arthritis. In gout, monosodium urate or uric acid crystals are found in the joint cavity. Another condition in which calcium pyrophosphate dihydrate crystals are found in the joints sometimes is referred to as *pseudogout* or *chondrocalcinosis*. A brief discussion of pseudogout is provided in the section on rheumatic diseases in the elderly.

Gout

Gout is actually a group of diseases known as the *gout syndrome*. It includes acute gouty arthritis with recurrent attacks of severe articular and periarticular inflammation; tophi or the accumulation of crystalline deposits in articular surfaces, bones, soft tissue, and cartilage; gouty nephropathy or renal impairment; and uric acid kidney stones.

The term *primary gout* is used to designate cases in which the cause of the disorder is unknown or an inborn error in metabolism and is characterized primarily by hyperuricemia and gout. Primary gout is predominantly a disease of men, with a peak incidence in the fourth or sixth decade. In secondary gout, the cause of the hyperuricemia is known but the gout is not the main disorder. Asymptomatic hyperuricemia is a laboratory finding and not a disease. Most persons with hyperuricemia do not develop gout.

Pathogenesis. The pathogenesis of gout resides in an elevation of the serum uric acid levels. Uric acid is the end product of purine (adenine and guanine from DNA and RNA) metabolism.³⁵ Two pathways are involved in purine synthesis: (1) a de novo pathway, in which purines are synthesized from nonpurine precursors, and (2) the salvage pathway, in which purine bases are recaptured from the breakdown of nucleic acids derived from exogenous (dietary) or endogenous sources. The elevation of uric acid and the subsequent development of gout can result from (1) overproduction of purines, (2) decreased salvage of free purine bases, (3) augmented breakdown of nucleic acids as a result of increased cell turnover, or (4) decreased urinary excretion of uric acid. Primary gout, which constitutes 90% of cases, results from enzyme defects that result in an overproduction of uric acid, inadequate elimination of uric acid by the kidney, or a combination of the two. In most cases, the reason is unknown. In secondary gout the hyperuricemia may be caused by increased breakdown of nucleic acids production, as occurs with rapid tumor cell lysis during treatment for lymphoma or leukemia. Other cases of secondary gout result from chronic renal disease. Some of the diuretics, including the thiazides, can interfere with the excretion of uric acid.

An attack of gout occurs when monosodium urate crystals precipitate in the joint and initiate an inflammatory response. Synovial fluid is a poorer solvent for uric acid than plasma, and uric acid crystals are even less soluble at temperatures below 37°C.12 Crystal deposition usually occurs in peripheral areas of the body, such as the great toe, where the temperatures are cooler than other parts of the body. With prolonged hyperuricemia, crystals and microtophi accumulate in the synovial lining cells and in the joint cartilage. The released crystals are chemotactic to leukocytes and also activate complement, leading to inflammation and destructive changes to the cartilage and subchondral bone. Repeated attacks of acute arthritis eventually lead to chronic arthritis and the formation of tophi. Tophi are large, hard nodules that have irregular surfaces and contain crystalline deposits of monosodium urate¹² (Fig. 43-14). They are found most commonly in the synovium, olecranon bursa, Achilles tendon, subchondral bone, and extensor surface of the forearm and may be mistaken for rheumatoid nodules. Tophi usually do not appear until 10 years or more after the first gout attack. This stage of gout, called chronic tophaceous gout, is characterized by more frequent and prolonged attacks, which often are polyarticular.

Clinical Manifestations. The typical acute attack of gout is monoarticular and usually affects the first metatarsophalangeal joint. The tarsal joints, insteps, ankles, heels, knees, wrists, fingers, and elbows also may be initial sites of involvement. Acute gout often begins at night and may be precipitated by excessive exercise, certain medications, foods, alcohol, or dieting. The onset of pain typically is abrupt, and redness and swelling are observed. The attack may last for days or weeks. Pain may be severe enough to be aggravated even by the weight of a bed sheet covering the affected area.

In the early stages of gout after the initial attack has subsided, the person is asymptomatic, and joint abnormalities are not evident. This is referred to as *intercritical gout*. After the first attack, it may be months or years before another attack. As attacks recur with increased frequency, joint changes occur and become permanent.

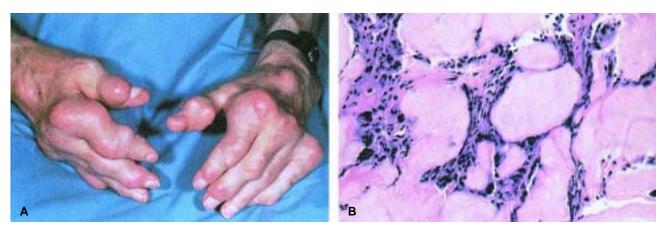


FIGURE 43-14 Gout. (**A**) Gouty tophi project from the fingers as rubbery nodules. (**B**) A section from a trophus shows extracellular masses of urate crystals with accompanying foreign-body giant cells. (Rubin E., Farber J.L. [1999]. *Pathology* [3rd ed., p. 1404]. Philadelphia: Lippincott Williams & Wilkins)

Diagnosis and Treatment. Although hyperuricemia is the biochemical hallmark of gout, the presence of hyperuricemia cannot be equated with gout because many persons with this condition never develop gout. A definitive diagnosis of gout can be made only when monosodium urate crystals are in the synovial fluid or in tissue sections of tophaceous deposits. Synovial fluid analysis is useful in excluding other conditions, such as septic arthritis, pseudogout, and RA.³⁶ Diagnostic methods also include measures to determine if the disorder is related to overproduction or to underexcretion of uric acid.

The objectives for treatment of gout include the termination and prevention of the acute attacks of gouty arthritis and the correction of hyperuricemia, with consequent inhibition of further precipitation of sodium urate and absorption of urate crystal deposits already in the tissues. Some changes in lifestyle may be needed, such as maintenance of ideal weight, moderation in alcohol consumption, and avoiding purinerich foods, such as liver, kidney, sardines, anchovies, and sweetbreads, particularly by persons with excessive tophaceous deposits.

Pharmacologic management of acute gout is directed toward reducing joint inflammation. Hyperuricemia and related problems of tophi, joint destruction, and renal problems are treated after the acute inflammatory process has subsided. NSAIDs, particularly indomethacin and ibuprofen, are used for treating acute gouty arthritis. Alternative therapies include colchicine and intra-articular deposition of corticosteroids. Treatment with colchicine is used early in the acute stage. Colchicine produces its anti-inflammatory effects by inhibition of leukocyte migration and phagocytosis.

Treatment of hyperuricemia is aimed at maintaining normal uric acid levels and is lifelong. Two classes of drugs may be used. Allopurinol prevents the production of uric acid, and uricosuric drugs, such as probenecid or sulfinpyrazone, may be used to prevent the tubular reabsorption of urate and increase its excretion.³⁰ Prophylactic colchicine or NSAIDs may be used between gout attacks. If the uric acid level is normal and the person has not had recurrent attacks of gout, the use of these medications may be discontinued.³⁵ **In summary**, rheumatoid arthritis is a chronic systemic inflammatory disorder affecting multiple joints. Women are affected more frequently than men. Joint involvement, which is symmetric, begins with inflammatory changes of the synovium and formation of a destructive granulation tissue called *pannus* that leads to joint instability and eventual deformity. Systemic manifestations include weakness, anorexia, weight loss, and low-grade fever. Extra-articular features include rheumatoid nodules and vasculitis.

Systemic lupus erythematosus is a chronic autoimmune disorder that affects multiple body systems. There is no known cause of SLE, but the disease may result from an immunoregulatory disturbance brought about by a combination of genetic, hormonal, and environmental factors. Some drugs have been shown to induce lupus, especially in the elderly. There is an exaggerated production of autoantibodies, which interact with antigens to produce an immune complex. These immune complexes produce an inflammatory response in affected tissues. Systemic sclerosis, often prefixed by the term *progressive*, is sometimes called *scleroderma*. In this disorder, the skin is thickened through fibrosis with an accompanying fixation to the subdermal structures, including the sheaths or fascia covering tendons and muscles.

The spondyloarthropathies affect the axial skeleton, particularly the spine. Inflammation develops at sites where ligaments insert into bone. Because they lack the RF, they are referred to as *seronegative spondyloarthropathies*. They include ankylosing spondylitis, reactive arthritis, and psoriatic arthritis. Ankylosing spondylitis is considered a prototype of this classification category. Bilateral sacroiliitis is the primary feature of ankylosing spondylitis. The disease spectrum ranges from asymptomatic sacroiliitis to a progressive disorder affecting many body systems. The cause remains unknown; however, a strong association between the HLA-B27 antigen and ankylosing spondylitis has been identified. Loss of motion in the spinal column is characteristic of the disease. Peripheral arthritis may occur in some persons. Osteoarthritis, the most common form of arthritis, is a localized condition affecting primarily the weight-bearing joints. Risk factors for OA progression include older age, OA in multiple joints, neuropathy, and for knees, obesity. The disorder is characterized by degeneration of the articular cartilage and subchondral bone. As cartilage ages, biochemical events such as collagen fatigue and fracture occur with less stress. Attempts at repair by increased matrix synthesis and cellular proliferation maintain the integrity of the cartilage until failure of reparative processes allows the degenerative changes to progress. Pain and stiffness are primary features of the disease. Inflammatory mediators (*e.g.*, prostaglandins) may increase the inflammatory and degenerative response.

Gout is a crystal-induced arthropathy. Acute attacks of arthritis occur with gout and are characterized by the presence of monosodium urate crystals in the joint. The disorder is accompanied by hyperuricemia, which results from overproduction of uric acid or from the reduced ability of the kidney to rid the body of excess uric acid. Management of acute gout is directed first toward the reduction of joint inflammation; then the hyperuricemia is treated. Hyperuricemia is treated with uricosuric agents, which prevent the tubular reabsorption of urate, or with medication that inhibits the production of uric acid.

RHEUMATIC DISEASES IN CHILDREN AND THE ELDERLY

Rheumatic Diseases in Children

Children can be affected with almost all of the rheumatic diseases. In addition to disease-specific differences, these conditions affect not only the child but the family. Growth and development require special attention. Rheumatic disorders of children include juvenile rheumatoid arthritis, systemic lupus erythematosus, and juvenile spondyloarthropathies.

Juvenile Rheumatoid Arthritis

Juvenile rheumatoid arthritis (JRA) is a chronic disease that affects approximately 60,000 to 200,000 children in the United States.¹⁸ It is characterized by synovitis and can influence epiphyseal growth by stimulating growth of the affected side. Generalized stunted growth also may occur.

Systemic onset (*i.e.*, Still's disease) affects approximately 20% of children with JRA.¹⁸ The symptoms of Still's disease include a daily intermittent high fever, which usually is accompanied by a rash, generalized lymphadenopathy, hepatosplenomegaly, leukocytosis, and anemia. Most of these children also have joint involvement. Systemic symptoms usually subside in 6 to 12 months. This form of JRA also can make an initial appearance in adulthood. Infections, heart disease, and adrenal insufficiency may cause death.

A second subgroup of JRA, pauciarticular arthritis, affects no more than four joints. This disease affects 55% to 75% of children with JRA. Pauciarticular arthritis affects two distinct groups. The first group generally consists of girls younger than 6 years of age with chronic uveitis. The results of ANA testing in this group usually are positive. The second group, characterized by late-onset arthritis, is made up mostly of boys. The HLA-B27 test results are positive in more than one half of this group. They are affected by sacroiliitis, and the arthritis usually occurs in the lower extremities.

The third subgroup of JRA, accounting for approximately 20% of the total, is polyarticular onset disease. It affects more than four joints during the first 6 months of the disease. This form of arthritis more closely resembles the adult form of the disease than the other two subgroups. RF sometimes is present and may indicate a more active disease process. Systemic features include a low-grade fever, weight loss, malaise, anemia, stunted growth, slight organomegaly (*e.g.*, hepatosplenomegaly), and adenopathy.¹⁸

The prognosis for most children with rheumatoid arthritis is good. NSAIDs are the first-line drugs used in treating JRA. Salicylates have been replaced by agents such as naproxen, ibuprofen, and ketoprofen. The second-line agent is low-dose methotrexate or, less often, sulfasalazine. Gold salts, hydroxychloroquine, and D-penicillamine rarely are used.³⁷ Other aspects of treatment of children with JRA are similar to those used for the adult with rheumatoid arthritis. Children are encouraged to lead as normal a life as possible.

Juvenile Spondyloarthropathies

Ankylosing spondylitis, reactive arthritis, psoriatic arthritis, and spondyloarthropathies associated with ulcerative colitis and regional enteritis can affect children and adults. In children, spondyloarthritis manifests in peripheral joints first, mimicking pauciarticular JRA, with no evidence of sacroiliac or spine involvement for months to years after onset. The spondyloarthropathies are more common in boys and commonly occur in children who have a positive family history. HLA-B27 typing is helpful in diagnosing the disease in children because of the unusual presentation of the disease.

Management of the disease involves physical therapy, education, and attention to school and growth and development issues. Medication includes the use of salicylates or other NSAIDs such as tolmetin or indomethacin. More severe disease or symptoms may require systemic corticosteroids.¹⁸

Rheumatic Diseases in the Elderly

Arthritis is the most common complaint of elderly persons. The pain, stiffness, and muscle weakness affect daily life, often threatening independence and quality of life. Symptoms of the rheumatic diseases also can have an indirect effect and even threaten the duration of life for the elderly. The weakness and gait disturbance that often accompany the rheumatic diseases can contribute to falls and fractures, causing suffering, increased health care costs, further loss of independence, and the potential for a decreased life span.

There are differences in the manifestations, diagnosis, and treatment of some of the rheumatic diseases in the elderly. Older patients often have multiple problems complicating diagnosis and management. The diagnosis of an elderly patient with a musculoskeletal problem must consider a wide variety of disorders that usually are regarded as outside the range of typical rheumatic disease. Among these are metastatic malignancy, multiple myeloma, musculoskeletal disorders accompanying endocrine or metabolic disorders, orthopedic conditions, and neurologic disease.

Osteoarthritis is by far the most common form of arthritis among the elderly. It is the greatest cause of disability and limitation of activity in older populations. The prevalence of rheumatoid arthritis increases with advancing age, at least until 75 years of age.³⁸

Polymyalgia Rheumatica

Of the various forms of rheumatic disease affecting the elderly, polymyalgia rheumatica is one of the more difficult to diagnose and one of the most important to identify. It is a common syndrome of older patients, rarely occurring before age 50 and usually after age 60 years. Elderly women are especially at risk.

The clinical manifestations of polymyalgia rheumatica include pain and stiffness of the shoulder and pelvic girdle areas. The onset can be abrupt, with the patient going to bed feeling well and awakening with pain and stiffness in the neck, shoulders, and hips. These symptoms may be accompanied by fever, malaise, and weight loss. Because of the shoulder and pelvic area weakness, persons with disorder often have trouble combing their hair, putting on a coat, and getting out of a chair.

A certain percentage of patients with polymyalgia rheumatica also have giant cell arteritis (also called temporal arteritis). The two conditions are considered to represent different manifestations of the same disease. Giant cell arteritis, a form of systemic vasculitis, is a systemic inflammatory disease of large and medium-size arteries (see Chapter 15). It predominantly affects branches of arteries originating from the aortic arch, including the superficial temporal, vertebral, ophthalmic, and posterior ciliary arteries. The disorder often is insidious in onset and may be heralded by the sudden onset of headache, tenderness over the artery, swelling and redness of the overlying skin, blurred vision or diplopia, and facial pain.

The diagnosis of polymyalgia rheumatica is based on the pain and stiffness persisting for at least 1 month and an elevated ESR. The diagnosis is confirmed when the symptoms respond dramatically to a small dose of prednisone, a corticosteroid. For patients with an elevated ESR, the diagnosis usually is based on a 3-day trial of prednisone treatment.³⁸ People with polymyalgia rheumatica typically exhibit striking clinical improvement approximately the second day. Treatment with NSAIDs provides relief for some patients, but most require continuing therapy with prednisone, with gradual reduction of the dose over the course of 1.5 to 2 years, using the patient's symptoms as the primary guide. Treatment of persons with giant cell arteritis requires use of high-dose prednisone to prevent loss of vision.

Pseudogout

As part of the tissue-aging process, OA develops with associated cartilage degeneration. Calcium pyrophosphate crystals are shed into the joint cavity. These crystals may produce a low-grade chronic inflammation—the chronic pseudogout syndrome. The accumulation of calcium pyrophosphate and related crystalline deposits in articular cartilage is common in the elderly. There are no medications that can remove the crystals from the joints. Although it may be asymptomatic, presence of the crystals may contribute to more rapid cartilage deterioration. This condition may coexist with severe OA.

In summary, rheumatic diseases that affect children can be similar to the adult diseases, but there also are manifestations unique to the younger population. JRA is a chronic disease, characterized by synovitis, that can influence epiphyseal growth. One subgroup of the JRA (Still's disease) presents with systemic manifestations that include a daily intermittent high fever, rash, generalized lymphadenopathy, hepatosplenomegaly, leukocytosis, and anemia. A second subgroup, pauciarticular JRA, affects no more than four joints. The third subgroup of JRA affects more than four joints during the first 6 months of the disease and is similar to the adult form of the disease. Managing rheumatic diseases in children requires a team approach to address issues of the family, school, growth and development, and coping strategies and requires a comprehensive disease management program.

Arthritis is the most common complaint of the elderly population. The pain, stiffness, and muscle weakness affect daily life, often threatening independence and quality of life. There is a difference in the manifestations, diagnosis, and treatment of some of the rheumatic diseases in the elderly compared with those in the younger population. One form of rheumatic disease that has a predilection for the elderly is polymyalgia rheumatica. A certain percentage of patients with polymyalgia rheumatica also have giant cell arteritis, frequently with involvement of the ophthalmic arteries.

REVIEW QUESTIONS

Describe the roles of osteoclasts and osteoblasts in terms of bone remodeling and relate to the pathogenesis of osteoporosis.

Explain how factors that affect bone mass during childhood and early adult life influence the risk of osteoporosis, and relate factors such as diet, exercise, and gonadal hormones to the risk of the osteoporosis.

Describe the pathogenesis and manifestations of osteomalacia and rickets.

- Characterize the cause and manifestations of Paget's disease.
- Describe the difficulty in defining the term arthritis.
- Compare rheumatoid arthritis and osteoarthritis in terms of pathogenesis, joint involvement, level of inflammation, and local and systemic manifestations.
- Compare rheumatoid arthritis and osteoarthritis in terms of pathogenesis, joint pathology, and treatment.

Describe the immunologic process that occurs in systemic lupus erythematosus.

Contrast and compare ankylosing spondylitis, reactive arthritis, and psoriatic arthritis in terms of cause, pathogenesis, and clinical manifestations.

Describe the clinical manifestations, diagnostic measures, and methods used in the treatment of gouty arthritis.

List three types of juvenile rheumatoid arthritis and differentiate among their major characteristics.

Characterize the manifestations of polymyalgia rheumatica, a common musculoskeletal disorder affecting the elderly.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- 1. Manolagas S.C., Jilka R.L. (1995). Bone marrow, cytokines, and bone remodeling. *New England Journal of Medicine* 332, 305–310.
- Burns K., Kumar V. (2003). The musculoskeletal system. In Kumar V., Cotran R.S., Robbins S.L. (Eds.), *Basic pathology* (7th ed., pp. 755–718). Philadelphia: W.B. Saunders.
- Klibanski A. (Panel and Conference Chair). (2000). Osteoporosis prevention, diagnosis and therapy: Consensus statement. Bethesda, MD: National Institutes of Health. Available online: http://consensus.nih. gov/cons/111/111_statement.htm.
- Riggs B.L., Melton L.J. (1992). The prevention and treatment of osteoporosis. New England Journal of Medicine 327, 620–627.
- Gambert S.R., Schulz B.M., Hamdy B.C. (1995). Osteoporosis: Clinical features, prevention, and treatment. *Endocrinology and Metabolic Clinics of North America* 24, 317–371.
- Tebas P., Powerly W.G., Claxton S., et al. (2000). Accelerated bone mineral loss in HIV-infected patients receiving potent antiretroviral therapy. *AIDS* 14 (4), F63–F67.
- Dueck C.A., Matt K.S., Manore M.M., et al. (1996). Treatment of athletic amenorrhea with a diet and training intervention program. *International Journal of Sport Nutrition* 6 (2), 24–40.
- Otis C.L., Drinkwater B., Johnson M., et al. (1997). American College of Sports Medicine position stand: The female athlete triad. *Medicine* and Science in Sports and Exercise 29 (5), i–ix.
- 9. Hobart J.A., Smucker D.R. (2000). The female athlete triad. *American Family Physician* 61, 3357–3367.
- Weinstein L., Ullery B. (2000). Age, weight, and estrogen use determine need for osteoporosis screen. *American Journal of Obstetrics and Gynecology* 183, 547–549.
- Bukata S.V., Rosier R.N. (2000). Diagnosis and treatment of osteoporosis. *Current Opinion in Orthopedics* 11, 336–340.
- Shiller A.L., Teitlbaum S.L. (1999). Bones and joints. In Rubin E, Farber J.L. (Eds.), *Pathology* (3rd ed., pp. 1365–1406). Philadelphia: Lippincott Williams & Wilkins.
- Bishop M. (1999). Rickets today—Children still need milk and sunshine. New England Journal of Medicine 341, 602–604.
- Centers for Disease Control. (2001). Severe malnutrition among young children—Georgia, January 1997–June 1999. Morbidity and Mortality Weekly Report 50 (12), 224–227.
- Bouillon R. (1998). The many faces of rickets. New England Journal of Medicine 131, 935–942.
- Hamdy R.C. (1990). Paget's disease of bone. *Hospital Practice* 25 (10), 33–41.
- 17. Sadovsky R. (1997). Paget's disease of the bone: Bisphosphonate treatment. *American Family Physician* 55, 1400–1401.
- Klippel J.R. (Ed.). (2001). Primer on the rheumatic diseases (12th ed.). Atlanta: Arthritis Foundation.
- Harris E.D. (2001). Clinical features of rheumatoid arthritis. In Ruddy S., Harris E.D., Sledge C.B. (Eds.), *Textbook of rheumatology* (6th ed., pp. 967–1000). Philadelphia: W.B. Saunders.

- Green M., Marzo-Ortega H., McGonagle D., et al. (1999). Persistence of mild, early inflammatory arthritis: The importance of disease duration, rheumatoid factor, and the shared epitope. *Arthritis and Rheumatism* 42, 2184–2188.
- 21. Klippel J.H., Dieppe P.A. (1998). *Rheumatology* (2nd ed.). St. Louis: Mosby.
- 22. Irvine S., Munro R., Porter D. (1999). Early referral, diagnosis, and treatment of rheumatoid arthritis: Evidence for changing medical practice. *Annals of Rheumatic Disease* 58, 510–513.
- Arnette F.C., Edsworthy S.M., Bloch D.A., et al. (1988). The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis and Rheumatism* 31, 315–324.
- 24. Weaver A.L. (1999). The evolving therapy of rheumatoid arthritis. In *Today's challenges of rheumatoid arthritis patient management* (pp. 7–15). *University of Nebraska Medical Center CME proceedings*. Omaha, NE: Impact Unlimited, Inc.
- 25. Jones R. E., Moreland L.W. (1999). Tumor necrosis factor inhibitors for rheumatoid arthritis. *Bulletin on the Rheumatic Diseases* 48 (3).
- Mitchel R.N., Kumar V. (2003). Diseases of immunity. In Kumar V., Cotran R.S., Robbins S.L. (Eds.), *Basic pathology* (7th ed., pp. 130–143). Philadelphia: W.B. Saunders.
- McCowan C.B. (1998). Systemic lupus erythematosus. Journal of the American Academy of Nurse Practitioners 10, 225–231.
- Pigg J.S., Bancroft D.A. (2000). Management of patients with rheumatic diseases. In Smeltzer S.C., Bare B.C. (Eds.), Brunner and Suddarth's textbook of medical-surgical nursing (9th ed., pp. 1405–1433). Philadelphia: Lippincott Williams & Wilkins.
- Steen V.D., Medsger T.A. (2000). Severe organ involvement in systemic sclerosis with diffuse scleroderma. *Arthritis and Rheumatism* 43, 2437–2444.
- Hellman D.B., Stone J.H. (2002). Arthritis and musculoskeletal disorders. In Tierney L.M., McPhee S.J., Papadakis M.A. (Eds.), *Current medical diagnosis and treatment* (41st ed., pp. 833–889). New York. Lange Medical Books/McGraw-Hill.
- Schumacher H.R. (1998). Reactive arthritis. Rheumatic Disease Clinics of North America 24, 261–273.
- 32. Sigal L.H. (2001). Update on reactive arthritis. Bulletin on the Rheumatic Diseases 50 (4).
- 33. Felson D.T. (1998). Preventing knee and hip osteoarthritis. *Bulletin on the Rheumatic Diseases* 47 (7).
- Puppione A.A. (1999). Management strategies for older adults with osteoarthritis: How to promote and maintain function. *Journal of the American Academy of Nurse Practitioners* 11, 167–171.
- 35. Agudelo C.A., Wise K.M. (2000). Crystal-associated arthritis in the elderly. *Rheumatic Disease Clinics of North America* 26, 527–546.
- Segal J.B., Albert D. (1999). Diagnosis of crystal-induced arthritis by synovial fluid examination for crystals: Lessons from an imperfect test. *Arthritis Care and Research* 12, 376–380.
- Lomater C., Gerloni V., Gattinara M., et al. (2000). Systemic onset juvenile idiopathic arthritis: A retrospective study of 80 consecutive patients followed for 10 years. *Journal of Rheumatology* 27, 491–496.
- Yaziu Y., Paget S.A. (2000). Elderly-onset rheumatoid arthritis. Rheumatic Disease Clinics of North America 26, 517–526.

14 Alt

CHAPTER

Alterations in Skin Function and Integrity

Structure of the Skin Layers of the Skin **Epidermis** Dermis Subcutaneous Tissue Innervation and Blood Supply **Manifestations of Skin Disorders** Lesions, Rashes, and Vascular Disorders **Pruritus** Variations in Black Skin Skin Damage Caused by Ultraviolet Radiation **Ultraviolet Rays** Sunburn **Drug-Induced Photosensitivity** Sunscreens and Other Protective Measures Primary Disorders of the Skin **Infectious Processes Fungal Infections Bacterial Infections Viral Infections** Acne and Rosacea **Acne Vulgaris** Acne Conglobata Rosacea Atopic Eczema and Nummular Eczema Urticaria **Drug-Induced Skin Eruptions Papulosquamous Dermatoses Psoriasis Pityriasis Rosea Lichen Planus** Lichen Simplex Chronicus

Nevi and Skin Cancers Nevi Skin Cancer Malignant Melanoma Basal Cell Carcinoma Squamous Cell Carcinoma Age-Related Skin Manifestations Skin Disorders in Infants and Children Skin Disorders in Infants Skin Manifestations of Common Infectious Diseases Skin Manifestations and Disorders in the Elderly Normal Age-Related Changes Common Skin Lesions

he skin is one of the most versatile organs of the body. It forms a protective surface that prevents many harmful substances, including microorganisms, from entering the body; it retards the loss of body fluids; it helps regulate body temperature; and it houses sensory receptors, contains immune system cells, synthesizes chemicals, and excretes small quantities of waste products. The skin is also unique in that the manifestations of disease or injury are immediately observable. It also provides a sensitive reflection of many internal disorders. A number of systemic diseases are manifested by skin disorders (*e.g.*, rash associated with systemic lupus erythematosus and jaundice caused by liver disease).

STRUCTURE OF THE SKIN

The skin is one of the largest organs of the body and accounts for about 7% of total body weight in the average adult. The skin, which is also called the *integument*, which simply means "covering," serves as an interface between the internal and external environments.^{1,2} The skin is tough, yet pliable, allowing it to endure the effects of any number of external agents. Although the skin may become bruised, lacerated, burned, or infected, it has remarkable properties that allow for a continuous cycle of healing, shedding, and cell regeneration.

Layers of the Skin

The skin is composed of three rather distinct layers: the epidermis, dermis, and subcutaneous tissues. The layer of subcutaneous tissue binds the dermis to the underlying tissues of the body (Fig. 44-1). The basal lamina (basement membrane) is a layer of intercellular and extracellular matrices that serves as an interface between the dermis and the epidermis. It provides for adhesion of the dermis to the epidermis and serves as a selective filter for molecules moving between the two layers. It is also a major site of immunoglobulin and complement deposition in skin disease. The basal lamina is involved in skin disorders that cause bullae or blister formation.

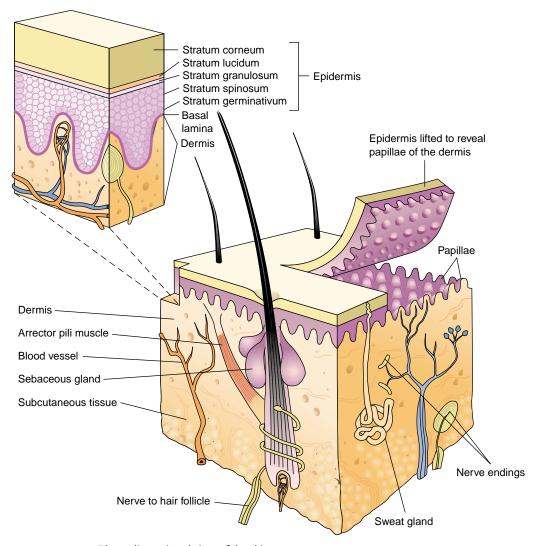
Epidermis

The epidermis contains the keratinocytes, which produce a fibrous protein called *keratin*, which is essential to the protective function of skin. Because of its high keratin content, the outer layer of the epidermis has a rough, horny texture. In addition to the keratinocytes, the epidermis has three other types of cells that arise from its basal layer: melanocytes that pro-

duce a pigment called *melanin*, which is responsible for skin color; Merkel's cells that provide sensory information; and Langerhans' cells that link the epidermis to the immune system. The epidermis contains openings for two types of glands: sweat glands, which produce watery secretions, and sebaceous glands, which produce an oily secretion called *sebum*.

Keratinocytes. The keratinocytes, or keratin-forming cells, are the major cells of the epidermis. They develop into five distinct layers, or strata, as they divide and mature: the stratum germinativum, the stratum spinosum, the stratum granulosum, the stratum lucidum, and the stratum corneum.

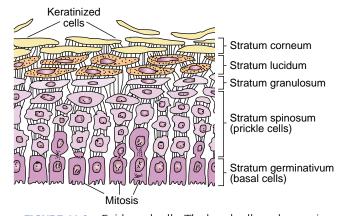
The deepest layer, the *stratum germinativum* or *stratum basale*, consists of a single layer of basal cells that are attached to the basal lamina. The basal cells are the only epidermal cells that are mitotically active. All cells of the epidermis arise from this layer. As new cells form in the basal layer, the older cells change shape and are pushed outward (Fig. 44-2). As these cells near the surface, they die and their cytoplasm is converted to keratin. It normally takes 3 to 4 weeks for the epidermis to replicate itself. This cell turnover is greatly accelerated in diseases such as psoriasis.



SKIN STRUCTURES AND THE MANIFESTATIONS OF SKIN DISORDERS

- The skin has two layers, an outer epidermis and an inner dermis, separated by a basement membrane, all of which can contribute to the development and symptomatology of skin disorders.
- The epidermis, which is avascular, is composed of four to five layers of stratified squamous keratinized epithelial cells that are formed in the deepest layer of the epidermis and migrate to the skin surface to replace cells that are lost during normal skin shedding. The papulosquamous dermatoses, such as psoriasis, involve increased epidermal cell turnover with marked thickening of the epidermis. The avascular layers of the epidermis also serve as a site for superficial fungal infections.
- The basement membrane is a thin adhesive layer that cements the epidermis to the dermis. This is the layer involved in blister formation.
- The dermis is a connective tissue layer that separates the epidermis from the underlying subcutaneous fat layer. It contains the blood vessels that produce hyperemic skin responses and nerve fibers that are the source of pain, discomfort, and itch associated with skin disorders.
- The Langerhans' cells of the epidermis, which bind antigen; the dermal dendrocytes of the dermis, which have phagocytic properties; and immune cells (T cells and mast cells), which are found in the dermis, contribute to the antigen-antibody responses affecting the skin.
- Sebaceous glands in the skin produce an oily secretion called sebum, which is secreted into the hair follicles on the skin. The hair follicle and sebaceous gland form the pilosebaceous unit, which is the site of acne lesions.

The second layer, the *stratum spinosum*, is formed as cells from the basal cell layer move outward toward the skin surface. The stratum spinosum is two to four layers thick. The cells of this layer are commonly referred to as *prickle cells* because they develop a spiny appearance as their cell borders interact. The third layer, the *stratum granulosum* is only a few cells thick; it is composed of flatter cells containing protein granules called *keratohyalin granules*. The *stratum lucidum*, which lies just superficial to the stratum granulosum, is a thin, transparent layer mostly confined to the palms of the hands and soles of the feet. It consists of transitional cells that retain some of the functions of living skin cells from the layers below but otherwise resemble the cells of the stratum corneum.



■ FIGURE 44-2 ■ Epidermal cells. The basal cells undergo mitosis, producing keratinocytes that change their size and shape as they move upward, replacing cells that are lost during normal cell shedding.

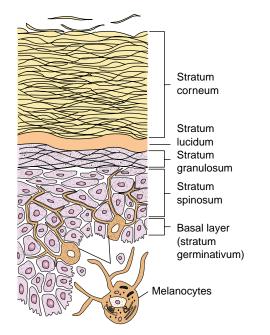
The top or surface layer of the epidermis is the *stratum corneum*. It is made up of stratified layers of dead keratinized cells that are constantly shedding. The stratum corneum contains the most cell layers and the largest cells of any zone of the epidermis. It ranges from 15 layers thick in areas such as the face to 25 layers or more on the arm. Specialized areas, such as the palms of the hands or soles of the feet, have 100 or more layers.

Melanocytes. The melanocytes are pigment-synthesizing cells that are located at or in the basal layer. They function to produce pigment granules called *melanin*, the black or brown substance that gives skin its color. The melanocytes have long, cytoplasm-filled extensions that extend between the keratinocytes. Although the melanocytes remain in the basal layer, melanin is transferred to the keratinocytes through these extensions (Fig. 44-3). Each melanocyte is capable of supplying several keratinocytes with melanin. The primary function of melanin is to protect the skin from harmful ultraviolet sun rays. Exposure to the sun's ultraviolet rays increases the production of melanin, causing tanning to occur. The amount of melanin in the keratinocytes determines a person's skin color. All people have relatively few or no melanocytes in the epidermis of the palms of the hands or soles of the feet.

The ability to synthesize melanin depends on the ability of the melanocytes to produce an enzyme called *tyrosinase*, which converts the amino acid tyrosine to a precursor of melanin. A genetic lack of this enzyme results in a clinical condition called *albinism*. Persons with this disorder lack pigmentation in the skin, hair, and iris of the eye.

Langerhans' Cells. Langerhans' cells are star-shaped macrophages of the immune system (see Chapter 8). They arise from the bone marrow and migrate to the epidermis, where they help activate the immune system. Their slender dendritic, or threadlike processes, extend through the keratinocytes in the epidermis, forming a more or less continuous network recognizing foreign antigens (Fig. 44-4). The Langerhans' cells bind antigen to their surface, process it, and then, bearing the processed antigen, migrate from the epidermis into lymphatic vessels and then into regional lymph nodes. They are the only

838 Unit Eleven: Alterations in the Skeletal and Integumentary Systems

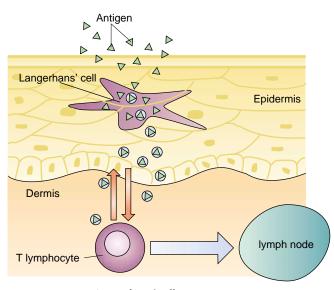


■ FIGURE 44-3 ■ Melanocytes. The melanocytes, which are located in the basal layer of the skin, produce melanin pigment granules that give skin its color. The melanocytes have threadlike cytoplasmic-filled extensions that are used in passing the pigment granules to the keratinocytes.

immune cells in the skin known to be capable of antigenpresentation and therefore may be responsible for allergic reactions affecting the skin.

Dermis

The dermis is the connective tissue layer that separates the epidermis from the subcutaneous fat layer. It supports the epidermis and serves as its primary source of nutrition. The two layers of the dermis, the papillary dermis and the reticular dermis,



[■] FIGURE 44-4 ■ Langerhans' cells.

are composed of cells, fibers, ground substances, nerves, and blood vessels. The hair and glandular structures are embedded in this layer and continue through the epidermis.

Papillary Dermis. The papillary dermis is a thin, superficial layer that lies adjacent to the epidermis. It consists of collagen fibers and ground substance. This layer is densely covered with conical projections called *dermal papillae* (see Fig. 44-1). The basal cells of the epidermis project into the papillary dermis, forming *rete ridges*. It is believed that the dense structure of the dermal papillae serves to minimize the separation of the dermis and the epidermis. Dermal papillae contain capillary venules that nourish the epidermal layers of the skin. Lymph vessels and nerve tissue also are found in this layer.

Reticular Dermis. The reticular dermis is the thicker area of the dermis and forms the bulk of the dermal layer. The reticular dermis is characterized by a complex meshwork of three-dimensional collagen bundles interconnected with large elastic fibers and ground substance, a viscid gel that is rich in mucopolysaccharides. The collagen fibers are oriented parallel to the body's surface in any given area. Collagen bundles may be organized lengthwise, as on the abdomen, or in round clusters, as in the heel. The direction of surgical incisions is often determined by this organizational pattern.

The reticular dermis also contains dendritic cells with threadlike projections, called *dermal dendrocytes*. Dermal dendrocytes, which have phagocytic properties, are believed to possess antigen-presenting functions and play an important part in the immune function of the skin. Immune cells found in the dermis include macrophages, T cells, mast cells, and fibroblasts. The major type of T-cell-mediated immune response in the skin is delayed-type hypersensitivity (see Chapter 10). The mast cells play a prominent role in IgE-mediated hypersensitivity responses.

Subcutaneous Tissue

The subcutaneous tissue layer consists primarily of loose connective and fatty tissues that lend support to the vascular and neural structures supplying the outer layers of the skin. There is controversy about whether the subcutaneous tissue should be considered an actual layer of the skin. Because the eccrine glands and deep hair follicles extend to this layer and several skin diseases involve the subcutaneous tissue, the subcutaneous tissue may be considered part of the skin.

Innervation and Blood Supply

The innervation of the skin is complex. The skin, with its accessory structures, serves as an organ for receiving sensory information from the environment. The dermis is well supplied with sensory neurons as well as nerves that supply the blood vessels, sweat glands, and arrector pili muscles. The receptors for touch, pressure, heat, cold, and pain are widely distributed in the dermis. The papillary layer of the dermis is supplied with free nerve endings that serve as nociceptors (*i.e.*, pain receptors) and thermoreceptors. The dermis also contains encapsulated pressure-sensitive receptors that detect pressure and touch.

The arterial vessels that nourish the skin form two plexuses (*i.e.*, collection of blood vessels), one located between the papillary and reticular layers of the dermis and the other be-

tween the dermis and the subcutaneous tissue layer. Capillary flow that arises from vessels in this plexus extends up and nourishes the overlaying epidermis by diffusion. Blood leaves the skin by way of small veins that accompany the subcutaneous vessels. The lymphatic system of the skin, which aids in combating certain skin infections, also is limited to the dermis.

Most of the skin's blood vessels are under sympathetic nervous system control. The sweat glands are innervated by cholinergic fibers but controlled by the sympathetic nervous system. Likewise, the sympathetic nervous system controls the arrector pili (pilomotor) muscles that cause elevation of hairs on the skin. Contraction of these muscles tends to cause the skin to dimple, producing "goose bumps."

In summary, the skin, which forms the major barrier between the internal organs and the external environment, is primarily an organ of protection. The skin is composed of three layers, the epidermis, dermis, and subcutaneous tissues. The skin also houses a variety of appendages, including hair, nails, and sebaceous and sweat glands. The epidermis, the outermost layer of the skin, contains five layers, or strata. The major cells of the epidermis are the keratinocytes, melanocytes, Langerhans' cells, and Merkel's cells. The stratum germinativum, or basal layer, is the source of the cells in all five layers of the epidermis. The keratinocytes, which are the major cells of the epidermis, are transformed from viable keratinocytes to dead keratin as they move from the innermost layer of the epidermis (i.e., stratum germinativum) to the outermost layer (i.e., stratum corneum). The melanocytes are pigment-synthesizing cells that give skin its color.

The dermis provides the epidermis with support and nutrition and is the source of blood vessels, nerves, and skin appendages. Sensory receptors for touch, pressure, heat, cold, and pain are widely distributed in the dermis. Both the epidermis and dermis participate in the immune functions of the skin; the Langerhans' cells of the epidermis process foreign antigens for presentation to immune cells (T cells, macrophages, mast cells, and fibroblasts) in the dermis. The subcutaneous tissue layer consists primarily of fat and connective tissues that lend support to the vascular and neural structures supplying the outer layers of the skin.

MANIFESTATIONS OF SKIN DISORDERS

Skin disorders are manifested by a variety of primary lesions and rashes (Fig. 44-5). Commonly, secondary lesions result from overtreatment, scratching, and infection that accompanies the primary skin disorders.

Lesions, Rashes, and Vascular Disorders

Rashes are temporary eruptions of the skin, such as those associated with childhood diseases, heat, diaper irritation, or drug-induced reactions. The term *lesion* refers to a traumatic or pathologic loss of normal tissue continuity, structure, or function. The components of a rash sometimes are referred to as *lesions*. Rashes and lesions may range in size from a fraction of a millimeter (*e.g.*, the pinpoint spots of petechiae) to many

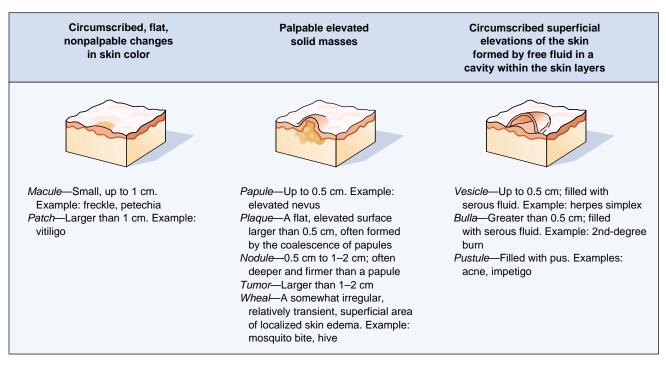


FIGURE 44-5 Primary lesions may arise from previously normal skin. Authorities vary somewhat in their definitions of skin lesions by size. Dimensions given should be considered approximate. (Bates B.B. [1995]. *A guide to physical examination and history taking* [6th ed.]. Philadelphia: J.B. Lippincott)

centimeters (*e.g.*, decubitus ulcer, or pressure sore). They may be blanched (white), erythematous (reddened), hemorrhagic or purpuric (containing blood), or pigmented. Repeated rubbing and scratching can lead to lichenification (thickened and roughened skin characterized by prominent skin markings caused by repeated scratching or rubbing) or excoriation (lesion caused by breakage of the epidermis, producing a raw linear area). Skin lesions may occur as primary lesions arising in previously normal skin, or they may develop as secondary lesions resulting from other disease conditions.

A *blister* is a vesicle or fluid-filled papule. Blisters of mechanical origin form from the friction caused by repeated rubbing on a single area of the skin. Friction blisters most commonly occur on the palmar and plantar surfaces of the hands and feet where the skin is thick enough to form a bleb. Blisters also develop from first-degree and second-degree partial-thickness burns. Histologically, there is degeneration of epidermal cells and a disruption of intercellular junctions that causes the layers of the skin to separate. As a result, fluid accumulates, and a noticeable bleb forms on the skin surface.

A *callus* is a hyperkeratotic plaque of skin caused by chronic pressure or friction. It represents a hyperplasia of the dead keratinized cells that make up the cornified or horny layer of the skin. Increased cohesion between cells results in hyperkeratosis and decreased skin shedding. *Corns* are small, wellcircumscribed, conical keratinous thickenings of the skin. They usually appear on the toes from rubbing or ill-fitting shoes. The actual corn may be either hard with a central horny core or soft, as commonly seen between the toes. They may appear on the hands as an occupational hazard. Corns on the feet often are painful, whereas corns on the hands may be asymptomatic.

Telangiectasis are dilated superficial blood vessels, capillaries, or terminal arteries that appear either red or bluish. They can appear by themselves or as a part of other skin disorders such as rosacea or basal cell carcinoma.

Pruritus

Pruritus, or the sensation of itch, is a symptom common to many skin disorders. Generalized itching in the absence of a primary skin disease may be symptomatic of other organ disorders, such as chronic renal disease, diabetes, or biliary disease. Warmth, touch, and vibration also can act locally to trigger the itch phenomenon.

Itch is mediated by cutaneous receptors. Substances such as histamine, bradykinin, substance P, and bile salts act locally to stimulate the itch receptors. Prostaglandins are modulators of the itch response, lowering the threshold for other mediators. One type of itch, sometimes referred to as *central itch*, is perceived as occurring on the skin, but originates in the central nervous system (CNS).³ For example, the pain reliever morphine promotes itch by acting on central opioid receptors in the CNS.

Scratching, the well-known response to itch, is a neurologic reflex that to varying degrees can be controlled by the individual. Although scratching may temporarily relieve itch, many types of itch are not easily localized and are not relieved by scratching. In many people, excoriations and thickened papular areas develop at the site of repeated scratching or rubbing.

Variations in Black Skin

Some skin disorders common to African Americans are not commonly found in European Americans. Similarly, some skin disorders, such as skin cancers, affect light-skinned persons more commonly than dark-skinned persons. Because of these differences, serious skin disorders may be overlooked, and normal variations in darker skin may be mistaken for anomalies. Skin color is determined by the melanin produced by the melanocytes. Although the number of melanosomes in dark and white skin are the same, black skin produces more melanin and produces it faster than does white skin. Because of their skin color, blacks are better protected against skin cancer and the premature wrinkling and aging of the skin that occurs with sun exposure.

Some conditions common in people with black skin are too much or too little color. Areas of the skin may darken after injury, such as a cut or scrape, or after disease conditions, such as acne.⁴ These darkened areas may take many months or years to fade. Dry or "ashy" skin also can be a problem for people with black skin. It often is uncomfortable, and it also is easily noticed because it gives the skin an ashen, or grayish, appearance. Although using a moisturizer may help relieve the discomfort, it may cause a worsening of acne in predisposed persons.

Normal variations in skin structure and skin tones often make evaluation of dark skin difficult (Table 44-1). The darker pigmentation can make skin pallor, cyanosis, and erythema more difficult to observe. Therefore, verbal histories must be relied on to assess skin changes. The verbal history should include clients' descriptions of their normal skin tone. Changes in skin color, in particular hypopigmentation and hyperpigmentation, often accompany ethnic skin disorders and are very important signs to observe for when diagnosing skin conditions.

TABLE 44-1Common NormalVariations in Dark Skin

Variation	Appearance
Futcher (Voigt's) line	Demarcation between darkly pigmented and lightly pigmented skin in upper arm; follows spinal nerve distribution; com- mon in black and Japanese populations
Midline hypo- pigmentation	Line or band of hypopigmentation over the sternum, dark or faint, lessens with age; common in Latin American and black populations
Nail pigmentation	Linear dark bands down nails or diffuse nail pigmentation, brown, blue or blue-black
Oral pigmentation	Blue to blue-gray pigmentation of oral mucosa; gingivae also affected
Palmar changes	Hyperpigmented creases, small hyperkera- totic papules, and tiny pits in creases
Plantar changes	Hyperpigmented macules, can be multiple with patchy distribution, irregular borders, and variance in color

(Developed from information in Rosen T., Martin S. [1981]. Atlas of black dermatology. Boston: Little, Brown)

In summary, skin lesions and rashes are the most common manifestations of skin disorders. Rashes are temporary skin eruptions. Lesions result from traumatic or pathologic loss of the normal continuity, structure, or function of the skin. Lesions may be vascular in origin; they may occur as primary lesions in previously normal skin; or they may develop as secondary lesions resulting from primary lesions. Blisters, calluses, and corns result from rubbing, pressure, and frictional forces applied to the skin.

Pruritus and dry skin are symptoms common to many skin disorders. Scratching because of pruritus can lead to excoriation, infection, and other complications.

Normal variations in black skin often make evaluation difficult and result in some disorders being overlooked. Changes in color, especially hypopigmentation or hyperpigmentation, often accompany the skin disorders of dark-skinned people.

SKIN DAMAGE CAUSED BY ULTRAVIOLET RADIATION

The skin is the protective shield against harmful ultraviolet rays from the sun. Skin cancers and other skin disorders such as early wrinkling and aging of the skin have been attributed to the damaging effects of sunlight.

Ultraviolet Rays

The earth's sunlight is measured in wavelengths ranging from approximately 290 nm in the ultraviolet region up to approximately 2500 nm in the infrared region. Ultraviolet (UV) radiation is divided into three types: UVC, UVB, and UVA. UVC rays are short (100 to 289 nm) and do not pass through the earth's atmosphere. However, they can be produced artificially and are damaging to the eyes. UVB rays are 290 to 320 nm. These are the rays that are primarily responsible for nearly all the skin effects of sunlight. They are more commonly referred to as sunburn rays. UVA rays are 321 to 400 nm. These rays, which can pass through window glass, are more commonly referred to as suntanning rays. In general, it takes approximately 1000 times more UVA to match the untoward effects of UVB. Nonetheless, UVA contributes to many skin alterations. Artificial sources of UVA, such as tanning salons, may produce the same effects as UVB.5

With UV exposure, skin cells release vasoactive and injurious chemicals, resulting in vasodilation and sunburn. Melanin in the stratum corneum absorbs UV radiation as a means of preventing destruction of the lower skin layers. The skin responds to UV exposure with an increase in melanin content. Components of the immune system in the skin, especially Langerhans' cells, also respond to UV radiation. The number of immune cells is decreased, and their activity is lessened by UV exposure.⁶ It is thought that the immune cells are important in removing sun-damaged cells with malignant potential.

Sunburn

Sunburn is caused by excessive exposure of the epidermal and dermal layers of the skin to UV radiation, resulting in an erythematous inflammatory reaction. Sunburn ranges from mild to severe. A mild sunburn consists of various degrees of skin redness. Inflammation, vesicle eruption, weakness, chills, fever, malaise, and pain accompany more severe forms of sunburn. Scaling and peeling follow any overexposure to sunlight. Black skin also burns and may appear grayish or gray-black.

Severe sunburns are treated with wet Burow's solution soaks and topical creams and lotions to limit inflammation and pain.⁷ Extensive second- and third-degree burns may require hospitalization and specialized burn care techniques.

Drug-Induced Photosensitivity

Some drugs are classified as photosensitive drugs because they produce an exaggerated response to ultraviolet light. Examples include some of the anti-infective agents (sulfonamides, tetra-cyclines, nalidixic acid), antihistamines (cyproheptadine, diphenhydramine), antipsychotic agents (phenothiazines, haloperidol), diuretics (thiazides, acetazolamide, amiloride), and nonsteroidal anti-inflammatory drugs (phenylbutazone, ketoprofen, naproxen).⁸ Severe sunburn can result when persons taking these drugs are exposed to sunlight.

Drug-induced photosensitivity, such as UVA photosensitivity induced by the psoralens, may be used in treating skin conditions, such as psoriasis, that respond well to ultraviolet radiation exposure. Because an increased incidence of cancerous lesions has been reported in people who have been treated with these agents, their use requires caution and careful surveillance.

Sunscreens and Other Protective Measures

The ultraviolet rays of sunlight or other sources can be either completely or partially blocked from the skin surface by sunscreens. The U.S. Food and Drug Administration (FDA) requires a *sun protection factor* (SPF) rating on all commercial preparations based on their ability to obstruct ultraviolet radiation absorption. The ratings usually are on a scale of 1 to 30+; higher ratings block more sunlight.⁹ Products with a higher SPF screen out more UVB rays, which are responsible for acute sun damage.

There are two primary types of sunscreens available on the market—chemical (soluble) agents and physical (insoluble) agents.⁹ Chemical agents (*e.g.*, para-aminobenzoic acid [PABA]) protect the skin from absorbing sunlight, and physical agents (*e.g.*, micronized titanium dioxide and microfine zinc) work by reflecting sunlight.

Other protective measures include knowledge about sunlight and how to protect the skin. Shielding the skin with protective clothing and hats or head coverings helps decrease ultraviolet radiation exposure.

In summary, there has been an alarming increase in skin cancers since the early 1980s, and repeated exposure to the ultraviolet rays of the sun has been implicated as its principal cause. Solar and artificial sources of radiation contribute to the amount of radiation to which human beings are exposed. Sunburn, which is caused by excessive exposure to ultraviolet radiation, is an erythematous inflammatory reaction, ranging from mild to severe. Photosensitive drugs can also produce an exaggerated response to ultraviolet light when they are taken in combination with sun exposure. Sunscreens are protective agents that work by either reflecting sunlight or preventing its absorption.

PRIMARY DISORDERS OF THE SKIN

Primary skin disorders are those originating in the skin. They include infectious processes, acne and rosacea, papulosquamous dermatoses, allergic disorders and drug reactions, and arthropod infestations. Although most of these disorders are not life threatening, they can affect the quality of life.

Infectious Processes

The skin is subject to invasion by a number of microorganisms, including fungi, bacteria, and viruses. Normally, the skin flora, sebum, immune responses, and other protective mechanisms guard the skin against infection. Depending on the virulence of the infecting agent and the competence of the host's resistance, infections may result.

Fungal Infections

Fungi are free-living, saprophytic, plantlike organisms, certain strains of which are considered part of the normal skin flora.¹⁰ Fungal or mycotic infections of the skin are traditionally classified as superficial or deep. The superficial mycoses, more commonly known as *tinea* or *ringworm*, invade only the superficial keratinized tissue (skin, hair, and nails). Deep fungal infections involve the epidermis, dermis, and subcutis. Infections that typically are superficial may exhibit deep involvement in immunosuppressed individuals.

The fungi that cause superficial mycoses live on the dead keratinized cells of the epidermis. They emit an enzyme that enables them to digest keratin, which results in superficial skin scaling, nail disintegration, or hair breakage, depending on the location of the infection. Deeper reactions involving vesicles, erythema, and infiltration are caused by the inflammation that results from exotoxins liberated by the fungus. Fungi also are capable of producing an allergic or immune response.

Superficial Fungal Infections. Superficial fungal infections affect various parts of the body, with the lesions varying according to site and fungal species. Tinea can affect the body (tinea corporis), face and neck (tinea faciei), scalp (tinea capitis), hands (tinea manus), feet (tinea pedis), or nails (tinea unguium).

Tinea corporis (ringworm of the body) can be caused by any of the fungi. Although tinea corporis affects all ages, children seem most prone to infection. Transmission is most commonly from kittens, puppies, and other children who have infections. The lesions vary, depending on the fungal agent. The most common types of lesions are oval or circular patches on exposed skin surfaces and the trunk, back, or buttocks (Fig. 44-6). Less common are foot and groin infections. The lesion begins as a red papule and enlarges, often with a central clearing. Patches have raised red borders consisting of vesicles, papules, or pustules. The borders are sharply defined, but lesions may coalesce. Pruritus, a mild burning sensation, and erythema frequently accompany the skin lesion.

Tinea capitis (ringworm of the scalp) occurs in two forms: primary (noninflammatory) and secondary (inflammatory). Depending on the invading fungus, the lesions of the *noninflammatory type* can vary from grayish, round, hairless patches to balding spots or black dots on the head. The lesions vary in size and are most commonly seen on the back of the



■ FIGURE 44-6 ■ Tinea of the body caused by *Microsporum canis*. (Sauer G.C., Hall J.C. [1996]. *Manual of skin diseases* [7th ed.]. Philadelphia: Lippincott-Raven)

head (Fig. 44-7). Mild erythema, crust, or scale may be present. The individual usually is asymptomatic, although pruritus may exist. Children between 3 and 14 years of age are primarily affected, although an increasing number of adults are receiving diagnoses of the infection. The *inflammatory type* of tinea capitis has a rapid onset, and lesions usually are localized to one area. The initial lesion consists of a pustular, scaly, round patch with broken hairs. A secondary bacterial infection is common and may lead to a painful, circumscribed, boggy, and indurated lesion called a *kerion*. The highest incidence is among children and farmers who work with infected animals.

Tinea pedis (athlete's foot, or ringworm of the feet) is a common dermatosis primarily affecting the spaces between the toes, the soles of the feet, or the sides of the feet (Fig. 44-8). The lesions vary from a mildly scaling lesion to a painful, exudative, erosive, inflamed lesion with fissuring. Lesions often are accompanied by pruritus, pain, and foul odor. Some persons are prone to chronic tinea pedis, whereas others have a



■ FIGURE 44-7 ■ Tinea of the scalp caused by Microsporum audouinii. (Sauer G.C., Hall J.C. [1996]. Manual of skin diseases [7th ed.]. Philadelphia: Lippincott-Raven)



■ FIGURE 44-8 ■ Chronic tinea of sole of the foot caused by *Tri-chophyton rubrum*. (Schering Corp.) (Sauer G.C., Hall J.C. [1996]. *Manual of skin diseases* [7th ed.]. Philadelphia: Lippincott-Raven)

milder form that is exacerbated during hot weather or when the feet are exposed to moisture or occlusive shoes. *Tinea manus* (ringworm of the hands) usually is a secondary infection with tinea pedis as the primary infection. In contrast to other skin disorders such as contact dermatitis and psoriasis, which affect both hands, tinea manus usually occurs only on one hand. The characteristic lesion is a blister on the palm or finger surrounded by erythema. Chronic lesions are scaly and dry. Cracking and fissuring may occur. The lesions may spread to the plantar surfaces of the hand. If chronic, tinea manus may lead to tinea of the fingernails.

Tinea unguium is a dermatophyte infection (onychomycosis) of the nails. Toenails are involved more commonly than fingernails. Toenail infection is common in persons prone to chronic infections of tinea pedis. Often, the infection in the toenails becomes a ready source for future infections of the foot. It may begin from a crushing injury to a toenail or from the spread of tinea pedis. The infection often begins at the tip of the nail, where the fungus digests the nail keratin. Initially, the nail appears opaque, white, or silvery (Fig. 44-9). The nail



■ **FIGURE 44-9** Tinea of the fingernail caused by *Trichophyton rubrum*. (Duke Laboratories, Inc.) (Sauer G.C., Hall J.C. [1996]. *Manual of skin diseases* [7th ed.]. Philadelphia: Lippincott-Raven)

then turns yellow or brown. The condition often remains unchanged for years. During this time it may involve only one or two nails and may produce little or no discomfort. Gradually, the nail thickens and cracks as the infection spreads. Permanent discoloration and distortion result as the nail separates from the underlying epidermis. Less common forms of tinea unguium are superficial white onychomycosis, in which areas of the nails become powdery white and erode, and proximal subungual onychomycosis (PSO), in which there is rapid invasion of the nail, leaving it white with no additional thickening of the nail. Although it is one of the less common forms of tinea unguium, PSO has increased among people with acquired immunodeficiency syndrome (AIDS).

Chapter 44: Alterations in Skin Function and Integrity

Diagnosis of superficial fungal infections is primarily done by microscopic examination of skin scrapings for fungal spores, the reproducing bodies of fungi. Potassium hydroxide (KOH) preparations are used to prepare slides of skin scrapings.¹¹ KOH disintegrates human tissue and leaves behind the threadlike filaments, called *hyphae*, that grow from the fungal spores. Cultures also may be done.

Superficial fungal infections may be treated with topical or systemic antifungal agents. Tinea treatment usually follows diagnosis confirmed by KOH preparation or culture, particularly if a systemic agent is to be used. Topical agents, both prescription and over-the-counter preparations, are commonly used in the treatment of tinea infections; however, outcome success often is limited because of the lengthy duration of treatment, poor compliance, and high rates of relapse at specific body sites.

The oral systemic antifungal agents include griseofulvin, the azoles, and the allylamines.¹¹ Griseofulvin is a fungicidal agent derived from a species of Penicillium that is used only in the treatment of dermatophytoses. It acts by binding to the keratin of newly forming skin, protecting the skin from new infection. Because its action is to prevent new infection, it must be administered for 2 to 6 weeks to allow for skin replacement. The azoles (e.g., ketoconazole, itraconazole, and fluconazole) are a group of synthetic antifungal drugs that act by inhibiting the fungal enzymes needed for the synthesis of ergosterol, which is an essential part of fungal cell membranes. Terbinafine, a synthetic allylamine, acts by interrupting ergosterol synthesis, causing the accumulation of a metabolite that is toxic to the fungus. In contrast to griseofulvin, the synthetic agents are fungicidal (i.e., kill the fungus) and thus are more effective in shorter treatment periods.¹⁰ Some of the oral agents can produce serious side effects, such as hepatic toxicity, or interact adversely with other medications being taken. A number of the synthetic fungicides (e.g., ketoconazole, miconazole, clotrimazole, and terbinafine) are available as topical preparations and produce less severe side effects. Topical corticosteroids may be used in conjunction with antifungal agents to relieve itching and erythema secondary to inflammation.

Candidal Infections. Candidiasis (moniliasis) is a fungal infection caused by *Candida albicans.* This yeastlike fungus is a normal inhabitant of the gastrointestinal tract, mouth, and vagina (see Chapter 35). The skin problems result from the release of irritating toxins on the skin surface. Some persons are predisposed to candidal infections by conditions such as diabetes mellitus, antibiotic therapy, pregnancy, use of birth control pills, poor nutrition, and immunosuppressive diseases.

Oral candidiasis may be the first sign of infection with human immunodeficiency virus (HIV).

C. albicans thrives in warm, moist intertriginous areas of the body. The rash is red with well-defined borders. Patches erode the epidermis, and there is scaling. Mild to severe itching and burning often accompany the infection. Severe forms of infection may involve pustules or vesiculopustules. In addition to microscopic analysis, a candidal infection often can be differentiated from a tinea infection by the presence of satellite lesions. These satellite lesions are maculopapular and are found outside the clearly demarcated borders of the candidal infection. Satellite lesions often are diagnostic of diaper rash complicated by *Candida*. The appearance of candidal infections varies according to the site (Table 44-2).

Diagnosis usually is based on microscopic examination of skin or mucous membrane scrapings placed in KOH solution. Depending on the site of infection and extent of involvement, topical and oral antifungal agents may be used in treatment.

Bacterial Infections

Bacteria are considered normal flora of the skin. Most bacteria are not pathogenic, but when pathogenic bacteria invade the skin, superficial or systemic infections may develop. Bacterial skin infections are commonly classified as primary or secondary infections. Primary infections are superficial skin infections such as impetigo or ecthyma. Secondary infections consist of deeper cutaneous infections, such as infected ulcers. Diagnosis usually is based on cultures taken from the infected site. Treatment measures include antibiotic therapy and measures to promote comfort and prevent the spread of infection.

Impetigo. Impetigo is a common superficial bacterial infection caused by *staphylococci* or *group* A β -*hemolytic streptococci* (GABHS), or both. It is common among infants and young children, although older children and adults occasionally con-

TABLE 44-2 Candidal Infections: Locations and Appearance of Lesions

Location	Appearance
Breasts, groin, axillae, anus, umbilicus, toe or fingerwebs Vagina	Red lesions with well-defined bor- ders and presence of satellite le- sions; lesions may be dry or moist Red, oozing lesions with sharply
-	defined borders and inflamed vagina; cervix may be covered with moist, white plaque; cheesy, foul-smelling discharge; presence of pruritus and burning
Glans penis (balanitis)	Red lesions with sharply defined borders; penis may be covered with white plaque; presence of pruritus and burning
Mouth (thrush)	Creamy white flakes on a red, in- flamed mucous membrane; papil- lae on tongue may be enlarged
Nails	Red, painful swelling around nail bed; common in persons who often have their hands in water

tract the disease. Impetigo initially appears as a small vesicle or pustule or as a large bulla on the face or elsewhere on the body. As the primary lesion ruptures, it leaves a denuded area that discharges a honey-colored serous liquid that hardens on the skin surface and dries as a honey-colored crust with a "stuckon" appearance (Fig. 44-10). New vesicles erupt within hours. Pruritus often accompanies the lesions, and the skin excoriations that result from scratching multiply the infection sites. A possible complication of untreated GABHS impetigo is poststreptococcal glomerulonephritis (see Chapter 23). Topical mupirocin, which has few side effects, may be effective for limited disease. If the area is large or if there is concern about complications, systemic antibiotics are used.

Ecthyma is an ulcerative form of impetigo, usually secondary to minor trauma. It is caused by GABHS, *Staphylococcus aureus*, or *Pseudomonas*. It frequently occurs on the buttocks and thighs of children (Fig. 44-11). The lesions are similar to those of impetigo. A vesicle or pustule ruptures, leaving a skin erosion or ulcer that weeps and dries to a crusted patch, often resulting in scar formation. With extensive ecthyma, there is a low-grade fever and extension of the infection to other organs. Treatment usually involves the use of systemic antibiotics.

Viral Infections

Viruses are intracellular pathogens that rely on live cells of the host for reproduction. They have no organized cell structure but consist of a DNA or RNA core surrounded by a protein coat. The viruses seen in skin lesion disorders tend to be DNAcontaining viruses. Viruses invade the keratinocyte, begin to reproduce, and cause cellular proliferation or cellular death. The rapid increase in viral skin diseases has been attributed to the use of corticosteroid drugs, which have immunosuppressive qualities, and the use of antibiotics, which alter the bacterial flora of the skin. As the number of bacterial infections



■ FIGURE 44-10 ■ Impetigo of the face. (Abner Kurten, Folia Dermatologica. No. 2. Geigy Pharmaceuticals.) (Sauer G.C., Hall J.C. [1996]. Manual of skin diseases [7th ed.]. Philadelphia: Lippincott-Raven)



■ FIGURE 44-11 ■ Ecthyma on the buttocks of a 13-year-old boy. (Glaxo-Wellcome Co.) (Sauer G.C., Hall J.C. [1996]. *Manual of skin diseases* [7th ed.]. Philadelphia: Lippincott-Raven)

has decreased, there has been a proportional rise in viral skin diseases.

Verrucae. Verrucae, or warts, are common, benign papillomas caused by DNA-containing human papillomaviruses (HPV). The lesions are circumscribed, symmetrical epidermal neoplasms that are often elevated above the skin and often appear papillary. Histologically, there is an irregular thickening of the stratum spinosum and greatly increased thickening of the stratum corneum.

There are more than 50 types of HPVs found on the skin and mucous membranes of humans that cause several different kinds of warts, including skin warts and genital warts.¹² The skin warts, caused by HPV types 1, 2, 3, and 4, usually are not precancerous. HPV transmission usually occurs through breaks in skin integrity. They are known as common warts, flat warts, and plantar warts. *Common warts* are papillary growths, slightly raised above the skin surface, and varying in size from pin-head sized to large clusters of pea-sized tumors (Fig. 44-12). They are seen most commonly on the hands. Common hand warts can be transmitted by biting the cuticles surrounding the nail. *Flat*

Chapter 44: Alterations in Skin Function and Integrity

warts are small flat tumors that are often barely visible but can occur in clusters of 10 or more. They are commonly seen on the forehead and the dorsum of the hand. Plantar warts are flat to slightly raised painful growths that extend deep into the skin. They are frequently transmitted to the abraded, softened heels of children in gym showers or swimming areas. Genital warts are discussed in Chapter 35. Genital warts are sexually transmitted; some types of HPV may increase the risk of genital (cervix, vulva, and penis) cancers.

Treatment usually is directed at inducing a "wart-free" period without producing scarring. Warts resolve spontaneously when immunity to the virus develops. The immune response may be delayed for years. Removal is usually done by applying a keratolytic agent, such as salicylic acid, that breaks down the wart tissue, or by freezing with liquid nitrogen. Various types of laser surgery, electrosurgery, and the use of cytotoxic or antiviral therapy also have been successful in wart eradication.

Herpes Simplex. Herpes simplex virus (HSV) infections of the skin and mucous membrane (*i.e.*, cold sore or fever blister) are common. Two types of herpesviruses infect humans: type 1 and type 2. HSV-1 usually is confined to the oropharynx, and the organism is spread by respiratory droplets or by direct contact with infected saliva. Genital herpes usually is caused by HSV-2 (see Chapter 35), although HSV-1 also can cause genital herpes.

Infection with HSV-1 may present as a primary or recurrent infection. Primary HSV-1 infections usually are asymptomatic. Symptomatic disease occurs most frequently in young children (1 to 5 years of age). Symptoms include fever, sore throat, painful vesicles, and ulcers of the tongue, palate, gingiva, buccal mucosa, and lips. Primary infection results in the production of antibodies to the virus so that recurrent infections are more localized and less severe. After an initial infection, the herpesvirus persists in the trigeminal and other dorsal root ganglia in the latent state. It is likely that many adults were exposed to HSV-1 during childhood and therefore have antibodies to the virus.

The recurrent lesions of HSV-1 usually begin with a burning or tingling sensation. Vesicles and erythema follow and progress to pustules, ulcers, and crusts before healing (Fig. 44-13). The lesion is most common on the lips, face, and mouth. Pain



■ FIGURE 44-12 Common and periungual warts. (Reed & Carnrick Pharmaceuticals.) (Sauer G.C., Hall J.C. [1996]. *Manual of skin diseases* [7th ed.]. Philadelphia: Lippincott-Raven)



■ FIGURE 44-13 ■ Recurrent herpes simplex of the face. (Dermik Laboratories, Inc.) (Sauer G.C., Hall J.C. [1996]. *Manual of skin diseases* [7th ed.]. Philadelphia: Lippincott-Raven)

is common, and healing takes place within 10 to 14 days. Precipitating factors may be stress, sunlight exposure, menses, or injury. Individuals who are immunocompromised may have severe attacks.

There is no cure for oropharyngeal herpes simplex; most treatment measures are palliative. Penciclovir cream, a topical antiviral agent, applied at the first symptom may be used to reduce the duration of an attack. Application of over-the-counter topical preparations containing antihistamines, antipruritics, and anesthetic agents along with aspirin or acetaminophen may be used to relieve pain. Oral antiviral drugs that inhibit herpesvirus replication may be used prophylactically to prevent recurrences. Sunscreen preparations applied to the lips can prevent sun-induced herpes simplex.

Herpes Zoster. Herpes zoster (shingles) is an acute, localized vesicular eruption distributed over a dermatomal segment of the skin. It is caused by the same herpesvirus, varicella-zoster, that causes chickenpox. It is believed to be the result of reactivation of a latent varicella-zoster virus that was dormant in the sensory dorsal root ganglia since a childhood infection. During an episode of herpes zoster, the reactivated virus travels from the ganglia to the skin of the corresponding dermatome. Although herpes zoster is not as contagious as chickenpox, the reactivated virus can be transmitted to nonimmune contacts.

The incidence of herpes zoster increases with age; it occurs 8 to 10 times more frequently in persons older than 60 years than in younger persons.¹³ The normal age-related decrease in cell-mediated immunity is thought to account for the increased viral activation in this age group.¹⁴ Other persons at increased risk because of impaired cell-mediated immunity are those with conditions such as HIV infection and certain malignancies, chronic corticosteroid users, and those undergoing cancer chemotherapy and radiation therapy.

The lesions of herpes zoster typically are preceded by a prodrome consisting of a burning pain, tingling sensation, extreme sensitivity of the skin to touch, and pruritus along the affected dermatome. This may be present for 1 to 3 days or longer before the appearance of the rash. During this time, the pain may be mistaken for a number of other conditions, such as heart disease, pleurisy, various musculoskeletal disorders, or gastrointestinal disorders.

The rash appears as an eruption of vesicles with erythematous bases that are restricted to skin areas supplied by sensory neurons of a single or associated group of dorsal root ganglia (Fig 44-14). In immunosuppressed persons, the lesions may extend beyond the dermatome. Eruptions usually are unilateral in the thoracic region, trunk, or face. New crops of vesicles erupt for 3 to 5 days along the nerve pathway. The vesicles dry, form crusts, and eventually fall off. The lesions usually clear in 2 to 3 weeks.

Serious complications can accompany eruptions. Eye involvement can result in permanent blindness and occurs in a large percentage of cases involving the ophthalmic division of the trigeminal nerve (see Chapter 38). Pain can persist for several months after the rash disappears. Postherpetic neuralgia, which is pain that persists longer than 1 to 3 months after the resolution of the rash, is an important complication of herpes zoster (see Chapter 38).¹⁴

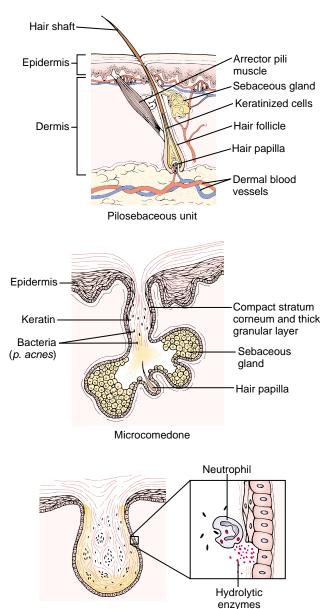
The treatment of choice for herpes zoster is the administration of an antiviral agent. The treatment is most effective when



started within 72 hours of rash development.^{15,16} Narcotic analgesics, tricyclic antidepressants or anticonvulsant drugs, and nerve blocks may be used for the management of herpetic pain. Oral corticosteroids sometimes are used to reduce the inflammation that may contribute to the pain.

Acne and Rosacea

Acne is a disorder of the pilosebaceous unit (hair follicle and sebaceous gland). The hair follicle is a tubular invagination of the epidermis in which hair is produced. The sebaceous glands empty into the hair follicle, and the pilosebaceous unit opens to the skin surface by means of a widely dilated opening called a pore (Fig. 44-15). The sebaceous glands produce a complex lipid mixture called sebum, from the Latin word meaning *tallow* or *grease*. Sebum consists of a mixture of free fatty acids, triglycerides, diglycerides, monoglycerides, sterol esters, wax es-



Open comedone

■ FIGURE 44-15 ■ The pathogenesis of acne vulgaris. The pilosebaceous unit (A) and the development of a microcomedone (whitehead) (B), and an open comedone with an accumulation of melanin (C).

ters, and squalene. Sebum production occurs through what is called a *holocrine process*, in which the sebaceous gland cells that produce the sebum are completely broken down and their lipid contents are emptied through the sebaceous duct into the hair follicle. The amount of sebum produced depends on two factors: the size of the sebaceous gland and the rate of sebaceous cell proliferation. The sebaceous glands are largest on the face, scalp, and scrotum, but are present in all areas of the skin except for the soles of the feet and palms of the hands. Sebaceous cell proliferation and sebum production are uniquely responsive to direct hormonal stimulation by androgens. In men, tes-

Acne lesions consist of comedones (whiteheads and blackheads), papules, pustules, nodules, and, in severe cases, cysts. Whiteheads are pale, slightly elevated papules with no visible orifice. Blackheads are plugs of material that accumulate in sebaceous glands that open to the skin surface. The color of blackheads results from melanin that has moved into the sebaceous glands from adjoining epidermal cells. Papules are raised areas less than 5 mm in diameter. Pustules have a central core of purulent material. Nodules are larger than 5 mm in diameter and may become suppurative or hemorrhagic. Suppurative nodules often are referred to as cysts because of their resemblance to inflamed epidermal cysts. Acne lesions are divided into noninflammatory and inflammatory lesions. Noninflammatory acne consists primarily of comedones. Inflammatory acne consists of papules, pustules, nodules, and cysts. The inflammatory lesions are believed to develop from the escape of sebum into the dermis and the irritating effects of the fatty acids contained in the sebum.

Two types of acne occur during different stages of the life cycle: acne vulgaris, which is the most common form among adolescents and young adults, and acne conglobata, which develops later in life.¹⁷ Other types of acne occur in association with various etiologic agents and influences.

Acne Vulgaris

Acne vulgaris is a common skin condition of adolescence and young adults. The condition is so common during adolescence that is often regarded as a normal part of the maturing process. The acne vulgaris lesions, which consist of comedones (whiteheads and blackheads) or inflammatory lesions (pustules, nodules, or cysts), are found primarily on the face and neck and to a lesser extent, on the back, chest, and shoulders (Fig. 44-16).

The cause of acne vulgaris remains unknown. Several factors are believed to contribute to acne, including (1) the influence of androgens on sebaceous cell activity, (2) increased proliferation of the keratinizing epidermal cells that form the sebaceous cells, (3) increased sebum production in relation to the severity of the disease, (4) decreased amounts of linoleic acid in the sebum, and (5) the presence of the *Propionibacterium acnes*, the microorganism responsible for inflammatory stage of the disorder.^{17,18} These factors probably are interrelated. Increased androgen production results in increased sebaceous cell activity, with a resultant plugging of the pilosebaceous ducts. The excessive sebum provides a medium for the growth of *P. acnes*. The *P. acnes* organism contains lipases that break down the free fatty acids that produce the acne inflammation.

Treatment of acne focuses on clearing up existing lesions, preventing new lesions from forming, and limiting scar formation.^{19,20} Milder forms of acne usually respond to proper cleansing and application of topical keratolytic agents that act chemically to break down keratin, loosen comedones, and exert a peeling effect on the skin. The treatment measures for moderate to severe acne are directed at correcting the defect in epidermal cell proliferation, decreasing sebaceous gland activity, reducing the *P. acnes* population, and limiting the inflammatory process. Often, a combination of keratolytic and antibacterial agents is used. Topically applied antibiotics do not affect existing lesions but prevent future lesions by decreasing the amount of *P. acnes*

848 Unit Eleven: Alterations in the Skeletal and Integumentary Systems



■ FIGURE 44-16 ■ (A) Acne of the face and (B) acne of the chest. (Hall J.C. [1999]. Sauer's manual of skin diseases [8th ed., p. 118]. Philadelphia: Lippincott Williams & Wilkins)

on the skin, thereby reducing subsequent inflammation formed from the presence of sebaceous fatty acid metabolites.

Severe cases of acne are managed with topical vitamin A, systemic antibiotics, and oral retinoids or acid form of vitamin A. The action of topical vitamin A (tretinoin) has been attributed to decreased cohesiveness of epidermal cells and increased epidermal cell turnover. This is thought to result in increased extrusion of open comedones and transformation of closed comedones into open ones. The oral retinoids (isotretinoin) have revolutionized the treatment of recalcitrant cases of acne and cystic acnes. Although the exact mode of action is unknown, isotretinoin decreases sebaceous gland activity, prevents new comedones from forming, reduces the *P. acnes* count through sebum reduction, and has an anti-inflammatory effect. Because of its many side effects, it is used only in persons with severe acne. Isotretinoin is a known teratogen and should not be used in women who are pregnant or may become pregnant.

Acne Conglobata

Acne conglobata occurs later in life and is a chronic form of acne. Comedones, papules, pustules, nodules, abscesses, cysts, and scars occur on the back, buttocks, and chest. Lesions occur to a lesser extent on the abdomen, shoulders, neck, face, upper arms, and thighs.¹⁷ The comedones have multiple openings. Their discharge is odoriferous, serous, and purulent or mucoid. Healing leaves deep keloidal lesions. Affected persons have anemia with increased white blood cell counts, sedimentation rates, and neutrophil counts. The treatment is difficult and stringent. It often includes debridement, systemic corticosteroid therapy, oral retinoids, and systemic antibiotics.

Rosacea

Rosacea, formerly called *acne rosacea*, is a chronic acneform type eruption of the butterfly area of the face that occurs in middle-age and older adults. It is easily confused with acne and may coexist with it. In the early stage of development, there are repeated episodes of blushing, eventually progressing to a persistent erythema on the nose and cheeks that sometimes extends to the forehead and chin (Fig. 44-17). This stage often occurs before 20 years of age. As the person ages, the erythema persists, and telangiectasia with or without acneiform components (*e.g.*, comedones, pustules, nodules) develops. After years of affliction, acne rosacea may develop into an irregular bullous hyperplasia (thickening) of the nose, known as *rhinophyma*. The sebaceous follicles and openings enlarge, and the skin color changes to a purple-red. Rhinophyma occurs mainly in men.¹⁷

The cause of rosacea remains unknown. It is seen more commonly in fair-skinned persons and has been called "the curse of the Celts." The capillaries in the involved facial areas display an increased reactivity to heat and other vascular stimulating agents such as highly spiced food and alcohol.

Treatment measures are similar to those used for acne vulgaris. Rhinophyma can be treated surgically. Because of increased capillary reactivity, people with rosacea are instructed to avoid heat, sunlight, hot liquids, highly seasoned foods, and alcohol.

Atopic Eczema and Nummular Eczema

Atopic eczema (atopic dermatitis) is a common skin disorder that occurs in two clinical forms: infantile and adult.^{21,22} It is associated with a type I hypersensitivity reaction (see Chap-



■ FIGURE 44-17 ■ Chronic rosacea with rhinophyma. (Hoechst Marion Roussel Pharmaceuticals, Inc.) (Sauer G.C., Hall J.C. [1996]. *Manual of skin diseases* [7th ed.]. Philadelphia: Lippincott-Raven)

ter 10). There usually is a family history of asthma, hay fever, or atopic dermatitis. The infantile form is characterized by vesicle formation, oozing, and crusting with excoriations. It usually begins in the cheeks and may progress to involve the scalp, arms, trunk, and legs (Fig. 44-18). The skin of the cheeks may be paler, with extra creases under the eyes, called *Dennie-Morgan folds*. There is marked follicle involvement in persons with black skin. Lesions may be hypopigmented or hyperpigmented or both on a black-skinned person. The infantile form usually becomes milder as the child ages, often disappearing by the age of 15 years. Adolescents and adults usually have dry, leathery, and hyperpigmented or hypopigmented lesions located in the antecubital and popliteal areas. These may spread to the neck, hands, feet, eyelids, and behind the ears. Itching may be severe with both forms. Secondary in-

fections are common. Treatment of atopic eczema is designed to target the underlying abnormalities such as dryness, pruritus, superinfection, and inflammation. It involves allergen control, basic skin care, and medications. Because dry skin and pruritus often exacerbate the condition, hydration of the skin is essential to treating atopic dermitis. Mild or healing lesions may be treated with lotions containing a mild antipruritic agent. Acute weeping lesions are treated with soothing lotions, soaks, or wet dressings. Topical corticosteroids provide an effective form of treatment but can cause local and systemic side effects. Because of their side effects, systemic corticosteroids usually are reserved for severe cases.



FIGURE 44-18 Atopic eczema on an infant's face and on a wrist. (Dome Chemicals.) (Sauer G.C., Hall J.C. [1996]. *Manual of skin diseases* [7th ed.]. Philadelphia: Lippincott-Raven)

The lesions of nummular eczema (discoid eczema) are coinshaped (nummular) papulovesicular patches mainly involving the arms and legs (Fig. 44-19). Lichenification and secondary bacterial infections are common. It is not unusual for the initial lesions seemingly to heal, followed by a secondary outbreak of mirror-image lesions on the opposite side of the body. Most nummular eczema is chronic, with weeks to years between exacerbations. Exacerbations occur more frequently in the cold winter months. The exact cause of nummular eczema is unknown. There usually is a history of asthma, hay fever, or atopic dermatitis. Ingestion of iodides and bromides usually aggravates the condition. Treatment is palliative. Frequent bathing, foods rich in iodides and bromides, and stress should be reduced, whereas the environmental humidity should be increased. Topical corticosteroids, coal tar preparations, and ultraviolet light treatments are prescribed as necessary.

Urticaria

Urticaria, or hives, is characterized by edematous plaques, called *wheals*, that are accompanied by intense itching. Wheals typically appear as raised pink or red areas surrounded by a paler halo. They blanch with pressure and vary in size from a few millimeters to centimeters. Thicker lesions that result from massive transudation of fluid into the dermis or subcutaneous tissue are referred to as *angioedema*. Although angioedema lesions can occur on any skin surface, they typically involve the larynx, causing hoarseness or sore throat, or mucosal surface of the gastrointestinal tract, causing abdominal pain.

Histamine, released from mast cells, is the most common mediator of urticaria. It causes hyperpermeability of the microvessels of the skin and surrounding tissue, allowing fluid to leak into the tissues, causing edema and wheal formation.^{23,24} A variety of immunologic, nonimmunologic, physical, and chemical stimuli can cause urticaria.

Urticaria can be acute or chronic. Daily or almost daily episodes of urticaria persisting for longer than 6 weeks are considered to be chronic. The most common causes of acute urticaria are foods or drinks, medications, or exposure to pollens or chemicals. Food is the most common cause of acute urticaria in children. Although nonsteroidal anti-inflammatory drugs, including aspirin, do not normally cause urticaria, they may exacerbate pre-existing disease.



■ FIGURE 44-19 ■ Nummular eczema of the buttocks. (Johnson & Johnson.) (Sauer G.C., Hall J.C. [1996]. *Manual of skin diseases* [7th ed.]. Philadelphia: Lippincott-Raven)

Chronic urticaria affects primarily adults and is twice as common in women as in men. Usually its cause cannot be determined despite extensive laboratory tests. Some forms of chronic urticaria are associated with histamine-releasing autoantibodies. In rare cases, urticaria is a manifestation of underlying disease, such as certain cancers, collagen diseases, and hepatitis. There is an association between chronic urticaria and autoimmune thyroid disease (*e.g.*, Hashimoto's thyroiditis, Graves' disease, toxic multinodular thyroiditis). A hereditary deficiency of a C1 (complement 1) inhibitor also can cause urticaria and angioedema.

Physical urticarias constitute another form of chronic urticaria.³⁵ Physical urticarias are intermittent, usually last less than 2 hours, are produced by appropriate stimuli, have distinctive appearances and locations, and are seen most frequently in young adults. Dermographism, or skin writing, is one form of physical urticaria in which wheals appear in response to simple rubbing of the skin (Fig. 44-20). The wheals follow the pattern of the scratch or rubbing, appearing within 10 minutes, and dissolving completely within 20 minutes. Other types of physical urticaria are cholinergic (*i.e.*, exercise-induced), cold, delayed pressure, solar (*i.e.*, sunlight), aquagenic (*i.e.*, water), vibratory, and external (localized heat-induced). Appropriate challenge tests (*e.g.*, application of an ice cube to the skin to initiate development of cold urticaria) are used to differentiate physical urticaria from chronic urticaria due to other causes.

Most types of urticaria are treated with antihistamines: drugs that block histamine type 1 (H_1) and less frequently, H_1 in combination with histamine type 2 (H_2). They control urticaria by inhibiting vasodilation and escape of fluid into the surrounding tissues. Severe urticaria and angioedema are treated with epinephrine. Oral corticosteroids may be used in the treatment of refractory urticaria.

Drug-Induced Skin Eruptions

Almost any drug can cause a localized or generalized skin eruption. Topical drugs usually are responsible for a localized contact dermatitis type of rash, whereas systemic drugs cause generalized skin lesions. Most drug eruptions are morbilliform (*i.e.*, measles-like) or exanthematous. Drug eruptions usually occur within 1 to 2 weeks after starting therapy. Some drug reactions progress to more severe skin lesions, necessitating prompt medical attention. Drug-induced skin reactions mimic almost all other skin lesions described in this chapter.

The diagnosis of a drug sensitivity depends almost entirely on accurate reporting by the person because the lesions from drug sensitization vary greatly. Treatment of mild cases is aimed at eliminating the offending drug while treating the symptoms. In mild cases, the lesions usually clear within 7 to 14 days after withdrawal of the offending drug. Severe drug eruptions often require systemic corticosteroid therapy and antihistamines.

Three types of bullous skin manifestations that result from drug reactions end in epidermal skin detachment: erythema multiforme minor, Stevens-Johnson syndrome (erythema multiforme major), and toxic epidermal necrolysis. The latter two are rare occurrences, but they can be life threatening. Both usually are caused by sensitivity to such drugs as sulfonamides and anticonvulsants. Although erythema multiforme minor may be drug induced, it more frequently occurs after infections, especially with herpes simplex. It is self-limiting, with a small amount of skin detachment at the lesion sites.

The lesions of erythema multiforme minor and Stevens-Johnson syndrome are similar. The primary lesion of both is a round, erythematous papule, resembling an insect bite. Within hours to days, these lesions change into several different patterns. The individual lesions may enlarge and coalesce, producing small plaques, or they may change to concentric zones of color appearing as "target" or "iris" lesions (Fig. 44-21). The outermost rings of the target lesions usually are erythematous; the central portion usually is opaque white, yellow, or gray (dusky). In the center, small blisters on the dusky purpuric macules may form, giving them their characteristic target-like appearance. Although there is wide distribution of lesions over the body surface area, there is a propensity for them to occur on the face and trunk. With Stevens-Johnson syndrome, there is more skin detachment.



■ FIGURE 44-20 ■ Dermographism on a patient's back. (Dermik Laboratories, Inc.) (Sauer G.C., Hall J.C. [1996]. *Manual of skin diseases* [7th ed.]. Philadelphia: Lippincott-Raven)



■ FIGURE 44-21 ■ Erythema-multiforme-like eruption on the patient's arm. Notice the dusky, target-like appearance. (Dermik Laboratories, Inc.) (Sauer G.C., Hall J.C. [1996]. *Manual of skin diseases* [7th ed.]. Philadelphia: Lippincott-Raven)

Chapter 44: Alterations in Skin Function and Integrity

Toxic epidermal necrolysis is the most serious and lifethreatening drug reaction. The person experiences a prodromal period of malaise, low-grade fever, and sore throat. Within a few days, widespread erythema and large, flaccid bullae appear, followed by the loss of the epidermis. This leaves a denuded and painful dermis. The skin surrounding large denuded areas may have the typical target-like lesions seen with Stevens-Johnson syndrome. The skin separates easily from the dermis with lateral pressure. The epithelium of mucosal surfaces, especially the mouth and eyes, may be involved.

These three types of bullous skin eruptions are seemingly quite similar. The diagnostic boundary for erythema multiforme minor is that it usually occurs after herpes simplex infection and is self-limiting. Precise diagnostic boundaries between Stevens-Johnson syndrome and toxic epidermal necrolysis have not been established. However, it is generally agreed that cases involving less than 10% of the body surface area are called *Stevens-Johnson syndrome*, and detachment of more than 30% of the epidermis is labeled *toxic epidermal necrolysis*.^{25,26} The mortality rate associated with Stevens-Johnson syndrome is less than 5%, and that for toxic epidermal necrolysis is 30% or greater.²⁶

The skin detachment associated with these drug reactions is different from the desquamation (*i.e.*, peeling) discussed with other skin disorders. For example, with scarlet fever there is peeling of the stratum corneum, the dead keratinized layer. In the bullous disorders discussed here, there is full-thickness detachment (*i.e.*, peeling of the entire epidermis down to the dermis). This leaves the person vulnerable to multiple problems, such as loss of body fluid and thermal control, nutritional deficits, and electrolyte imbalance.

Treatment of erythema multiforme minor and less severe cases of Stevens-Johnson syndrome includes relief of symptoms using compresses, antipruritic drugs, and topical anesthetics. Corticosteroid therapy may be indicated in moderate cases, although its use is a matter of controversy. For severe cases of Stevens-Johnson syndrome or toxic epidermal necrolysis, hospitalization is required for fluid replacement, antibiotics, respiratory care, analgesics, and moist dressings. When large areas of skin are detached, the care is similar to that provided to patients with thermal burns.

Papulosquamous Dermatoses

Papulosquamous dermatoses are a group of skin disorders characterized by scaling papules and plaques. Among the major papulosquamous diseases are psoriasis, pityriasis rosea, and lichen planus.

Psoriasis

Psoriasis is a common skin disease characterized by circumscribed red, thickened plaques with an overlying silvery-white scale. Psoriasis occurs worldwide, although the incidence is lower in warmer, sunnier climates. In the United States, it affects 2.6% or 6 million Americans.²⁷ The average age of onset is in the third decade; its prevalence increases with age. Approximately one third of patients have a genetic history, indicating a hereditary factor. Childhood onset of the disease is more strongly associated with a family history than psoriasis occurring in adults older than 30 years.²⁸ There appears to be an association between psoriasis and arthritis. Psoriatic arthritis occurs in 5% to 7% of persons with psoriasis (see Chapter 43). The cause of psoriasis remains poorly understood and is probably multifactorial. There is evidence of a genetic component. The more severe the disease, the greater is the likelihood of a familial background. Environmental factors may also play a role. A variety of stimuli, such as physical injury, infections, use of certain drugs, and photosensitivity, may precipitate the development or exacerbation of lesions in people who are predisposed to the disease. The reaction of the skin to an original trauma of any type is called the *Köbner reaction*.

Histologically, psoriasis is characterized by increased epidermal cell turnover with marked epidermal thickening, a process called *hyperkeratosis*. The migration time of the keratinocyte from the basal cell layer to the stratum corneum decreases from the normal 14 days to approximately 4 to 7 days. The granular layer of the epidermis is thinned or absent. There is also an accompanying thinning of the epidermal cell layer that overlies the tips of the dermal papillae (suprapapillary plate), and the blood vessels within dermal papillae become tortuous and dilated. These capillary beds show permanent damage even when the disease is in remission or has resolved. The close proximity of the vessels in the dermal papillae to the hyperkeratotic scale accounts for multiple, minute bleeding points that are seen when the scale is lifted.¹²

The lesions of plaque-type psoriasis may occur anywhere on the skin but most often involve the elbows, knees, and scalp (Fig. 44-22). The primary lesions are papules that vary in shape. The papules form thick red plaques with a silvery scale. In darker-skinned persons, the plaques may appear purple. There may be excoriation, thickening, or oozing from the lesions. A differential diagnostic finding is that the plaques bleed from minute points when removed.

There is no cure for psoriasis. The goal of treatment is to suppress the signs and symptoms of the disease: hyperkeratosis, epidermal inflammation, and abnormal keratinocyte



■ FIGURE 44-22 ■ Psoriasis on the elbows of a 17-year-old girl. (Roche Laboratories.) (Sauer G.C., Hall J.C. [1996]. *Manual of skin diseases* [7th ed.]. Philadelphia: Lippincott-Raven)

differentiation.^{27,28} Treatment measures are divided into topical and systemic approaches. Psoriasis has long been treated topically with keratolytic agents, coal tar products, and anthralin.¹⁷ Topical and systemic corticosteroids may be used, depending on the severity of the disease. Severe, generalized psoriasis may be treated with agents such as methotrexate. Methotrexate, which is used for cancer treatment, is an antimetabolite that inhibits DNA synthesis and prevents cell mitosis. Phototherapy after the administration of a psoralen has proven to be effective in many severe cases. Methoxsalen, a psoralen, exerts its actions when exposed to UVA radiation in 320to 400-nm wavelengths. Methoxsalen is given orally before UVA exposure. Activated by the UVA energy, methoxsalen inhibits DNA synthesis, thereby preventing cell mitosis and decreasing the hyperkeratosis that occurs with psoriasis. The combination treatment regimen of psoralen and UVA is known by the acronym PUVA.

Pityriasis Rosea

Pityriasis rosea is a rash that primarily affects young adults. The origin of the rash is unknown, but it is believed to be caused by an infective agent. Numerous viruses have been investigated, but no conclusive evidence has been found. The incidence is highest in winter. Cases occur in clusters and among persons who are in close contact with each other, indicating an infectious spread. However, there are no data to support communicability. It may be an immune response to any number of agents.

The characteristic lesion is an oval macule or papule with surrounding erythema (Fig. 44-23). The lesion spreads with central clearing, much like tinea corporis. This initial lesion is a solitary lesion called the *herald patch* and is usually on the trunk or neck. As the lesion enlarges and begins to fade away (2 to 10 days), successive crops of lesions appear on the trunk and neck. The lesions on the back have a characteristic "Christmas tree" pattern. The extremities, face, and scalp may be involved. Mild to severe pruritus may occur.

The disease is self-limited and usually disappears within 6 to 8 weeks. Treatment measures are palliative and include topical steroids, antihistamines, and colloid baths. Systemic corticosteroids may be indicated in severe cases.



■ FIGURE 44-23 ■ Pityriasis rosea of the thighs. (Syntex Laboratories.) (Sauer G.C., Hall J.C. [1996]. *Manual of skin diseases* [7th ed.]. Philadelphia: Lippincott-Raven)

Lichen Planus

The term *lichen* is of Greek origin and means "tree moss." The term is applied to skin disorders characterized by small (2 to 10 mm), flat-topped papules with irregular, angulated borders (Fig. 44-24).¹⁷ Lichen planus is a relatively common chronic, pruritic disease. It involves inflammation and papular eruption of the skin and mucous membranes. There are variations in the pattern of lesions (*e.g.*, annular, linear) and differences in the sites (*e.g.*, mucous membranes, genitalia, nails, scalp). The characteristic lesion is a purple, polygonal papule covered with a shiny, white, lacelike pattern. The lesions appear on the wrist, ankles, and trunk of the body. Most persons who have skin lesions also have oral lesions, appearing as milky white lacework on the buccal mucosa or tongue.

The etiology of lichen planus is unknown. There is increasing evidence of a cell-mediated response involving the epidermal-dermal junction with damage to the basal cell layer. Although the cause of most cases of lichen planus is unknown, some are linked to medication use or hepatitis C virus infection. The most common offending agents include gold, antimalarial agents, thiazide diuretics, beta blockers, nonsteroidal anti-inflammatory agents, quinidine, and angiotensin-converting enzyme inhibitors.²⁹

Diagnosis is based on the clinical appearance of the lesions and the histopathologic findings from a punch biopsy. For most persons, lichen planus is a self-limited disease. Treatment measures include discontinuation of all medications, followed by treatment with topical corticosteroids and occlusive dressings. Antipruritic agents are helpful in reducing itch. Systemic corticosteroids may be indicated in severe cases. Intralesional corticosteroid injections also may be used.



■ FIGURE 44-24 ■ Lichen planus of the dorsum of the hand and wrist. Notice the violacenous color of the papules and the linear Köbner's phenomenon. (Sauer G.C., Hall J.C. [1996]. *Manual of skin diseases* [7th ed.]. Philadelphia: Lippincott-Raven)

Lichen simplex chronicus is a localized lichenoid dermatitis. The term lichen simplex denotes that there was no known predisposing skin disorder in the affected person. It is characterized by the occurrence of itchy, reddened, thickened, and scaly patches of dry skin. Persons with the condition may have a single or, less frequently, multiple lesions. The lesions are seen most commonly at the nape of the neck, wrist, ankles, or anal area. The condition usually begins as a small pruritic patch, which culminates in a repetitive cycle of itching and scratching that develops into a chronic dermatosis. Because of the chronic itching and scratching, excoriations and lichenification with thickening of the skin develop, often giving the appearance of tree bark. Treatment consists of measures to decrease scratching of the area. A moderate-potency corticosteroid is often prescribed to decrease the itching and subsequent inflammatory process.

In summary, primary disorders of the skin include pigmentary skin disorders, infectious processes, inflammatory conditions, immune disorders, and allergic reactions. Pigmentary skin disorders include vitiligo, albinism, and melasma. Although the causes of the disorders vary, all involve changes in the amount of melanin produced by the melanocytes. These disorders appear in people of every skin type; however, the manifestations of the disorders vary among light-skinned and dark-skinned persons. Superficial fungal infections are called dermatophytoses and are commonly known as tinea or ringworm. Impetigo, which is caused by staphylococci or β-hemolytic streptococci, is the most common superficial bacterial infection. Viruses are responsible for verrucae (warts), herpes simplex type 1 lesions (cold sores or fever blisters), and herpes zoster (shingles). Noninfectious inflammatory skin conditions such as acne, lichen planus, psoriasis, and pityriasis rosea are of unknown origin. They usually are localized to the skin and are rarely associated with specific internal disease. Allergic skin responses involve the body's immune system and are caused by hypersensitivity reactions to allergens, environmental agents, drugs, and other substances.

NEVI AND SKIN CANCERS

Nevi

Nevi, or moles, are common congenital or acquired tumors of the skin that are benign. Almost all adults have nevi, some in greater numbers than others. Nevi can be pigmented or nonpigmented, flat or elevated, and hairy or nonhairy.

Nevocellular nevi are pigmented skin lesions resulting from proliferation of melanocytes in the epidermis or dermis. Nevocellular nevi are tan to deep brown, uniformly pigmented, small papules with well-defined and rounded borders. They are formed initially by melanocytes with their long dendritic extensions that are normally interspersed among the basal keratinocytes.¹² These melanocytes are transformed into round or oval melanin-containing cells that grow in nests or clusters along the dermal-epidermal junction. Because of their location, these lesions are called *junctional nevi* (Fig. 44-25). Eventually, most junctional nevi grow into the surrounding dermis as nests



■ **FIGURE 44-25** ■ Junctional nevi of the back of a 16-year-old patient. (Owen Laboratories, Inc.) (Sauer G.C., Hall J.C. [1996]. *Manual of skin diseases* [7th ed.]. Philadelphia: Lippincott-Raven)

or cords of cells. *Compound nevi* contain epidermal and dermal components. In older lesions, the epidermal nests may disappear entirely, leaving a *dermal nevi*. Compound and dermal nevi usually are more elevated than junctional nevi.

Another form of nevi, the *dysplastic nevi*, are important because of their capacity to transform to malignant melanomas. Although the association between dysplastic nevi and malignant melanoma was made more than 175 years ago, it was not until 1978 that the role of dysplastic nevi as a precursor of malignant melanoma was described in detail. Dysplastic nevi are larger than other nevi (often >5 mm in diameter). Their appearance is one of a flat, slightly raised plaque with a pebbly surface, or a target-like lesion with a darker, raised center and irregular border. They vary in shade from brown and red to flesh tones and have irregular borders. A person may have hundreds of these lesions. Unlike other moles or nevi, they occur on both sun-exposed and covered areas of the body. Dysplastic nevi have been documented in multiple members of families prone to the development of malignant melanoma.¹²

Because of the possibility of malignant transformation, any mole that undergoes a change warrants immediate medical attention. The changes to observe and report are changes in size, thickness, or color, and itching or bleeding.

Skin Cancer

There has been an alarming increase in skin cancers during the past several decades. Since the 1970s, the incidence rate of malignant melanoma, the most serious form of skin cancer, has increased significantly.³⁰ These increases are, on average, 4% per year, from 5.7 per 100,000 in 1973 to 13.8 per 100,000 in 1996, with approximately 47,700 new cases and 9600 deaths each year from melanoma. There also are approximately 1.3 million cases each year of highly curable basal cell and squamous cell cancers.

The rising incidence of skin cancer may be attributed primarily to increased sun exposure associated with societal and lifestyle shifts in the United States. The thinning of the ozone layer in the earth's stratosphere is thought to be an important

KEY CONCEPTS

SKIN CANCERS

- Increased and unprotected exposure to the ultraviolet rays of sunlight produces sunburn and increases the risk for development of skin cancer.
- The melanocytes, which protect against sunburn through increased production of melanin and suntanning, are particularly vulnerable to the adverse effects of unprotected exposure to ultraviolet light. Malignant melanoma, which is a malignant tumor of melanocytes, is a rapidly progressive and metastatic form of skin cancer.
- Basal cell carcinoma and squamous cell carcinoma, which also reflect the effects of increased sun exposure, are less aggressive forms of skin cancer and are more easily cured.

factor in this incidence rate. Society's emphasis on suntanning also is implicated. Persons have more leisure time and spend increasing amounts of time in the sun with uncovered skin.

Although the factors linking sun exposure to skin cancer are incompletely understood, both total cumulative exposure and altered patterns of exposure (in the case of melanoma) are strongly implicated. Basal cell and squamous cell carcinomas are associated with total cumulative exposure to ultraviolet radiation, whereas melanomas are associated with intense intermittent exposure. Thus, basal cell and squamous cell carcinomas occur more commonly on maximally sun-exposed parts of the body, such as the face and back of the hands and forearms. In contrast, melanomas occur most commonly in areas of the body that are exposed to the sun intermittently, such as the back in men and the lower legs in women. It is more common in persons with indoor occupations whose exposure to sun is limited to weekends and vacations.

Malignant Melanoma

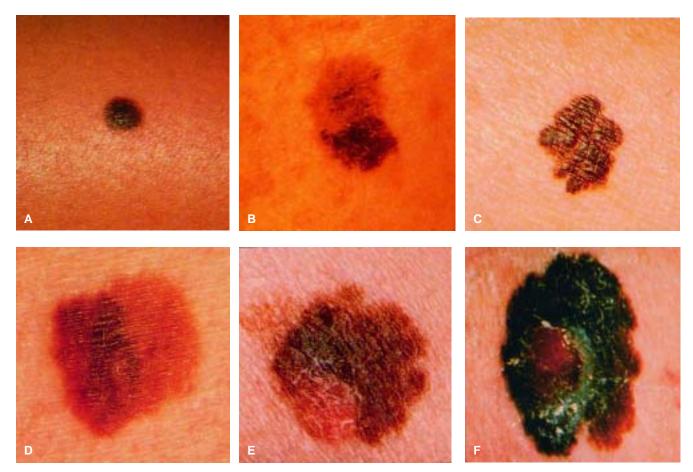
Malignant melanoma is a malignant tumor of the melanocytes. It is a rapidly progressing, metastatic form of cancer. The increased incidence of melanoma that has occurred during the past several decades has been attributed to an increase in sun exposure. The risk is greatest in fair-skinned people, particularly those with blond or red hair who sunburn and freckle easily. Fortunately, the increased risk of melanoma has been associated with a concomitant increase in the 5-year survival rate, from approximately 40% in the 1940s to 90% at present.³¹ Public health screening measures, early diagnosis, increased knowledge of precursor lesions, and greater public knowledge of the disease may account for earlier intervention.

Severe, blistering sunburns in early childhood and intermittent intense sun exposures (trips to sunny climates) contribute to increased susceptibility to melanoma in young and middleage adults. Roughly 90% of malignant melanomas in whites occur on sun-exposed skin. However, in African Americans and Asians, roughly 67% occur on non–sun-exposed areas, such as mucous membranes and subungual, palmar, and plantar surfaces.³² Although sun exposure remains a significant risk factor for melanoma, other potential risk factors have been identified, including atypical mole/dysplastic nevus syndrome, immunosuppression, prior PUVA therapy, and exposure to ultraviolet light at tanning salons. Using statistical analysis, it has been determined that six factors independently influence the risk for development of malignant melanoma: family history of malignant melanoma, presence of blond or red hair, presence of marked freckling on the upper back, history of three or more blistering sunburns before 20 years of age, history of 3 or more years of an outdoor job as a teenager, and presence of actinic keratosis. Persons with two of these risk factors had a 3.5-fold increased risk of malignant melanoma, and those with three or more risk factors had a 20-fold increased risk.³¹

Malignant melanomas differ in size and shape. Usually, they are slightly raised and black or brown. Borders are irregular, and surfaces are uneven. Most seem to arise from pre-existing nevi or new molelike growths (Fig. 44-26). There may be surrounding erythema, inflammation, and tenderness. Periodically, melanomas ulcerate and bleed. Dark melanomas are often mottled with shades of red, blue, and white. These three colors represent three concurrent processes: melanoma growth (blue), inflammation and the body's attempt to localize and destroy the tumor (red), and scar tissue formation (white). Malignant melanomas can appear anywhere on the body. Although they frequently are found on sun-exposed areas, sun exposure alone does not account for their development. In men, they are found frequently on the trunk, head, neck, and arms; in women, they commonly are found on the legs.

Four types of melanomas have been identified: superficial spreading, nodular, lentigo maligna, and acral lentiginous.³⁰ Superficial spreading melanoma is characterized by a raised-edged nevus with lateral growth. It has a disorderly appearance in color and outline. This lesion tends to have biphasic growth, horizontally and vertically. It typically ulcerates and bleeds with growth. This type of lesion accounts for 70% of all melanomas and is most prevalent in persons who sunburn easily and have intermittent sun exposure. Nodular melanomas, which account for 15% to 30% of melanomas, are raised domeshaped lesions that can occur anywhere on the body. They are commonly a uniform blue-black color and tend to look like blood blisters. Nodular melanomas tend to rapidly invade the dermis from the start with no apparent horizontal growth phase. Lentigo maligna melanomas, which account for 4% to 10% of all melanomas, are slow growing, flat nevi that occur primarily on sun-exposed areas of elderly persons. Untreated lentigo maligna tends to exhibit horizontal and radial growth for many years before it invades the dermis to become lentigo maligna melanoma. Acral lentiginous melanoma, which accounts for 2% to 8% of melanomas, occurs primarily on the palms of the hands, soles of the feet, nail beds, and mucous membranes. It has the appearance of lentigo maligna. Unlike other types of melanomas, it has a similar incidence in all ethnic groups.

Because virtually all the known risks of melanoma are related to susceptibility and magnitude of ultraviolet light exposure, protection from the sun's rays plays a critical role in the prevention of malignant melanoma. Early detection is critical with malignant melanoma. Regular self-examination of the total skin surface in front of a mirror under good lighting provides a method for early detection. It requires that a person undress completely and examine all areas of the body using a full



■ FIGURE 44-26 ■ (A) Normal mole with even, round contour and sharply defined borders. (B) Changes in appearance of a mole: *asymmetry*. (C) Changes in appearance of a mole: *border irregularity*. (D) Changes in appearance of a mole: *color and uneven pigmentation*. (E) Changes in the appearance of a mole: *diameter greater than 6 mm*. (F) Changes in the surface of a mole: *scaliness, oozing, and bleeding*. (American Cancer Society. [1995]. *What you should know about melanoma*. Dallas: American Cancer Society)

mirror, handheld mirror, and handheld hair dryer (to examine the scalp). An *ABCD* rule has been developed to aid in early diagnosis and timely treatment of malignant melanoma.³¹ The acronym stands for *asymmetry*, *b*order irregularity, *c*olor variegation, and *d*iameter greater than 0.6 cm (pencil eraser size). People should be taught to watch for these changes in existing nevi or the development of new nevi, as well as other alterations such as bleeding or itching.

Diagnosis of melanoma is based on biopsy findings from a suspect lesion. Treatment is usually surgical excision, the extent of which is determined by the thickness of the lesion, invasion of the deeper skin layers, and spread to regional lymph nodes. Deep and wide excisions, with possible use of skin grafts, are used. Current cancer treatment, such as immunotherapy and chemotherapy, is indicated when the disease becomes systemic. Interferon alfa-2b is approved by the FDA for treatment of melanoma.^{30,31}

The prognosis for malignant melanoma varies. It depends on factors such as tumor thickness (measured in millimeters), anatomic site, type of lesion, and levels of invasion (degree of penetration in the anatomic layers of the skin). Tumor thickness is an important factor in determining prognosis in persons with malignant melanoma.

Basal Cell Carcinoma

Basal cell carcinoma is the most common skin cancer in humans, accounting for 75% of all nonmelanoma skin cancers.³⁴ Like other skin cancers, basal cell carcinoma has increased in incidence during the past several decades. Basal cell carcinoma usually occurs in persons who were exposed to great amounts of sunlight. The incidence is twice as high among men as women and greatest in the 55- to 75-year-old age group.

Basal cell carcinoma usually is a nonmetastasizing tumor that extends wide and deep if left untreated. These tumors are seen most frequently on sun-exposed areas of the body, such as the head and neck, but do occur on other skin surfaces that were not exposed to the sun (Fig. 44-27). Although there are several histologic types of basal cell carcinoma, nodular ulcerative and superficial basal cell carcinomas are the most frequently occurring types. *Nodular ulcerative basal cell carcinoma* is the most common type.³⁴ It is a nodulocystic structure that begins as a small, flesh-colored or pink, smooth, translucent nodule that enlarges with time. Telangiectatic vessels frequently are seen beneath the surface. Over the years, a central depression forms that progresses to an ulcer surrounded by the original shiny, waxy border. Basal cell carcinoma in darker-skinned persons usually is darkly pigmented



■ FIGURE 44-27 ■ Basal cell carcinoma and wrinkling of the hand. (Syntex Laboratories.) (Sauer G.C., Hall J.C. [1996]. *Manual of skin diseases* [7th ed.]. Philadelphia: Lippincott-Raven)

and frequently misdiagnosed as other skin diseases, including melanoma.

The second most common form is *superficial basal cell carcinoma*, which is seen most often on the chest or back. It begins as a flat, nonpalpable, erythematous plaque. The red, scaly areas slowly enlarge, with nodular borders and telangiectatic bases. This type of skin cancer is difficult to diagnose because it mimics other dermatologic problems.

All suspected basal cell carcinomas should undergo biopsy for diagnosis. The treatment depends on the site and extent of the lesion. The most important treatment goal is complete elimination of the lesion. Also important is the maintenance of function and optimal cosmetic effect. Curettage with electrodesiccation, surgical excision, irradiation, and chemosurgery are effective in removing all cancerous cells. Patients should be checked at regular intervals for recurrences.

Squamous Cell Carcinoma

Squamous cell carcinomas are malignant tumors of the outer epidermis. They are commonly found on the sun-exposed area of the skin of people with fair complexions.³⁴ Metastasis is more common with squamous cell carcinoma than with basal cell carcinoma.

The mechanisms of squamous cell carcinoma development are unclear. Most squamous cell cancers occur in sun-exposed areas of the skin, and persons who spend much time outdoors, have lighter skin, and live in lower latitudes are more affected. The increase in the incidence of squamous cell carcinomas is consistent with increased ultraviolet radiation exposure. Other suspected causes include exposure to arsenic (*i.e.*, Bowen's disease), gamma radiation, tars, and oils.

There are two types of squamous cell carcinoma: intraepidermal and invasive. *Intraepidermal squamous cell carcinoma* remains confined to the epidermis for a long time. However, at some unpredictable time, it may penetrate the basement membrane to the dermis and metastasize to the regional lymph nodes. It then converts to *invasive squamous cell carcinoma*. The invasive type can develop from intraepidermal carcinoma or from a premalignant lesion (*e.g.*, actinic keratoses). It may be slow growing or fast growing with metastasis.



■ FIGURE 44-28 ■ Squamous cell carcinoma of the chin. (Syntex Laboratories, Westwood Pharmaceuticals.) (Sauer G.C., Hall J.C. [1996]. *Manual of skin diseases* [7th ed.]. Philadelphia: Lippincott-Raven)

Squamous cell carcinoma is a red-scaling, keratotic, slightly elevated lesion with an irregular border, usually with a shallow chronic ulcer (Fig. 44-28). Later lesions grow outward, show large ulcerations, and have persistent crusts and raised, erythematous borders. The lesions characteristically occur on the nose, forehead, helixes of the ears, lower lip, and back of the hands. In blacks, the lesions may appear as hyperpigmented nodules and occur more frequently on non–sunexposed areas.

Treatment measures are aimed at the removal of all cancerous tissue using methods such as electrosurgery, excision surgery, chemosurgery, or radiation therapy. After treatment, the person is observed for the remainder of his or her life for signs of recurrence. The recurrence rate is roughly 50%, with a 70% metastatic rate.³⁵

In summary, nevi are moles that usually are benign. Because they may undergo cancerous transformation, any mole that undergoes a change warrants immediate medical attention. There has been an alarming increase in skin cancers during the past few decades. Repeated exposure to the ultraviolet rays of the sun has been implicated as the principal cause of skin cancer. Neoplasms of the skin include malignant melanoma, basal cell carcinoma, and squamous cell carcinoma. Malignant melanoma is a malignant tumor of the melanocytes. It is a rapidly progressing, metastatic form of cancer. The most important clinical sign is the change in size, shape, and color of a pigmented skin lesion, such as a mole. Early detection through skin self-examination is critical. Squamous cell carcinoma and basal cell carcinoma are of epidermal origin. Basal cell carcinomas are the most common form of skin cancer among whites. They are slow-growing tumors that rarely metastasize. Squamous cell carcinoma is common in pale-skinned elderly persons. The two types of squamous cell carcinoma are intraepidermal and invasive. Intraepidermal squamous cell carcinoma remains confined to the epidermis for a long time. Invasive squamous cell carcinoma can develop from intraepidermal carcinoma or from premalignant lesions such as actinic keratoses.

Many skin problems occur more commonly in certain age groups. Because of aging changes, infants, children, and elderly persons tend to have different skin problems.

Skin Disorders in Infants and Children

Skin disorders in infants and children may be different from those in older children and adults. Certain skin disorders such as diaper rash and cradle cap are seen only in infants. Other conditions such as measles and chickenpox are more common in school age children.

Skin Disorders in Infants

Infancy connotes the image of perfect, unblemished skin. For the most part, this is true. However, several congenital skin lesions, such as mongolian spots, hemangiomas, and nevi, are associated with the early neonatal period.

Vascular and Pigmented Birthmarks. Pigmented and vascular lesions comprise most birthmarks.³⁶ Pigmented birthmarks represent abnormal migration or proliferation of melanocytes. Mongolian spots are caused by selective pigmentation. They usually occur on the buttocks or sacral area and are seen commonly in Asians and blacks. Nevi or moles are small, tan to brown, uniformly pigmented solid macules. Nevocellular nevi are formed initially from aggregates of melanocytes and keratinocytes along the dermal-epidermal border. Congenital melanocytic nevi are collections of melanocytes that are present at birth or develop within the first year of life. They present as macular, papular, or plaquelike pigmented lesions of various shades of brown, with a black or blue focus. The texture of the lesions varies, and they may be with or without hair. They usually are found on the hands, shoulders, buttocks, entire arm, or trunk of the body. Some involve large areas of the body in garment-like fashion. They usually grow proportionately with the child. Congenital melanocytic nevi are clinically significant because of their association with malignant melanoma.

Vascular birthmarks are cutaneous anomalies of angiogenesis and vascular development.³⁶ Two types of vascular birthmarks commonly are seen in infants and small children: bright red raised strawberry hemangiomas and flat, reddish-purple port-wine stains.

The strawberry hemangiomas begin as small, red lesions that are noticed shortly after birth. Hemangiomas are benign vascular tumors produced by proliferation of the endothelial cells. They are seen in approximately 5% to 10% of 1-year-old children.37 Female infants are three times as likely as male infants to have hemangiomas, and there is an increased incidence in premature infants. Approximately 35% of these lesions are present at birth, and the remainder develop within a few weeks after birth. Hemangiomas typically undergo an early period of proliferation during which they enlarge, followed by a period of slow involution in which the growth is reversed, and finally complete resolution. Most strawberry hemangiomas disappear before 5 to 7 years of age without leaving an appreciable scar. Hemangiomas can occur anywhere in the body. Hemangiomas of the airway can be life threatening. Ulceration, the most frequent complication, can be painful and carries the risk of infection, hemorrhage, and scarring.37



■ FIGURE 44-29 ■ Port-wine stain on the face of a boy. (Ortho Dermatology Corp.) (Sauer G.C., Hall J.C. [1996]. Manual of skin diseases [7th ed.]. Philadelphia: Lippincott-Raven)

Port-wine stains are pink or red patches that can occur anywhere on the body and are very noticeable (Fig. 44-29). They represent slow-growing capillary malformations that grow proportionately with the child and persist throughout life. Port-wine stains usually are confined to the skin but may be associated with vascular malformations of the eye or leptomeninges over the cortex, leading to cognitive disorders, seizures, and other neurologic deficits. Cover-up cosmetics are used in an attempt to conceal their disfiguring effects. Laser surgery has been used effectively in the treatment of port-wine stains.

Diaper Rash. The appearance of diaper rash ranges from simple (*i.e.*, widely distributed macules on the buttocks and anogenital areas) to severe (i.e., beefy, red, excoriated skin surfaces in the diaper area).³⁸ It results from a combination of ammonia and other breakdown products of urine. The treatment includes measures to minimize or prevent skin wetness. It includes frequent diaper changes with careful cleaning of the irritated area to remove the waste products. This is particularly important in hot weather. Exposing the irritated area to air is helpful. Secondary candidal (*i.e.*, yeast) infections may occur and require additional treatment (Fig. 44-30).

Prickly Heat. Prickly heat (heat rash) results from constant maceration of the skin because of prolonged exposure to a warm, humid environment. Maceration leads to midepidermal obstruction and rupture of the sweat glands. Although commonly seen during infancy, prickly heat may occur at any age. The treatment includes removing excessive clothing, cooling the skin with warm water baths, drying the skin with powders, and avoiding hot, humid environments.

Cradle Cap. Cradle cap is a greasy crust or scale formation on the scalp. It usually is attributed to infrequent and inadequate washing of the scalp. Cradle cap is treated by mild shampooing and gentle combing to remove the scales. Sometimes oil can be left on the head for minutes to several hours, softening



■ FIGURE 44-30 ■ Candida intertrigo after a course of oral antibiotics in a 1-year-old child. (Owen Laboratories, Inc.) (Sauer G.C., Hall J.C. [1996]. Manual of skin diseases [7th ed.]. Philadelphia: Lippincott-Raven)

the scales before scrubbing. Other emulsifying ointments or creams may be helpful in difficult cases. The scalp may need to be rubbed firmly to remove the buildup of keratinized cells.

Skin Manifestations of Common Infectious Diseases

Infectious childhood diseases that produce rashes include roseola infantum, rubella, rubeola, and varicella. Although these diseases are seen less frequently because of successful immunization programs and the use of antibiotics, they still occur. Immunization has greatly reduced the incidence of rubella and rubeola. Varicella vaccine is also available.

Roseola Infantum. Roseola infantum (i.e., exanthema subitum) is a contagious viral disease of infants and young children, most frequently between 6 and 18 months of age. It is caused by human herpesvirus-6 and produces a characteristic maculopapular rash covering the trunk and spreading to the appendages. The rash is preceded by an abrupt onset of high fever $(\leq 105^{\circ} \text{F})$, inflamed tympanic membranes, and coldlike symptoms usually lasting 3 to 4 days. These symptoms improve at approximately the same time the rash appears. Unlike rubella, no cervical or postauricular lymph node adenopathy occurs. Roseola infantum frequently is mistaken for rubella. Rubella usually can be excluded by the age of the child and the absence of lymph node adenopathy. In general, rubella does not develop in children younger than 6 to 9 months of age because they retain some maternal antibodies. Blood antibody titers may be taken to determine the actual diagnosis. In most cases, there are no long-term effects associated with this disease.

Rubella. Rubella (*i.e.*, 3-day measles or German measles) is a childhood disease caused by the rubella virus. It is characterized by a diffuse, punctate, macular rash that begins on the trunk and spreads to the arms and legs. Mild febrile states occur; usually the fever is less than 100°F. Postauricular, sub-occipital, and cervical lymph node adenopathy is common. Coldlike symptoms (*e.g.*, nasal congestion and cough) usually accompany the disease.

Rubella usually has no long-lasting sequelae; however, the transmission of the disease to pregnant women early in their

gestation periods may result in congenital rubella syndrome. Among the clinical signs of congenital rubella syndrome are cataracts, microcephaly, mental retardation, deafness, patent ductus arteriosus, glaucoma, purpura, and bone defects.

Rubeola. Rubeola (measles, hard measles, 7-day measles) is an acute, highly communicable viral disease caused by Morbillivirus. The characteristic rash is macular and blotchy; sometimes the macules become confluent. The rubeola rash usually begins on the face and spreads to the appendages. There are several accompanying symptoms: a fever of 100°F or greater, Koplik's spots (*i.e.*, small, irregular red spots with a bluish-white speck in the center) on the buccal mucosa, and mild to severe photosensitivity. The patient commonly has coldlike symptoms, general malaise, and myalgia. In severe cases, the macules may hemorrhage into the skin tissue or onto the outer body surface. This form is called *hemorrhagic measles*. The course of measles is more severe in infants, adults, and malnourished children. There may be severe complications, including otitis media, pneumonia, and encephalitis. Antibody titers are determined for a conclusive diagnosis of rubeola.

Varicella. Varicella (chickenpox) is a common communicable childhood disease. It is caused by the varicella-zoster virus, which also is the agent in herpes zoster (shingles). The characteristic skin lesion occurs in three stages: macule, vesicle, and granular scab. The macular stage is characterized by development within hours of macules over the trunk of the body, spreading to the limbs, buccal mucosa, scalp, axillae, upper respiratory tract, and conjunctiva. During the second stage, the macules form vesicles with depressed centers. The vesicles break open, and a scab forms during the third stage. Crops of lesions occur successively, so that all three forms of the lesion usually are visible by the third day of the illness.

Mild to extreme pruritus accompanies the lesions, which can lead to scratching and subsequent development of secondary bacterial infections. Chickenpox also is accompanied by coldlike symptoms and sometimes photosensitivity. Mild febrile states usually occur, typically beginning 24 hours before lesion outbreak. Side effects, such as pneumonia, septic complications, and encephalitis, are rare.

Varicella in adults may be more severe, with a prolonged recovery rate and greater chances for development of varicella pneumonitis or encephalitis. Immunocompromised persons may experience a chronic, painful type.

🙀 Skin Disorders in the Elderly

Elderly persons experience a variety of age-related skin disorders and exacerbations of earlier skin problems. Aging skin is believed to involve a complex process of actinic (solar) damage, normal aging, and hormonal influences.³⁹ Actinic changes primarily involve increased occurrence of lesions on sunexposed surfaces of the body.

Normal Age-Related Changes

Normal aging consists of changes that occur on areas of the body that have not been exposed to the sun. They include thinning of the dermis and the epidermis, diminution in subcutaneous tissue, a decrease in and thickening of blood vessels, and a decrease in the number of melanocytes, Langerhans' cells,

and Merkel's cells. The keratinocytes shrink, but the number of dead keratinized cells at the surface increases. This results in less padding and thinner skin, with color and elasticity changes. The skin also loses its resistance to environmental and mechanical trauma. Tissue repair takes longer.

With aging, there is also less hair and nail growth, and there is permanent hair pigment loss. There is normally less hormonally influenced sebaceous gland activity, although the glands in the facial skin may increase in size. The skin in most persons older than 70 years becomes dry, rough, scaly, and itchy. When there is no underlying pathology, it is called *senile* pruritus. Itching and dryness often become worse during the winter, when the need for home heating lowers the humidity.

Common Skin Lesions

The most common skin lesions in the elderly are skin tags, keratoses, lentigines, and vascular lesions. Most are actinic manifestations; they occur as a result of exposure to sun and weather throughout the years.

Skin Tags. Skin tags are soft, brown or flesh-colored papules. They occur on any skin surface but most frequently the neck, axilla, and intertriginous areas. They range in size from a pinhead to the size of a pea. Skin tags have the normal texture of the skin. They are benign and can be removed with scissors or electrodesiccation for cosmetic purposes.

Keratoses. A keratosis is a horny growth or an abnormal growth of the keratinocytes. A seborrheic keratosis (i.e., seborrheic wart) is a benign, sharply circumscribed, wartlike lesion that has a stuck-on appearance (Fig. 44-31). They vary in size up to several centimeters. They are usually round or oval, tan, brown, or black lesions. Less pigmented ones may appear yellow or pink. Keratoses can be found on the face or trunk as a solitary lesion or sometimes by the hundreds. Seborrheic keratoses are benign, but they must be watched for changes in color, texture, or size, which may indicate malignant transformation to squamous cell carcinoma.

Actinic keratoses are the most common premalignant skin lesions that develop on sun-exposed areas. The lesions usually are less than 1 cm in diameter and appear as dry, brown scaly areas, often with a reddish tinge. Actinic keratoses often are multiple and more easily felt than seen (Fig. 44-32). They often are indistinguishable from squamous cell carcinoma without biopsy. A hyperkeratotic form also exists that is more prominent and palpable. Often, there is a weathered appearance of the surrounding skin. Slight changes, such as enlargement or ulceration, may indicate malignant transformation. Most actinic keratoses are treated with 5-fluorouracil cream, which erodes the lesions. Roughly 20% of actinic keratoses convert to squamous cell carcinomas.

Lentigines. A lentigo is a well-bordered brown to black macule, usually less than 1 cm in diameter. Solar lentigines are tan to brown, benign spots on sun-exposed areas. They are commonly referred to as liver spots. Creams and lotions containing hydroquinone (e.g., Eldoquin, Solaquin) may be used temporarily to bleach the spots. These agents inhibit the synthesis of new pigment without destroying existing pigment. Higher concentrations are available by prescription. Successful treatment depends on avoiding sun exposure and consistent use of sunscreens. Liquid nitrogen applications have been successful in eradicating senile lentigines.

Lentigo maligna (i.e., Hutchinson's freckle) is a slowly progressive (≤20 years) preneoplastic disorder of melanocytes. It occurs on sun-exposed areas, particularly the face. The lesion is a pigmented macule with a well-defined border and grows to 5 cm or sometimes larger. As it grows throughout the years, it may become slightly raised and wartlike. If untreated, a true malignant melanoma often develops. Surgery, curettage, and cryotherapy have been effective at removing the lentigines. Careful monitoring for conversion to melanoma is important.

Vascular Lesions. Vascular lesions are vascular tumors with chronically dilated blood vessels. The small blood vessels lie in



FIGURE 44-31 Large seborrheic keratoses on the hand of an 84-year-old woman. (Sauer G.C., Hall J.C. [1996]. Manual of skin diseases [7th ed.]. Philadelphia: Lippincott-Raven)



FIGURE 44-32 Multiple actinic keratoses of the face of an 80-year-old man. (Dermik Laboratories, Inc.) (Sauer G.C., Hall J.C. [1996]. Manual of skin diseases [7th ed.]. Philadelphia: Lippincott-Raven)

the middle to upper dermis. *Senile angiomas* (cherry angiomas) are smooth, cherry-red or purple, dome-shaped papules. They usually are found on the trunk. *Telangiectases* are single dilated blood vessels, capillaries, or terminal arteries that appear on areas exposed to sun or harsh weather, such as the cheeks and the nose. The lesions can become large and disfiguring. Pulsed dye lasers have been effective in removing them. *Venous lakes* are small, dark blue, slightly raised papules that have a lakelike appearance. They occur on exposed body parts, particularly the backs of the hands, ears, and lips. They are smooth and compressible. Venous lakes can be removed by electrosurgery, laser therapy, or surgical excision if a person desires.

In summary, some skin problems occur in specific age groups. Common in infants are diaper rash, prickly heat, and cradle cap. Infectious childhood diseases that are characterized by rashes include roseola infantum, rubella, rubeola, and varicella.

Changes in skin that occur with aging involve a complex process of actinic damage, normal aging, and hormonal influences. With aging, there is thinning of the dermis and the epidermis, diminution in subcutaneous tissue, lessening and thickening of blood vessels, and a slowing of hair and nail growth. Dry skin is common among the elderly, becoming worse during the winter months. Among the skin lesions seen in the elderly are skin tags, keratoses, lentigines, and vascular skin lesions.

REVIEW QUESTIONS

Contrast and compare the following lesions: macule, patch, papule, plaque, nodule, pustule, blister, callus, and corn.

Describe the three types of ultraviolet radiation and relate them to sunburn, aging skin changes, and the development of skin cancer.

Relate the behavior of fungi to the production of superficial skin lesions associated with tinea or ringworm.

State the cause and describe the appearance of impetigo and ecthyma.

Compare the viral causes, manifestations, and treatments of verrucae, herpes simplex, and herpes zoster lesions.

Describe the pathogenesis of acne vulgaris and relate it to measures used in treating the disorder.

Describe the differences and similarities between erythema multiforme minor, Stevens-Johnson syndrome, and toxic epidermal necrolysis.

Define the term papulosquamous and use the term to describe the lesions associated with psoriasis, pityriasis rosea, and lichen planus.

Describe the origin of nevi and state their relationship to skin cancers.

Compare the appearance and outcome of basal cell carcinoma, squamous cell carcinoma, and malignant melanoma.

Describe the distinguishing features of rashes associated with the common infectious childhood diseases: roseola infantum, rubeola, rubella, chickenpox, and scarlet fever.

Describe the appearance of skin tags, keratoses, lentigines, and vascular lesions that are commonly seen in the elderly.

connection-

Visit the Connection site at **connection.lww.com/go/porth** for links to chapter-related resources on the Internet.

REFERENCES

- Harrist T.J., Schapiro B., Quinn T.R., et al. (1999). The skin. In Rubin E., Farber J.L. (Eds.), *Pathology* (3rd ed., pp. 1237–1298). Philadelphia: Lippincott Williams & Wilkins.
- Watson K.R. (2000). Structure of the skin. In Hall J.C. (Ed.), Sauer's manual of skin diseases (8th ed., pp. 1–5). Philadelphia: Lippincott Williams & Wilkins.
- 3. Greaves M.H. (1998). In Champion R.H., Burton J.L., Burns D.A., et al. (Eds.), *Textbook of dermatology* (6th ed., pp. 617–618). Oxford: Blackwell Science.
- 4. American Academy of Dermatology. (1987). Black skin. [On-line]. Available: http://www.aad.org/pamphlets/ black.html.
- Diffey B.L. (1992). Human exposure to ultraviolet violet radiation. In Marks R., Plewig G. (Eds.), *The environmental threat to the skin* (pp. 3–9). London: Martin Dunitz.
- Gilchrest B.A., Eller M.S., Geller A.C., et al. (1999). The pathogenesis of melanoma induced by ultraviolet radiation. *New England Journal of Medicine* 340, 1341–1348.
- Hall J.C. (Ed.). (2000). Sauer's manual of skin diseases (8th ed., p. 295). Philadelphia: Lippincott Williams & Wilkins.
- 8. Drug Facts and Comparisons Staff. (2000). *Drug facts and comparisons* 2000 (54th ed., p. 1714). St. Louis: Facts and Comparisons.
- 9. Epstein J., Kaplan L., Levine N. (2000). The value of sunscreens. *Patient Care* 34 (11), 103–107.
- Brooks G.F., Butel J.S., Ornston L.N. (1995). Medical microbiology (20th ed., pp. 531–536). Norwalk, CT: Appleton & Lange.
- Noble S.L., Forbes R.C. (1998). Diagnosis and management of common tinea infections. *American Family Physician* 58, 177.
- Murphy G.F. (2003). The skin. In Kumar V., Cotran R.S., Robbins S.L. (Eds.), *Basic pathology* (7th ed., pp. 789–807). Philadelphia: W.B. Saunders.
- Gilden D.H., Kleinschmidt-DeMasters B.K., LaGuardia J.J., et al. (2000). Neurologic complications of reactivation of varicella-zoster virus. *New England Journal of Medicine* 342, 635–645.
- Stankus S.J., Dlugopolski M., Packer D. (2000). Management of herpes zoster (shingles) and postherpetic neuralgia. *American Family Phy*sician 61, 2437–2444, 2447–2448.
- 15. Tyring S.K. (1996). Early treatment of herpes zoster. *Hospital Practice* 31, 137–144.
- Landow K. (2000). Acute and chronic herpes zoster. An ancient scourge yields to timely therapy. *Postgraduate Medicine* 107 (7), 107–118.
- Hall J.C. (Ed.). (2000). Sauer's manual of skin diseases (8th ed., pp. 114– 126, 127–144). Philadelphia: Lippincott Williams & Wilkins.
- Usantine R.P., Quan M.A., Strick R. (1998). Acne vulgaris: A treatment update. *Hospital Practice* 33 (2), 111–127.
- 19. Krowchuk D.P. (2000). Managing acne in adolescents. *Pediatric Clinics of North America* 47, 841–857.
- Leyden J.J. (1997). Therapy for acne vulgaris. New England Journal of Medicine 336, 1156–1163.
- 21. Correale C.E., Walker C., Craig T.J. (1999). Atopic dermatitis: A review of diagnosis and treatment. *American Family Physician* 60, 1191–1210.

Chapter 44: Alterations in Skin Function and Integrity

- 22. Kristal L., Klein P.A. (2000). Atopic dermatitis in infants and children. *Pediatric Clinics of North America* 47, 877–894.
- Scott C.B., Moloney M.F. (1996). Physical urticaria. Nurse Practitioner 21 (11), 42–59.
- Greaves M.W. (1995). Chronic urticaria. New England Journal of Medicine 332, 1767–1772.
- Bastuji-Garin S., Rzany B., Stern R.S., et al. (1993). Clinical classification of cases of toxic epidermal necrolysis, Stevens-Johnson syndrome, and erythema multiforme. *Archives of Dermatology* 129, 92–96.
- Roujeau J.C., Stern R.S. (1994). Severe adverse cutaneous reactions to drugs. New England Journal of Medicine 331, 1272–1284.
- 27. Koo J.Y. (1999). Current consensus and update on psoriasis therapy: A perspective from the U.S. *Journal of Dermatology* 26, 723–733.
- 28. Camisa C. (1994). Psoriasis (pp. 3, 30, 31, 55). Boston: Blackwell.
- Katta R. (2000). Lichen planus. American Family Physician 61, 3319– 3324, 3327–3328.
- Urist M.M., Heslin M.J., Miller D.M. (2001). Malignant melanoma. In Lenhard R.E., Osteen R.T., Gansler T. (Eds.), *Clinical oncology* (pp. 553–563). Atlanta: American Cancer Society.

- Rigel D.S., Carucci J.A. (2000). Malignant melanoma: Prevention, early detection, and treatment in the 21st century. CA: A Cancer Journal for Clinicians 50, 215–236.
- 32. Halder R.M., Bridgeman-Shah S. (1995). Skin cancer in African Americans. *Cancer* 75, 667–673.
- Jerant A.F., Johnson J.T., Sheridan C.M., et al. (2000). Early detection and treatment of cancer. American Family Physician 62, 357–368, 375–376, 381–382.
- Preston D.S., Stern R.S. (1992). Nonmelanoma cancers of the skin. New England Journal of Medicine 327, 1649–1662.
- Frankel D.H. (1992). Squamous cell carcinoma of the skin. Hospital Practice 27, 99–106.
- Dohil M.A., Baugh W.P., Eichenfield L.F. (2000). Vascular and pigmented birthmarks. *Pediatric Clinics of North America* 47, 783–810.
- Drolet B.A., Esterly N.B., Frieden I.J. (1999). Hemangiomas in children. New England Journal of Medicine 341, 173–181.
- Kazaks E.L., Lane A.T. (2000). Diaper dermatitis. *Pediatric Clinics of* North America 47, 909–918.
- Bolognia J.L. (1995). Aging skin. American Journal of Medicine 98, 998–1038.