Clinical Gynecologic Endocrinology and Infertility

Eighth Edition

MARC A. FRITZ and LEON SPEROFF

Wolters Kluwer Health | Lippincott Williams & Wilkins

Clinical Gynecologic Endocrinology and Infertility

Clinical Gynecologic Endocrinology and Infertility

Eighth Edition

Marc A. Fritz, MD

The University of North Carolina at Chapel Hill

Leon Speroff, MD

Oregon Health & Science University

Wolters Kluwer Health | Lippincott Williams & Wilkins

Philadelphia • Baltimore • New York • London
Buenos Aires • Hong Kong • Sydney • Tokyo

Acquisitions Editor: Sonya Seigafuse
Product Manager: Nicole Walz
Vendor Manager: Alicia Jackson
Senior Manufacturing Manager: Benjamin Rivera
Marketing Manager: Kim Schonberger
Design Coordinator: Doug Smock
Production Service: SPi Technologies

Two Commerce Square
2001 Market Street
Philadelphia, PA 19103 USA
LWW.com

Printed in the USA.

Library of Congress Cataloging-in-Publication Data
Fritz, Marc A.
Clinical gynecologic endocrinology and infertility / Marc A. Fritz, Leon Speroff. – 8th ed.
p. ; cm.
Speroff's name appears first in previous editions.
Includes bibliographical references and index.
ISBN 978-0-7817-7968-5
1. Endocrine gynecology. 2. Infertility–Endocrine aspects. 3. Contraception. I. Speroff, Leon, 1935- II. Title.
[DNLM: 1. Endocrine System Diseases. 2. Genital Diseases, Female. 3. Contraception. 4. Infertility, Female. 5. Reproductive Techniques, Assisted. WP 505]

RG159.S62 2011
618.1–dc22

2010031879

To purchase additional copies of this book, call our customer service department at (800) 638-3030 or fax orders to (301) 223-2320. International customers should call (301) 223-2300.

Visit Lippincott Williams & Wilkins on the Internet: at LWW.com. Lippincott Williams & Wilkins customer service representatives are available from 8:30 am to 6 pm, EST.

10 9 8 7 6 5 4 3 2 1

Preface

Thirty-eight years ago, I was walking along a corridor at Yale and encountered Bob Glass coming toward me. He stopped and said, “Nate (Nathan Kase) and I are writing a book. Are you interested in joining us?” “You bet!” I said, and a year later, the first edition of our book appeared, typed by me on a Royal portable typewriter, 273 pages long, $17 in price. Not too long ago, I was talking with one of my colleagues and mentioned that I had just looked at the first edition of our book, and it looked to me like a grade school primer. My colleague looked me in the eye and said, “That’s why I liked it!”

Many years and several editions later, I was standing on a New York City corner, waiting to cross the street. Without warning, a thought struck me like a thunderbolt. It froze me on that corner, and when the light changed, everyone crossed without me. The thought was: What an enormous obligation it is to not let anything written in our book lead to improper care of a patient. I had to get it right!

How did I get here? My grandparents, father, and my uncle were Macedonian mountain peasants in the northern limit of Greece. One day in 1911, my grandfather just walked away and disappeared. For the next 10 years, the village helped my grandmother take care of her two boys. Then one day in 1921, a hand-delivered telegram from my grandfather came for my grandmother, saying: “I’m in Sofia, Bulgaria. Come and join me.” He had been in the U.S. for those 10 years, working on railroad construction.

He neglected to send any money. My grandmother and her two sons walked 200 miles to Bulgaria. It wasn’t as difficult as it sounds because they walked for 2 months from village to village, and the villagers gave them food and a place to sleep. They found my grandfather in a hotel, with thousands of dollars saved from his work in America. The plan was to purchase a farm, but somehow they were cheated out of half their money, with nothing to show for it. My father, now age 18, said, “If you can make that much money in America, let’s go.”

They came through Ellis Island to Ohio, where a close family friend had arranged a job for my grandfather in the steel factory in Lorain. With the last of their money, they bought a 24-acre farm. My early years were spent on that farm, speaking Macedonian, and little English.

By now, you are wondering, “What is the point of this story?” The point is that if you would have told me when I was a boy how my life would turn out, I never would have believed you. If you had told me that someday I would be writing a Preface in the Eighth Edition of a large, medical book, I would have been incredulous. Because of my early life, I never take anything for granted. I am deeply appreciative for all that has happened throughout my career, especially this book.

For many years and through multiple editions in eleven languages, this book has opened doors and made friends for me and my family in numerous countries. I am thankful for an experience that has always been heart-warming and uplifting.

Now it is time to pass the torch. Marc Fritz came to Oregon in 1981 for his fellowship in reproductive endocrinology. Early on, I asked Marc to write a review on the regulation of the menstrual cycle. Finished with the task, he carefully put the manuscript in my office mailbox on a Friday afternoon when I wasn’t around, anticipating a critical destruction

of his work. He didn't have long to wait. I called him that Sunday afternoon, but with a message he was not expecting. I congratulated him on his work, telling him how impressed I was with his ability to articulate the science in a clear and conceptual fashion. Some of those sentences are still in Chapter 6.

This book may be the only remaining medical book of its size to be single-authored, a reason why the writing style is consistent from beginning to end. The writing style of this book has always been a major factor in its success, a style that avoided medical jargon and never feared to make relevant clinical conclusions and recommendations based on up-to-date medical knowledge. Marc Fritz and I have been good friends since 1981, and his writings taught me that he matched me in a compulsion to get it right and his efforts to be clinically relevant. For these reasons, Marc was an obvious and natural choice to become the senior author of this book.

Now I am a professor emeritus, riding my tractor, playing softball, fly-fishing, and still writing. My best wishes to all for good health and a happy, rewarding life.

Leon Speroff
Portland, Oregon

Twenty-nine years ago, my fellowship training in Reproductive Endocrinology with Leon Speroff began. I was Leon's first fellow, arriving in Portland, Oregon in 1981, eager, excited, and determined. The next 2 years, in many ways, shaped all of those that have followed.

Leon's story about the first manuscript I wrote as a fellow is quite true. When first assigned the task, I was again eager, excited, and determined, but also uncertain. I spent countless hours searching the stacks and copying articles at the medical library, organizing and synopsizing the literature on 4 x 6 index cards, and typing the paper on an Underwood portable typewriter at the kitchen table. I had done my best with the assignment, but was not at all sure it would be good enough. As it turned out, that review on the menstrual cycle, published as a Modern Trends article in *Fertility and Sterility* in 1982, became the foundation for a lifelong friendship.

During my residency, this book (then in its second edition, with 433 total pages) was a constant companion. I read and re-read it from cover to cover, and found in its pages my passion, career path, and my teacher. Had I been told then that someday I would be writing a preface for the Eighth Edition of this book, I would have been incredulous. In the years that followed my training in Oregon, I became just as familiar with each subsequent edition. To me, reading the book was very much like having a conversation with Leon, or like listening to him lecture; always clear, logical, and practical, with a personal touch.

When Leon invited me to co-author the Seventh Edition of this book, I experienced many of the same emotions I had when preparing that first manuscript as a fellow; it was only natural. It was a joy to work closely together again in preparing the previous edition, and this one. I am truly honored to become the senior author of this book, and to be entrusted with its future. I understand the responsibility and am truly grateful for the opportunity. Understandably, Leon views it as passing the torch. Understandably, I view it as closing a perfect circle.

Marc A. Fritz
Chapel Hill, North Carolina

P.S. The cover colors of the Eighth Edition are those of Tulane University. The symbol on the cover is the Macedonian Star, from the days of Philip of Macedon and Alexander the Great.

Contents

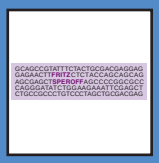
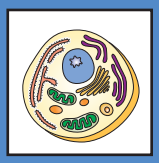
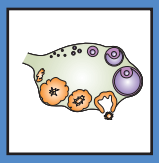
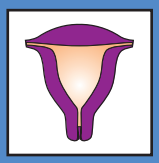
Section II
Clinical Endocrinology

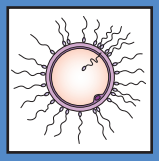
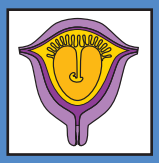
Section IV
Infertility

SECTION

I

REPRODUCTIVE PHYSIOLOGY

1 Molecular Biology for Clinicians

GCAGCCGTATTTCTACTGCGACGAGGAG
GAGAACTT**FRITZ**CTCTACCAGCAGCAG
AGCGAGCT**SPEROFF**AGCCCCGGCGCC
CAGGGATATCTGGAAGAAATTCGAGCT
CTGCCGCCCTGTCCCTAGCTGCGACGAG

The above DNA sequence is obviously a mutant. But the fact that we can recognize this cryptogram as a nucleotide sequence and diagnose a mutant change illustrates the incredible progress made in the understanding of human biology. Molecular biology is the subspecialty of science devoted to understanding the structure and function of the genome, the full complement of DNA (deoxyribonucleic acid), the macromolecule that contains all the hereditary information.

The Austrian monk Gregor Mendel studied his garden of peas for much of his life at his monastery and was the first to express the principles of heredity in the 1860s. He described dominant and recessive traits and the "laws" of transmission governing the homozygous and heterozygous inheritance of these traits. Mendel's theories remained unknown until 1900, when they were discovered. Unfortunately, Mendel died 16 years before recognition of his work. But how far we have come in only 150 years, mostly in the last 50 years!

The pairing and splitting of chromosomes at cell division was proposed in 1903, but it was not until 1946 that Edward Tatum and Joshua Lederberg at Yale University demonstrated in bacteria that DNA carried hereditary information. James Watson and Francis Crick, working at the Cavendish Laboratories in Cambridge, proposed in 1953 the structure of DNA by creating a model based on the parameters provided by Maurice Wilkins and Rosalind Franklin obtained with x-ray crystallography. Crick, Watson, and Wilkins received the Nobel Prize in 1962; Franklin died in 1958, and the Nobel Prize is not awarded posthumously.

DNA replication involves many enzyme systems. DNA polymerase was isolated in 1958, and ribonucleic acid (RNA) polymerase in 1960. In 1978, Werner Arber, Hamilton Smith, and Daniel Nathans received the Nobel Prize for their discovery, in the 1960s, of the enzymes for joining or cutting DNA. The use of ligase and restriction endonuclease enzymes permitted the production of recombinant DNA molecules, first accomplished by Paul Berg at Stanford University in 1972.

E.M. Southern of Edinburgh University developed in 1975 the technique to transfer (to blot) DNA from agarose gels onto nitrocellulose filters, enabling DNA fragments to

be joined with radiolabeled RNA probes and thus isolated. The cloning of genes or DNA fragments followed the breakthrough discovery that plasmids carrying foreign DNA molecules could be inserted into bacteria, leading to the replication of the foreign DNA.

The genome is the complete set of DNA in an organism. A gene is a contiguous region of DNA that can encode a protein product and contains within its area regulatory sequences that regulate its expression. The study of the functions and interactions of all the genes in the genome required a new designation, "genomics," coined in 1987 with a journal of that name.

We have entered the age of molecular biology. It won't be long before endocrine problems will be explained, diagnosed, and treated at the molecular level. Soon the traditional hormone assays will be a medical practice of the past. The power of molecular biology will touch us all, and the many contributions of molecular biology will be perceived throughout this book. But unfortunately, molecular biology has its own language, a language that is almost unintelligible to the uninitiated. We offer this chapter as a guide to molecular medicine.

To begin a clinical book with a chapter on molecular biology and a chapter on biochemistry only serves to emphasize that competent clinical judgment is founded on a groundwork of basic knowledge. On the other hand, clinical practice does not require a technical and sophisticated proficiency in a basic science. The purpose of these first two chapters, therefore, is not to present an intensive course in a basic science, but rather to review the most important principles and information necessary for the development of the physiological and clinical concepts to follow. It is further intended that certain details, which we all have difficulty remembering, will be available in these chapters for reference.

The Chromosomes

We are ***eukaryotes,*** organisms with cells having a true nucleus bounded by a nuclear membrane, with multiplication by mitosis. Bacteria are ***prokaryotes,*** organisms without a true nucleus, with reproduction by cell division. With the exception of DNA within mitochondria, all of our DNA is packaged in a nucleus surrounded by a nuclear membrane. Mitochondria are believed to be descendants of primitive bacteria engulfed by our ancestors, and they still contain some important genes. Because ova are rich in mitochondria, diseases due to mitochondrial genes (for example, Leber's optic neuropathy) are transmitted by the mother. The mitochondria in sperm are eliminated during fertilization.

Chromosomes are packages of genetic material, consisting of a DNA molecule (which contains many genes) to which are attached large numbers of proteins that maintain chromosome structure and play a role in gene expression. Human somatic cells contain 46 chromosomes, 22 pairs of autosomes, and one pair of sex chromosomes. All somatic cells are diploid—23 pairs of chromosomes. Only gametes are haploid, with 22 autosome chromosomes and one sex chromosome. The chromosomes vary in size, ranging from 50 million to 250 million base pairs. Chromosome 1 contains the most genes (2,968), and the Y chromosome has the smallest number (231). All contain a pinched portion called a centromere, which divides the chromosome into two arms, the shorter p arm and the longer q arm. The two members of any pair of autosomes are homologous, one homologue derived from each parent. The number of chromosomes does not indicate the level of evolutionary sophistication and complexity; the dog has 78 chromosomes and the carp has 104!

A single gene is a unit of DNA within a chromosome that can be activated to transcribe a specific RNA. The location of a gene on a particular chromosome is designated its locus. Because there are 22 pairs of autosomes, most genes exist in pairs. The pairs are homozygous when similar and heterozygous when dissimilar. Only 2% of the human genome

consists of genes that encode protein synthesis. In 1952, Alfred Hershey and Martha Chase confirmed that DNA is the source of genetic transmission. Their report is famously known as the blender experiment, in which labeled DNA in bacteriophages infected bacteria and blender agitation separated the phage particles, demonstrating that the radioactive tracer was present only in the bacterial cells.

The usual human karyotype is an arrangement of the chromosomes into pairs, usually after proteolytic treatment and Giemsa staining to produce characteristic banding patterns, allowing a blueprint useful for location. The staining characteristics divide each arm into regions, and each region into bands that are numbered from the centromere outward. A given point on a chromosome is designated by the following order: chromosome number, arm symbol (p for short arm, q for long arm), region number, and band number. For example, 7q31.1 is the location for the cystic fibrosis gene.

Mitosis

All eukaryotes, from yeasts to humans, undergo similar cell division and multiplication. The process of nuclear division in all somatic cells is called mitosis, during which each chromosome divides into two. For normal growth and development, the entire genomic information must be faithfully reproduced in every cell.

Mitosis consists of the following stages:

Interphase

During this phase, all normal cell activity occurs except active division. It is during this stage that the inactive X chromosome (the Barr body or the sex chromatin) can be seen in female cells.

Prophase

As division begins, the chromosomes condense, and the two chromatids become visible. The nuclear membrane disappears. The centriole is an organelle outside the nucleus that forms the spindles for cell division; the centriole duplicates itself, and the two centrioles migrate to opposite poles of the cell.

Metaphase

The chromosomes migrate to the center of the cell, forming a line designated the equatorial plate. The chromosomes are now maximally condensed. The spindle, microtubules of protein that radiate from the centrioles and attach to the centromeres, is formed.

Anaphase

Division occurs in the longitudinal plane of the centromeres. The two new chromatids move to opposite sides of the cell drawn by contraction of the spindles.

Telophase

Division of the cytoplasm begins in the equatorial plane, ending with the formation of two complete cell membranes. The two groups of chromosomes are surrounded by nuclear membranes forming new nuclei. Each strand of DNA serves as a template, and the DNA content of the cell doubles.

Meiosis

Meiosis is the cell division that forms the gametes, each with a haploid number of chromosomes. Meiosis has two purposes: ***reduction*** of the chromosome number and ***recombination*** to transmit genetic information. In meiosis I, homologous chromosomes pair and split apart. Meiosis II is similar to mitosis as the already divided chromosomes split and segregate into new cells.

The First Meiotic Division (Meiosis I)

Prophase

Lepotene: Condensation of the chromosomes.

Zygotene: Pairing of homologous chromosomes (synapsis).

Pachytene: Each pair of chromosomes thickens to form four strands. This is the stage in which ***crossing over*** or ***recombination*** can occur (DNA exchange of homologous segments between two of the four strands). Chiasmata are the places of contact where crossovers occur (and can be visualized). This movement of blocks of DNA is a method for creating genetic diversity. On the other hand, genetic diseases can result from the insertion of sequences during gametogenesis. Transpositional recombination, utilizing enzymes that recognize specific nucleic acid sequences, allows the insertion of a genetic element into any region of a chromosome. This is a method used by viruses (such as the human immunodeficiency virus) to transform host cells.

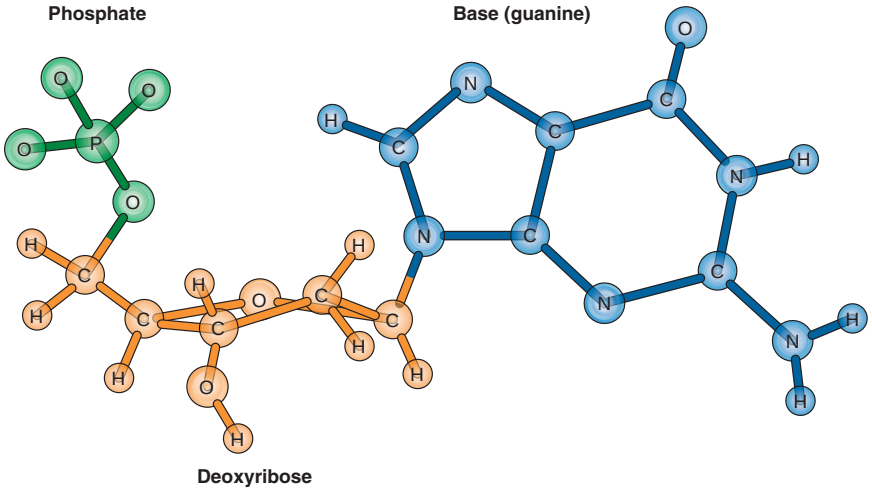
Diplotene: Longitudinal separation of each chromosome.

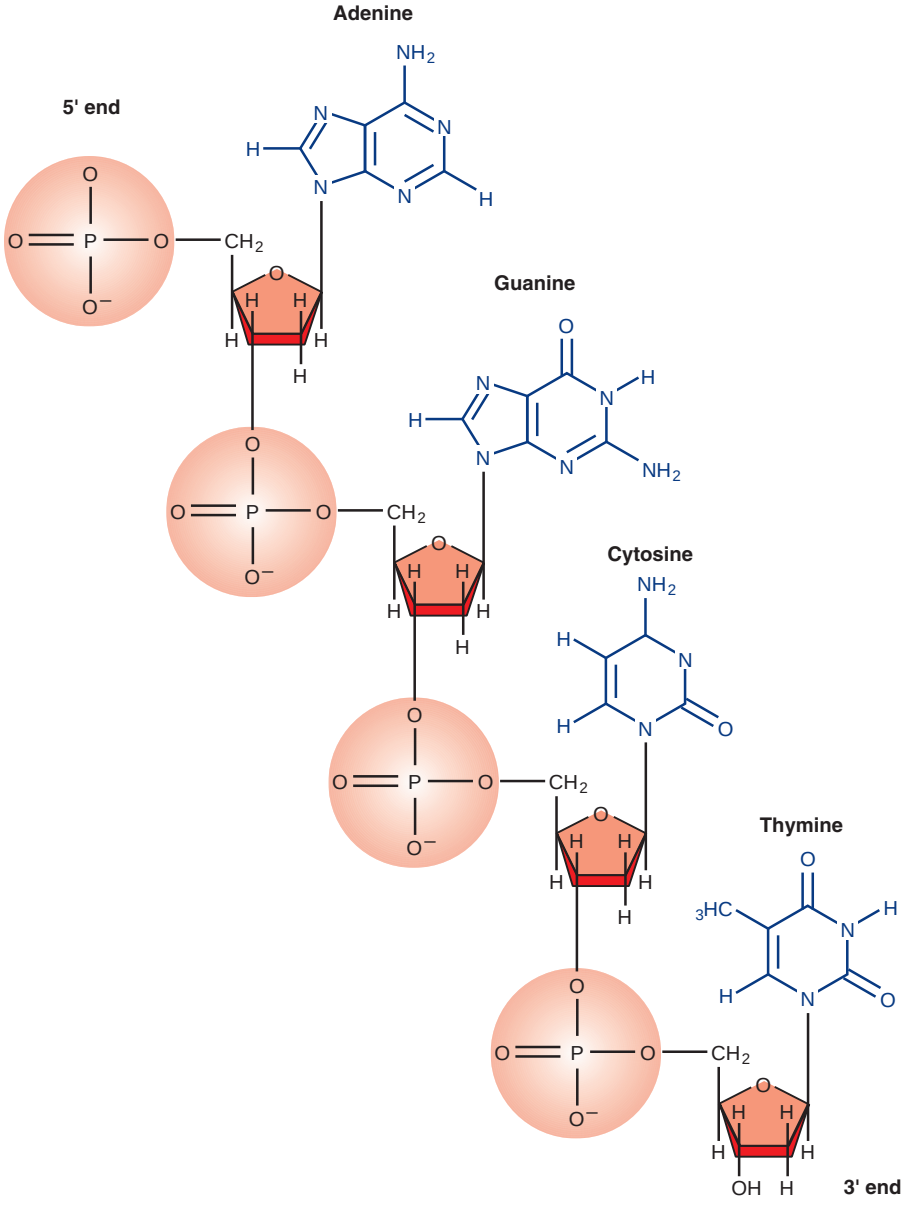
Metaphase, Anaphase, and Telophase of Meiosis I

The nuclear membrane disappears, and the chromosomes move to the center of the cell. One member of each pair goes to each pole, and the cells divide. Meiosis I is often referred to as reduction division because each new product now has the haploid chromosome number. It is during the first meiotic division that mendelian inheritance occurs. Crossovers that occur prior to metaphase result in new combinations of genetic material, both favorable and unfavorable.

The Second Meiotic Division (Meiosis II)

The second division follows the first without DNA replication. In the oocyte, meiosis II occurs after fertilization. The end result is the production of four haploid cells.


Phosphate
Base (guanine)
Deoxyribose


Adenine
5' end
Guanine
Cytosine
Thymine
3' end

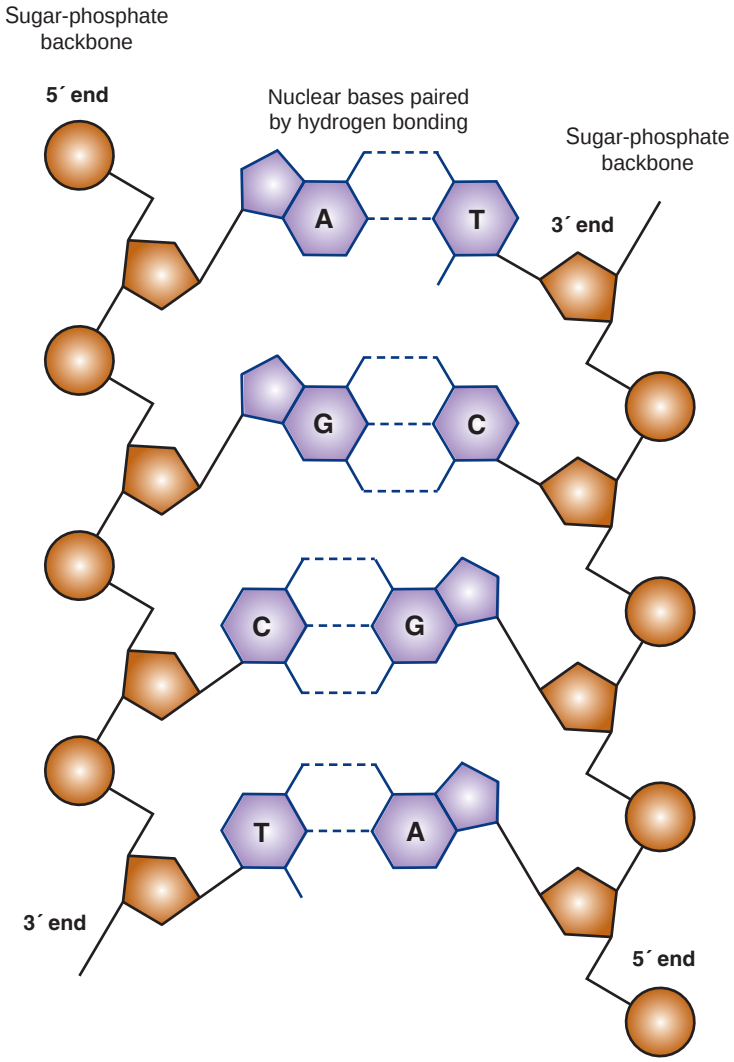
The Structure and Function of DNA

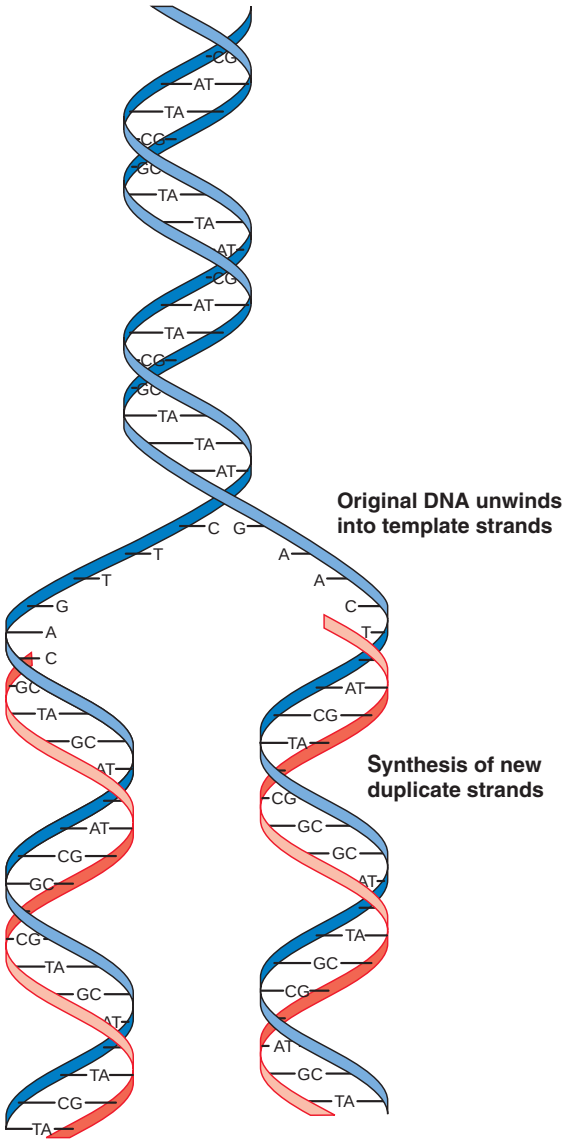
DNA is the material of the gene responsible for coding the genetic message as transmitted through specific proteins. Thus, it is the most important molecule of life and the fundamental mechanism for evolution. Genes are segments of DNA that code for specific proteins, together with flanking and intervening sequences that serve controlling and regulating functions. Each molecule of DNA has a deoxyribose backbone, identical repeating groups of deoxyribose sugar linked through phosphodiester bonds. Each deoxyribose is attached in order (giving individuality and specificity) to one of four nucleic acids, the nuclear bases:

A purine—adenine or guanine.

A pyrimidine—thymine or cytosine.

A nucleotide is the basic building block of DNA. It consists of three major components: the deoxyribose sugar, a phosphate group, and a nucleic acid base. The phosphate-sugar





linkages are asymmetric; the phosphorus is linked to the 5-carbon of one sugar and to the 3-carbon of the following sugar. Thus, one end is the 5′ (5 prime) end and the other the 3′ (3 prime) end. By convention, DNA and its nuclear acid sequences are written from left to right, from the 5′ end to the 3′ end, the direction of the transcription process. The 5′ end leads to the formation of the amino end of the protein; the 3′ end forms the carboxy end of the protein.

DNA consists of two deoxyribose strands twisted around each other clockwise in a double helix, with the nucleic acids on the inside and the nuclear bases paired by hydrogen bonding, adenine with thymine and cytosine with guanine. RNA differs from DNA in that it is single stranded, its sugar moiety is ribose, and it substitutes uracil for thymine. Phoebus Levene, a Russian immigrant to the United States, working at the Rockefeller Institute of Medical Research from 1905 until he died in 1940, identified the components of DNA (discovering and naming the ribose and deoxyribose sugars) and was the first to suggest the nucleotide structure, research that provided the foundation for the later delineation of the DNA's significance.

How can a cell's DNA, which stretched out measures nearly 2m long, fit into a cell? Watson and Crick figured this out when they proposed a tightly coiled two-stranded helix, the double helix. Like the centimeter is a measure of length, the base pair (bp) is the unit of measure for DNA. The base pair is either adenine-guanine or cytosine-thymine, the nucleic acid of one chain paired with the facing nucleic acid of the other chain. A fragment of DNA, therefore, is measured by the number of base pairs, e.g., a 4,800-bp fragment (a 4.8 kb fragment). It is estimated that we have nearly 3 billion bp of DNA, only a small portion of which actually codes out for proteins.

DNA does not exist within the cell as a naked molecule. The nucleotide chains wind about a core of proteins (histones) to form a nucleosome. The nucleosomes become condensed into many bands, the bands that are recognized in karyotype preparations. This condensation is another important mechanism for packing the long DNA structure into a cell. Many other proteins are associated with DNA, important for both structure and function.

The process of DNA replication begins with a separation of the double-stranded DNA helix, initiated at multiple steps by enzyme action. As the original DNA unwinds into template strands, DNA polymerase catalyzes the synthesis of new duplicate strands, which re-form a double helix with each of the original strands (this is called replication). Each daughter molecule, therefore, contains one of the parental strands. It is estimated that the original DNA molecule present in the fertilized zygote must be copied approximately 10^{15} times during the course of a human lifetime. Rapidity and accuracy are essential. By combining precision with error correction systems, errors that affect the function of the gene's protein are surprisingly rare.

The 20 Amino Acids in Proteins

Amino Acid	*Three-Letter Abbreviation*	*Single-Letter Code*
Glycine	Gly	G
Alanine	Ala	A
Valine	Val	V
Isoleucine	Ile	I
Leucine	Leu	L
Serine	Ser	S
Threonine	Thr	T
Proline	Pro	P
Aspartic acid	Asp	D
Glutamic acid	Glu	E
Lysine	Lys	K
Arginine	Arg	R
Asparagine	Asn	N
Glutamine	Gln	Q
Cysteine	Cys	C
Methionine	Met	M
Tryptophan	Trp	W
Phenylalanine	Phe	F
Tyrosine	Tyr	Y
Histidine	His	H

The ***homeobox*** is a DNA sequence, highly conserved throughout evolution, that encodes a series of 60 amino acids, called a homeodomain. Homeodomain protein products function as transcription factors by binding to DNA. The homeobox influences specific tissue functions that are critical for growth and development of the embryo.

The Human Genome

The genome for each species consists of the complete set of DNA sequences on all the chromosomes. There are nearly 3 billion bps in each haploid human genome; in the double-stranded helix DNA, there are 6 billion nucleotides, and there are an estimated 20,000 to 25,000 genes, the smallest functional unit of inherited information. Genes account for only about 2% of human DNA. Although enormously complex at first glance, the entire genetic language is written with only four letters: A, C, G, and T (U in RNA). Furthermore, the language is limited to only three-letter words, codons. Finally, the entire genetic message is fragmented into the 23 pairs of chromosomes. With four nucleotides, reading groups of three, there are 64 possible combinations. Essentially all living organisms use this code. The genome changes only by new combinations derived from parents or by mutation.

The mRNA Genetic Code

First Position (5′ end)	*Second Position* U	C	A	G	*Third Position (3′ end)*
U	Phe	Ser	Tyr	Cys	**U**
	Phe	Ser	Tyr	Cys	**C**
	Leu	Ser	Stop	Stop	**A**
	Leu	Ser	Stop	Trp	**G**
C	Leu	Pro	His	Arg	**U**
	Leu	Pro	His	Arg	**C**
	Leu	Pro	Gln	Arg	**A**
	Leu	Pro	Gln	Arg	**G**
A	Ile	Thr	Asn	Ser	**U**
	Ile	Thr	Asn	Ser	**C**
	Ile	Thr	Lys	Arg	**A**
	Met	Thr	Lys	Arg	**G**
G	Val	Ala	Asp	Gly	**U**
	Val	Ala	Asp	Gly	**C**
	Val	Ala	Glu	Gly	**A**
	Val	Ala	glu	Gly	**G**

Reading across the first row of the table, the codon UUU specifies Phenylalanine, the codon UCU specifies Serine, the codon UAU specifies Tyrosine, and the codon UGU specifies Cysteine. UAA, UAG, and UGA are stop codons.

Gene Structure and Function

The linear arrangement of many genes forms a chromosome. A gene is composed of a segment of DNA containing exons separated by introns, the coding and noncoding codons of nucleotides, respectively. Intron–exon patterns tend to be conserved during evolution. The alpha- and beta-globin genes are believed to have arisen 500 million years ago, with the introns in the same location as they are today.

Exon

The segment of a gene that yields a messenger RNA product that codes for a specific protein.

Intron

The segment of a gene not represented in mature RNA and, therefore, noncoding for protein, but capable of regulatory functions.

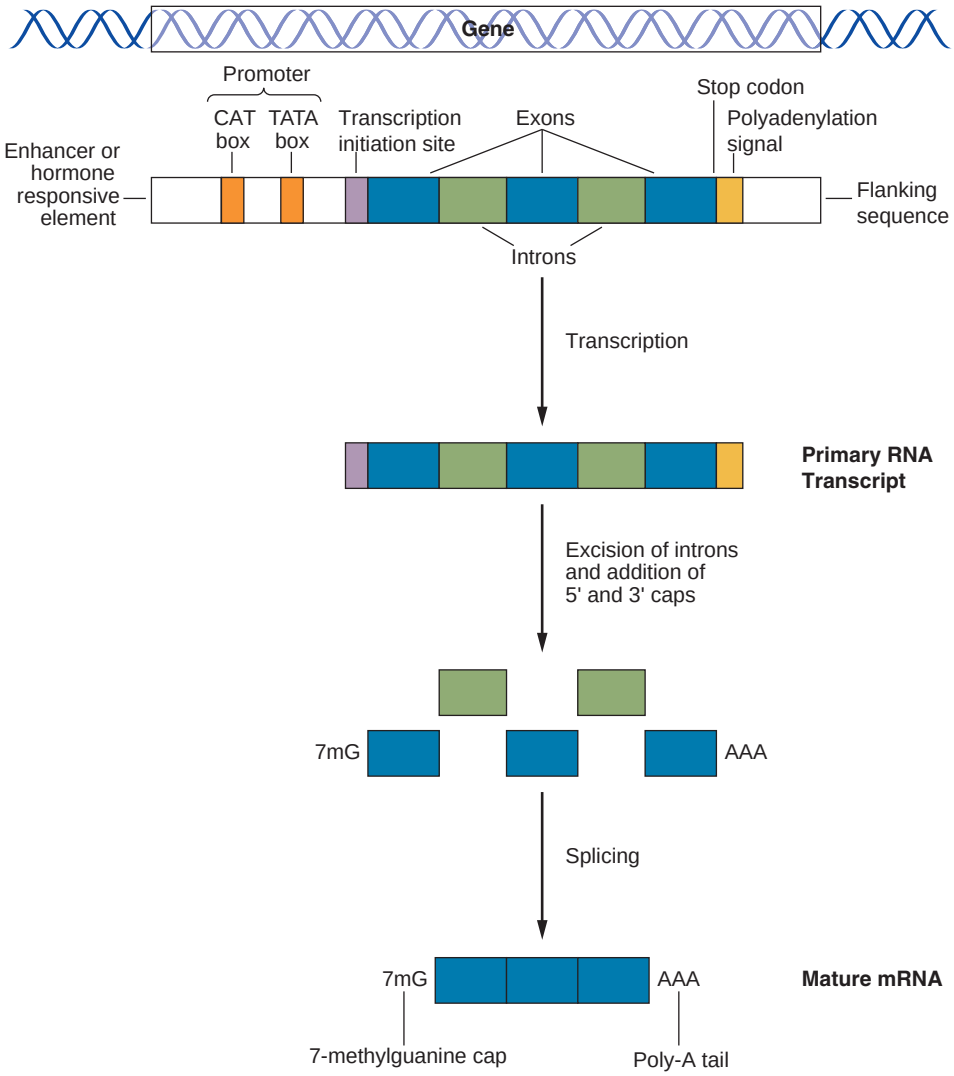
Codon

A sequence of three bases in RNA or DNA that codes for a specific amino acid; the triplet codon.

With some exceptions, it was believed that one gene yields only one protein. The one gene-one protein concept was first proposed in 1909 by Archibald Garrod and experimentally supported in the 1940s by George Beadle and Edward Tatum, American geneticists, by linking single gene mutations to single enzyme deficiencies. However, through a mechanism known as alternative splicing, the 20,000 to 25,000 human genes can produce more than 100,000 proteins. As noted above, the introns are not translated into protein products. Only the DNA sequences in the exons (the part that "exits" the nucleus) are transcribed into messenger RNA and then translated into proteins. However variations (alternative splicing) can produce related proteins.

Genes also include flanking sequences important for gene transcription. The area that will initiate DNA action (e.g., DNA binding to the hormone-receptor complex) is called an ***enhancer*** region. The actual area where transcription begins is the ***promoter*** region. Only a few relatively short nucleotide sequences are promoters, such as the T-A-T-A-A sequence, or TATA box, and the C-C-A-A-T sequence, or CAT box. The promoter sites (the binding sites for RNA polymerase and numerous cofactors) are usually near the start of the coding region of the gene. Enhancer sites are larger than promoter sites and can be located anywhere, even far from the gene, but usually are in the 5′ flanking end. At the 3′ end, a coding sequence is usually present for the polyadenine (poly-A) tail common for most messenger RNA molecules.

The enhancer sites bind proteins (regulatory proteins) that serve as signals to regulate gene expression by either increasing or repressing the binding of RNA polymerase in the pro-



moter region. This is one method of creating unique cellular functions. For example, a hormone target tissue can respond to the hormone because it contains a specific receptor protein that, on binding to the hormone, will bind to a DNA enhancer site. Specific proteins (called transcription factors) bind to enhancer sites and activate transcription. The regulation of gene transcription usually involves DNA sequences in the 5′ flanking upstream region of a gene.

Three codons (UAG, UAA, UGA) are called ***stop codons,*** because they specify a stop to translation of RNA into protein (like a period at the end of a sentence). By contrast, an ***open reading frame*** is a long series of base pairs between two stop codons; therefore, an open reading frame encodes the amino acid sequence of the protein product. Finding and identifying an open reading frame is an important step in analyzing DNA sequences because such a long sequence is usually encountered only in an active gene.

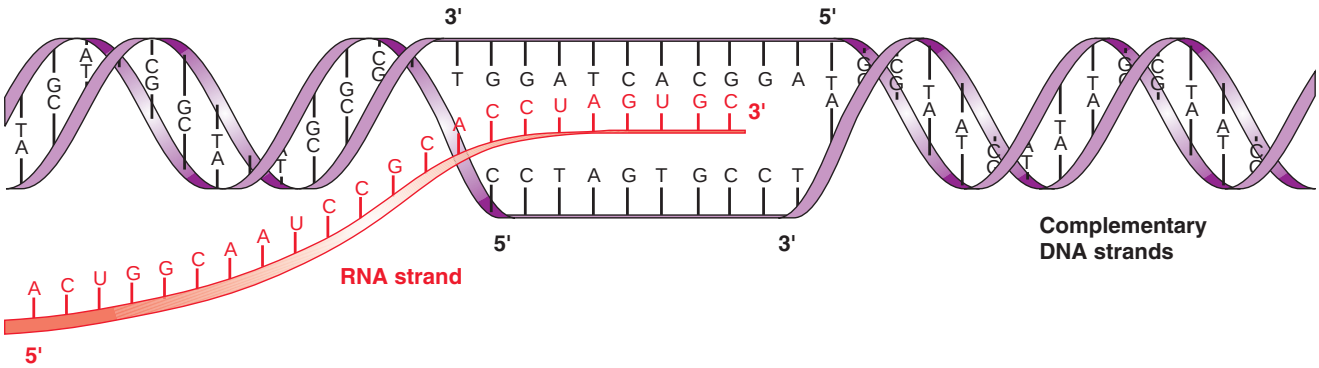
Gene expression is composed of the following steps: transcription of DNA to RNA, RNA processing to produce functional messenger RNA by splicing out introns, translation of messenger RNA on a ribosome to a peptide chain, and protein structural processing to the functional form.

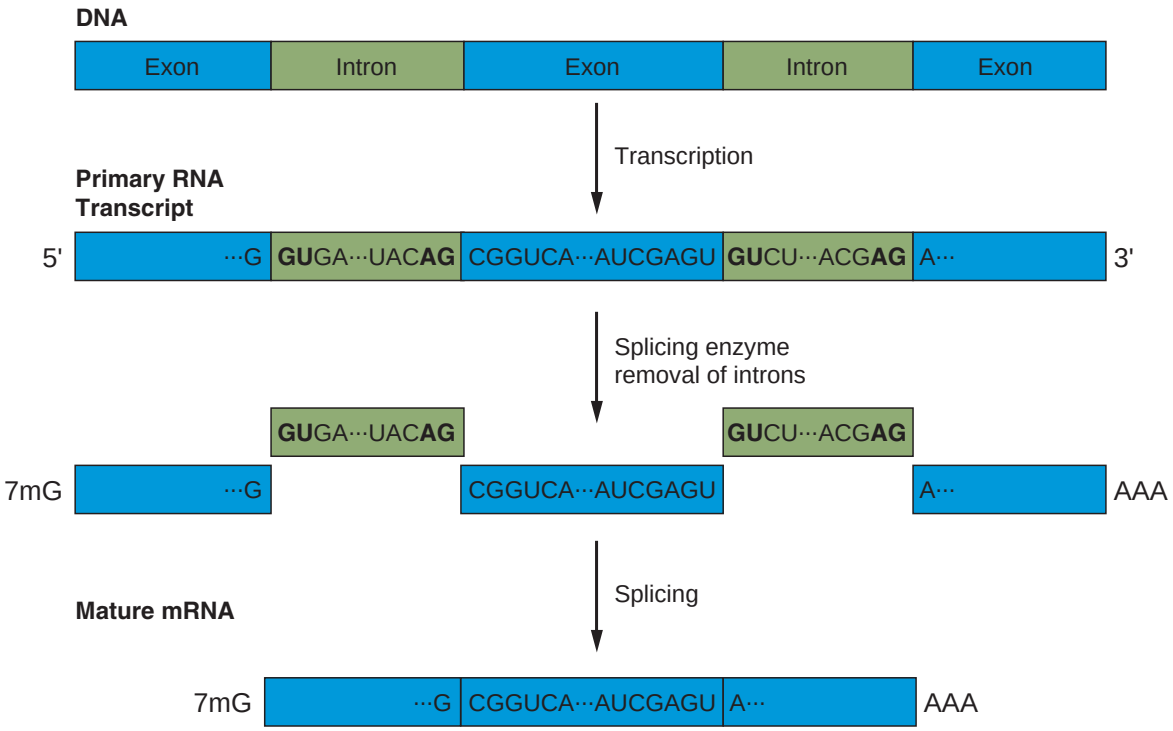
Transcription

Transcription is the synthesis of single-stranded messenger RNA from a gene (double-stranded DNA). The amino acid sequence of the protein is coded in the DNA by codons; a single amino acid is coded by each codon, a triplet of three nucleic acid bases. RNA polymerase constructs the messenger RNA by reading the DNA strand (the "antisense" strand) that is complementary to the RNA; thus, the RNA is an exact copy of the other DNA strand (the "sense" strand), which is also called the complementary strand of the DNA molecule (remember, important differences are that thymine in DNA is replaced by uracil, and ribose replaces deoxyribose in RNA).

Molecular complementarity is both a difficult and a simple concept to grasp. The simple aspect is the concept of one thing being like another. The difficult part is the necessity to understand and visualize that the complementary molecule is not identical to its template, but more like the place where the template goes in, and the complementary molecule goes out. Thus, the strands of the double helix are not identical. Each DNA strand has a complementary structure, in a sense, one positive template and one negative template, each specifying the other. Each strand, therefore, serves as a template for its complementary DNA (in the process of replication) or complementary RNA (in the process of transcription). Thus, messenger RNA is synthesized from the negative template, the "antisense" strand, so that it will have the same structure as the positive template, the "sense" strand. Molecular biologists have to think in three dimensions!

Transcription is initiated at the upstream start site, the 5′ untranslated flanking region where the two strands of the double helix come apart. The process continues downstream, copying one of the strands until a specific codon is reached, which provides a stop message. RNA synthesis continues with the addition of a long chain of adenines, the poly-A tail; this is the 3′ untranslated region that is believed to stabilize RNA by preventing degradation. After transcription from a gene, the RNA moves into the cytoplasm where the intron regions are excised, and the exons are joined together ***(RNA splicing)*** to produce a complete, mature RNA molecule. The start and end of each exon and intron have sequences that, when copied onto the RNA, signal an enzyme to remove the intervening parts. Almost all introns begin with GU and end with AG (GT and AG in the DNA intron). Introns are of varying lengths; a single intron can be longer than the final RNA product. The mature RNA molecule has an addition at one end ("capping," by the addition of a modified nucleotide, 7-methyl guanosine) to protect against ribonucleases (RNases) and at the other end, a polyadenine tail (the poly-A tail) is added (in addition to a stabilizing factor, perhaps a signal to direct exit from the nucleus). Both ends are untranslated in the ribosomes.



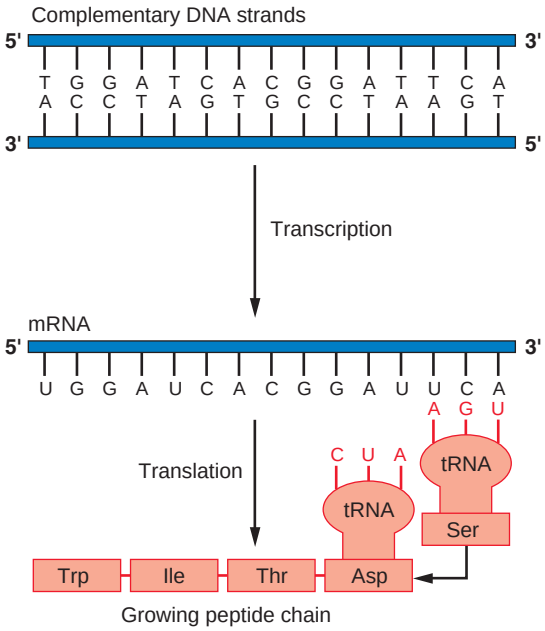


Transcription Factors

Transcription factors are proteins that bind to regulatory elements in DNA (enhancers and promoters) and thus influence gene expression. The steroid hormone receptors are transcription factors. Gene transcription and the formation of messenger RNA can be either stimulated or inhibited through direct interactions with DNA. Transcription factors can further interact with other factors (coactivators and corepressors, also called adapter proteins) to produce cooperative effects. The activity of these proteins can also be affected by phosphorylation triggered by signals from cell-surface receptors (often growth factors). MicroRNAs are non-protein-coding small RNAs of 18 to 25 nucleotides that are transcribed from genes and regulate target gene expression at both the transcriptional and translational levels; over 1,500 mRNAs have been identified. An important concept is to view the final result of hormonal activity and gene expression as a reflection of ***cellular context***, the nature and activity of transcription factors as influenced by specific intracellular adapter proteins. This explains how similar agents (and similar transcription factors, e.g., the estrogen receptor) can have different actions in different tissues.

Translation

The messenger RNA travels from the chromosome on which it was synthesized to a ribosome in the cytoplasm, where it directs the assembly of amino acids into proteins (translation). Each cell has a characteristic but dynamic and always changing **proteome**, the collection of proteins unique for that cell. Amino acids are brought into the process by specific transfer RNA molecules. The specific sequence of three bases at one end of the transfer RNA is complementary to the codon coding for the specific amino acid. Binding of this area to the messenger RNA codon places the specific amino acid at the other end into the proper sequence for the protein. The amino acids are placed one at a time as the transfer RNA molecules read the RNA template, beginning at the amino acid end (the 5′ end) and finishing at the carboxy end (the 3′ end). The process begins at the first AUG triplet and continues until a stop codon (UAA, UAG, or UGA) is reached, whereupon the messenger



RNA falls off the ribosome and degenerates. The specific linear sequence of amino acids is specified by the genetic coding; in turn, this sequence determines the 3-dimensional form of the protein, the folded structure necessary for function. There are only 20,000 to 25,000 protein-encoding human genes, but there are more than 100,000 proteins. Obviously a single gene can produce more than one protein.

The final expression of a gene may not end with the translation process. Further (posttranslational) processing of proteins occurs, such as glycosylation (the gonadotropins) or proteolytic cleavage (conversion of pro-opiomelanocortin to ACTH). These are referred to as "epigenetic" modifications.

The mechanisms that produce proteins from genes are similar throughout the biologic world. This means that important knowledge regarding human function can be gained by studying simple organisms, and microbes can be engineered to produce human proteins.

Mutations

Many genes exist in various forms, called alleles. A change in DNA sequence that results in a change in protein structure or function that is harmful constitutes a mutation. Substitution refers to a change in a single nucleic acid base. A substitution in a codon can result in the incorporation of the wrong amino acid into a protein, leading to a change or loss in function. Insertion or deletion of amino acids into the final protein product can result from improper RNA splicing. Because of great redundancy in the genetic code (many triplet codons code out for the same amino acid, and there are only 20 amino acids), not all substitutions produce an effect. A clinical example of a single base substitution (point mutation) is the sickle mutation, in which thymine is substituted for adenine in the beta-globin gene. If homologous regions of DNA are misaligned, unequal crossover can occur, resulting in deletions and insertions (additions). A "nonsense mutation" refers to a single base substitution that produces a stop codon, truncating the protein product. Deletions and insertions can involve single bases, up to entire exons, or genes or several genes. Recombination or exchange of genetic material usually occurs in meiosis. Even a change at the junction of a coding and noncoding region can lead to abnormal messenger RNA.

Chromosomal Abnormalities

Numerical Abnormalities

Numerical abnormalities usually are due to nondisjunction, a failure of separation at anaphase, either during mitotic division or during meiosis. ***Aneuploidy*** is a chromosome number that is not an exact multiple of the haploid number, e.g., monosomy (45,X Turner syndrome) or trisomy (trisomy 13 Patau syndrome, trisomy 18 Edwards syndrome, trisomy 21 Down syndrome, 47,XXY Klinefelter syndrome). ***Mosaicism*** indicates one or more cell lines with a different karyotype, usually arising from nondisjunction during early mitosis (failure of two paired chromosomes to separate). ***Polyploidy,*** multiples of the haploid number of chromosomes, is a significant cause of spontaneous miscarriage.

Structural Abnormalities

Structural abnormalities are usually due to chromosomal breaks induced by radiation, drugs, or viruses. The resulting abnormality depends on the rearrangement of the broken pieces. Thus, in a ***translocation*** there is interchange of material between two or more nonhomologous chromosomes. A balanced translocation is associated with neither gain nor loss of genetic material, and such an individual is a translocation carrier.

Single-Gene Defects

Single-gene defects are due to mutations in specific genes. These mutations are transmitted according to mendelian inheritance: autosomal dominant, autosomal recessive, X-linked recessive, and rarely X-linked dominant. In addition, single-gene disorders can be transmitted by mitochondrial gene inheritance, maternal or paternal imprinting, disomy (inheriting both pairs of a chromosome from one parent), and excessive repeats (a phenomenon in which more than the usual repeats of 3 bps occurs).

Autosomal Dominance

Transmission is not linked to the sex of an individual, and homozygous and heterozygous children are affected (only one allele needs to be abnormal). With two heterozygous parents, each child has a 75% risk of being affected. With one heterozygous parent, each child has a 50% risk of being affected. The effect is subject to variable expression. Examples of autosomal-dominant conditions include Huntington disease, neurofibromatosis, and Marfan syndrome. The effect of an abnormal dominant gene is influenced by **penetrance**, the degree the dominant gene is expressed. Complete penetrance, as opposed to incomplete penetrance, means that the gene is always expressed and always produces a recognizable phenotype.

Autosomal Recessive

These conditions are phenotypically expressed only in homozygotes (both alleles must be abnormal). With heterozygote parents, each child has a 25% risk of being affected and a 50% chance of being a carrier. Examples of autosomal-recessive conditions are cystic fibrosis, sickle cell disease, and adrenal hyperplasia due to a deficiency in 21-hydroxylase.

X-Linked Recessive Inheritance

An affected father can transmit the condition only to daughters. Only homozygous females are affected when the condition is recessive. Each son of a female carrier has a 50% chance of being affected and each daughter a 50% chance of being a carrier. Red–green color blindness and hemophilia A are examples.

Genomic Imprinting

Genomic imprinting indicates persisting influences on genome function by the male and female parental contributions. For example, placental development is controlled mostly by paternally derived genes. Thus, a hydatidiform mole has a normal karyotype, but all of its chromosomes are derived from the father. Placental structures are absent in ovarian teratomas, tumors that contain only maternally derived chromosomes. Experiments in nature and animal experiments indicate that the maternal contribution to the genome is more important for embryonic development. In certain autosomal-recessive conditions, the expression, severity, and age of onset are influenced by the gender of the parent providing the mutant gene or chromosome. Angelman syndrome, the Prader-Willi syndrome, and the Beckwith-Wiedemann syndrome are examples of human disorders associated with genomic imprinting.

Epigenetics is the study of changes in gene expression not directly caused by DNA sequence changes. Mechanisms include modifications of DNA without changing the DNA sequence, such as methylation of DNA to turn off gene expression, altering the histone proteins that are responsible for the overall structure of chromatin (affecting transmission during cell replication), the production of new forms of RNA, and alterations in cellular proteins that influence gene expression. Each of these mechanisms can be transmitted to offspring and be responsible for imprinting. This is another method by which species can adapt, another evolutionary pathway.

Techniques of Molecular Biology

An enzyme that breaks the phosphodiester bonds and cuts the DNA molecule into fragments is an endonuclease; a ***restriction enzyme*** (restriction endonuclease) cuts only at sites with specific nucleic acid sequences. Restriction enzymes were discovered in bacteria in which they form a defense mechanism to cut (and thus inactivate) any foreign DNA (from invading viruses) introduced into the bacterial cell. As part of this protection mechanism, bacteria also contain methylases that methylate recognition sites in native DNA, directing the action of the restriction enzyme to the nonmethylated foreign DNA. Different bacteria have different restriction enzymes with specific action sites. Restriction enzymes are available that cut DNA into pieces (restriction fragments), ranging from many small fragments to a few large pieces, depending on the number of nucleotides in the recognition sequence. The enzymes are named for the organism and strain from which they are derived. The combination of restriction fragments, the merger of two cut pieces of DNA, yields ***recombinant DNA***.

DNA polymerase is an enzyme that brings single nucleotides into a DNA molecule. A DNA polymerase can form DNA only in the presence of a DNA template; the synthesized DNA will be complementary to the template. RNA polymerase can make RNA also only in the presence of a DNA template.

A deoxyribonuclease ***(DNAase)*** can remove nucleotides. By combining DNAase treatment with DNA polymerase action, radiolabeled nucleotides can be introduced into a DNA molecule, producing a ***DNA probe.*** A DNA probe can be compared with the antibody used in immunoassays. The antibody is specific and recognizes the hormone against which it is formed. The DNA probe specifically detects a sequence of DNA.

Reverse transcriptase is DNA polymerase that is RNA dependent. It is called reverse transcriptase because the flow of information is from RNA to DNA, the reverse of the usual direction of flow. This enzyme permits the copying of essentially any RNA molecule into single-stranded DNA; such DNA is called ***complementary DNA*** because it is a mirror image of the messenger RNA. Complementary DNA probes are limited by their reading only the exons (remember that introns are excised from RNA), and thus these probes read only large areas.

DNA and RNA are charged molecules and, therefore, will migrate in an electrical field. Fragments can be analyzed by gel (agarose or polyacrylamide) electrophoresis, the largest fragments migrating the slowest. By convention, the gels are read from top to bottom, with the smallest fragments at the bottom.

Southern Blot Analysis

DNA is first denatured to separate the two strands, digested by restriction enzymes to produce smaller fragments that are loaded into an electrophoresis gel. The Southern blot method, named after its inventor E.M. Southern, determines the fragment sizes. The fragments are separated by electrophoresis; the smaller the fragment the faster the migration. The electrophoresis gel is placed over a thick piece of filter paper with its ends dipped in a high-salt solution. A special membrane (nitrocellulose or nylon) is placed over the gel, and over this is placed a stack of paper towels compressed by a weight. The salt solution rises by wick action into the filter paper; it moves by capillary action through the gel, carrying the DNA with it. The DNA is carried to the membrane to which it binds. The salt solution keeps moving and is absorbed by the paper towels. The nitrocellulose or nylon membrane thus creates a replica of the original electrophoresis pattern. The DNA is fixed to the membrane either by high-temperature baking or by ultraviolet light. Specific labeled DNA or RNA probes then can be introduced for hybridization. ***Hybridization*** means that a specific probe anneals to its complementary sequence. The fragments with this sequence are then identified by autoradiography. Fluorescent probes can be utilized that can be detected after activation by a laser beam, allowing qualitative and quantitative assessment by a computer. Southern blotting is still a necessary procedure, but it is often replaced by the faster and more sensitive technique of polymerase chain reaction.

Northern blotting detects RNA sequences, Northern because RNA is the opposite image of DNA. Extracted RNA is separated by electrophoresis and transferred to a membrane as in Southern blotting for hybridization with probes (complementary DNA). Northern blotting would be used, for example, to determine whether hormone stimulation of a specific protein in a tissue is mediated by messenger RNA, i.e., gene expression. Northern blotting can also be replaced by polymerase chain reaction using a reverse transcriptase enzyme.

Electrophoresis to separate and quantitate proteins is called ***Western blotting***, and antibodies are used for the hybridization identification process. Like Northern blotting, Western blotting tests gene expression, not just the presence of a gene. The terms *Northern* and *Western* represent intentional witticisms (a rare event in science) in response to Southern blotting. Hybridization without electrophoresis by placing a drop of the cell extract directly on filter paper is called ***dot or slot blotting.***

Hybridization

When two complementary strands of DNA reassociate, the process is called hybridization. Hybridization allows a specific area of the DNA to be studied using a radiolabeled DNA probe that is specific (a complementary sequence). The membrane produced after Southern blotting is first treated to block nonspecific binding sites. The membrane is then treated (hybridized) with the labeled probe. The location of bound probe is then identified by autoradiography (for radiolabeled probes) or by colorimetric methods. The sequence of the probe, therefore, determines the sequence at the site of binding. Whenever two products are complementary, hybridization occurs. Thus, complementary DNA can be hybridized to its template messenger RNA.

In situ hybridization is the technique in which labeled DNA or RNA probes are placed directly on a slide of tissue or cells. A piece of cloned DNA labeled with a fluorescent marker can be utilized; the method is referred to as ***FISH*** (fluorescence in situ hybridization). The region corresponding to the cloned DNA lights up under fluorescent illumination unless the region has been deleted from one of the chromosomes. Several microdeletion syndromes have been discovered with the FISH technique, e.g., the Prader-Willi syndrome.

Microarray Chip Technology

This method detects gene expression, testing thousands of genes simultaneously. Complementary cloned DNA is hybridized with labeled complementary DNA prepared from the tissue to be studied. If that tissue expresses a gene, the labeled signal is easily observed. The production of specific gene chips allows this technique to search for mutations and polymorphisms. The microarray gene chip is a physical array of DNA that can simultaneously identify thousands of unique mRNA products in heterogeneous samples. Chips have been developed that include the entire human genome and can assess more than 500,000 single-nucleotide polymorphisms (SNP-chips) in one sample. Thus, a DNA chip can contain all of the 20,000 to 25,000 protein-encoding human genes. This highly automated process can indicate, although not quantitate, the differential expression of genes in response to various stimuli or conditions. For example, gene expression can be compared before and after hormonal treatment. The method can also detect gene deletions, duplications, and alterations in a technique called comparative genomic hybridization, using a reference DNA sample for hybridization instead of complementary DNA.

Polymerase Chain Reaction (PCR)

The polymerase chain reaction (PCR) is a technique to amplify (relatively quickly) small fragments or areas of DNA into quantities large enough to be analyzed with electrophoresis and blotting methods. This technique produces enormous numbers of copies of a specific DNA sequence without resorting to cloning. The sequence to be amplified must be known. Specific markers (synthesized short sequences of DNA corresponding to each end of the sequence to be studied) are selected that will delineate the region of DNA to be amplified. These flanking sequences are called primers. The DNA sample, the primers, and an excess of free single nucleotides are incubated with a DNA polymerase.

The first step involves separating DNA into its single strands by denaturation with heat (92°C); then the temperature is lowered (40°C), causing the primers to stick (anneal) to

their complementary regions on the DNA. The temperature is raised to 62°C, and DNA polymerase then synthesizes a new strand beginning and ending at the primers, forming a new double-stranded DNA. Repeating the cycle many times (by alternating the reaction temperature) amplifies the amount of DNA available for study (more than 1 million-fold); the increase occurs exponentially. Thus, DNA can be analyzed from a single cell, and genes can be visualized by blotting without labeled probes, all in a thermocycler, the machine that carries out the entire procedure.

Because the process requires alternate heating and cooling, a DNA polymerase resistant to heat is an advantage in that periodic replenishment is not necessary. This problem was solved with the discovery of DNA polymerase (*Taq* polymerase) in a microorganism (*Thermophilus aquaticus*) that is a thermophile (a hot-water microbe) found in an out-of-the-way Yellowstone National Park hot spring called Mushroom Pool. This high-temperature polymerase allows automation of the process.

Reverse transcriptase PCR is used to amplify small amounts of a specific RNA. The starting template is an RNA molecule, which is converted to its complementary by reverse transcriptase, an enzyme originally discovered in retroviruses. The new DNA is then processed by the standard PCR procedure.

The technique of polymerase chain reaction has made possible the study of incredibly small amounts of DNA from virtually any tissue or body fluid. Down syndrome can be diagnosed from the few fetal cells obtained from maternal blood. Especially impressive is the amplification of small amounts of degraded DNA from extinct and rare species preserved in museums. DNA from fossils has been amplified and sequenced (e.g., from an 18-million-year-old magnolia plant). The method also makes it possible to identify a gene by its expression of messenger RNA. The RNA is the template for amplification by first converting it into complementary DNA. Polymerase chain reaction is used to detect microbes, providing results within hours even in the presence of antimicrobial drugs. This method can detect bacteria that cannot be isolated by culture techniques.

Real-time PCR is now routinely used in clinical laboratories. The method is faster than conventional PCR, an hour or less, using fluorescent probes. The light emitted by the probes is immediately displayed on a graph, allowing almost an instant quantification of the measured DNA.

Chromatin Immunoprecipitation (ChIP)

Chromatin immunoprecipitation localizes a protein to its binding site on DNA. The technique uses an antibody that is specific to the protein of interest. The DNA that is isolated by the antibody-protein-DNA complex can then be studied by PCR. The specific region of the genome can be determined by using a DNA microarray, a method known as ChIP-chip, ChIP-on-chip.

Cloning DNA

Cloning means isolating a gene and making copies of it. A DNA library is a collection of DNA molecules derived from cloning methods. A complementary DNA library is the DNA counterpart of all of the messenger RNA isolated from a particular cell or tissue. Complementary DNAs have been produced for more than 70% of the human and mouse genes. By starting with messenger RNA, the search for the gene of interest can be focused (instead of

searching the entire genome). Such a library is made using reverse transcriptase. The DNA molecules then can be inserted into an appropriate vector (described below) and replicate molecules can be produced. Using probes, the complementary DNA can be selected that matches the gene of interest (keeping in mind that complementary DNA only includes the exons of a gene). Cloning the DNA simply means the production of many identical copies of a specified fragment of DNA. Cloning can also be performed using the polymerase chain reaction. As indicated above, complementary DNA cloning focuses on the DNA counterpart of messenger RNA; genomic DNA cloning, using a restriction endonuclease, copies the DNA in genes. Cloning also can be used to make multiple copies of probes or unknown DNA fragments.

If the amino acid sequence is unknown, one can work backward. Knowing the specific protein product, antibodies can be produced against the protein. When complementary DNA is inserted into certain vectors, production of the protein can be identified with the antibodies; thus, the DNA fragment will be isolated.

A vector is an entity in which foreign DNA can be inserted. The vector plus the foreign DNA are inserted into a host cell; the host cell produces both the vector and the foreign DNA. The first vectors were bacterial plasmids, circular DNA molecules (minichromosomes) that coexist in the cytoplasm with the bacterial chromosomal DNA. Most noteworthy, they carry genes that code for antibiotic resistance. This enables the bacterial cells that contain the plasmid to be selected by appropriate antibiotic treatment. Plasmid vectors have also been developed that allow selection by color. A variety of bacterial strains have been developed, each for a specific use.

Disruption of the plasmid DNA with restriction enzymes, followed by incorporation of foreign DNA with DNA ligase, produces plasmid DNA molecules (recombinant DNA containing the foreign DNA) that can be replicated. Plasmid vectors can incorporate foreign DNA fragments up to 10 kb in size. Digestion of recovered plasmids with restriction enzymes releases the desired DNA fragment, which can then be recovered by electrophoresis.

Other vectors exist. Bacteriophages (or phages) are viruses that infect and replicate within bacteria. Phage vectors can incorporate larger DNA inserts, up to 20 kb. Cloning DNA with phage vectors follows the same basic design as with plasmids. Larger fragments of foreign DNA are cloned with cosmid vectors, artificially produced combinations of phage and plasmid vectors. Very large fragments, up to 1,000 kb, can be cloned using yeast artificial chromosomes. This method can work with whole genes.

Basic Steps for Cloning

1. Choose a DNA source: either genomic DNA or complementary DNA.
2. Fragment the DNA by restriction endonucleases.
3. Insert the fragments into vectors.
4. Introduce the vectors into bacteria.
5. Collect the cloned DNA propagated in the bacteria to form a library.
6. Screen the library for the desired sequence. Possible methods include the use of complementary nucleotide probes for fragments that hybridize or the detection of a specific protein produced with antibodies to the protein or by assaying the function of the protein.

Transgenic Animals

Transgenic animals are produced by inserting cloned genes or complementary DNA into bacteria, yeast, worms, fish, frogs, and mice. In mammals, the cloned gene or cDNA can be injected into a pronucleus of a fertilized egg where it will be integrated into the chromosomes. The eggs are transferred to the uterus of a recipient mother who will deliver progeny with the foreign DNA. Altered biologic behaviors and functions are then attributed to the foreign DNA. Transgenic animals provide animal models for inherited diseases and malignant tumors and provide a means to carry out experiments in gene therapy. The transfer of new or altered genes is an important method to study gene function. Transgenic plants can even be developed to produce new pharmaceuticals, and the introduction of genes conferring resistance to insects may solve the problem of insecticide contamination.

Knockout Animal Models

Animal models for the function of a gene employ the method of "knocking out" a specific gene. In a straightforward, but important demonstration, it can be determined whether a specific gene and its protein are essential for life, or for a function (such as pregnancy). This is usually accomplished by having foreign DNA introduced into an animal by a vector to replace endogenous DNA. A similar result can be achieved by interfering with specific messenger RNAs, although complete loss of a specific gene expression is not achieved because new cell division reactivates gene expression.

The Identification of Genes

To clone an entire gene whose protein product is known, a complementary DNA library is produced. The specific DNA fragment is identified by linking it to the protein. Once identified, the total gene can be screened using the identified complementary DNA, indicating the introns and exons. Another strategy is to synthesize an oligonucleotide probe, basing the sequence on the known amino acid sequence in the protein product (from the peptide sequence, the DNA sequence that codes for that protein can be predicted). This method can be used with just a relatively small piece of the peptide. As more and more genes are cloned, the codon frequency for particular amino acids is established. Complementary DNA can be cloned without producing a library by using the polymerase chain reaction to amplify complementary DNA made from messenger RNA by reverse transcriptase. Overlapping sequences of the genome can be cloned, using a piece of DNA from each succeeding product, to work across a chromosome in a systematic manner to search for a gene; this is called ***chromosome walking***.

The entire sequencing process can be performed by a computer, even searching for open reading fames. Once the sequence of a DNA fragment has been identified, the computer can utilize DNA and protein databases to predict sequence, recognition site, protein translation, and homology with known sequences. The scientist can then select restriction fragment sizes for cloning. Once a gene has been analyzed, it has to be compared with the gene in the disease state. If a mutation is of large size, this can be detected by Southern blotting. Minor alterations require comparisons in DNA sequences, which is possible by using polymerase chain amplification to produce specific gene sequences in amounts readily studied.

A gene can be localized to a specific chromosome when its protein product is unknown by studies involving chromosome rearrangements and linkage analysis. Specific diseases

are associated with karyotypic changes. Thus, the specific chromosome can be targeted for gene localization. Linkage analysis utilizes restriction fragment length polymorphisms.

DNA Polymorphism

Southern blotting reveals specific patterns of bands that reflect the varying lengths of the DNA fragments produced by restriction enzyme action. A specific site can exhibit a mutation by having a different pattern (a different length of the DNA fragment on Southern blotting due to sequence difference). These differences in DNA sequences are called **restriction fragment length polymorphisms**. Single-nucleotide polymorphisms, referred to as SNPs), or simply polymorphisms, are usually a benign and common variations. The human genome contains about 4.5 million polymorphisms, and more than 3 million have been identified. Although 99.9% of DNA sequences in humans are the same, SNPs can serve as genetic markers for a medically important gene. A polymorphism is governed by the mendelian regulations of inheritance, and if by chance a polymorphism is identified in a patient with a specific disease, the transmission of the disease can be studied. The polymorphism, which is linked to the disease by chance, can be used to study the inheritance of a disease when the genes are unknown. The polymorphism is like a flag that marks specific areas of chromosomes. This method of study requires DNA from at least one affected individual and a sufficient number of family members to trace the polymorphism, either by Southern blotting (for long sequences) or with the polymerase chain reaction (best for short sequences). Correlation of genetic markers (polymorphisms) and phenotype also employs ***haplotypes*** (similar to a polymorphism but a longer sequence of nucleotides, even a set of several polymorphisms). Efforts using genetic technologies to link SNPS and haplotypes to clinical problems and inherited traits are known as ***genome-wide association studies***.

Minisatellites are a form of polymorphisms. Genes concentrate in random areas along chromosomes separated by long sequences of noncoding DNA. Minisatellites are noncoding areas of DNA that repeat in variable numbers, so-called variable number ***tandem repeat sequences***, distributed throughout the length of every human chromosome. These areas can be followed by DNA probes, providing a "fingerprint" for specific individuals. This uniqueness is applied in forensic medicine. Microsatellites, as the name implies, are smaller than minisatellites. Usually, microsatellites consist of repetitions of only two nucleotides. DNA polymorphisms now number in the thousands and allow genetic mapping with great precision. The mapping of SNP variations in humans (the human haplotype map) is underway.

The Human Genome Project

All of the human genes are collectively known as the genome. Begun in 1990, the goal of the international Human Genome Project was to sequence the human genome, a goal that was reached in draft form in 2001 and 99% of the actual sequence in 2003, more than 2 years ahead of schedule, 50 years after Watson and Crick's publication. The number of protein-encoding genes (20,000 to 25,000) is smaller than previous estimates. Less than 2% of the human genome with nearly 3 billion nucleotides codes for proteins; therefore, the remainder is a rich resource for evolutionary historians and a target for inducing genetic changes. The number of genes in specific chromosomes varies; the chromosomes that are affected by trisomy, chromosomes 13, 18, and 21, have the fewest genes.

Information on the human genome project is available at:

http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml

Fortunately the Internet emerged in time to accommodate the enormous data generated by studying the human genome. The National Center for Biotechnology Information (NCBI) maintains a website that provides databases as well as the Online Mendelian Inheritance in Man site developed by Johns Hopkins University as a guide to human genes and inherited disorders:

http://www.ncbi.nlm.nig.gov/projects/genome/guide/human

The *GeneTests* web site is also recommended, a genetic information source funded by NIH. This site includes reviews of specific diseases, an international directory of genetic testing laboratories and prenatal diagnosis clinics, plus educational materials for teaching:

http://www.genetests.org

The next wave of information will come from projects like the ENCODE project and the 1000 Genomes project. ENCODE is a research consortium organized by the National Human Genome Research Institute of NIH to develop an encyclopedia of the functional elements in the human genome (http://www.genome.gov/ENCODE). The 1000 Genomes project will sequence the genomes of 1,000 people from around the world to map DNA variations with a fine resolution (http://www.1000genones.org/page.php).

The human genome map can serve as a foundation for locating sites for heritable factors that predispose for diseases, mutations that cause disease, and for integrating the genetic sequencing with biologic functions. Soon, we will be able to have in our possession a personal record of our individual complete genetic blueprint.

The chromosomal locations of genes responsible for hormone production have been mapped. From the cloned DNA sequences, the amino acid sequences can be predicted. Every protein product of a gene represents a potential diagnostic or therapeutic target. And of course, inherited disorders will be subject to characterization and, eventually, gene therapy. However, even after a gene has been identified and genetically mapped, its full characterization remains a difficult and time-consuming task. Full understanding of disorders that involve the interactions of multiple genes will be even more complicated.

Genomics and Proteomics

Genomics refers to the entire process involved in the Human Genome Project, the complete description of the genetic sequence and the study of gene expression, especially using the microarray technique with gene chips. Genomics, however, will not tell the whole story. The protein products of gene expression are altered in the translation process and also by posttranslational modifications such as glycosylation, methylation, and phosphorylation. The complete story, therefore, requires proteomics, the study of the biologically functional end products, the proteins of a cell or a tissue. Both genomics and proteomics are required for understanding physiology, diagnosing diseases, and for the design of new drugs.

Protein identification requires separation of the proteins by electrophoresis, digestion of large proteins into smaller proteins, measurement of the amino acid content by mass spectrophotometry, and specific identification of proteins by comparison with computerized

databases. The protein mass profiles of normal and abnormal cells can then be compared. Post-translational alterations can be detected by comparing specific proteins to known proteins. Protein chips are now available, comparable to DNA chips, that can be used to study protein changes.

Metabolomics is the study of small molecules within cells, the metabolic products that characterize cells. Crude cellular extracts are subjected to chromatography, spectrometry or nuclear magnetic resonance measurements to create characteristic patterns, chemical fingerprints, associated with normal or abnormal cells.

The amount of data generated by genomics, proteomics, and metabolomics is enormous, impossible to manage without computer databases. **Bioinformatics** is the science dedicated to the development of computer techniques to best organize and analyze growing wealth of information, creating efficient search and retrieval systems. This allows the data to be available to laboratories scattered throughout the world, the study of the information with sophisticated mathematical analysis, and worldwide sharing of data.

Clinical Applications

The challenge for modern medicine is to make clinical sense of the massive collection of data generated by genomics and proteomics. Understanding the function of genes and proteins will undoubtedly be an accelerating feature of human progress. Knowing an individual's genetic susceptibilities would allow medical diagnosis and treatment to be personalized to the ultimate degree. Nevertheless, this is a complicated new era and not always straight-forward. For example, a genetic susceptibility may carry with it only a small percentage increase in risk for a disease; at what point is intervention merited in terms of benefit and cost, not only for an individual, but for society? In addition, gene activity can be turned on and off in response to environmental factors. From the bench to the bedside may be slower than anticipated, but there is no doubt that we are on the verge of translating genomic theory into clinical practice.

The molecular diagnosis of genetic disorders requires only a small sample of DNA, obtainable from any cells that are nucleated, such as white blood cells or epithelial cells. Polymerase chain reaction carried out by automatic machinery allows speedy DNA diagnosis with material amplified from a single cell. This is an important advantage in prenatal genetic analysis and in preimplantation sexing and diagnosis. PCR makes it possible to perform DNA diagnosis from a single cell removed from embryos fertilized in vitro.

Molecular diagnosis is limited by the prevalence of heterogenic genetic changes. In other words, many disorders involve different mutations in different people. In contrast, some (like sickle cell disease) always involve the same change. With cystic fibrosis, 70% of patients (of northern European ancestry) have the same 3-base deletion, whereas the remaining 30% have an extremely heterogenic collection of mutations. Molecular diagnosis is further challenged by the need not only to find a subtle change in a gene but also to distinguish important changes from benign variations (polymorphisms). Ingenious PCR-based methods have been developed for rapid screening and detection of mutations. The significance of detected mutations requires segregation of the mutation with an identified disease in a family.

At least one type of growth hormone deficiency is inherited in an autosomal-recessive pattern. The cloning of growth hormone DNA complementary to its messenger RNA permitted localization of the growth hormone gene. The growth hormone gene is in a cluster

that also includes the gene for human placental lactogen. This cluster of genes contains multiple units of DNA that are homologous and prone to recombination, which leads to deletion on one chromosome and duplication on another. Similar mechanisms operate for other protein products governed by genes in clusters, such as the globins. The first report of a clinical diagnosis based on DNA hybridization was that of the prenatal diagnosis of alpha-thalassemia in the Department of Obstetrics and Gynecology at the University of California in San Francisco.

The commercial production of proteins from cloned genes inserted into bacteria is rapidly increasing. The production of insulin (the first) and growth hormone are good examples. Glycosylation does not occur in bacterial systems, and therefore the commercial production of recombinant glycoproteins requires a mammalian cell line for the process. This has been accomplished, and recombinant gonadotropins are now available. The gene for gonadotropin-releasing hormone on the short arm of chromosome 8 has been isolated and cloned. Molecular technology was important in the characterization of inhibin, the ovarian follicular hormone that inhibits follicle-stimulating hormone (FSH) secretion. The inhibin gene has been sequenced and found to be homologous to the gene for antimüllerian hormone. The alpha-subunit common to gonadotropins, thyroid-stimulating hormone, and human chorionic gonadotropin (hCG) has been traced to a gene that has been isolated, sequenced, and localized on chromosome 6.

The human genome contains many genes with the potential to cause cancer. Other genes have the ability to block malignant growth. Cancer is a genetic disease in that tumors can be said to be clonal; all the cells are genetically related (although subsequent genetic alterations can yield cytogenetically different cells in tumors). ***Oncogenes***, discovered in tumor viruses, are genes that transform cells from normal to abnormal growth by encoding proteins that are involved in signal transduction, specifically the transmission of growth-regulatory messages. There are many oncogenes and many different pathways of action, all of which result in a proliferative state. The mutations that activate these genes lead either to protein activity independent of incoming signals or to activity at the wrong place at the wrong time. The bottom line is the turning on (by an altered oncogene) of persistent growth. But this single change is probably not sufficient to produce a tumor. Tumors usually involve alterations in many oncogenes, as well as in anti-oncogenes.

Anti-oncogenes in normal cells are growth-suppressing genes that must be inactivated before tumors can grow. Thus, an inherited susceptibility for cancer can also result from a mutation in tumor suppressor genes. Although activation of an oncogene is a dominant effect, tumor suppressor mutations are recessive and can be carried and transmitted, but are not active as long as pairing occurs with a normal anti-oncogene.

Cancer, therefore, is a genetic disease, but regulation of normal growth involves a complex system that takes a long time to overcome. This involves alterations in many genes, eventually yielding a tumor with a heterogeneity that determines sensitivity to various treatment modalities. During this time period, the technology of recombinant DNA may be able to achieve diagnosis sufficiently early to yield cures. Knowing the specific oncogene involved in a given tumor also offers therapeutic possibilities. For example, an antimetabolite can be attached to an antibody for an oncogene, targeting the cancer cells. Monoclonal antibodies have been developed that affect the protein products of specific oncogenes.

Molecular biology is changing both diagnosis and therapy. Viral and bacterial DNA can be identified. The automated PCR process can produce electrophoretic patterns that can be read automatically. With this technique, a single human papillomavirus DNA molecule can be detected among 10,000 or more human cells. Hundreds of genetic tests are now in clinical use.

Pharmacogenomics is the study of genetic variations in pharmacodynamics that can explain different responses to the same dose of a drug. For example, different polymorphisms involved in enzyme synthesis can affect metabolism and clearance. Genetic testing offers the potential to predict drug pharmacokinetics in an individual, which would allow the selection of an appropriate dose to minimize side effects and maximize efficacy.

Faulty endogenous protein production can be corrected by replacing the problematic mechanism. There are two strategies: foreign cells that produce the missing protein could be introduced, or the faulty gene could be replaced (or more accurately, adding a complementary-corrected DNA). Thus, recessive single-gene disorders are potentially amenable to gene therapy, as are acquired diseases such as cancer and infections. Gene therapy is broadly defined as the enlistment of the patient's own cellular machinery to produce a therapeutic agent. A gene delivered to a cell can either replace a defective or missing gene or produce a protein with a desired effect. However, this is a field in its infancy.

Specific guidelines for gene therapy have been developed requiring several levels of review. One class of human therapy is the use of retroviral vectors to transfer marker genes into cultured human cells that are returned to patients of origin. For example, this allows tracking of tumor-infiltrating lymphocytes, donor hepatocytes, or killer T cells that are specific for the human immunodeficiency virus. These transferred genes can also be crafted to provide a function in patients with single-gene inherited disorders. Another class of therapy involves the transfer of genes that encode for factors that destroy tumor cells, such as tumor necrosis factor or interleukin. Retroviral vectors are viruses that have been altered so that no viral proteins can be made by cells infected by the vectors. Thus, viral replication and spread are prevented, but gene transfer into replicating cells can take place. Other transfer methods being developed include the use of adenovirus vectors and specifically targeted plasmid DNA.

A Final Thought

Molecular progress is inexorable. The future will see preventive medicine by prediction. By knowing an individual's genetic constitution, appropriate and intensive screening can be directed to predisposed conditions. This kind of knowledge will also require social and political considerations. It is not far-fetched to envision marriages and pregnancies avoided because of a bad match of genetic predispositions. Society is developing guidelines regarding the use of this information: by individuals, by employers, by health organizations, and by the government. Scientific progress must be matched by public and professional education to appropriately manage this knowledge.

References

Collins FS, Green ED, Guttmacher AE, Guyer MS, on behalf of the U.S. National Human Genome Research Institute, A vision for the future of genomics research. A blueprint for the genomic era, *Nature* 422:835, 2003.

Croce CM, Oncogenes and cancer, *New Engl J Med* 358:502, 2008.

Espy MG, Real-time PCR in clinical microbiology: applications for routine laboratory testing, *Clin Microbiol Rev* 19:165, 2006.

Hardy B-J, Séguin B, Goodsaid F, Jimenez-Sanchez G, Singer PA, Daar A, The next steps for genomic medicine: challenges and opportunities for the developing world, *Nat Rev Genet Suppl* 1:S23, October, 2008.

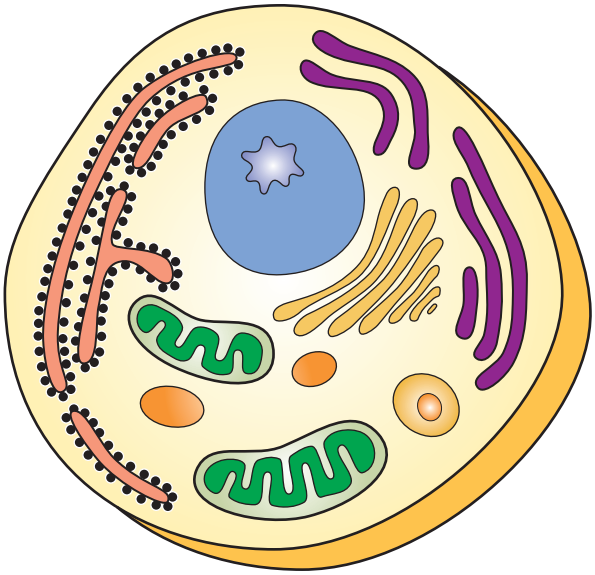
International HapMap Consortium, A haplotype map of the human genome, *Nature* 437:1299, 2005.

International Human Genome Sequencing Consortium, Finishing the euchromatic sequence of the human genome, *Nature* 431:931, 2004.

Pearson TA, Manolio TA, How to interpret a genome-wide association study, *JAMA* 299:1335, 2008.

2

Hormone Biosynthesis, Metabolism, and Mechanism of Action



The classical definition of a hormone is a substance that is produced in a special tissue, where it is released into the bloodstream, and travels to distant responsive cells in which the hormone exerts its characteristic effects. What was once thought of as a simple voyage is now appreciated as an odyssey that becomes more complex as new facets of the journey are unraveled in research laboratories throughout the world. Indeed, the notion that hormones are products only of special tissues has been challenged.

Complex hormones and hormone receptors have been discovered in primitive, unicellular organisms, suggesting that endocrine glands are a late development of evolution. The widespread capability of cells to make hormones explains the puzzling discoveries of hormones in strange places, such as gastrointestinal hormones in the brain, reproductive hormones in intestinal secretions, and the ability of cancers to unexpectedly make hormones. Hormones and neurotransmitters were and are a means of communication. Only when animals evolved into complex organisms did special glands develop to produce hormones that could be used in a more sophisticated fashion. Furthermore, hormones must have appeared even before plants and animals diverged because there are many plant substances similar to hormones and hormone receptors. Therefore, it is not surprising that, because every cell contains the genes necessary for hormonal expression, cancer cells, because of their dedifferentiation, can uncover gene expression and, in inappropriate locations and at inappropriate times, make hormones.

Hormones, therefore, are substances that provide a means of communication and should now be viewed broadly as chemical regulatory and signaling agents. The classic endocrine hormones travel through the bloodstream to distant sites, but cellular communication is also necessary at local sites. Paracrine, autocrine, and intracrine depict a more immediate form of communication.

Paracrine Communication

Intercellular communication involving the local diffusion of regulating substances from a cell to nearby (contiguous) cells.

Autocrine Communication

Intracellular communication whereby a single cell produces regulating substances that in turn act upon receptors on or within the same cell.

Intracrine Communication

This form of intracellular communication occurs when unsecreted substances bind to intracellular receptors; in other words, a regulating factor acts within the cell that secretes it.

Let us follow an estradiol molecule throughout its career and in so doing gain an overview of how hormones are formed, how hormones work, and how hormones are metabolized. Estradiol begins its lifespan with its synthesis in a cell specially suited for this task. For this biosynthesis to take place, the proper enzyme capability must be present along with the proper precursors. In the adult human female the principal sources of estradiol are the granulosa cells of the developing follicle and the corpus luteum. These cells possess the ability to turn on steroidogenesis in response to specific stimuli. The stimulating agents are the gonadotropins, follicle-stimulating hormone (FSH) and luteinizing hormone (LH). The initial step in the process that will give rise to estradiol is the transmission of the message from the stimulating agents to the steroid-producing mechanisms within the cells.

Messages that stimulate steroidogenesis must be transmitted through the cell membrane. This is necessary because gonadotropins, being large glycopeptides, do not ordinarily enter cells but must communicate with the cell by joining with specific receptors on the cell membrane. In so doing they activate a sequence of communication. A considerable amount of investigation has been devoted to determining the methods by which this communication takes place. E. W. Sutherland, Jr., received the Nobel Prize in 1971 for proposing the concept of a second messenger.

Gonadotropin, the first messenger, activates an enzyme in the cell membrane called adenylate cyclase. This enzyme transmits the message by catalyzing the production of a second messenger within the cell, cyclic adenosine 3′,5′-monophosphate (cyclic AMP). The message passes from gonadotropin to cyclic AMP, much like a baton in a relay race.

Cyclic AMP, the second messenger, initiates the process of steroidogenesis, leading to the synthesis and secretion of the hormone estradiol. This notion of message transmission has grown more and more complex with the appreciation of physiologic concepts, such as the heterogeneity of peptide hormones, the up- and down-regulation of cell membrane receptors, the regulation of adenylate cyclase activity, and the important roles for autocrine and paracrine regulating factors.

Secretion of estradiol into the bloodstream directly follows its synthesis. Once in the bloodstream, estradiol exists in two forms, bound and free. A majority of the hormone is bound to protein carriers, albumin and sex steroid hormone-binding globulin. The biologic activity of a hormone is limited by binding in the blood, thereby avoiding extreme or sudden reactions. In addition, binding prevents unduly rapid metabolism, allowing the hormone to exist for the length of time necessary to ensure a biologic effect. This reservoir-like mechanism avoids peaks and valleys in hormone levels and allows a more steady state of hormone action.

The biologic and metabolic effects of a hormone are determined by a cell's ability to receive and retain the hormone. The estradiol that is not bound to a protein, but floats freely in the bloodstream, readily enters cells by rapid diffusion. For estradiol to produce its effect, however, it must be grasped by a receptor within the cell. The job of the receptor is to aid in the transmission of the hormone's message to nuclear gene transcription. The result is production of messenger RNA leading to protein synthesis and a cellular response characteristic of the hormone.

Once estradiol has accomplished its mission, it is eventually released back into the bloodstream. It is possible that estradiol can perform its duty several times before being cleared from the circulation by metabolism. On the other hand, many molecules are metabolized without ever having the chance to produce an effect. Unlike estradiol, other hormones, such as testosterone, can either work directly or are metabolized and altered within the cell in which an effect is produced. In the latter case, a metabolite is released into the bloodstream as an inactive compound. Clearance of steroids from the blood varies according to the structure of the molecules.

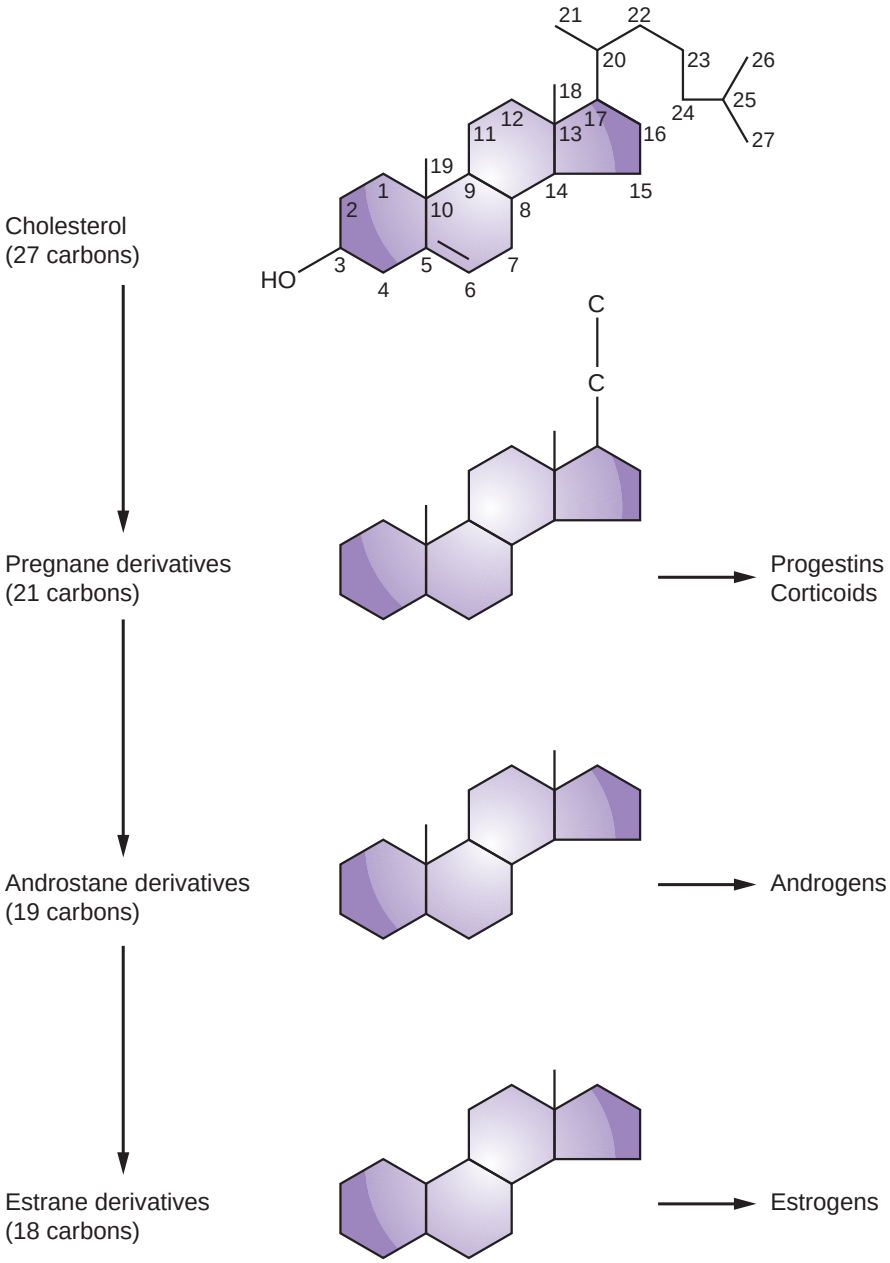
Cells that are capable of clearing estradiol from the circulation accomplish this by biochemical means (conversion to estrone and estriol, moderately effective and very weak estrogens, respectively) and conjugation to products that are water soluble and excreted in the urine and bile (sulfo and glucuro conjugates).

Thus, a steroid hormone has a varied career packed into a short lifetime. In this chapter we will review the important segments of this lifespan in greater detail.

Steroid Hormone Nomenclature

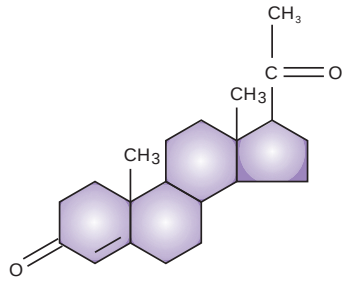
All steroid hormones are of basically similar structure with relatively minor chemical differences leading to striking alterations in biochemical activity. The basic structure is the perhydrocyclopentanephenanthrene molecule. It is composed of three 6-carbon rings and one 5-carbon ring. One 6-carbon ring is benzene, two of the 6-carbon rings are naphthalene, and three 6-carbon rings are phenanthrene; add a cyclopentane (5-carbon ring), and you have the perhydrocyclopentanephenanthrene structure of the steroid nucleus.

The sex steroids are divided into three main groups according to the number of carbon atoms they possess. The 21-carbon series includes the corticoids as well as the progestins, and the basic structure is the ***pregnane*** nucleus. The 19-carbon series includes all the androgens and is based on the ***androstane*** nucleus, whereas the estrogens are 18-carbon steroids based on the ***estrane*** nucleus.

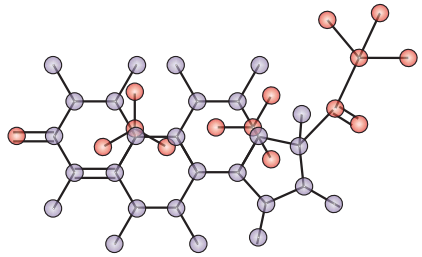


There are six centers of asymmetry on the basic ring structure, and there are 64 possible isomers. Almost all naturally occurring and active steroids are nearly flat, and substituents below and above the plane of the ring are designated alpha (α) (dotted line) and beta (β) (solid line), respectively. Changes in the position of only one substituent can lead to inactive isomers. For example, 17-epitestosterone is considerably weaker than testosterone; the only difference being a hydroxyl group in the α position at C-17 rather than in the β position.

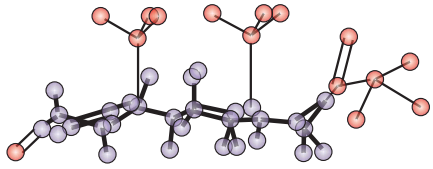
Progesterone



Top View

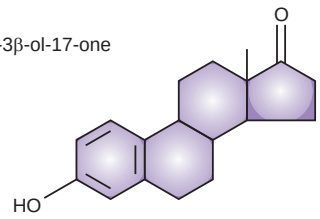


Side View

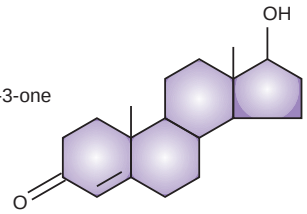


The convention of naming steroids uses the number of carbon atoms to designate the basic name (e.g., pregnane, androstane, or estrane). The basic name is preceded by numbers that indicate the position of double bonds, and the name is altered as follows to indicate one, two, or three double bonds: -ene, -diene, and -triene. Following the basic name, hydroxyl groups are indicated by the number of the carbon attachment, and one, two, or

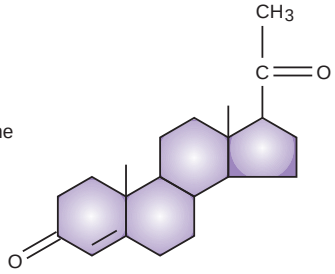
Estrone
1,3,5(10)-Estratriene-3β-ol-17-one



Testosterone
4-Androstene-17β-ol-3-one



Progesterone
4-Pregnene-3,20-dione



three hydroxyl groups are designated -ol, -diol, or -triol. Ketone groups are listed last with numbers of carbon attachments, and one, two, or three groups designated -one, -dione, or -trione. Special designations include dehydro, elimination of two hydrogens; deoxy, elimination of oxygen; nor, elimination of carbon; delta or Δ, location of double bond.

Lipoproteins and Cholesterol

Cholesterol is the basic building block in steroidogenesis. All steroid-producing organs except the placenta can synthesize cholesterol from acetate. Progestins, androgens, and estrogens, therefore, can be synthesized in situ in the various ovarian tissue compartments from the 2-carbon acetate molecule via cholesterol as the common steroid precursor. However, in situ synthesis cannot meet the demand, and, therefore, the major resource is blood cholesterol that enters the ovarian cells and can be inserted into the biosynthetic pathway or stored in esterified form for later use. The cellular entry of cholesterol is mediated via a cell membrane receptor for low-density lipoprotein (LDL), the bloodstream carrier for cholesterol.

Lipoproteins are large molecules that facilitate the transport of nonpolar fats in a polar solvent, the blood plasma. There are five major categories of lipoproteins according to their charge and density (flotation during ultracentrifugation). They are derived from each other in the following cascade of decreasing size and increasing density.

Chylomicrons

Large, cholesterol (10%)- and triglyceride (90%)-carrying particles formed in the intestine after a fatty meal.

Very Low-Density Lipoproteins (VLDL)

Also carry cholesterol, but mostly triglyceride; more dense than chylomicrons.

Intermediate-Density Lipoproteins (IDL)

Formed (for a transient existence) with the removal of some of the triglyceride from the interior of VLDL particles.

Low-Density Lipoproteins (LDL)

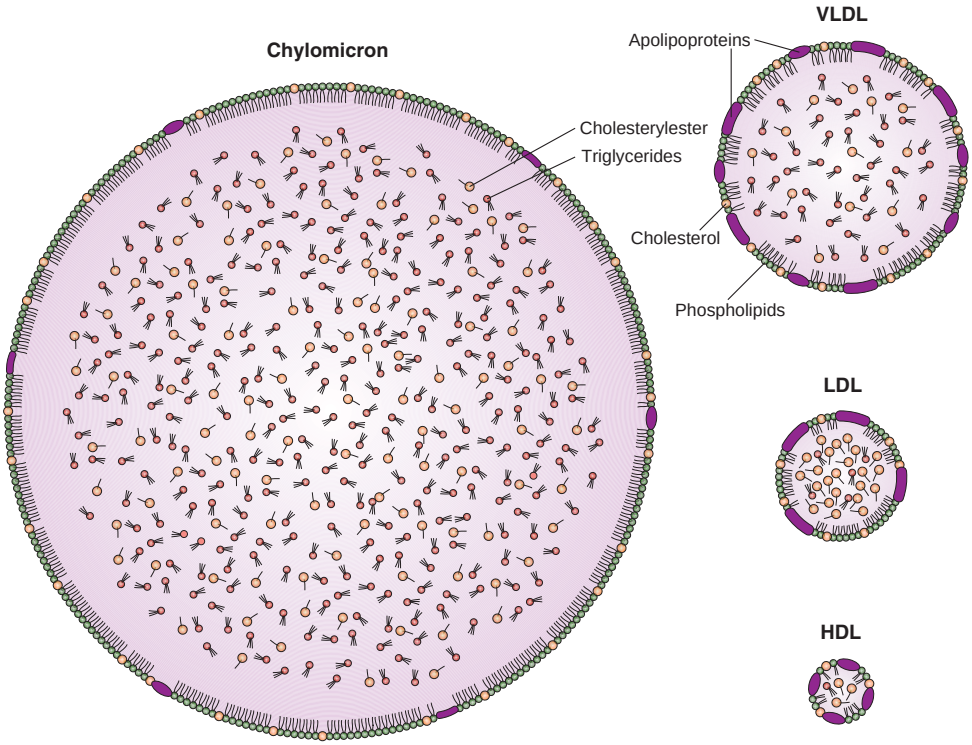
The end products of VLDL catabolism, formed after further removal of triglyceride leaving approximately 50% cholesterol; the major carriers (two-thirds) of cholesterol in the plasma and thus a strong relationship exists between elevated LDL levels and cardiovascular disease.

High-Density Lipoproteins (HDL)

The smallest and most dense of the lipoproteins with the highest protein and phospholipid content; HDL levels are inversely associated with atherosclerosis (high levels are protective). HDL can be further separated into a lighter fraction (HDL_2) and a denser fraction (HDL_3).

The lipoproteins contain four ingredients: (1) cholesterol in two forms: free cholesterol on the surface of the spherical lipoprotein molecule, and esterified cholesterol in the molecule's interior; (2) triglycerides in the interior of the sphere; (3) phospholipid; and (4) protein in electrically charged substances on the surface of the sphere and responsible for miscibility with plasma and water. The surface proteins, called ***apoproteins,*** constitute the sites that bind to the lipoprotein receptor molecules on the cell surfaces. The principal surface protein of LDL is apoprotein B, and apoprotein A-1 is the principal apoprotein of HDL.

Lipids for peripheral tissues are provided by the secretion of VLDL by the liver. Triglycerides are liberated from VLDL by lipoprotein lipase located in the capillary endothelial cells as well as a lipase enzyme located on the endothelial cells in liver sinusoids. In this process, the surface components (free cholesterol, phospholipids, and apoproteins) are transferred to HDL. Finally, the VLDL is converted to LDL, which plays the important role of transporting cholesterol to cells throughout the body. The hepatic lipase enzyme is sensitive to sex steroid changes: suppression by estrogen and stimulation by androgens.



LDL is removed from the blood by cellular receptors that recognize one of the surface apoproteins. The lipoprotein bound to the cell membrane receptor is internalized and degradated. Intracellular levels of cholesterol are partly regulated by the up- and down-regulation of cell membrane LDL receptors. When these LDL receptors are saturated or deficient, LDL is taken up by "scavenger" cells (most likely derived from macrophages) in other tissues, notably the arterial intima. Thus, these cells can become the nidus for atherosclerotic plaques.

HDL is secreted by the liver and intestine or is a product of the degradation of VLDL. Cholesteryl ester molecules move to form a core in a small spherical particle, the HDL_3 particle. These particles accept additional free cholesterol, perhaps mediated by receptors that recognize apoprotein A-1. With uptake of cholesterol, the particle size increases to form HDL_2, the fraction that reflects changes in diet and hormones. HDL_3 levels remain relatively stable.

The protein moieties of the lipoprotein particles are strongly related to the risk of cardiovascular disease, and genetic abnormalities in their synthesis or structure can result in atherogenic conditions. The lipoproteins are a major reason for the disparity in atherosclerosis risk between men and women. Throughout adulthood, the blood HDL-cholesterol level is about 10 mg/dL higher in women, and this difference continues through the postmenopausal years. Total and LDL-cholesterol levels are lower in premenopausal women than in men, but after menopause they rise rapidly.

The protective nature of HDL is due to its ability to pick up free cholesterol from cells or other circulating lipoproteins. This lipid-rich HDL is known as HDL_3, which is then converted to the larger, less dense particle, HDL_2. Thus, HDL converts lipid-rich scavenger cells (macrophages residing in arterial walls) back to their low-lipid state and carries the excess cholesterol to sites (mainly liver) where it can be metabolized. Another method by which HDL removes cholesterol from the body focuses on the uptake of free cholesterol from cell membranes. The free cholesterol is esterified and moves to the core of the HDL particle. Thus, HDL can remove cholesterol by delivering cholesterol to sites for utilization (steroid-producing cells) or metabolism and excretion (liver).

For good cardiovascular health, the blood concentration of cholesterol must be kept low, and its escape from the bloodstream must be prevented. The problem of cholesterol transport is solved by esterifying the cholesterol and packaging the ester within the cores of plasma lipoproteins. The delivery of cholesterol to cells is in turn solved by lipoprotein receptors. After binding the lipoprotein with its package of esterified cholesterol, the complex is delivered into the cell by receptor-mediated endocytosis (discussed later in this chapter), in which the lysosomes liberate cholesterol for use by the cell.

Major protection against atherosclerosis depends on the high affinity of the receptor for LDL and the ability of the receptor to recycle multiple times, thus allowing large amounts of cholesterol to be delivered while maintaining a healthy low blood level of LDL. Cells can control their uptake of cholesterol by increasing or decreasing the number of LDL receptors according to the intracellular cholesterol levels. Thus, a high-cholesterol diet influences the liver to reduce the number of LDL receptors on its cells, causing an elevated blood level of LDL. Statins protect against atherosclerosis by reducing cholesterol biosynthesis, increasing LDL receptors in the liver, and lowering circulating levels of LDL-cholesterol.

Steroidogenesis

The overall steroid biosynthesis pathway shown in the figure is based primarily on the pioneering work of Kenneth J. Ryan and his coworkers.[1,2] These pathways follow a fundamental pattern displayed by all steroid-producing endocrine organs. As a result, it should be no surprise that the normal human ovary produces all three classes of sex steroids: estrogens, progestins, and androgens. The importance of ovarian androgens is appreciated, not only as obligate precursors to estrogens, but also as clinically important secretory products. The ovary differs from the testis in its fundamental complement of critical enzymes and, hence, its distribution of secretory products. The ovary is distinguished from the adrenal gland in that it is deficient in 21-hydroxylase and 11β-hydroxylase reactions. Glucocorticoids and mineralocorticoids, therefore, are not produced in normal ovarian tissue.

During steroidogenesis, the number of carbon atoms in cholesterol or any other steroid molecule can be reduced but never increased. The following reactions can take place:

1. Cleavage of a side chain (desmolase reaction).
2. Conversion of hydroxyl groups into ketones or ketones into hydroxyl groups (dehydrogenase reactions).
3. Addition of OH group (hydroxylation reaction).
4. Creation of double bonds (removal of hydrogen).
5. Addition of hydrogen to reduce double bonds (saturation).

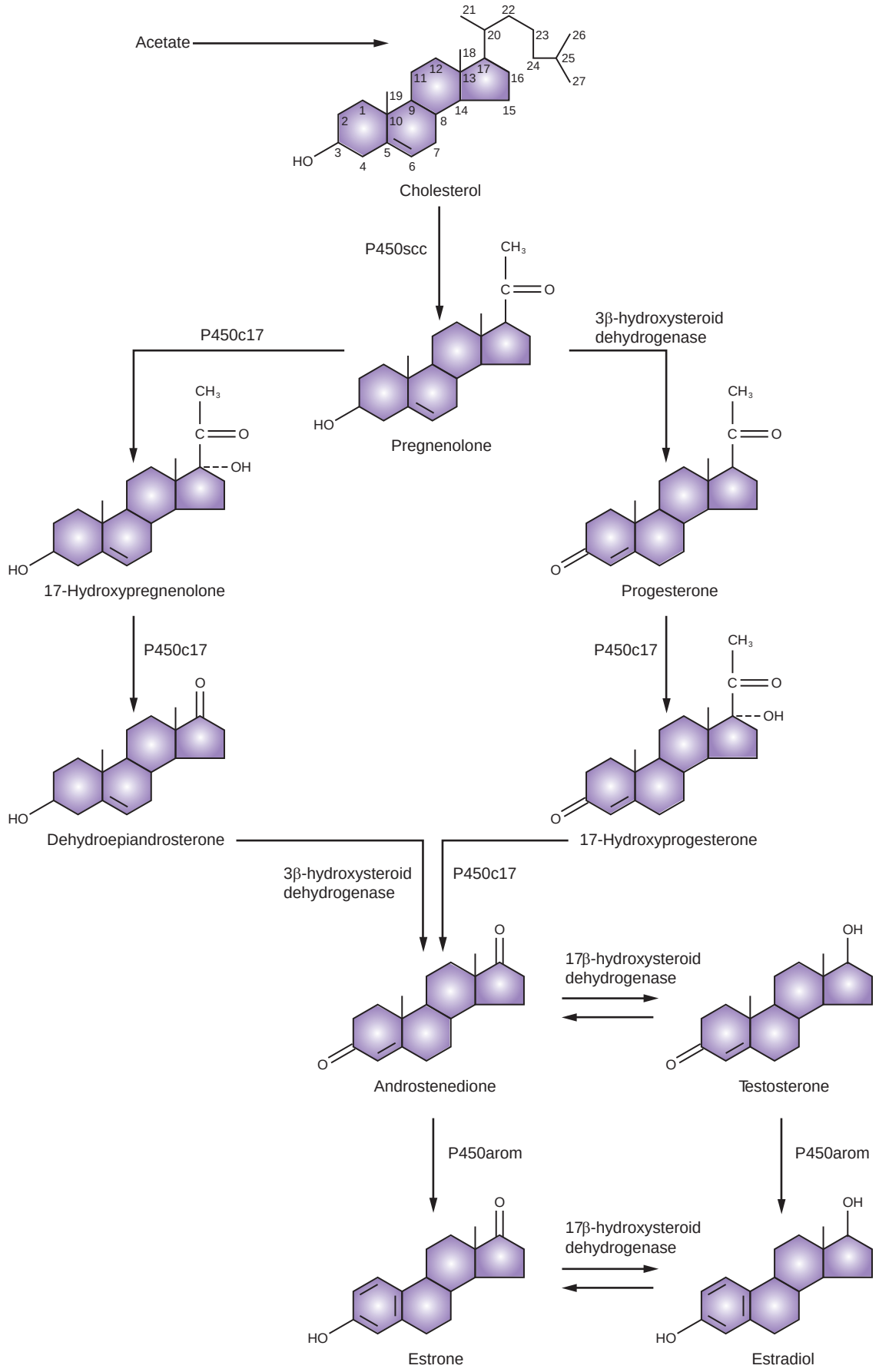
The traditional view of steroidogenesis was that each step was mediated by many enzymes, with differences from tissue to tissue. A fundamental simplicity to the system emerged when the responsible complementary DNAs and genes were cloned.[3–5]

Steroidogenic enzymes are either dehydrogenases or members of the cytochrome P450 group of oxidases. Cytochrome P450 is a generic term for a family of oxidative enzymes, termed 450 because of a pigment (450) absorbance shift when reduced. P450 enzymes can metabolize many substrates; e.g., in the liver, P450 enzymes metabolize toxins and environmental pollutants. The human genome contains genes for 57 cytochrome P450 enzymes, 7 in mitochondria and 50 in the endoplasmic reticulum (the major site for metabolic clearance). The following distinct P450 enzymes are identified with steroidogenesis: P450scc is the cholesterol side chain cleavage enzyme; P450c11 mediates 11-hydroxylase, 18-hydroxylase, and 19-methyloxidase; P450c17 mediates 17-hydroxylase and 17,20-lyase; P450c21 mediates 21-hydroxylase; and P450arom mediates aromatization of androgens to estrogens. Marked differences in the exon–intron organization of the P450 genes are compatible with an ancient origin; thus, the superfamily of P450 genes diverged more than 1.5 billion years ago.

Enzyme	*Cellular Location*	*Reactions*
P450scc	Mitochondria	Cholesterol side chain cleavage
P450c11	Mitochondria	11-hydroxylase 18-hydroxylase 19-methyloxidase
P450c17	Endoplasmic reticulum	17-hydroxylase, 17,20-lyase
P450c21	Endoplasmic reticulum	21-hydroxylase
P450arom	Endoplasmic reticulum	Aromatase

The structural knowledge of the P450 enzymes that has been derived from amino acid and nucleotide sequencing studies demonstrated that all the steps between cholesterol and pregnenolone were mediated by a single protein, P450scc, bound to the inner mitochondrial membrane. Cloning data indicate the presence of a single, unique *P450scc* gene on chromosome 15. These experiments indicated that multiple steps did not require multiple enzymes. Differing activity in different tissues may reflect posttranslational modifications. In addition, these genes contain tissue-specific promoter sequences, which is another reason that regulatory mechanisms can differ in different tissues (e.g., placenta and ovary). P450scc mutations are very rare, producing impaired steroidogenesis in both the adrenal glands and the gonads, causing abnormal phenotypes and adrenal failure.[6]

Conversion of cholesterol to pregnenolone involves hydroxylation at the carbon 20 and 22 positions, with subsequent cleavage of the side chain. Conversion of cholesterol to pregnenolone by P450scc takes place within the mitochondria. It is one of the principal effects of tropic hormone stimulation, which also causes the uptake of the cholesterol substrate for


Acetate
Cholesterol
P450scc
Pregnenolone
P450c17
3β-hydroxysteroid dehydrogenase
17-Hydroxypregnenolone
Progesterone
P450c17
P450c17
Dehydroepiandrosterone
17-Hydroxyprogesterone
3β-hydroxysteroid dehydrogenase
P450c17
17β-hydroxysteroid dehydrogenase
Androstenedione
Testosterone
P450arom
P450arom
17β-hydroxysteroid dehydrogenase
Estrone
Estradiol

this step in the ovary. The tropic hormones from the anterior pituitary bind to the cell surface receptor of the G protein system, activate adenylate cyclase, and increase intracellular cyclic AMP. Cyclic AMP activity leads to gene transcription that encodes the steroidogenic enzymes and accessory proteins. In a process that is faster than gene transcription, cyclic AMP stimulates the hydrolysis of cholesteryl esters and the transport of free cholesterol to the mitochondria.

The cholesterol used for steroid synthesis is derived from circulating low-density lipoproteins (LDL), followed by the mobilization and transport of intracellular stores.[5, 7, 8] LDL cholesterol esters are incorporated into the cell by tropic hormone stimulation of endocytosis via clathrin-coated pits (a mechanism discussed later in this chapter). Cholesterol is stored in the cell in the ester form or as free cholesterol. Indeed, the rate-limiting step in steroidogenesis is the transfer of cholesterol from the outer mitochondrial membrane to the inner mitochondrial membrane where fully active P450scc waits for substrate. The rate-limiting transfer of hydrophobic cholesterol through the aqueous space between the outer and inner mitochondrial membranes is mediated by protein activation stimulated by the tropic hormone. Long-term, chronic steroidogenesis requires gene transcription and protein synthesis, but short-term, acute responses occur independently of new RNA synthesis, although protein synthesis is still necessary, specifically the proteins that regulate cholesterol transfer across the mitochondrial membrane.

Several proteins have been characterized and proposed as regulators of acute intracellular cholesterol transfer. Sterol carrier protein 2 (SCP2) is able to bind and transfer cholesterol between compartments within a cell. Another candidate is a small molecule, steroidogenesis activator polypeptide (SAP), and still another is peripheral benzodiazepine receptor (PBR), which affects cholesterol flux through a pore structure. But the most studied and favored protein as a regulator of acute cholesterol transfer is ***steroidogenic acute regulator (StAR) protein***.[9–13] StAR messenger RNA and proteins are induced concomitantly with acute steroidogenesis in response to cyclic AMP stimulation. StAR protein increases steroid production and is imported and localized in the mitochondria. But most impressively, congenital lipoid adrenal hyperplasia (an autosomal-recessive disorder) is a failure in adrenal and gonadal steroidogenesis due to a mutation in the *StAR* gene that results in premature stop codons.[14, 15] With this mutation, a low level of steroidogenesis is possible, even permitting feminization at puberty, but continuing tropic hormonal stimulation results in an accumulation of intracellular lipid deposits that destroy steroidogenic capability.[16] StAR is required for adrenal and gonadal steroidogenesis, and, therefore, is necessary for normal male sexual differentiation.

StAR mediates the transport of cholesterol into mitochondria in adrenal and gonadal steroidogenesis, but not in the placenta and brain. StAR moves cholesterol from the outer mitochondrial membrane to the inner mitochondrial membrane where it can enter the steroidogenic pathway by being converted to prenenolone. A group of proteins structurally related to StAR have been identified, designated StARD4, StARD5, and StARD6. StARD4 is responsible for binding free cholesterol as it is produced in the cytoplasm and transporting it to the outer mitochondrial membrane.[12] Because steroid-producing cells do not store large amounts of hormones, acute increases in secretion depend on this system to produce rapid synthesis.

StAR is synthesized in a precursor form as a 285-amino acid protein that has a 25-residue sequence cleaved from the NH_2-terminal after transport into mitochondria.[17] The mutant forms of StAR undergo premature truncation that prevents this proteolytic cleavage. Mutations of the *StAR* gene, located on chromosome 8p11.2, are the only inherited disorder of steroidogenesis not caused by a defect in one of the steroidogenic enzymes. The absence of StAR expression in placenta and brain indicates the presence of different mechanisms for cholesterol transport in those tissues.

Once pregnenolone is formed, further steroid synthesis in the ovary can proceed by one of two pathways, either via Δ^5-3β-hydroxysteroids or via the Δ^4-3-ketone pathway. The first

(the Δ^5 pathway) proceeds by way of pregnenolone and dehydroepiandrosterone (DHEA) and the second (the Δ^4 pathway) via progesterone and 17α-hydroxyprogesterone.

The conversion of pregnenolone to progesterone involves two steps: the 3β-hydroxysteroid dehydrogenase and Δ^{4-5} isomerase reactions that convert the 3-hydroxyl group to a ketone and transfer the double bond from the 5–6 position to the 4–5 position. The 3β-hydroxysteroid dehydrogenase enzyme catalyzes both the dehydrogenation and isomerization reactions, and exists in two forms (type I and type II), encoded by two separate genes on chromosome 1 (the type I gene is expressed in the placenta, breast, and other non-glandular tissues, the type II gene is expressed in the gonads and the adrenal glands). Once the Δ^{4-5} ketone is formed, progesterone is hydroxylated at the 17 position to form 17α-hydroxyprogesterone. 17α-Hydroxyprogesterone is the immediate precursor of the C-19 (19 carbons) series of androgens in this pathway. By peroxide formation at C-20, followed by epoxidation of the C-17, C-20 carbons, the side chain is split off, forming androstenedione. The 17-ketone may be reduced to a 17β-hydroxyl to form testosterone by the 17β-hydroxysteroid dehydrogenase reaction. Both C-19 steroids (androstenedione and testosterone) can be converted to corresponding C-18 phenolic steroid estrogens (estrone and estradiol) by microsomal reactions in a process referred to as aromatization. This process includes hydroxylation of the angular 19-methyl group, followed by oxidation, loss of the 19-carbon as formaldehyde, and ring A aromatization (dehydrogenation).

As an alternative, pregnenolone can be directly converted to the Δ^5-3β-hydroxy C-19 steroid, dehydroepiandrosterone (DHEA), by 17α-hydroxylation followed by cleavage of the side chain. With formation of the Δ^4-3-ketone, DHEA is converted into androstenedione. The four reactions involved in converting pregnenolone and progesterone to their 17-hydroxylated products are mediated by a single enzyme, P450c17, bound to smooth endoplasmic reticulum, regulated by a gene on chromosome 10q24.3. 17-Hydroxylase and 17,20-lyase were traditionally regarded as separate enzymes. These two different functions of a single enzyme, P450c17, are not genetic or structural but represent the effect of posttranslational influencing factors.[18] In the adrenal gland pathway to cortisol, very little 17,20-lyase activity is expressed. In the ovarian theca cells, the testicular Leydig cells, and the adrenal reticularis, both 17-hydroxylase and 17,20-lyase activities are expressed, directing the steroidogenic pathway via dehydroepiandrosterone (DHEA). In the corpus luteum, the principal pathway is via progesterone.

Characterization of the P450c21 protein and gene cloning indicate that the 21-hydroxylase gene, *CYP21*, is located on chromosome 6p21.3. An inactive pseudogene, *CYP21P*, is nearby. Many of the mutations that affect *CYP21* and cause congenital adrenal hyperplasia are gene conversions involving recombinations between *CYP21* and inactivating mutations in *CYP21P*.

Aromatization is mediated by P450arom found in the endoplasmic reticulum.[19, 20] Aromatase cytochrome P450 is derived from chromosome 15q21.1, at a site designated as the *CYP19A1* (cytochrome P450, family 19, subfamily A, polypeptide 1) gene, denoting oxidation of the C-19 methyl group. Aromatization in different tissues with different substrates is the result of the single P450arom enzyme encoded by the single gene. A specific haplotype of genetic polymorphisms in *CYP19* may be linked to endometrial cancer, a known consequence of excessive estrogenic stimulation of the endometrium.[21] Aromatase deficiency because of an inactivating mutation of *CYP19A1* is very rare; only a handful of cases have been reported.[22] Affected females present at birth with virilization because the placenta cannot convert fetal adrenal androgens to estrogens; thus, maternal virilization during the pregnancy is usually also present.

Aromatase transcription is regulated by several promoter sites that respond to cytokines, cyclic nucleotides, gonadotropins, glucocorticoids, and growth factors.[23] Tissue-specific expression is regulated by tissue-specific promoters at the 5′ end of the gene. Thus, this gene has alternative promoters that allow the extremes of highly regulated expression in the

ovary in response to cyclic AMP and gonadotropins, expression in adipose tissue stimulated by prostaglandin E_2, and nonregulated expression in the placenta and adipose. Very specific inhibitors of P450arom have been developed, called "aromatase inhibitors," that allow intense blockage of estrogen production, with clinical applications that include the treatment of breast cancer (e.g., anastrozole and letrozole) and dysfunctional uterine bleeding. The aromatase complex also includes NADPH-cytochrome P450 reductase, a ubiquitous flavoprotein involved in reduction reactions.

The 17β-hydroxysteroid dehydrogenase and 5α-reductase reactions are mediated by non-P450 enzymes. The 17β-hydroxysteroid dehydrogenase is bound to the endoplasmic reticulum and the 5α-reductase to the nuclear membrane. The 17β-hydroxysteroid dehydrogenase enzymes convert estrone to estradiol, androstenedione to testosterone, and DHEA to androstenediol, and vice versa. Eight different isozymes have been cloned and characterized.[24] The type 1 enzyme is active in the placenta and granulosa cells, converting estrone to estradiol. The type 2 and 4 enzymes, found in many tissues, form androstenedione and estrone from testosterone and estradiol, respectively. The type 3 and 5 enzymes in the testis reduce androstenedione to testosterone. The type 6 enzyme may be found only in rodents, and the type 7 and 8 enzymes are widespread, but have limited activity. Thus types 1, 3, and 5 form active estrogens and androgens, whereas types 2 and 4 produce weaker products, a form of inactivation, important, for example, in protecting the fetus against testosterone and estradiol in the maternal circulation. The cell-specific production of each of these isoforms is a method for regulating the local concentration of estrogens and androgens.

The Two-Cell System

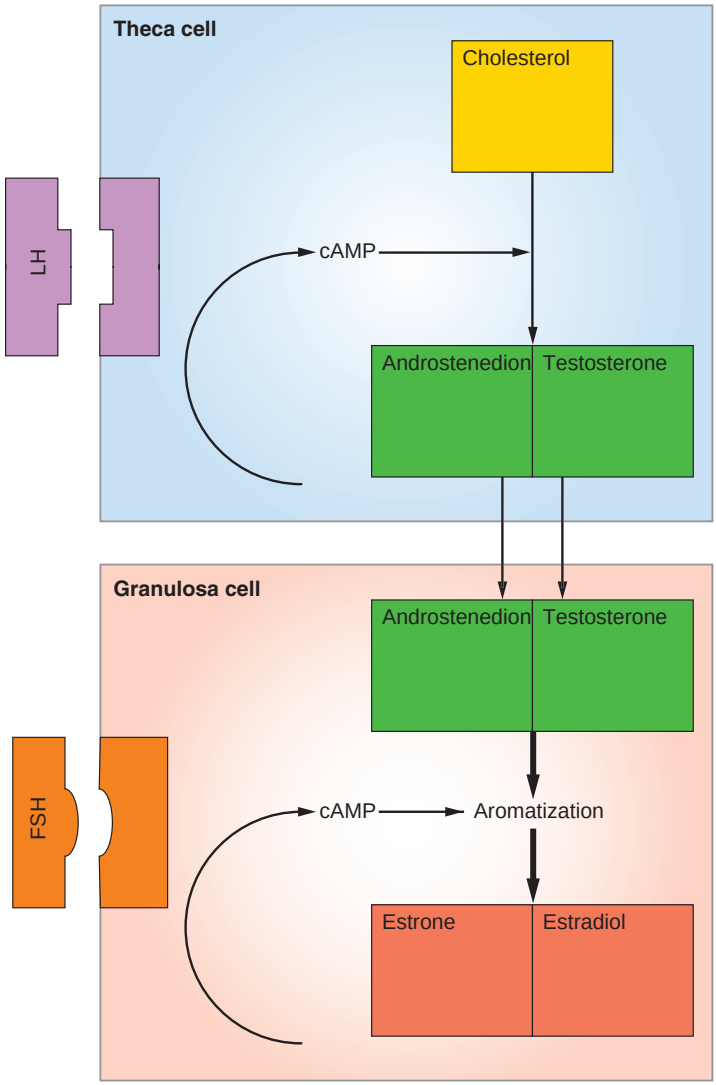
The two-cell system is a logical explanation of the events involved in ovarian follicular steroidogenesis.[25] This explanation, first proposed by Falck in 1959,[26] brings together information on the site of specific steroid production, along with the appearance and importance of hormone receptors. The following facts are important:

1. FSH receptors are present on the granulosa cells.
2. FSH receptors are induced by FSH itself.
3. LH receptors are present on the theca cells and initially absent on the granulosa cells, but, as the follicle grows, FSH induces the appearance of LH receptors on the granulosa cells.
4. FSH induces aromatase enzyme activity in granulosa cells.
5. The above actions are modulated by autocrine and paracrine factors secreted by the theca and granulosa cells.

These facts combine into the two-cell system to explain the sequence of events in ovarian follicular growth and steroidogenesis. The initial change from a primordial follicle to a preantral follicle is independent of hormones, and the stimulus governing this initial step in growth is unknown. Continued growth, however, depends on FSH stimulation. As the granulosa responds to FSH, proliferation and growth are associated with an increase in FSH receptors, a specific effect of FSH itself, but an action that is enhanced very significantly by the autocrine and paracrine peptides. The theca cells are characterized by steroidogenic activity in response to LH, specifically resulting in androgen production, by transcription of the P450scc, P450c17, and 3β-hydroxysteroid dehydrogenase genes. Aromatization of

androgens to estrogens is a distinct activity within the granulosa layer induced by FSH by activation of the P450arom gene. Androgens produced in the theca layer, therefore, must diffuse into the granulosa layer. In the granulosa layer they are converted to estrogens, and the increasing level of estradiol in the peripheral circulation reflects release of the estrogen back toward the theca layer and into blood vessels.

The theca and granulosa cells secrete peptides that operate as both autocrine and paracrine factors.[27] Insulin-like growth factor (IGF) is secreted by the theca and enhances the LH stimulation of androgen production in the theca cells as well as FSH-mediated aromatization in the granulosa. Evidence indicates that the endogenous insulin-like growth factor in the human ovarian follicle is IGF-II in both the granulosa and the theca cells.[28] Studies indicating activity of IGF-I with human ovarian tissue can be explained by the fact that both IGF-I and IGF-II activities can be mediated by the type I IGF receptor, which is structurally similar to the insulin receptor. The regulation of FSH receptors on granulosa cells is relatively complex. Although FSH increases the activity of its own receptor gene in a cyclic AMP-mediated mechanism, this action is influenced by inhibitory agents, such as epidermal growth factor, fibroblast growth factor, and even a gonadotropin-releasing hormone (GnRH)-like protein. Inhibin and activin are produced in the granulosa in response to FSH, and activin has the important autocrine role of enhancing FSH actions, especially the production of FSH receptors. Inhibin enhances LH stimulation of androgen synthesis in the



theca to provide a substrate for aromatization to estrogen in the granulosa, whereas activin suppresses androgen synthesis. This important paracrine regulation of androgen production in thecal cells by inhibin and activin, discussed in Chapter 6, is exerted primarily through modification of the expression of steroidogenic enzymes, especially P450c17.[29]

After ovulation, the dominance of the luteinized granulosa layer is dependent on preovulatory induction of an adequate number of LH receptors, and, therefore, dependent on adequate FSH action. Prior to ovulation the granulosa layer is characterized by aromatization activity and conversion of the theca androgens to estrogens, an FSH-mediated activity. After ovulation the granulosa layer secretes progesterone and estrogens directly into the bloodstream, an LH-mediated activity.

Granulosa and theca cells each have an androgen aromatase system that can be demonstrated in vitro. However, in vivo, the activity of the granulosa layer in the follicular phase is several hundred times greater than the activity of the theca layer, and, therefore, the granulosa is the main biosynthetic source of estrogen in the growing follicle.[30] Because granulosa cells lack P450c17, the rate of aromatization in the granulosa layer is directly related to and dependent on the androgen substrate made available by the theca cells. Hence, estrogen secretion by the follicle prior to ovulation is the result of combined LH and FSH stimulation of the two cell types, the theca and the granulosa. After ovulation, it is believed the two cell types continue to function as a two-cell system; luteal cells derived from theca produce androgens for aromatization into estrogens by luteal cells derived from granulosa.

	Free (Unbound (%))	*Albumin-Bound (%)*	*SHBG-Bound (%)*
Estrogen	1	30	69
Testosterone	1	30	69
DHEA	4	88	8
Androstenedione	7	85	8
Dihydrotestosterone	1	71	28

From Mendel[31]

Blood Transport of Steroids

While circulating in the blood, a majority of the principal sex steroids, estradiol and testosterone, is bound to a protein carrier, known as sex hormone-binding globulin (SHBG) produced mainly in the liver. Another 30% is loosely bound to albumin, leaving only about 1% unbound and free. A very small percentage also binds to corticosteroid-binding globulin. Hyperthyroidism, pregnancy, and estrogen administration all increase SHBG levels, whereas corticoids, androgens, progestins, growth hormone, insulin, and IGF-I decrease SHBG.

The circulating level of SHBG is inversely related to body weight, and, thus, significant weight gain can decrease SHBG and produce important changes in the unbound levels of the sex steroids. Another important mechanism for a reduction in circulating SHBG levels is insulin resistance and hyperinsulinemia.[32, 33] Thus, increased insulin levels in the circulation lower SHBG levels, and this may be the major mechanism that mediates the impact of increased body weight on SHBG. This relationship between the levels of insulin and SHBG is so strong that SHBG concentrations are a marker for hyperinsulinemic insulin resistance, and a low level of SHBG is a predictor for the development of type 2 diabetes mellitus.[34]

The distribution of body fat has a strong influence on SHBG levels. Android or central fat is located in the abdominal wall and visceral-mesenteric locations. This fat distribution

is associated with hyperinsulinemia, hyperandrogenism, and decreased levels of SHBG.[35] The common mechanism for these changes is probably the hyperinsulinemia.

SHBG is a glycoprotein that contains a single binding site for androgens and estrogens, even though it is a homodimer composed of two monomers. Its gene has been localized to the short arm (p12–13) of chromosome 17.[36] Genetic studies have revealed that the SHBG gene also encodes the androgen-binding protein present in the seminiferous tubules, synthesized by the Sertoli cells.[37, 38] Dimerization is believed to be necessary to form the single steroid-binding site. Specific chromosomal abnormalities with decreased or abnormal SHBG have not been reported. SHBG gene expression has now been identified in other tissues (brain, placenta, and endometrium), although a biologic significance has not been determined.

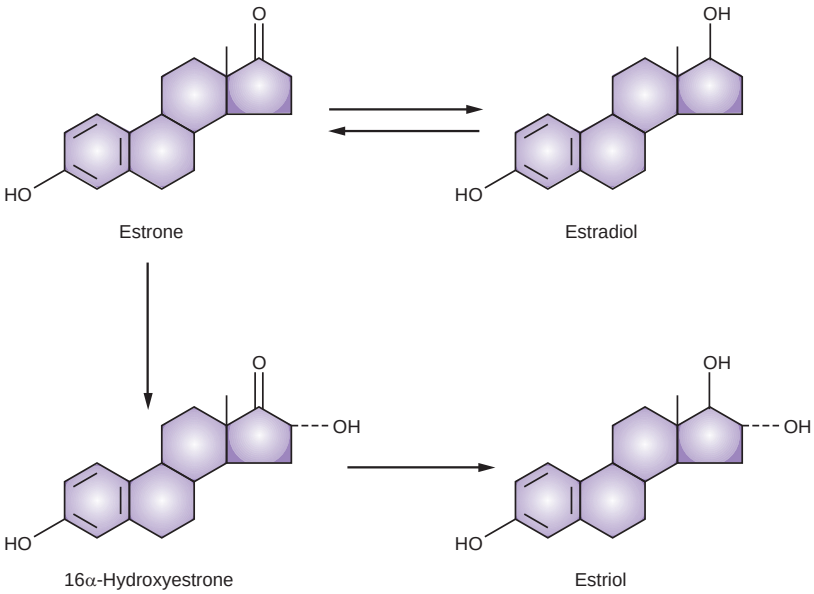
Transcortin, also called corticosteroid-binding globulin, is a plasma glycoprotein that binds cortisol, progesterone, deoxycorticosterone, corticosterone, and some of the other minor corticoid compounds. Normally about 75% of circulating cortisol is bound to transcortin, 15% is loosely bound to albumin, and 10% is unbound or free. Progesterone circulates in the following percentages: less than 2% unbound, 80% bound to albumin, 18% bound to transcortin, and less than 1% bound to SHBG. Binding in the circulation follows the law of mass action: the amount of the free, unbound hormone is in equilibrium with the bound hormone. Thus, the total binding capacity of a binding globulin will influence the amount that is free and unbound.

The biologic effects of the major sex steroids are largely determined by the unbound portion, known as the free hormone. In other words, the active hormone is unbound and free, whereas the bound hormone is relatively inactive. This concept is not without controversy. The hormone-protein complex may be involved in an active uptake process at the target cell plasma membrane.[39–41] The albumin-bound fraction of steroids may also be available for cellular action because this binding has low affinity. Because the concentration of albumin in plasma is manyfold greater than that of SHBG, the contribution of the albumin-bound fraction can be significant. Routine assays determine the total hormone concentration, bound plus free, and special steps are required to measure the active free level of testosterone, estradiol, and cortisol.

Estrogen Metabolism

Androgens are the precursors of estrogens. 17β-Hydroxysteroid dehydrogenase activity converts androstenedione to testosterone, which is not a major secretory product of the normal ovary. It is rapidly demethylated at the C-19 position and aromatized to estradiol, the major estrogen secreted by the human ovary. Estradiol also arises to a major degree from androstenedione via estrone, and estrone itself is secreted in significant daily amounts. Estriol is the peripheral metabolite of estrone and estradiol and not a secretory product of the ovary. The formation of estriol is typical of general metabolic "detoxification," conversion of biologically active material to less active forms.

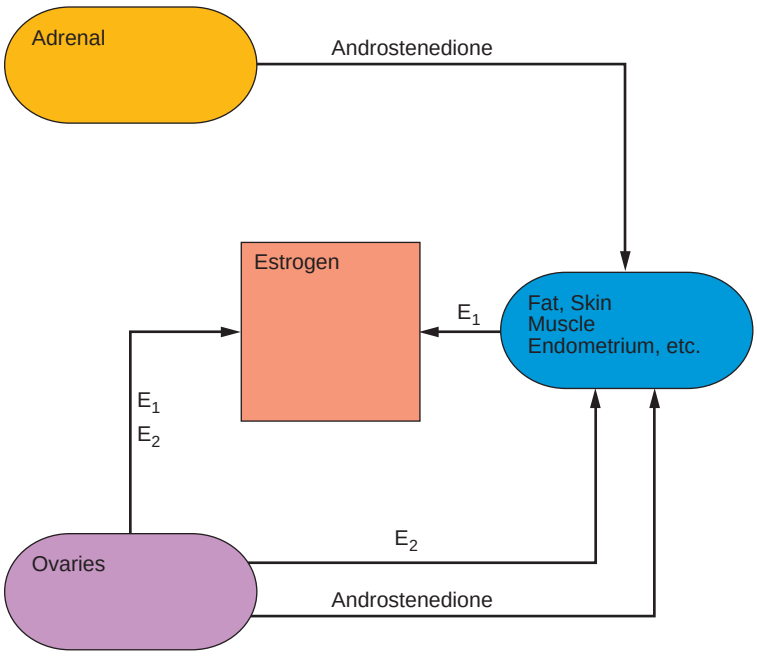
The conversion of steroids in peripheral tissues is not always a form of inactivation. Free androgens are peripherally converted to free estrogens, for example, in skin and adipose cells. The location of the adipose cells influences their activity. Women with central obesity (the abdominal area) produce more androgens.[42] The work of Siiteri and MacDonald[43] demonstrated that enough estrogen can be derived from circulating androgens to produce bleeding in the postmenopausal woman. In the female the adrenal gland remains the major source of circulating androgens, in particular androstenedione. In the male, almost all of the circulating estrogens are derived from peripheral conversion of androgens. The precursor



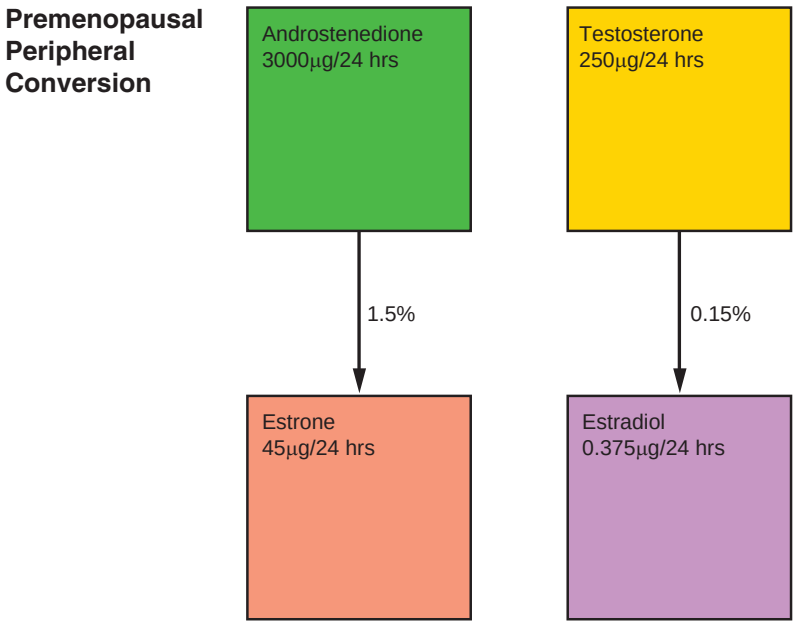
androgens consist principally of androstenedione, dehydroepiandrosterone, and dehydroepiandrosterone sulfate.

It can be seen, therefore, that the pattern of circulating steroids in the female is influenced by the activity of various processes outside the ovary. Because of the peripheral contribution to steroid levels, the term ***secretion rate*** is reserved for direct organ secretion, whereas ***production rate*** includes organ secretion plus peripheral contribution via conversion of precursors. The ***metabolic clearance rate (MCR)*** equals the volume of blood that is cleared of the hormone per unit of time. The ***blood production rate (PR)*** then equals the metabolic clearance rate multiplied by the concentration of the hormone in the blood.

MCR = Liters/Day
PR = MCR × Concentration (Liters/Day × Amount/Liter = Amount/Day)



In the normal nonpregnant female, estradiol is produced at the rate of 100–300 μg/day. The production of androstenedione is about 3 mg/day, and the peripheral conversion (about 1%) of androstenedione to estrone accounts for about 20–30% of the estrone produced per day. Because androstenedione is secreted in milligram amounts, even a small percent conversion to estrogen results in a significant contribution to estrogens, which exist and function in the circulation in picogram amounts. Thus, the circulating estrogens in the female are the sum of direct ovarian secretion of estradiol and estrone, plus peripheral conversion of C-19 precursors. Whereas estradiol is produced in μg amounts, it circulates and functions within cells in concentrations of pg/mL.

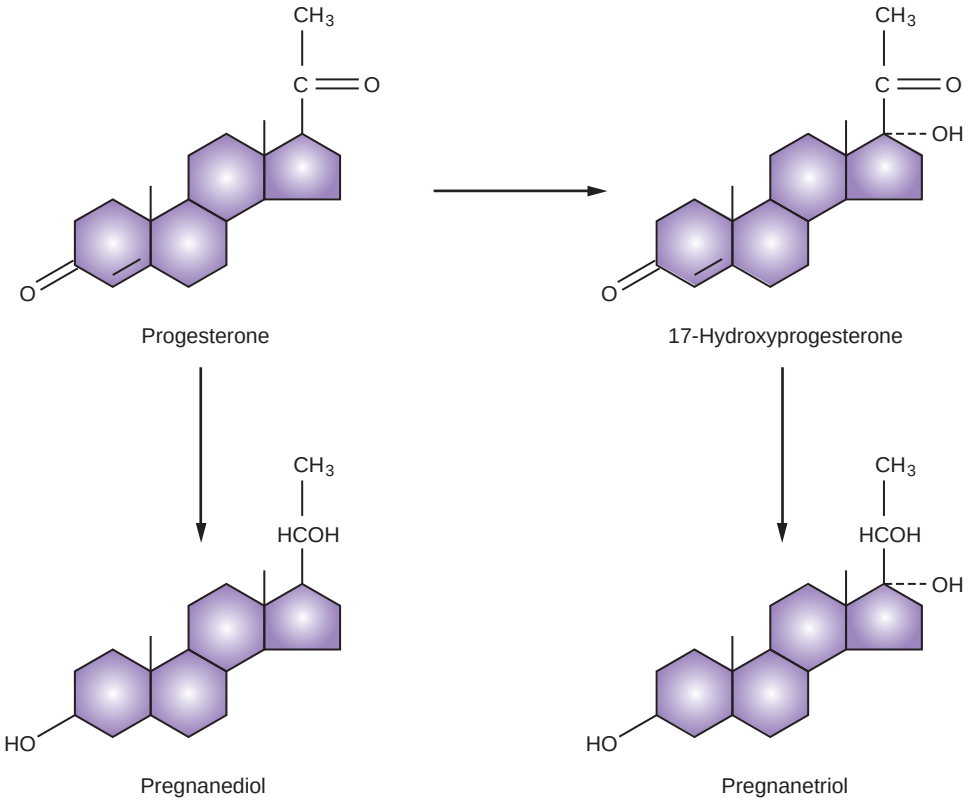


Progesterone Metabolism

Peripheral conversion of steroids to progesterone is not seen in the nonpregnant female; rather, the progesterone production rate is a combination of secretion from the adrenal and the ovaries. Including the small contribution from the adrenal, the blood production rate of progesterone in the preovulatory phase is less than 1 mg/day. During the luteal phase, production increases to 20–30 mg/day. The metabolic fate of progesterone, as expressed by its many excretion products, is more complex than estrogen. About 10–20% of progesterone is excreted as pregnanediol.

Pregnanediol glucuronide is present in the urine in concentrations less than 1 mg/day until ovulation. Postovulation pregnanediol excretion reaches a peak of 3–6 mg/day, which is maintained until 2 days prior to menses. The assay of pregnanediol in the urine now has little clinical use.

In the preovulatory phase in adult females, in all prepubertal females, and in the normal male, the blood levels of progesterone are at the lower limits of immunoassay sensitivity: less than 1 ng/mL. After ovulation, i.e., during the luteal phase, progesterone ranges from 3 to 15 ng/mL. In congenital adrenal hyperplasia, progesterone blood levels can be as high as 50 times above normal.



Pregnanetriol is the chief urinary metabolite of 17α-hydroxyprogesterone and has clinical significance in the adrenogenital syndrome, a syndrome of virilizing adrenal hyperplasia in which an enzyme defect results in accumulation of 17α-hydroxyprogesterone and increased excretion of pregnanetriol. The inherited abnormality in virilizing adrenal hyperplasia results in an inability to synthesize glucocorticoids. The hypothalamic-pituitary axis reacts to the low level of cortisol by elevated ACTH secretion in a homeostatic response to achieve normal levels of cortisol production. This stimulation induces a hyperplastic adrenal cortex that produces androgens as well as corticoid precursors in abnormal quantities. The plasma or serum assay of 17α-hydroxyprogesterone is a more sensitive and accurate index of this enzyme deficiency than measurement of pregnanetriol. Normally, the blood level of 17α-hydroxyprogesterone is less than 100 ng/dL, although after ovulation and during the luteal phase of a normal menstrual cycle, a peak of 200 ng/dL can be reached. In syndromes of adrenal hyperplasia, values can be 10–400 times normal.

Androgen Metabolism

The major androgen products of the ovary are dehydroepiandrosterone (DHEA) and androstenedione (and only a little testosterone), which are secreted mainly by stromal tissue derived from theca cells. With excessive accumulation of stromal tissue or in the presence of an androgen-producing tumor, testosterone becomes a significant secretory product. Occasionally, a nonfunctioning tumor can induce stromal proliferation and increased androgen production. The normal accumulation of stromal tissue at midcycle

results in a rise in circulating levels of androstenedione and testosterone at the time of ovulation.

The adrenal cortex produces three groups of steroid hormones: the glucocorticoids, the mineralocorticoids, and the sex steroids. The adrenal sex steroids represent intermediate byproducts in the synthesis of glucocorticoids and mineralocorticoids, and excessive secretion of the sex steroids occurs only with neoplastic cells or in association with enzyme deficiencies. Under normal circumstances, adrenal gland production of the sex steroids is less significant than gonadal production of androgens and estrogens. About one-half of the daily production of DHEA and androstenedione comes from the adrenal gland; the other half of androstenedione is secreted by the ovary, but the other half of DHEA is split almost equally between the ovary and peripheral tissues. The production rate of testosterone in the normal female is 0.2–0.3 mg/day, and approximately 50% arises from peripheral conversion of androstenedione (and a small amount from DHEA) to testosterone, whereas 25% is secreted by the ovary and 25% by the adrenal. The major androgens are excreted in the urine as 17-ketosteroids.

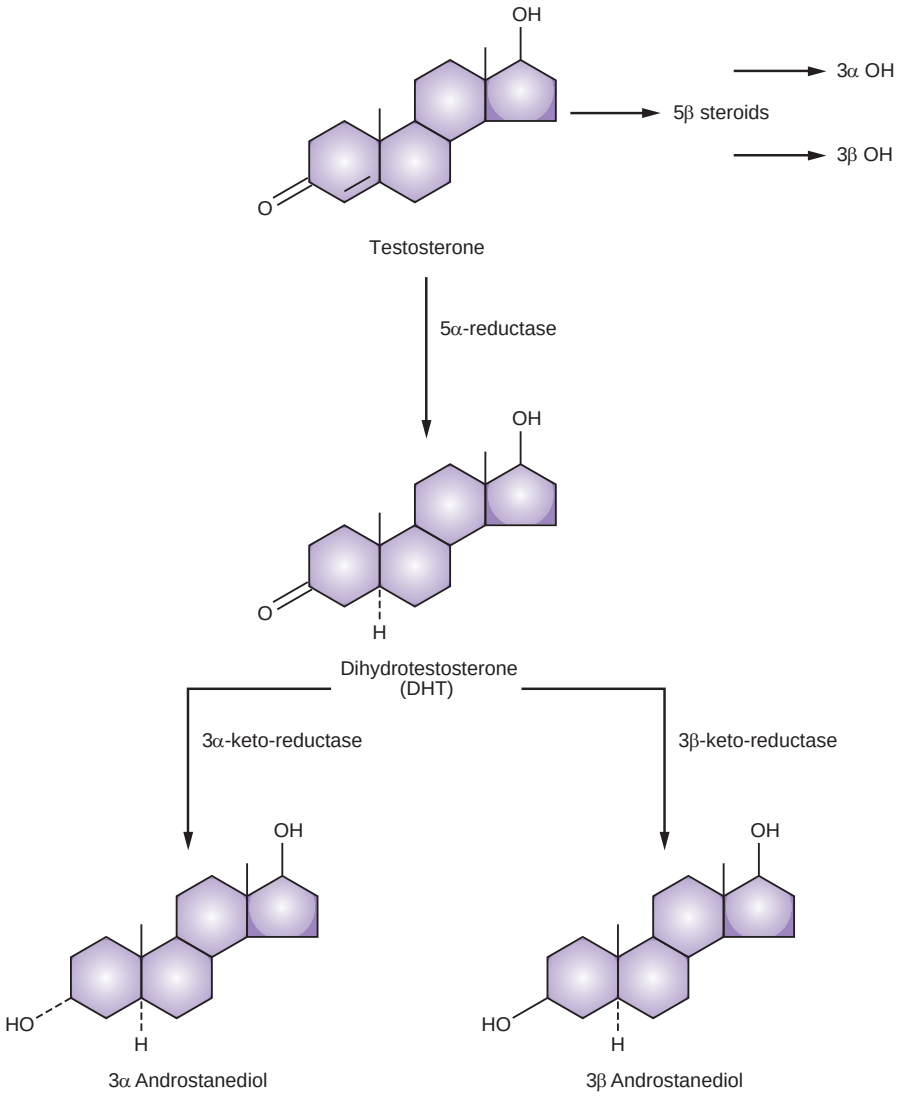
There is no circadian cycle of the major sex steroids in the female. However, short-term variations in the blood levels due to episodic secretion require multiple sampling for absolutely accurate assessment. ***Although frequent sampling is necessary for a high degree of accuracy, a random sample is sufficient for clinical purposes to determine whether a level is within a normal range***.

The testosterone-binding capacity is decreased by androgens; hence, the binding capacity in men is lower than that in normal women. The binding globulin level in women with increased androgen production is also depressed. Androgenic effects are dependent on the unbound fraction that can move freely from the vascular compartment into the target cells. Routine assays determine the total hormone concentration, bound plus free. Thus, a total testosterone concentration can be in the normal range in a woman who is hirsute or even virilized, but because the binding globulin level is depressed by the androgen effects, the percent free and active testosterone is elevated. The need for a specific assay for the free portion of testosterone can be questioned because the very presence of hirsutism or virilism indicates increased androgen effects. In the face of hirsutism, one can reliably interpret a normal testosterone level as compatible with decreased binding capacity and increased active free testosterone.

Both total and unbound testosterone are normal in only a few women with hirsutism. In these cases, the hirsutism, heretofore regarded as idiopathic, most likely results from excessive intracellular androgen effects (specifically increased intracellular conversion of testosterone to dihydrotestosterone).

Reduction of the Δ^4 unsaturation (an irreversible pathway) in testosterone is very significant, producing derivatives very different in their spatial configuration and activity. The 5β-derivatives are not androgenic, and this is not an important pathway; however, the 5α-derivative (a very active pathway) is extremely potent. Indeed, dihydrotestosterone (DHT), the 5α-derivative, is the principal androgenic hormone in a variety of target tissues and is formed within the target tissue itself.

In men, the majority of circulating DHT is derived from testosterone that enters a target cell and is converted by means of 5α-reductase to DHT. In women, because the production rate of androstenedione is greater than testosterone, blood DHT is primarily derived from androstenedione and partly from dehydroepiandrosterone.[44] Thus, in women, the skin production of DHT is predominantly influenced by androstenedione. DHT is by definition an intracrine hormone, formed and acting within target tissues.[45] The 5α-reductase enzyme exists in two forms, type I and II, each encoded by a separate gene, with the type I enzyme found in skin and the type II reductase predominantly expressed in reproductive tissues.[46]



DHT is largely metabolized intracellularly; hence, the blood DHT is only about one-tenth the level of circulating testosterone, and it is clear that testosterone is the major circulating androgen. In tissues sensitive to DHT (which includes hair follicles), only DHT enters the nucleus to provide the androgen message. DHT also can perform androgenic actions within cells that do not possess the ability to convert testosterone to DHT. DHT is further reduced by a 3α-keto-reductase to androstanediol, which is relatively inactive. The metabolite of androstanediol, 3α-androstanediol glucuronide, is the major metabolite of DHT and can be measured in the plasma, indicating the level of activity of target tissue conversion of testosterone to DHT.

Not all androgen-sensitive tissues require the prior conversion of testosterone to DHT. In the process of masculine differentiation, the development of the wolffian duct structures (epididymis, the vas deferens, and the seminal vesicle) is dependent on testosterone as the intracellular mediator, whereas development of the urogenital sinus and urogenital tubercle into the male external genitalia, urethra, and prostate requires the conversion of testosterone to DHT.[47] Muscle development is under the direct control of testosterone. Testosterone is also aromatized to a significant extent in the brain, liver, and breast; and in some circumstances (e.g., in the brain) androgenic messages can be transmitted via estrogen.

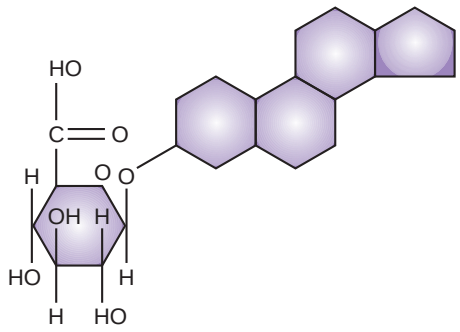
The Importance of Local Sex Hormone Production

The characteristics of sex steroid metabolism reviewed above contribute to an important clinical concept: *circulating levels of sex hormones do not always reflect concentrations in target cells*. In premenopausal women, target tissues synthesize and metabolize most of the testosterone produced. Thus, in women testosterone functions as a paracrine and intracrine hormone. In men, abundant secretion of testosterone creates circulating levels that are sufficient to allow testosterone to function as a classic hormone. In women, the same description applies to estradiol. Estradiol functions as a classical circulating hormone until menopause, after which both estradiol and testosterone activities are due to local target tissue synthesis, using precursors derived from the circulation. Clinical interventions after menopause, therefore, are directed to local hormone production; for example, the use of aromatase inhibitors to treat breast cancer.

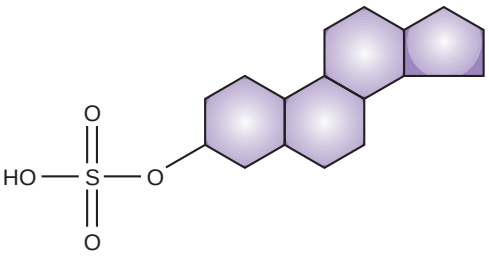
Excretion of Steroids

Active steroids and metabolites are excreted as sulfo and glucuro conjugates. Conjugation of a steroid converts a hydrophobic compound into a hydrophilic one and generally reduces or eliminates the activity of a steroid. This is not completely true, however, because hydrolysis of the ester linkage can occur in target tissues and restore the active form. Furthermore, estrogen conjugates can have biologic activity, and it is known that sulfated conjugates are actively secreted and may serve as precursors, present in the circulation in relatively high concentrations because of binding to serum proteins. Ordinarily, however, conjugation by liver and intestinal mucosa is a step in deactivation preliminary to, and essential for, excretion into urine and bile.

Glucosiduronate



Sulfate



Cellular Mechanism of Action

Hormones circulate in extremely low concentrations and, in order to respond with specific and effective actions, target cells require the presence of special mechanisms. There are two major types of hormone action at target tissues. One mediates the action of tropic hormones (peptide and glycoprotein hormones) with receptors at the cell membrane level. In contrast, the smaller steroid hormones enter cells readily, and the basic mechanism of action involves specific receptor molecules within the cells. It is the affinity, specificity, and activity of the receptors, together with the large concentration of receptors in cells, that allow a small amount of hormone to produce a biologic response.

The many different types of receptors can be organized into the following basic categories:

Intracellular Receptors

Receptors within cells lead to transcription activation. Examples include the receptors for estrogen and thyroid hormones.

G Protein Receptors

These receptors are composed of a single polypeptide chain that spans the cell membrane. Binding to a specific hormone leads to interaction with G proteins that, in turn, activate second messengers. Examples include receptors for tropic hormones, prostaglandins, light, and odors. The second messengers include the adenylate cyclase enzyme, the phospholipase system, and calcium ion changes.

Ion Gate Channels

These cell surface receptors are composed of multiple units, that after binding, open ion channels. The influx of ions changes the electrical activity of the cells. The best example of this type is the acetylcholine receptor.

Receptors with Intrinsic Enzyme Activity

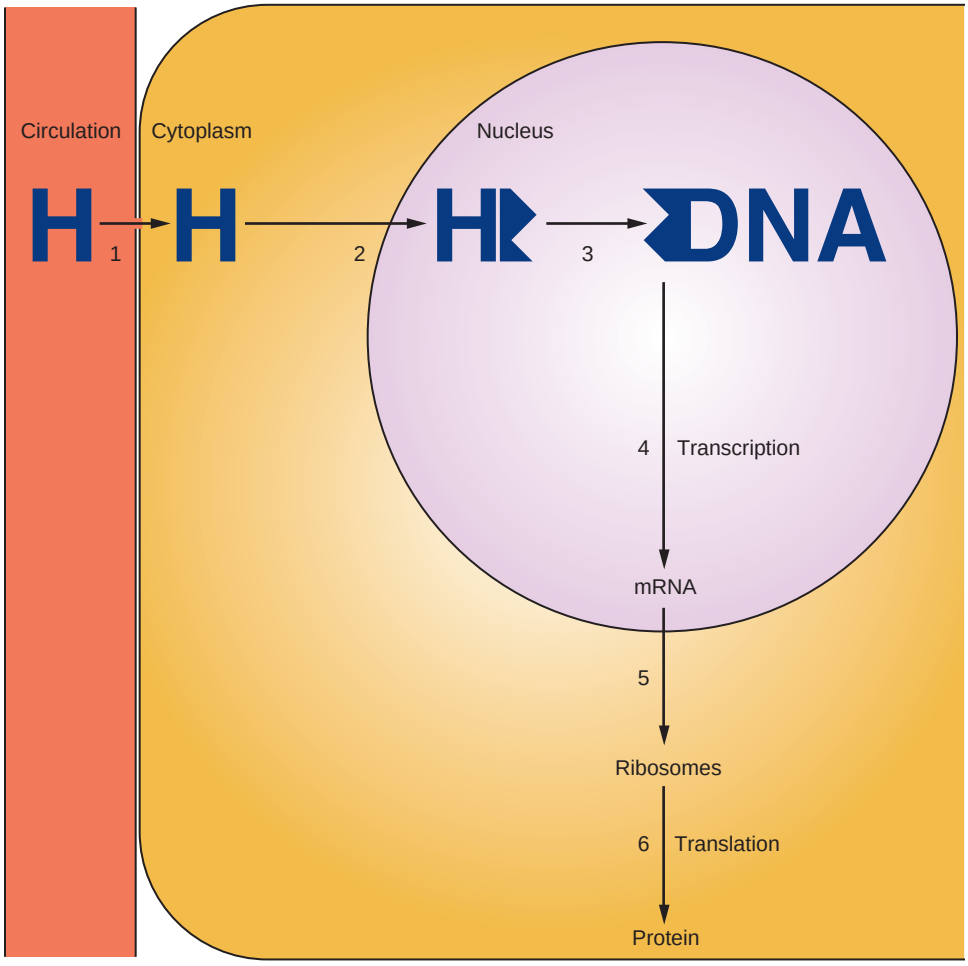
These transmembrane receptors have an intracellular component with tyrosine or serine kinase activity. Binding leads to receptor autophosphorylation and activity. Examples include the receptors for insulin and growth factors (tyrosine kinase) and the receptors for activin and inhibin (serine kinase).

The System of Internalization

Receptors that do not fit the above categories include the receptors for LDL, prolactin, growth hormone, and some of the growth factors. These receptors allow entry of their ligands into cells by the process of endocytosis (discussed later in this chapter).

Mechanism of Action for Steroid Hormones

The specificity of the reaction of tissues to sex steroid hormones is due to the presence of intracellular receptor proteins. Different types of tissues, such as liver, kidney, and uterus, respond in a similar manner. The mechanism includes: (1) steroid hormone diffusion across the cell membrane, (2) steroid hormone binding to a receptor protein, (3) interaction of a hormone-receptor complex with nuclear DNA, (4) synthesis of messenger RNA (mRNA), (5) transport of the mRNA to the ribosomes, and finally, (6) protein synthesis in the cytoplasm that results in specific cellular activity. The steroid hormone receptors primarily affect gene transcription, but also regulate posttranscriptional events and nongenomic events. Steroid receptors regulate gene transcription through multiple mechanisms, not all of which require direct interactions with DNA.



Each of the major classes of the sex steroid hormones, including estrogens, progestins, and androgens, act according to this general mechanism. Glucocorticoid and mineralocorticoid receptors, when in the unbound state, reside in the cytoplasm and move into the nucleus after hormone-receptor binding. Estrogens, progestins, and androgens are transferred across the nuclear membrane and bind to their receptors within the nucleus.

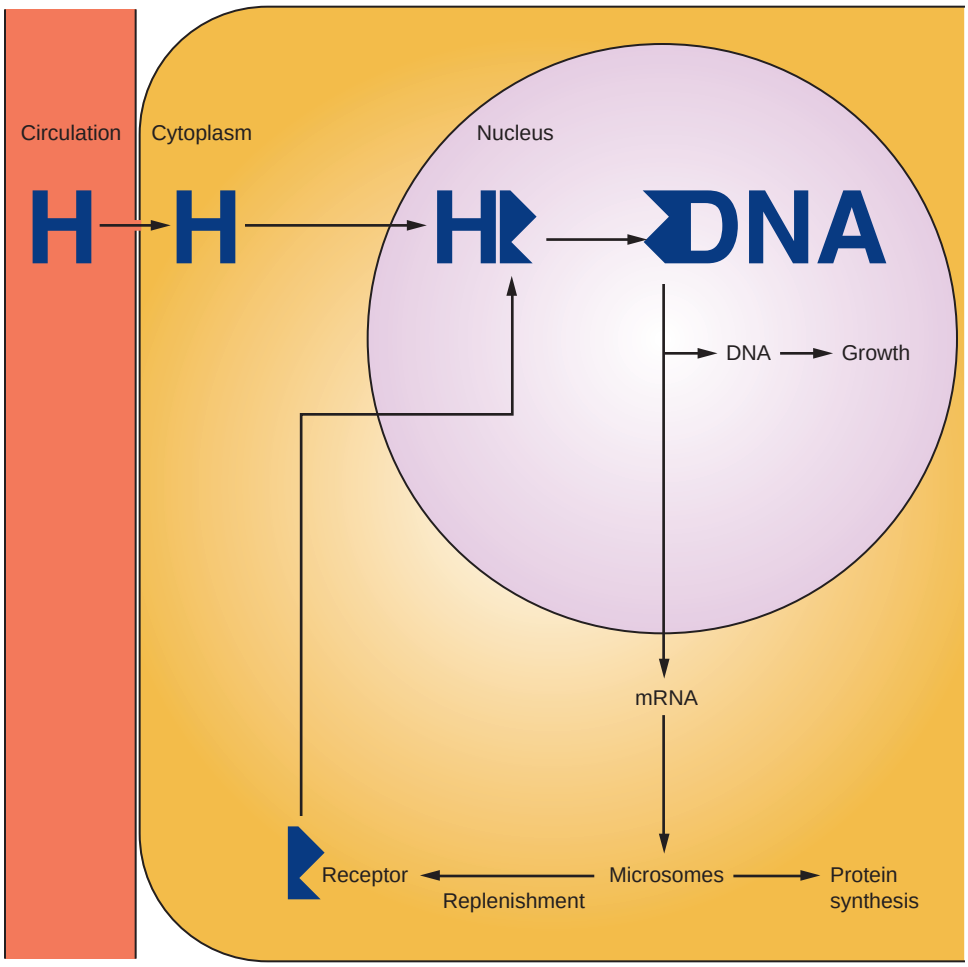
Steroid hormones are rapidly transported across the cell membrane by simple diffusion. The factors responsible for this transfer are unknown, but the concentration of free (unbound) hormone in the bloodstream seems to be an important and influential determinant of cellular function. Once in the cell, the sex steroid hormones bind to their individual receptors. During this process, ***transformation or activation*** of the receptor occurs. Transformation refers to a conformational change of the hormone-receptor complex revealing or producing a binding site that is necessary in order for the complex to bind to the chromatin. In the unbound state, the receptor is associated with heat shock proteins that stabilize and protect the receptor and maintain a conformational shape that keeps the DNA binding region in an inactive state. Activation of the receptor is driven by hormone binding that causes a dissociation of the receptor-heat shock protein complex.

The hormone-receptor complex binds to specific DNA sites (***hormone-responsive elements***) that are located upstream of the gene. The specific binding of the hormone-receptor complex with DNA results in RNA polymerase initiation of transcription. Transcription leads to translation, mRNA-mediated protein synthesis on the ribosomes. The principal action of steroid hormones is the regulation of intracellular protein synthesis by means of the receptor mechanism.

Biologic activity is maintained only while the nuclear site is occupied with the hormone-receptor complex. The dissociation rate of the hormone and its receptor as well as the half-life of the nuclear chromatin-bound complex are factors in the biologic response because the hormone response elements are abundant and, under normal conditions, are occupied only to a small extent.[48] Thus, an important clinical principle is the following: ***duration of exposure to a hormone is as important as dose***. One reason only small amounts of estrogen need be present in the circulation is the long half-life of the estrogen hormone-receptor complex. Indeed, a major factor in the potency differences among the various estrogens (estradiol, estrone, estriol) is the length of time the estrogen-receptor complex occupies the nucleus. The higher rate of dissociation with the weak estrogen (estriol) can be compensated for by continuous application to allow prolonged nuclear binding and activity. Cortisol and progesterone must circulate in large concentrations because their receptor complexes have short half-lives in the nucleus.

An important action of estrogen is the modification of its own and other steroid hormone activity by affecting receptor concentrations. Estrogen increases target tissue responsiveness to itself and to progestins and androgens by increasing the concentration of its own receptor and that of the intracellular progestin and androgen receptors. Progesterone and clomiphene, on the other hand, limit tissue response to estrogen by blocking this mechanism, thus decreasing over time the concentration of estrogen receptors. Small amounts of receptor depletion and small amounts of steroid in the blood activate the mechanism.

The synthesis of the sex steroid receptors obviously takes place in the cytoplasm, but with estrogen and progestin receptors, synthesis must be quickly followed by transportation into the nucleus. There is an amazingly extensive nuclear traffic.[49] The nuclear membrane contains 3,000 to 4,000 pores. A cell synthesizing DNA imports about one million histone molecules from the cytoplasm every 3 minutes. If the cell is growing rapidly, about three newly assembled ribosomes will be transported every minute in the other direction. The typical cell can synthesize 10,000 to 20,000 different proteins. How do they know where to go? The answer is that these proteins have localization signals. In the case of steroid hormone receptor proteins, the signal sequences are in the hinge region.



Estrogen and progestin receptors exit continuously from the nucleus to the cytoplasm and are actively transported back to the nucleus. This is a constant shuttle; diffusion into the cytoplasm is balanced by the active transport into the nucleus. This raises the possibility that some diseases are due to poor traffic control. This can be true of some acquired diseases as well, e.g., Reye's syndrome, an acquired disorder of mitochondrial enzyme function.

The fate of the hormone-receptor complex after gene activation is referred to as hormone-receptor ***processing***. In the case of estrogen receptors, processing involves the rapid degradation of receptors unbound with estrogen, and a much slower degradation of bound receptors after gene transcription. The rapid turnover of estrogen receptors has clinical significance. The continuous presence of estrogen is an important factor for continuing response.

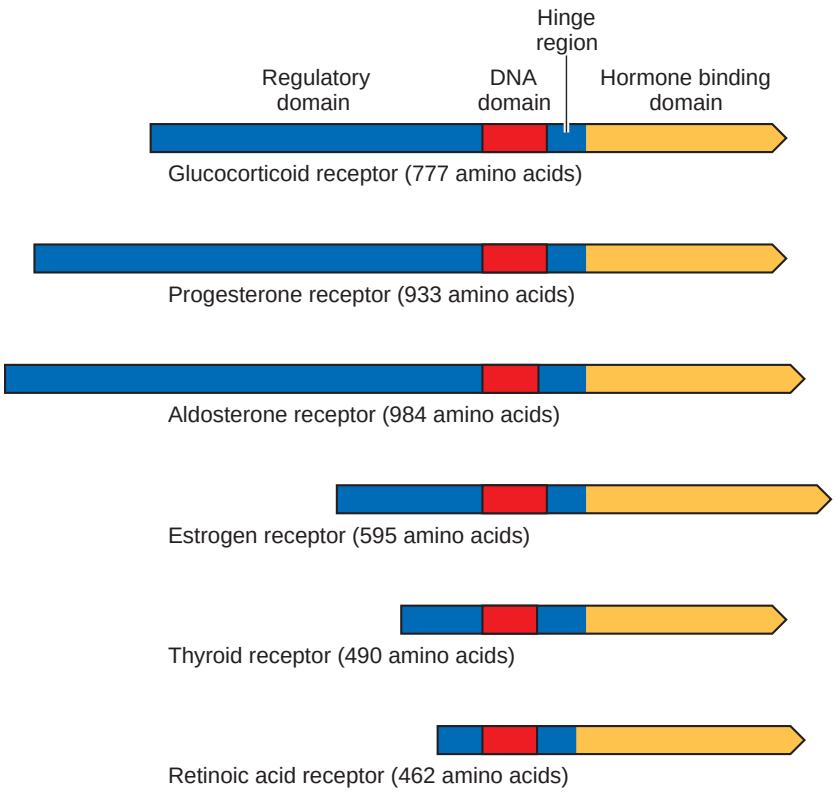
The best example of the importance of these factors is the difference between estradiol and estriol. Estriol has only 20–30% affinity for the estrogen receptor compared with estradiol; therefore, it is rapidly cleared from a cell. But if the effective concentration is kept equivalent to that of estradiol, it can produce a similar biologic response.[50] In pregnancy, where the concentration of estriol is very great, it can be an important hormone, not just a metabolite.

The depletion of estrogen receptors in the endometrium by progestational agents is the fundamental reason for adding progestins to estrogen treatment programs. The progestins

accelerate the turnover of preexisting receptors, and this is followed by inhibition of estrogen-induced receptor synthesis. Using monoclonal antibody immunocytochemistry, this action has been pinpointed to the interruption of transcription in estrogen-regulated genes. The mechanism is different for androgen antiestrogen effects. Androgens also decrease estrogen receptors within target tissues, especially in the uterus.[51, 52]

The Receptor Superfamily

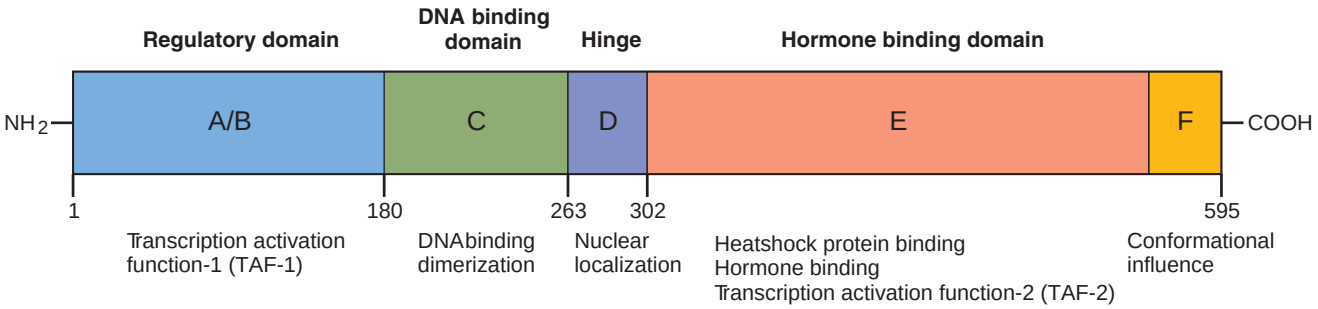
Recombinant DNA techniques have permitted the study of the gene sequences that code for the synthesis of nuclear receptors. Steroid hormone receptors share a common structure with the receptors for thyroid hormone, 1,25-dihydroxyvitamin D3, and retinoic acid; thus, these receptors are called a superfamily.[53, 54] Each receptor contains characteristic domains that are similar and interchangeable. Therefore, it is not surprising that the specific hormones can interact with more than one receptor in this family. Analysis of these receptors suggests a complex evolutionary history during which gene duplication and swapping between domains of different origins occurred. This family now includes hundreds of proteins, present in practically all species, from worms to insects to humans. Some are called ***orphan receptors*** because specific ligands for these proteins have not been identified, but the number of orphan receptors is gradually diminishing (deorphaning). It has been convincingly argued that the 6 steroid receptors originated in a common ancestral receptor gene.[55] The identification of steroid receptors in the sea lamprey dates the origin to over 450 million years ago, and the characterization of a receptor that functions like an estrogen receptor in the mollusk indicates that the ancient and initial sex steroid receptor was an estrogen receptor.[56] Knowledge of the complete human genome has confirmed that there are 48 nuclear receptors in the receptor superfamily.[54]



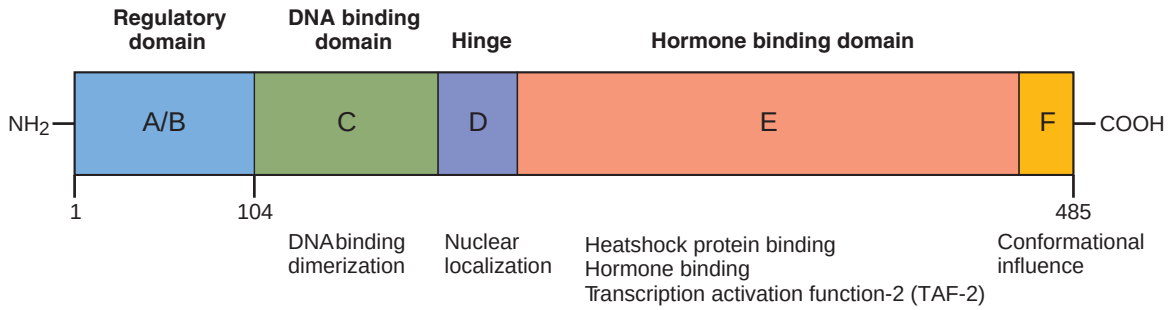
The Estrogen Receptors

Two estrogen receptors have been identified, designated as estrogen receptor-alpha (ER-α) and estrogen receptor-beta (ER-β).[57, 58] The estrogen receptor-α was discovered about 1960, and the amino acid sequence was reported in 1986.[59–61] The estrogen receptor-α is translated from a 6.8-kilobase mRNA derived from a gene that contains eight exons on the long arm of chromosome 6.[62] It has a molecular weight of approximately 66,000 with 595 amino acids. The receptor-α half-life is approximately 4–7 hours; thus the estrogen receptor-α is a protein with a rapid turnover. The more recently discovered estrogen receptor-β, a protein with 530 amino acids, is encoded by a gene localized to chromosome 14,q23.2, in close proximity to genes related to Alzheimer's disease.[63, 64] Multiple isoforms exist of ER-β, including five full-length forms.

The Estrogen Receptor-Alpha



The Estrogen Receptor-Beta



Orphan receptors have been identified that are related to the estrogen receptors. They have been designated as estrogen-related receptor (ERRα, ERRβ, and ERRγ). ERRα may be regulated by coactivator proteins and interacts with typical steroid signaling pathways.[65, 66] These orphan receptors are expressed in most tissues and may be involved in typical estrogen activities, such as the proliferation and differentiation of target cells in bone and in the breast. Nevertheless, they do not bind estrogens, and no endogenous ligand has been yet identified.

The story is further complicated with the recognition that members of the receptor superfamily are each associated with multiple isoforms.[67] This increases the number of possible signaling pathways in physiology and disease. In this discussion, we will mention only the most biologically important isoforms.

The estrogen receptors are divided into six regions in five domains, labeled A to F. The ER-β is 96% homologous in amino acid sequence with the alpha estrogen receptor in the DNA binding domain and 60% homologous in the hormone-binding domain. The full comparison is as follows:[63, 68, 69]

	ER-α and ER-β Homology (%)
The regulatory domain	18
The DNA-binding domain	96
The hinge	30
The hormone-binding domain	55
The F region	18

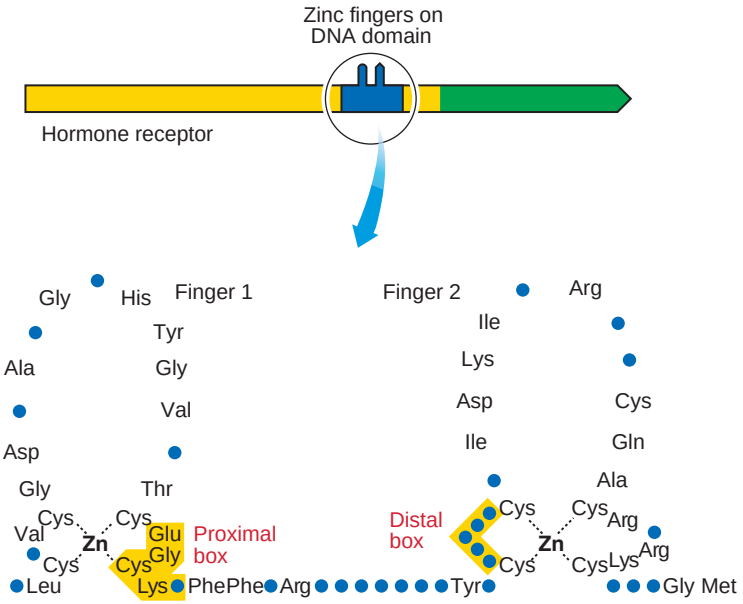
The hormone-binding characteristics of the ER-α and the ER-β are similar, indicating that they respond in a comparable manner to the same hormones.[70] Thus, both receptors bind to the estrogen response element with a similar affinity, and the affinity of estradiol for each receptor is similar. There are differences, however; for example, phytoestrogens have a greater affinity for ER-β than for ER-α. In other words, estrogenic agents demonstrate preferential binding for one or the other receptor. Different genetic messages can result not only because of differences in binding affinity, but also through variations in the mechanisms to be discussed, notably differences in conformational shape and cellular contexts. In addition, because the regulatory domains differ in the two receptors, the ability of ER-β to activate gene transcription by means of TAF-1 is impaired (discussed below).

A/B Region, The Regulatory Domain

The amino acid terminal is the most variable in the superfamily of receptors, ranging in size from 20 amino acids in the vitamin D receptor, to 600 amino acids in the mineralocorticoid receptor. In the ER-α, it contains several phosphorylation sites and the ***transcription activation function called TAF-1***. TAF-1 can stimulate transcription in the absence of hormone binding. The regulatory domain is considerably different in the two estrogen receptors; in ER-β, TAF-1 is either significantly modified or absent.

C Region, The DNA-Binding Domain

The middle domain binds to DNA and consists of 100 amino acids with nine cysteines in fixed positions, the two ***zinc fingers***. This domain is essential for activation of transcription. Hormone binding induces a conformational change in the three helices that allows binding to the hormone-responsive elements in the target gene. This domain is very similar for each member of the steroid and thyroid receptor superfamily; however, the genetic message is specific for the hormone that binds to the hormone-binding domain. The DNA-binding domain controls which gene will be regulated by the receptor and is responsible for target gene specificity and high-affinity DNA binding. The specificity of receptor binding to its hormone responsive element is determined by the zinc finger region, especially the first finger. The specific message can be changed by changing the amino acids in the base of the fingers. Substitutions of amino acids in the fingertips lead to loss of function. Functional specificity is localized to the second zinc finger in an area designated the d (distal) box. Different responses are due to the different genetic expression of each target cell (the unique activity of each cell's genetic constitution allows individual behavior).



D Region, The Hinge

The region between the DNA-binding domain and the hormone-binding domain contains a signal area that is important for the movement of the receptor to the nucleus following synthesis in the cytoplasm. This nuclear localization signal must be present for the estrogen receptor to remain within the nucleus in the absence of hormone. This region is also a site of rotation (hence the hinge designation) in achieving conformational change.

E Region, The Hormone-Binding Domain

The carboxy end of the estrogen receptor-α is the hormone-binding domain (for both estrogens and antiestrogens), consisting of 251 amino acids (residues 302–553). It consists of 12 helices with a folding pattern that forms a pocket where the hormones bind. The pocket is about 20% smaller in ER-β. In addition to hormone binding, this region contains the sites for cofactor binding, is responsible for ***dimerization,*** and harbors the ***transcription activation function called TAF-2***. This is also the site for binding by heat shock proteins (specifically hsp 90), and it is this binding to the heat shock proteins that prevents dimerization and DNA binding. In contrast to TAF-1 activity, TAF-2 depends on hormone binding for full activity.

F Region

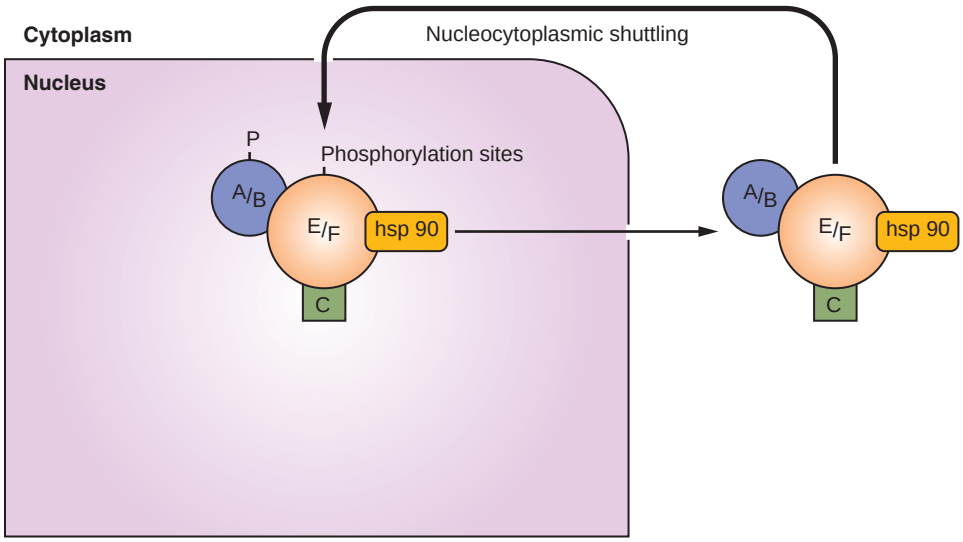
The F region of ER-α is a 42 amino acid C-terminal segment. This region modulates gene transcription by estrogen and antiestrogens, having a role that influences antiestrogen efficacy in suppressing estrogen-stimulated transcription.[71] The conformation of the receptor-ligand complex is different with estrogen and antiestrogens, and this conformation is different with and without the F region. The F region is not required for transcriptional response to estrogen; however, it affects the magnitude of ligand-bound receptor activity.

It is speculated that this region affects conformation in such a way that protein interactions are influenced. Thus, it is appropriate that the effects of the F domain vary according to cell type and protein context. The F region affects the activities of both TAF-1 and TAF-2, which is what one would expect if the effect is on conformation.[72]

Estrogen Receptor Mechanism of Action

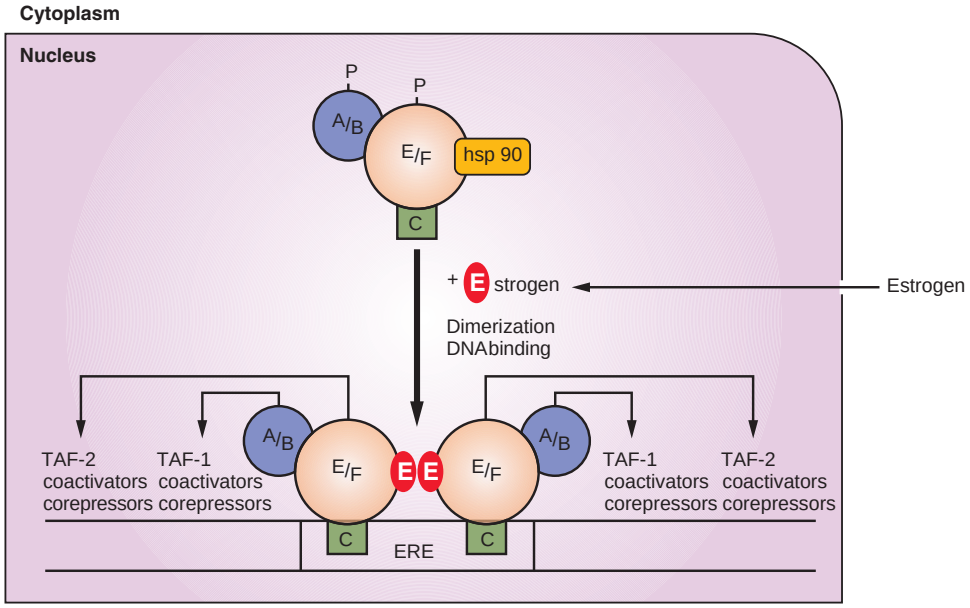
Ligand-Dependent Nuclear Activity

The steroid family receptors are predominantly in the nucleus even when not bound to a ligand, except for mineralocorticoid and glucocorticoid receptors where nuclear uptake depends on hormone binding. But the estrogen receptor does undergo what is called ***nucleocytoplasmic shuttling***. The estrogen receptor can diffuse out of the nucleus and be rapidly transported back in or undergo metabolism. When this shuttling is impaired, receptors are more rapidly degraded. Agents that inhibit dimerization (e.g., the pure estrogen antagonists) prevent nuclear translocation and thus increase cytoplasmic degradation.



In the absence of estrogen (the ligand), the receptor can become associated with the estrogen response element on a gene, a signal for a process that leads to its proteasome degradation via the ubiquitin pathway.[73] The receptor bound to its ligand, estrogen, undergoes the same process, but at a pace much slower than unliganded receptor, allowing time for gene transcription. This cyclic turnover allows the target cell to be very sensitive to the concentration of the ligand (estrogen) within the cell.

Prior to binding, the estrogen receptor is an inactive complex that includes a variety of chaperone proteins, including the heat shock proteins. Heat shock protein 90 appears to be a critical protein, and many of the others are associated with it. This heat shock protein is important not only for maintaining an inactive state, but also for causing proper folding for transport across membranes. “Activation” or “transformation” is the dissociation of heat shock protein 90.[74]



Imagine the unoccupied steroid receptor as a loosely packed, mobile protein complexed with heat shock proteins. The steroid family of receptors exists in this complex and cannot bind to DNA until union with a steroid hormone liberates the heat shock proteins and allows dimerization. The conformational change induced by hormone binding involves a dissociating process to form a tighter packing of the receptor. The hormone-binding domain contains helices that form a pocket.[75] After binding with a hormone (or with drugs engineered for this purpose), this pocket undergoes a conformational change that creates new surfaces with the potential to interact with coactivator and corepressor proteins. ***Conformational shape is an important factor in determining the exact message transmitted to the gene***. Conformational shape is slightly but significantly different with each ligand; estradiol, tamoxifen, and raloxifene each induce a distinct conformation that contributes to the ultimate message of agonism or antagonism.[76, 77] Tamoxifen and raloxifene, both TAF-2 antagonists, cause a steric repositioning, about a 90 degree rotation, of a helix (the TAF-2 helix) that then occupies the binding site of a coactivator in those tissues where such a coactivator is a requisite for TAF-2 activity. The weak estrogen activity of estriol is because of its altered conformational shape when combined with the estrogen receptor in comparison with estradiol.[78]

The hormone-binding domain of the estrogen receptors contains a cavity (the pocket) surrounded by a wedge-shaped structure, and it is the fit into this cavity that is so influential in the final genetic message. The size of this cavity on the estrogen receptor is relatively large, larger than the volume of an estradiol molecule, explaining the acceptance of a large variety of ligands. Thus, estradiol, tamoxifen, and raloxifene each bind in the same cavity within the hormone-binding domain, but the conformational shape with each is not identical.

Conformational shape is a major factor in determining the ability of a ligand and its receptor to interact with coactivators and corepressors. Conformational shapes are not simply either “on” or “off,” but intermediate conformations are possible providing a spectrum of agonist/antagonistic activity. The specific conformational shape of a receptor allows or prevents the recruitment of coactivators and corepressors that ultimately yield various biological reponses.

Members of the thyroid and retinoic acid receptor subfamily do not exist in inactive complexes with heat shock proteins. They can form dimers and bind to response elements in DNA, but without ligand, and they act as repressors of transcription.

Estrogen receptor mutants can be created that are unable to bind estradiol. These mutants can form dimers with natural estrogen receptor (wild type), and then bind to the estrogen response element, but they cannot activate transcription.[79] This indicates that transcription is dependent on the result after estradiol binding to the estrogen receptor, an estrogen-dependent structural change. Dimerization by itself is not sufficient to lead to transcription; neither is binding of the dimer to DNA sufficient.

Molecular modeling and physical energy calculations indicate that binding of estrogen with its receptor is not a simple key and lock mechanism. It involves conversion of the estrogen-receptor complex to a preferred geometry dictated to a major degree by the specific binding site of the receptor. The estrogenic response depends on the final bound conformation and the electronic properties of functional groups that contribute energy. The final transactivation function is dependent on these variables.

Estrogen, progesterone, androgen, and glucocorticoid receptors bind to their response elements as dimers, one molecule of hormone to each of the two units in the dimer. The estrogen receptor-α can form dimers with other alpha receptors (homodimers) or with an estrogen receptor-β (heterodimer). Similarly, the estrogen receptor-β can form homodimers or heterodimers with the alpha receptor. This creates the potential for many pathways for estrogen signaling, alternatives that are further increased by the possibility of utilizing various response elements in target genes. Cells that express only one of the estrogen receptors would respond to the homodimers; cells that express both could respond to a homodimer and a heterodimer.

The similar amino acid sequence of the DNA-binding domains in this family of receptors indicates evolutionary conservation of homologous segments. An important part of the conformational pattern consists of multiple cysteine-repeating units found in two structures, each held in a finger-like shape by a zinc ion, the so-called zinc fingers.[80] The zinc fingers on the various hormone receptors are not identical. These fingers of amino acids interact with similar complementary patterns in the DNA. Directed changes (experimental mutations) indicate that conservation of the cysteine residues is necessary for binding activity, as is the utilization of zinc.

The DNA binding domain is specific for an enhancer site (the hormone-responsive element) in the gene promoter, located in the 5′ flanking region. The activity of the hormone-responsive element requires the presence of the hormone-receptor complex. Thus, this region is the part of the gene to which the DNA-binding domain of the receptor binds. There are at least four different hormone-responsive elements, one for glucocorticoids/progesterone/androgen, one for estrogen, one for vitamin D3, and one for thyroid/retinoic acid.[81] These sites significantly differ only in the number of intervening nucleotides.

Binding of the hormone-receptor complex to its hormone-responsive element leads to many changes, only one of which is a conformational alteration in the DNA. Although the hormone-responsive elements for glucocorticoids, progesterone, and androgens mediate all of these hormonal responses, there are subtle differences in the binding sites, and there are additional sequences outside of the DNA-binding sites that influence activation by the three different hormones. The cloning of complementary DNAs for steroid receptors has revealed a large number of similar structures of unknown function. It is believed that the protein products of these sequences are involved in the regulation of transcription initiation that occurs at the TATA box.

There are three different RNA polymerases (designated I, II, and III), each dedicated to the transcription of a different set of genes with specific promoters (the site of polymerase initiation of transcription). ***Transcription factors are polypeptides, complexed with the polymerase enzyme, that modulate transcription either at the promoter site or at a***

sequence further upstream on the DNA.[82] The steroid hormone receptors, therefore, are transcription factors. The polymerase transcription factor complex can be developed in sequential fashion with recruitment of individual polypeptides, or transcription can result from interaction with a preformed complete complex. The effect can be either positive or negative, activation or repression.

In most cases, therefore, the steroid hormone receptor activates transcription in partnership with several groups of polypeptides[82]:

1. Other transcription factors—peptides that interact with the polymerase enzyme and DNA.

2. Coactivators and corepressors—peptides that interact with the TAF areas of the receptor, also called adaptor proteins or coregulators. Previously considered nuclear proteins, these regulators may also have functions within the cytoplasm.

3. Chromatin factors—structural organizational changes that allow an architecture appropriate for transcription response.

The steroid-receptor complex regulates the amount of mRNA transcripts emanating from target genes. The estrogen-occupied receptor binds to estrogen response elements in the 5′ flanking regions of estrogen-regulated genes, allowing efficient induction of RNA transcription. This can occur by direct binding to DNA and interaction with the estrogen response element or by protein interactions with ***coactivators*** between the estrogen receptor and DNA sites. ***Coactivators and corepressors are intracellular proteins, recruited by hormone receptors, that activate or suppress the TAF areas, by acting enzymatically either on the receptors or on DNA.***[83–88] Most of the genes regulated by estrogens respond within 1–2 hours after estrogen administration. Only a few respond within minutes. This time requirement may reflect the necessity to synthesize regulating proteins.[89] A large number (over 300—a current list is available at www.nursa.org) of coactivator and corepressor proteins have been identified and designated by code letters and numbers, suggesting that there is a process involved with these proteins, causing selection, activation, queuing, and coordination.[90] In general, corepressor proteins bind to hormone receptors in the absence of a ligand and suppress any basal transcription activity. Active investigation of this step will undoubtedly yield understanding of pathologic responses and new pharmacologic developments because it is already recognized that these proteins influence the phenotypes of human diseases.[91]

Just as coregulators can lead to different responses to the same hormones in different tissues, a new avenue of research indicates that posttranslational modifications of the coregulators can diversify the regulatory effects.[88, 90] The activity of a coregulator protein can be altered by modifications such as phosphorylation or methylation. The response to a specific hormone, therefore, can be complex, differing in various tissues as directed by the coregulators present in that tissue and how the coregulators are modified. The diversity from tissue to tissue is thus very complex, but even within a single cell impressive diversity is the norm. Because genomic differences among various species are amazingly small (1% or less), evolutionary differences in the way in which genes act, including the complex coregulatory story, make a major contribution to the phenotypic and behavioral differences in life forms, especially humans. In addition, coregulators provide another evolutionary reservoir that can yield adaptation to new environmental challenges and stresses.

The concentration of coactivators/corepressors can affect the cellular response, and this is another explanation for strong responses from small amounts of hormone. With a small amount of receptor but a large amount of coactivator/corepressor, the cell can be very responsive to a weak signal.

One of the aspects of activation, for example with the estrogen receptor, is an increase in affinity for estrogen. This is an action of estrogen, and it is greatest with estradiol and least with estriol. This action of estradiol, the ability of binding at one site to affect another site, is called ***cooperativity***. An increase in affinity is called positive cooperativity. The biologic advantage of positive cooperativity is that this increases the receptor's ability to respond to small changes in the concentration of the hormone. One of the antiestrogen actions of clomiphene is its property of negative cooperativity, the inhibition of the transition from a low-affinity to a high-affinity state. The relatively long duration of action exhibited by estradiol is due to the high-affinity state achieved by the receptor.

TAF (transcriptional activation function) is the part of the receptor that activates gene transcription after binding to DNA. Ligand binding produces a conformational change that allows TAFs to accomplish their tasks. TAF-1 can stimulate transcription in the absence of hormone when it is fused to DNA; however, it also promotes DNA binding in the intact receptor. TAF-2 is affected by the bound ligand, and the estrogen receptor depends on estrogen binding for full activity. TAF-2 consists of a number of dispersed elements that are brought together after estrogen binding. The activities of TAF-1 and TAF-2 vary according to the promoters in target cells. These areas can act independently or with one another. Indeed, the classic estrogen compounds (e.g., estradiol) produce a conformational shape that allows TAF-1 and TAF-2 to react in a synergistic fashion.

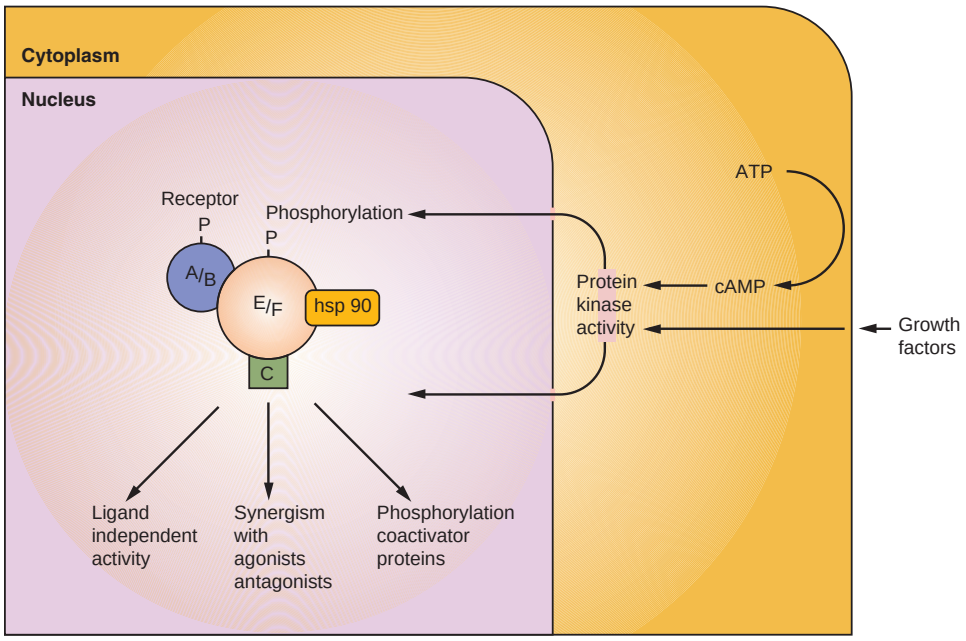
Thus the differential activities of the TAFs account for different activities in different cells. In addition to the binding of the dimerized steroid receptor to the DNA response element, steroid hormone activity is modulated by other pathways (other protein transcription factors and coactivators/corepressors) that influence transcription activation.[92, 93] ***This is an important concept: the concept of cellular context. The same hormone can produce different responses in different cells according to the cellular context of protein regulators, and responses can be altered by posttranslational modifications of the cogregulator proteins.***

Ligand-Independent Nuclear Activity

Phosphorylation of specific receptor sites is an important method of regulation, as well as phosphorylation of other peptides that influence gene transcription. Phosphorylation can be regulated by cell membrane receptors and ligand binding, thus establishing a method for cell membrane-bound ligands to communicate with steroid receptor genes.

Cyclic AMP and protein kinase A pathways increase transcriptional activity of the estrogen receptor by phosphorylation. In some cases phosphorylation modulates the activity of the receptor; in other cases, the phosphorylation regulates the activity of a specific peptide or coactivator/corepressor that, in turn, modulates the receptor. Phosphorylation follows steroid binding and occurs in both the cytoplasm and nucleus. Thus phosphorylation enhances activity of the steroid receptor complex.

Phosphorylation of the receptor increases the potency of the molecule to regulate transcription. Growth factors can stimulate protein kinase phosphorylation that can produce synergistic activation of genes or even ligand-independent activity. Epidermal growth factor (EGF), IGF-I, and transforming growth factor-alpha (TGF-α) can activate the estrogen receptor in the absence of estrogen, through the TAF-1 domain. This response to growth factors can be blocked by pure antiestrogens (suggesting that a strong antagonist locks the receptor in a conformation that resists ligand-independent pathways). The exact mechanism of growth factor activation is not known, but it is known that a steroid receptor can



be activated by means of a chemical signal (a phosphorylation cascade) originating at the plasma membrane. ***The recruitment of kinase activity is specific for specific ligands; thus not all ligands stimulate phosphorylation.***

Another explanation for strong responses from small amounts of steroids is a positive feedback relationship. Estrogen activates its receptor, gene expression stimulates growth factors (EFG, IGF-I, TGF-α, fibroblast growth factor), and the growth factors in an autocrine fashion further activate the estrogen receptor.[94] Ligand-independent activation of the estrogen receptor may be an important mechanism where estrogen levels are low, such as in the male.[95]

SUMMARY—Steps in the Steroid Hormone–Receptor Mechanism

1. Binding of the hormone to the hormone-binding domain that has been kept in an inactive state by various heat shock proteins.
2. Activation of the hormone-receptor complex, by conformational change, follows the dissociation of the heat shock proteins.
3. Dimerization of the complex.
4. Binding of the dimer to the hormone-responsive element on DNA at the zinc finger area of the DNA-binding domain.
5. Stimulation of transcription, mediated by transcription activation functions (TAFs), and influenced by the protein (other transcription factors and coactivators/corepressors) context of the cell, and by phosphorylation.

SUMMARY—Factors that Determine Biologic Activity

1. Affinity of the hormone for the hormone-binding domain of the receptor.
2. Target tissue differential expression of the receptor subtypes (e.g., ER-α and ER-β).
3. The concept of conformational shape: the structure of the ligand-receptor complex, with effects on two important activities: dimerization and the recruitment of regulating proteins.
4. The concept of cellular context: the differential expression of target tissue regulating proteins, coactivators and corepressors, and phosphorylation, yielding various biological responses.

Different Roles for ER-α and ER-β

Male and female mice have been developed that are homozygous for disruption of the estrogen receptor genes, "estrogen receptor-knockout mice."[68, 96, 97] Estrogen receptor α-deficient mice are known as αERKO or ERKO mice, ER-β deficient mice as βERKO or BERKO mice.[98]

Characteristics of ERKO and BERKO MICE

ERKO Mice	*BERKO Mice*
Normal lifespan	Normal lifespan
Anovulatory	Oligovulatory
Absent breast response at puberty	Normal breast glands and function
Normal G-U development but no adult response	Normal G-U development and normal adult response
Increase in visceral adiposity and insulin resistance	Normal body fat distribution and insulin secretion
Infertile males and females	Fertile males, subfertile females with reduced follicular growth

Spermatogenesis in the αERKO male is reduced and the testes undergo progressive atrophy, a result of a testicular role for estrogen, because gonadotropin levels and testicular steroidogenesis remain normal. Sexual mounting behavior is not altered, but intromission, ejaculation, and aggressive behaviors are reduced. Female mice with the alpha estrogen receptor gene disrupted do not ovulate, and the ovaries do not respond to gonadotropin stimulation. These female animals have high levels of estradiol, testosterone, and LH. FSH β-subunit synthesis is increased, but FSH secretion is at normal levels, indicating different sites of action for estrogen and inhibin. Uterine development is normal (due to a lack of testosterone in early life), but growth is impaired. Mammary gland ductal and alveolar development is absent. Female mice with absent alpha estrogen receptor activity do not display sexual receptive behaviors. This genetically engineered line of mice demonstrates essential activities for the alpha estrogen receptor. Relatively normal fetal and early development suggests that the beta estrogen receptor plays a primary role in

these functions. For example, the fetal adrenal gland expresses high levels of ER-β and low levels of ER-α.[99] However, nongenomic actions of estrogen are also possible and can explain some of the estrogenic responses in a knockout model. The results from estrogen receptor knockout mice as well as mice with disruption of the aromatase enzyme indicate that estrogen is essential for fertility, but not for the development of the reproductive tract or for survival.[100]

These genetic mice experiments also highlight the importance of estrogen in preventing the development of the metabolic syndrome. Knockout models for the estrogen receptors as well as the knockout model for the aromatase enzyme yield mice with hyperinsulinemia and increased visceral adiposity, with a reversal achieved by estrogen treatment.[20]

Differential expression of the alpha and beta receptors is present in various tissues (e.g., ER-β is the prevalent estrogen receptor in certain areas of the brain and the cardiovascular system) resulting in different and selective responses to specific estrogens.[97, 101, 102] Human granulosa cells from the ovarian follicle contain *only* ER-β mRNA; the human breast expresses both ER-α and ER-β, but ER-α is primarily involved in mammary development and function. Some parts of the rat brain contain only ER-β, others only ER-α, and some areas contain both receptors.[103] Target tissues that have been classically regarded as estrogen-sensitive (such as the uterus and the breast) express mainly ER-α. But, the knockout models have oversimplified the roles of ER-α and ER-β, at least in breast tissue; their roles are dynamic and changing, not a simple expression of always one or the other.

Two commonly used estrogens, 17β-estradiol and ethinyl estradiol (the estrogen component of steroid contraceptives) bind equally well to the alpha and beta estrogen receptors. However, ER-β plays a lesser role in those target tissues affected by these two estrogens, specifically the uterus, breast, bone, hypothalamus, and pituitary.

ER-β may have a regulatory role. In some tissues ER-β reduces ER-α-regulated gene transcription, even though in the absence of ER-α, ER-β can function as an estrogen receptor.[104] ER-β acts as a natural suppressor of estrogen (ER-α) activity in breast tissue, and that decreased concentrations of ER-β are associated with more aggressive tumors and reduced sensitivity to tamoxifen.[105, 106] The colon contains only ER-β, and the reduction in the risk of colonic cancer associated with postmenopausal estrogen therapy may reflect an antiproliferative activity of the beta receptor. Decreases in ER-β expression have been observed in cancers occurring in the endometrium ovary, colon, and prostate.[107]

The estrogen story is further complicated by the fact that the same estrogen binding to the alpha and beta receptors can produce opposite effects. For example, estradiol can stimulate gene transcription with ER-α at a given site of the estrogen response element, whereas estradiol inhibits gene transcription with ER-β in this same system.[108] In other tissues, the opposite scenario can occur with estradiol increasing ER-β expression. Different and unique messages, therefore, can be determined by the specific combination of (1) a particular estrogen, (2) the alpha or beta receptor, and (3) the targeted response element. ***To some degree, differences with ER-α and ER-β are influenced by activation of TAF-1 and TAF-2; agents that are capable of mixed estrogen agonism and antagonism produce agonistic messages via TAF-1 with ER-α, but because ER-β lacks a similar TAF-1, such agents can be pure antagonists in cells that respond only to ER-β.***[106] ER-α and ER-β affect the peptide context of a cell, especially coactivators and corepressors, differently. At least one component of this differing behavior is the fact that the two receptors do not bind to DNA in the same exact site, and the locations have different properties that could account for some of the differences in effects produced by each receptor; specifically the two receptors can each activate regions of a gene, but some regions selectively respond to one or the other.[109]

Ligand–Cell Membrane Extranuclear Receptor Activity

Not all actions of estrogen, and presumably all steroid hormones, are genomic, requiring gene transcription. Rapid cellular responses after estrogen stimulation are initiated by estrogen binding at the level of the cell wall membrane. These responses are traditionally associated with growth factors and G protein-coupled receptors. However, it is not appropriate to designate this activity as "nongenomic" because the cell membrane, estrogen-induced signaling, leads to both gene transcription and to events independent of transcription. This extranuclear pathway activates various protein kinases that can cause ion fluxes of calcium and potassium, modification of second messenger systems, and indirect effects on growth factors, transcription factors, and genetic promoters.[110] A good example of membrane-associated estrogen activity is the stimulation of endothelial nitric oxide synthase.[111] Putative membrane receptors have been reported to be both related to ER-α and different from the estrogen receptor. The expression of a truncated isoform of ER-α involved in acute activation of nitric oxide has been described in the caveoli of human vascular endothelial cells.[112] A G protein-coupled receptor localized to the endoplasmic reticulum has been identified that binds estrogen and affects intracellular functions.[113]

Ligand-Dependent, ERE-Independent Activity

We have described three pathways that mediate estrogen activity:

1. Ligand-Dependent Nuclear Activity: the classical mechanism involving the estrogen receptors with binding to DNA estrogen response elements.

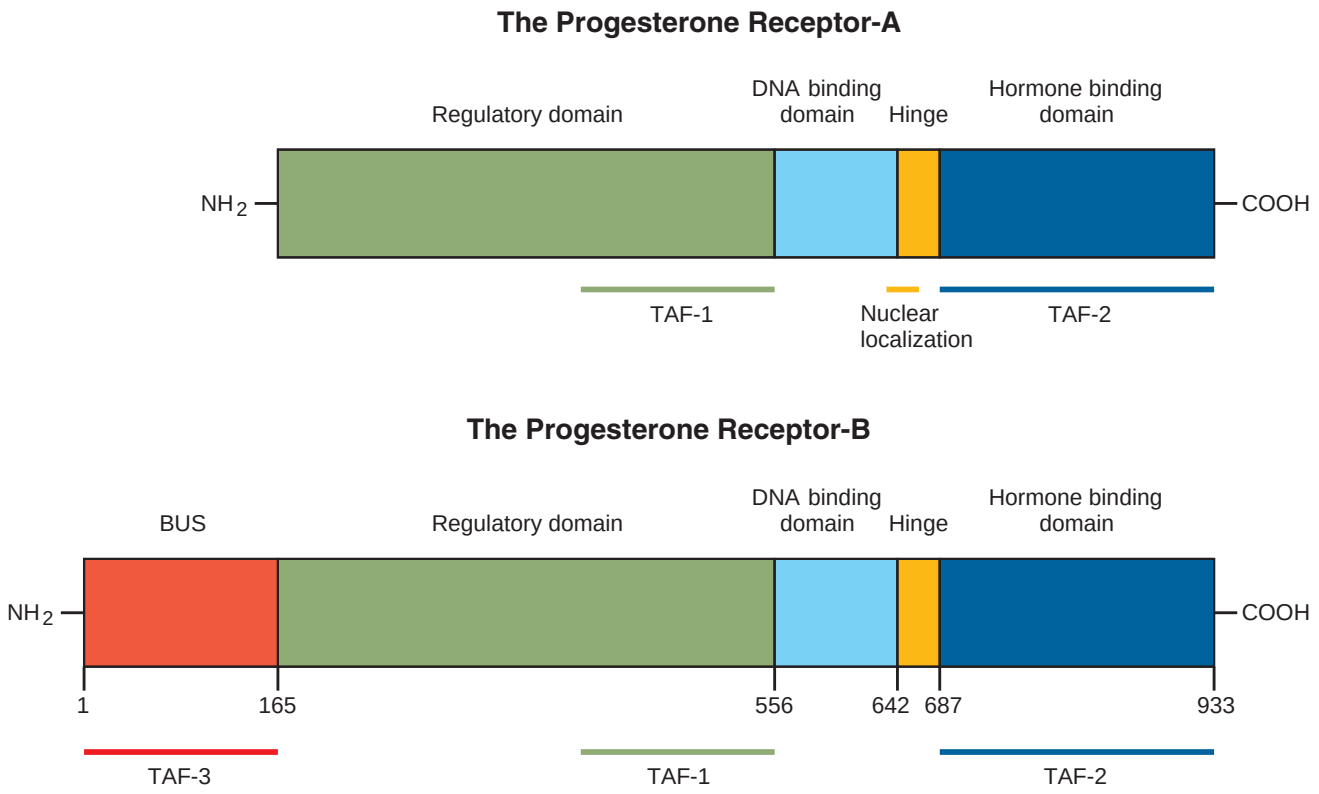
2. Ligand-Independent Nuclear Activity: activation of the estrogen receptor pathway through second messengers, involving phosphorylation of estrogen receptors and coregulator proteins.

3. Ligand–Cell Membrane Extranuclear Receptor Activity: rapid responses mediated by estrogen receptors in cell membranes.

There is at least one more pathway, evident from studies with mutant mice, the ligand-dependent, ERE-independent nuclear pathway.[114] In these animals, mutant forms of the estrogen receptor cannot bind to ERE, the estrogen response element in DNA, and yet some physiological actions of estrogen can be demonstrated, such as negative feedback inhibition of LH secretion, indicating hormonal mediation with its receptor of a cellular response, but not through the classical pathway.

The Progesterone Receptor

The progesterone receptor is induced by estrogens at the transcriptional level and decreased by progestins at both the transcriptional and translational levels (probably through receptor phosphorylation).[115, 116] The progesterone receptor (in a fashion similar to the estrogen receptor) has three major forms, designated the A, B, and C receptors.[117] The three isoforms are expressed by a single gene on chromosome 11 at q22–23; the three forms are a consequence of transcription from distinctly different promoters, in a complex system of transcription regulation.[118] Each form is associated with additional proteins, which



are important for folding of the polypeptide into a structure that allows hormone binding and receptor activity.[119] The molecular weight of A is 94,000 and B, 114,000, with 933 amino acids, 164 more than A. The B receptor has a unique upstream segment (128–165 amino acids, depending on the species) referred to as the B-upstream segment (BUS). The C receptor is the smallest and lacks the ability to initiate transcription, functioning in some tissues as an inhibitor of the B receptor.

On the progesterone receptor, TAF-1 is located in a 91-amino acid segment just upstream of the DNA-binding domain. TAF-2 is located in the hormone-binding domain. A fragment missing the hormone-binding domain activates transcription to levels comparable to full-length hormone-activated B receptors, and higher than that with the A receptor, thus beyond that of TAF-1 alone. In appropriate cells, therefore, BUS contains a third activation domain, TAF-3, and can autonomously activate transcription or it can synergize with the other TAFs.[120] In the absence of hormone binding, the C-terminal region of the progesterone receptor exerts an inhibitory effect on transcription.[121] Progesterone agonists induce a conformational change that overcomes the inherent inhibitory function within the carboxy tail of the receptor. Binding with a progesterone antagonist produces a structural change that allows the inhibitory actions to be maintained.

Progestational agents can elicit a variety of responses determined by target tissue production and activity of the two receptor forms with dimerization as AA and BB (homodimers) or AB (heterodimer). The progesterone receptors function in the mechanism shared by this superfamily of receptors: an unbound complex with heat shock proteins, hormone binding, dimerization, DNA binding to a progesterone response element, and modulation of transcription by phosphorylation and various proteins.[77, 122]

A and B are expressed in varying amounts in breast cancer and endometrial cancer cell lines. Studies indicate that the two receptors can be regulated independently; e.g., the relative levels differ in endometrium during the menstrual cycle.[123] Tissue specificity with the progesterone

receptor is influenced by which receptor and which dimer is active, and in addition, the transcriptional activities of A and B depend on target cell differences, especially in promoter context. ***However, in most cells, B is the positive regulator of progesterone-responsive genes, and A inhibits B activity***. Mutations within the carboxy terminus of B affect the transcriptional activity of B. But mutations in A have no effect on its transcriptional inhibitory activity. This indicates two separate pathways for transcription activation and repression by the progesterone receptor. Thus, repression of human estrogen receptor transcriptional activity (as well as glucocorticoid, mineralocorticoid, and androgen transcription) is dependent on the expression of A.[124, 125] The A and B progesterone receptors have different molecular functions, affecting different genes, and, therefore, target tissue response to progesterone will be influenced by the differential expression of each receptor and the ratio of their concentrations, as well as the target tissue context of adaptor proteins.[126, 127]

The broad activity of A in regard to all steroids suggests that A regulates inhibition of steroid hormone action wherever it is expressed. A does not form a heterodimer with the estrogen receptor. A does not prevent the estrogen receptor from binding with DNA. A does not change the structure of the estrogen receptor. Therefore, either A competes with the estrogen receptor for a critical protein; in this case A would inhibit the estrogen receptor only in cells that contain the critical factor, or the target is a critical protein, again an essential transcription activator.[119, 123]

Progesterone shares with estrogen (and probably all steroid hormones) the ability to exert activity at the cell membrane, independently of the progesterone receptor.[128] For example, progesterone or a progesterone metabolite can prevent uterine contractions by binding to the oxytocin G protein receptor in the cell membrane and inhibiting its function.[129]

PRKO mice (lacking both progesterone receptors) are unable to ovulate due to a failure to expel a mature oocyte in a fully developed follicle, specifically a failure in the LH-induced rupture of a follicle.[130] When only PR-A is deficient, ovulation is severely impaired, but not totally reduced, indicating that both receptors contribute to ovulation, but PR-A is essential for normal function.

Like estrogen, G protein-coupled receptors for progesterone have been identified, providing a pathway for progesterone activation of various signaling cascades involved in cellular functions, including gene expression.[131] A role for a membrane receptor has been proposed for progesterone's antiapoptotic actions in ovarian granulosa cells.[132]

The Androgen Receptor

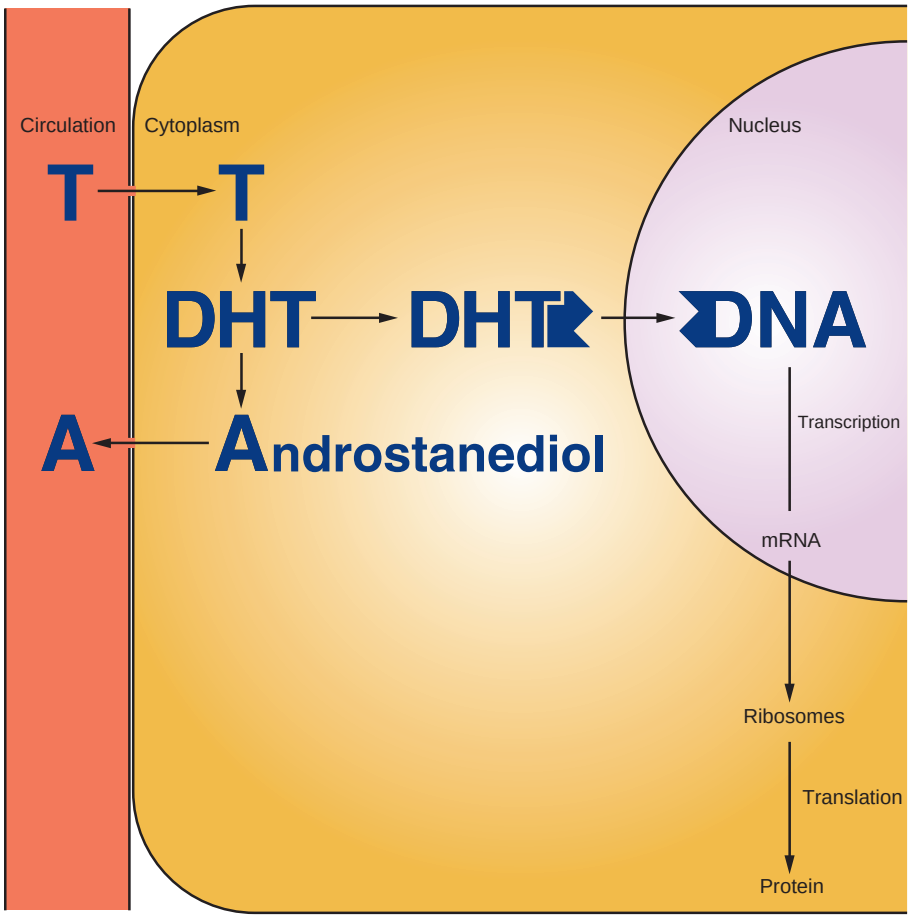
The cellular mechanism is more complex for androgens. Androgens can work in any one of three ways:

1. By intracellular conversion of testosterone to dihydrotestosterone (DHT), intracrine activity.

2. By testosterone itself, endocrine activity.

3. By intracellular conversion of testosterone to estradiol (aromatization), intracrine activity.

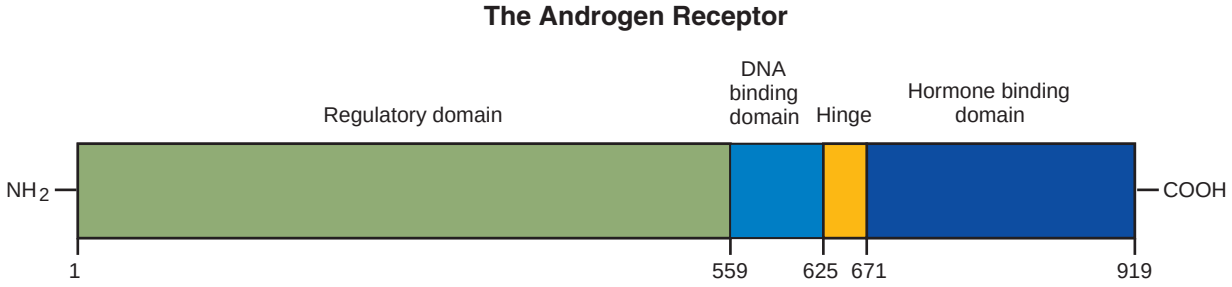
Tissues that exclusively operate via the testosterone pathway are the derivatives of the wolffian duct, whereas hair follicles and derivatives of the urogenital sinus and urogenital tubercle require the conversion of testosterone to DHT. The hypothalamus actively converts



androgens to estrogens; hence, aromatization may be necessary for certain androgen feedback messages in the brain.

In those cells that respond only to DHT, only DHT will be found within the nucleus activating messenger RNA production. Because testosterone and DHT bind to the same high-affinity androgen receptor, why is it necessary to have the DHT mechanism? One explanation is that this is a mechanism for amplifying androgen action, because the androgen receptor preferentially will bind DHT (greater affinity). The antiandrogens, including cyproterone acetate and spironolactone, bind to the androgen receptor with about 20% of the affinity of testosterone.[133] This weak affinity is characteristic of binding without activation of the biologic response.

The androgen receptor, like the progesterone receptor, exists as the full-length B form and a shorter A form.[134] It is likely that the A and B forms of the androgen receptor have functional differences. The amino acid sequence of the androgen receptor in the DNA-binding domain resembles that of the receptors for progesterone, mineralocorticoids, and glucocorticoids but most closely that of the progesterone receptor.[135] Androgens and progestins can cross react for their receptors but do so only when present in pharmacologic concentrations. Progestins compete not only for androgen receptors but also for the metabolic utilization of the 5α-reductase enzyme. The dihydroprogesterone that is produced, in turn, also competes with testosterone and DHT for the androgen receptor. A progestin, therefore, can act both as an antiandrogen and as an antiestrogen. Androgen-responsive gene expression can also be modified by estrogen; it has been known for years that androgens and estrogens can counteract each other's biologic responses. These responses of target tissues are determined by gene interactions with the hormone-receptor complexes, androgen with its



receptor and estrogen with its receptor. The ultimate biologic response reflects the balance of actions of the different hormones with their respective receptors, modified by various transcription regulators.

The syndrome of androgen insensitivity represents a congenital abnormality in the androgen intracellular receptor (several hundred unique mutations have been identified—www.androgendb.mcgill.ca).[136, 137] The androgen receptor gene is localized on the human X chromosome at Xq11-12, the only steroid hormone receptor to be located on the X chromosome.[138] Thus, androgen insensitivity is an X-linked disorder. Molecular studies of patients with androgen insensitivity have indicated a deletion of amino acids from the steroid-binding domain due to nucleotide alterations in the gene that encodes the androgen receptor.[139] What was once a confusing picture is now easily understood as a progressive increase in androgen receptor action. At one end, there is a complete absence of androgen binding—complete androgen insensitivity. In the middle is a spectrum of clinical presentations representing varying degrees of abnormal receptors and binding. At the other end, about 25% of infertile men with normal genitalia and normal family histories have azoospermia due to a receptor disorder.[140, 141] The androgen receptor also plays a role in motor neuron physiology, because a specific mutation in the androgen receptor is responsible for Kennedy's disease (X-linked spinobulbar muscular atrophy), a condition associated with motor neuron degeneration.[142]

Agonists and Antagonists

An agonist is a substance that stimulates a response. An antagonist inhibits the actions of an agonist. Agonistic activity follows receptor binding, which leads to stimulation of the message associated with that receptor. Antagonistic activity follows receptor binding and is characterized by blockage of the receptor message or nontransmission of the message. Most compounds used in this fashion that bind to hormone nuclear receptors have a mix of agonist and antagonist responses, depending on the tissue and hormonal milieu. Examples of antagonists include tamoxifen, mifepristone (RU 486), and the histamine receptor antagonists.

Short-Acting Antagonists

Short-acting antagonists, such as estriol, are actually a mixed combination of agonism and antagonism depending on time. Short-term estrogen responses can be elicited because estriol binds to the nuclear receptor, but long-term responses do not occur because this binding is short-lived. Antagonism results when estriol competes with estradiol for receptors. However, if a constant presence of the weak hormone, estriol, can be maintained, then long-term occupation is possible, and a potent estrogen response can be produced.

Long-Acting Antagonists

Clomiphene and tamoxifen are mixed estrogen agonists and antagonists. The endometrium is very sensitive to the agonistic response, whereas the breast is more sensitive to the antagonistic behavior. The antagonistic action is the result of nuclear receptor binding with an alteration in normal receptor-DNA processing and eventual depletion of hormone receptors.

Alteration of the GnRH molecule has produced both agonists and antagonists. GnRH is a decapeptide; antagonists have substitutions at multiple positions, while agonists have substitutions at the 6 or 10 positions. The GnRH agonist molecules first stimulate the pituitary gland to secrete gonadotropins, then because of the constant stimulation, down-regulation and desensitization of the cell membrane receptors occur, and gonadotropin secretion is literally turned off. The antagonist molecules bind to the cell membrane receptor and fail to transmit a message and thus are competitive inhibitors. Various GnRH agonists and antagonists are used to treat endometriosis, uterine leiomyomas, precocious puberty, cancer of the prostate gland, ovarian hyperandrogenism, and the premenstrual syndrome.

Physiologic Antagonists

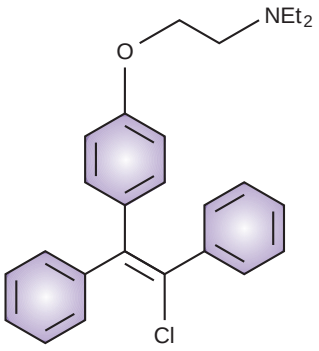
Strictly speaking, a progestin is not an estrogen antagonist. It modifies estrogen action by causing a depletion of estrogen receptors. There is also evidence that a progestin can inhibit transcription activation by the estrogen receptor.[143] In addition, progestins induce enzyme activity that converts the potent estradiol to the impotent estrone sulfate, which is then secreted from the cell.[144] Androgens block the actions of estrogen, also by depleting target tissues of estrogen receptors.

Antiestrogens

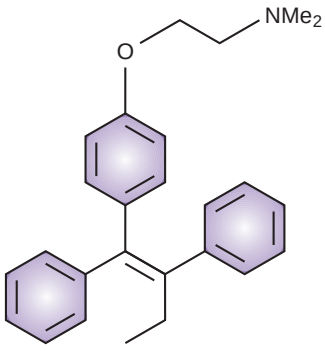
Currently, there are two groups of antiestrogens: pure antiestrogens and compounds with both agonistic and antagonistic activities. The mixed agonist–antagonist compounds include both the triphenylethylene derivatives (the nonsteroidal estrogen relatives such as clomiphene and tamoxifen) and the nonsteroidal sulfur-containing agents (the benzothiophenes, such as raloxifene). The pure antiestrogens have a bulky side chain that, with only a little imagination, can be pictured as an obstruction to appropriate conformational changes. An ideal antiestrogen would have the following properties:

1. A compound that would be a pure antagonist on proliferating breast carcinoma cells.
2. Development of resistance would be rare or require long exposure.
3. High affinity for the estrogen receptor so that therapeutic doses could be easily achieved.
4. No interference with the beneficial actions of estrogens.
5. No toxic or carcinogenic effects.

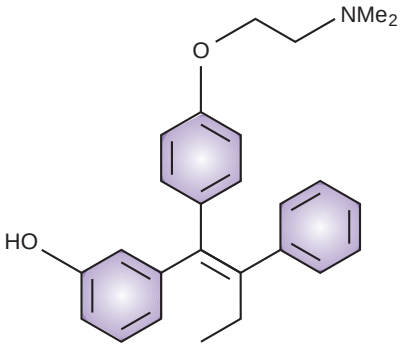
Triphenylethylene derivatives



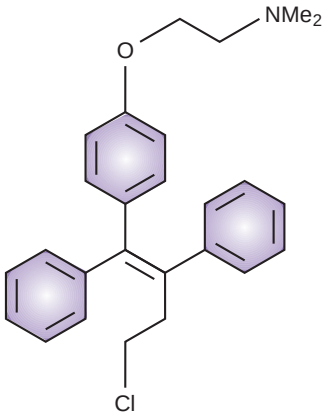
Clomiphene



Tamoxifen

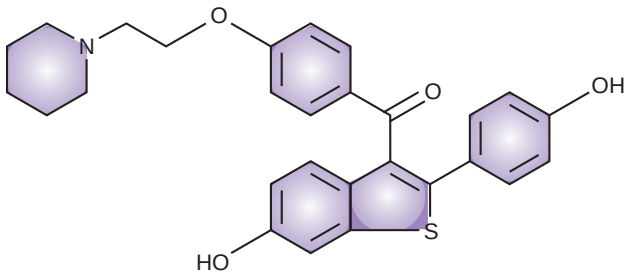


Droloxifene



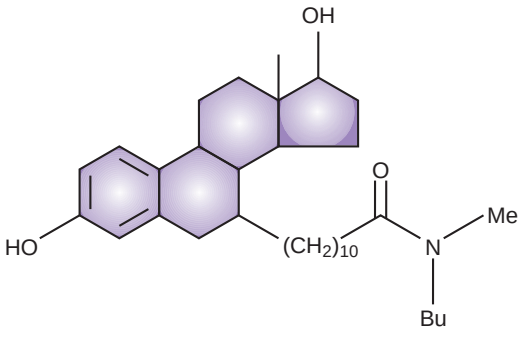
Toremifene

Benzothiphene derivative

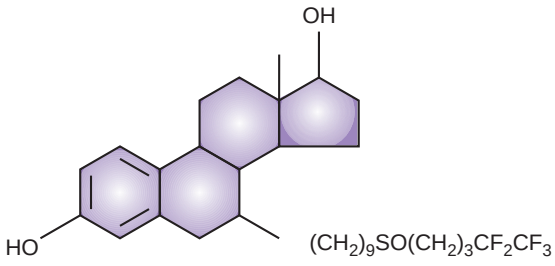


Raloxifene

Pure antiestrogens



ICI 164.384



Fulvestrant
(ICI 182,780)

The Antiestrogen Tamoxifen

Tamoxifen is very similar to clomiphene (in structure and actions), both being nonsteroidal compounds structurally related to diethylstilbestrol. Tamoxifen, in binding to the estrogen receptor, competitively inhibits estrogen binding. In vitro, the estrogen binding affinity for its receptor is 100–1,000 times greater than that of tamoxifen. Thus, tamoxifen must be present in a concentration 100–1,000 times greater than estrogen to maintain inhibition of breast cancer cells. In vitro studies demonstrated that this action was not cytocidal, but, rather, cytostatic (and thus its use must be long-term). The tamoxifen-estrogen receptor complex binds to DNA, but whether an agonistic, estrogenic message or an antagonistic, antiestrogenic message predominates is determined by what promoter elements (coactivators) are present in specific cell types. If the mechanism is cytostatic, why does a treatment period of 5 years provide protection against recurrent disease for at least 10 years? It has been suggested that exposure to tamoxifen sensitizes cells to the apoptotic effects of a woman's own estrogen levels.[145, 146]

There have been many clinical trials with adjuvant treatment of breast cancer with tamoxifen, and many are still ongoing.[147–149] Overall, the impact of tamoxifen treatment on breast cancer can be summarized as follows: Disease-free survival is prolonged. There is an increased survival after 5 years of treatment and 10 years of follow-up of approximately 26%, most evident in women over age 50. Response rates in advanced breast cancer are 30–35%, most marked in patients with tumors that are positive for estrogen receptors, reaching 75% in tumors highly positive for estrogen receptors.

Serum protein changes reflect the estrogenic (agonistic) action of tamoxifen. This includes decreases in antithrombin III, cholesterol, and LDL-cholesterol, while sex hormone-binding globulin (SHBG) levels increase (as do other binding globulins). The estrogenic activity of tamoxifen, 20 mg daily, is nearly as potent as 2 mg estradiol in lowering FSH levels in postmenopausal women, 26% versus 34% with estradiol.[150] The estrogenic actions of tamoxifen include the stimulation of progesterone receptor synthesis, an estrogen-like maintenance of bone, and estrogenic effects on the vaginal mucosa and the endometrium. Tamoxifen causes a decrease in antithrombin III, and there has been a small increase in the incidence of venous thromboembolism observed in tamoxifen-treated patients compared with controls.[151–153]

All too often, the antagonistic, antiestrogenic action of tamoxifen is featured, and the estrogenic, agonistic action is ignored. There is about a 4-fold increase in endometrial cancer occurring in women receiving tamoxifen treatment.[153–155] In addition, tamoxifen has been associated with major flare-ups in endometriosis. Tamoxifen, therefore, has a variety of side effects that indicate both estrogenic activity and antiestrogenic activity. How can tamoxifen be both an estrogen agonist and an estrogen antagonist?

Tamoxifen Mechanism of Action

TAF-1 and TAF-2 areas can both activate transcription, but TAF-2 activates transcription only when it is bound by estrogen. The individual transactivating abilities of TAF-1 and TAF-2 depend on the promoter and cell context. Tamoxifen's agonistic ability is due to activation of TAF-1; its antagonistic activity is due to competitive inhibition of the estrogen-dependent activation of TAF-2.

An estrogen-associated protein, a coactivator, binds to the right hand side of TAF-2. Estrogen binding induces binding of this protein, which then activates transcription. This protein

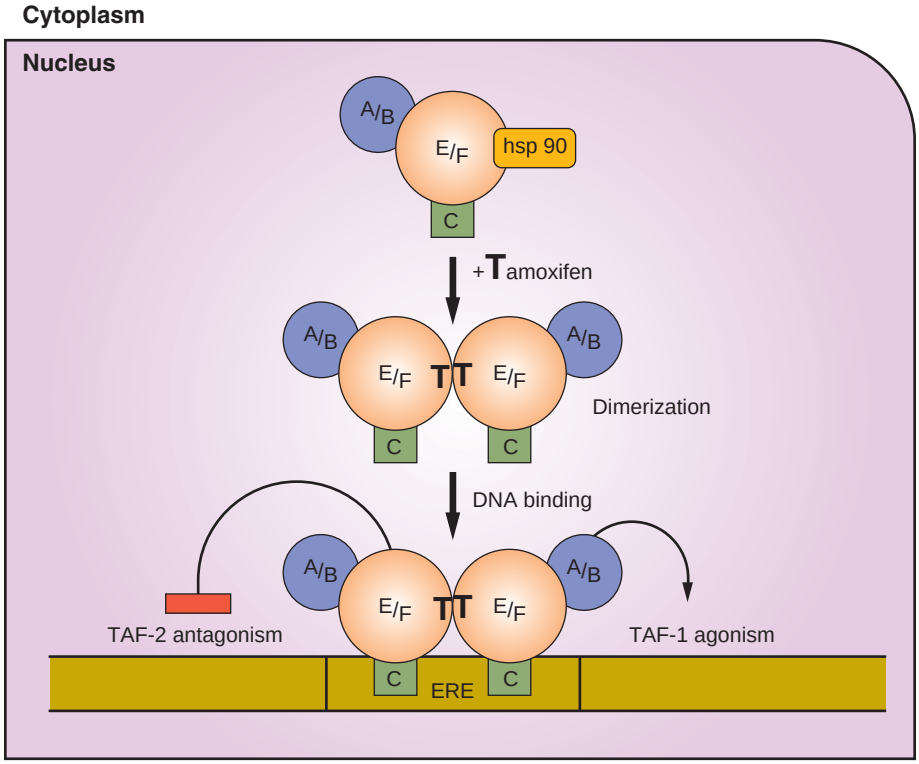
recognizes only an activated conformation of the estrogen receptor, the result of estrogen binding. Tamoxifen binding to the TAF-2 area does not activate this domain because, in at least one explanation, the conformational change does not allow binding of the estrogen-associated protein, the activating factor.[83,156] Antagonism of TAF-2 activity is further enhanced by the recruitment of corepressors after tamoxifen binds to the hormone-binding domain.[93]

The activity of TAF-2 is negligible in the presence of tamoxifen. In cells where TAF-1 and TAF-2 function independently of each other, tamoxifen would be chiefly an antagonist in cells where TAF-2 predominates, and an agonist where TAF-1 predominates, and in some cells a mixed activity is possible.[157]

The contact sites of estrogens and antiestrogens with the estrogen receptor are not identical.[158] When an antiestrogen binds to the estrogen receptor, the conformational changes that are induced alter the ability of the estrogen receptor-antiestrogen complex to modulate transcriptional activity. The relative agonist–antagonist activity is determined by the specific conformation achieved by the specific antiestrogen.

Even though tamoxifen can block estrogen-stimulated transcription of many genes, its degree of antagonistic activity varies among different animals, different cell types, and with different promoters within single cells. These differences are due to differences in the relative activities of the TAFs. Thus, the extent to which an antiestrogen inhibits an estrogen-mediated response depends on the degree to which that response is mediated by TAF-2 activity as opposed to TAF-1 activity, or mixed activity.[159] In some cell lines TAF-1 is dominant; in others, both are necessary. No cells have yet been identified where TAF-2 is dominant.

In most cell types, TAF-1 is too weak to activate transcription by itself, but, of course, there are well-known exceptions: endometrium, bone, and liver. In these tissues, the promoter context is right. Tamoxifen is a significant activator of estrogen receptor-mediated induction of promoters that are regulated by the TAF-1 site.

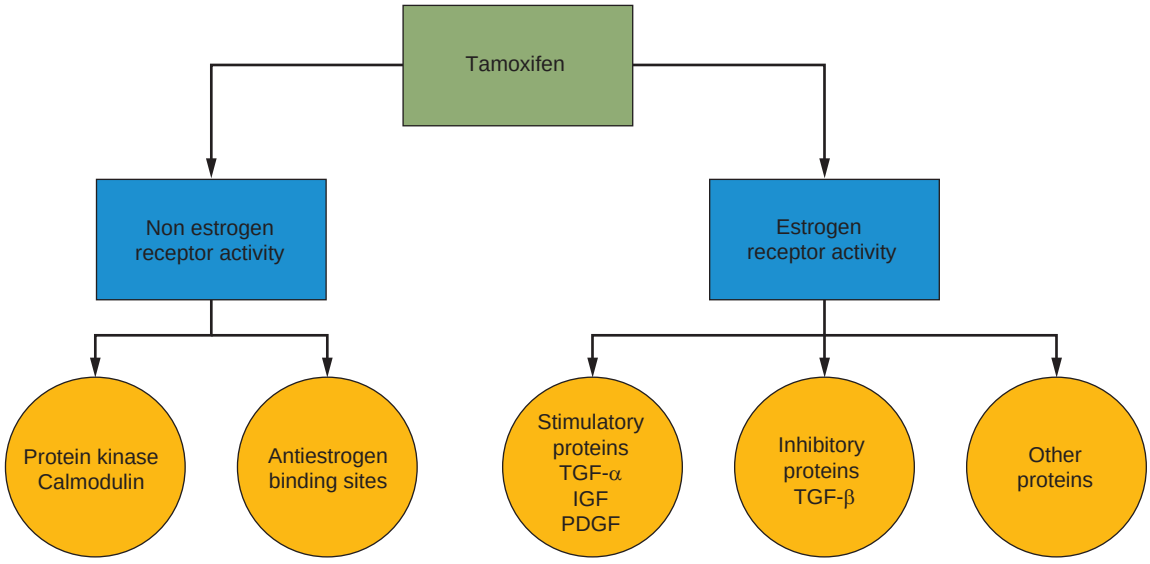


SUMMARY — The Response of Cells to Estrogens and Antiestrogens Depends on:

1. The nature of the estrogen receptor.
2. The estrogen response elements and nearby promoters.
3. The cell context of protein coactivators and corepressors.
4. The properties of the ligand.
5. Modulation by growth factors and agents that affect protein kinases and phosphorylation.

Tamoxifen Treatment of Breast Cancer

Tamoxifen treatment achieves its greatest effect (50% reduction in recurrent disease) in estrogen receptor-positive tumors, but it is also effective in estrogen receptor-negative tumors. Most importantly, it is now recognized that acquired resistance eventually develops. Therefore, there are two important questions. *Why is tamoxifen treatment effective with estrogen receptor-negative tumors? How does tamoxifen resistance develop?*



Efficacy of Tamoxifen with Estrogen Receptor-Negative Tumors

Besides binding to the estrogen receptor and providing competitive inhibition, tamoxifen has the following actions:

1. Tamoxifen and clomiphene inhibit protein kinase C activity (phosphorylation).
2. Tamoxifen inhibits calmodulin-dependent cyclic AMP phosphodiesterase, by binding to calmodulin.

3. Tamoxifen and estrogen have opposing effects on growth factors. 160, 161 Tamoxifen stimulates secretion of TGF-β in breast cancer cells as well as in fibroblasts and stromal cells, and TGF-β inhibits growth of breast cancer cells, whereas estrogen and insulin decrease the secretion of TGF-β in cancer cells. Tamoxifen decreases and estrogen increases IGF-I and IGF-II production in stromal fibroblasts.

Some of these actions (especially inhibition of protein kinase activity and stimulation of TGF-β production) occur independently of tamoxifen binding to the estrogen receptor, and thus, estrogen receptor-negative tumors can be affected by these actions; however, the overall impact of tamoxifen on recurrence or death in women with estrogen-receptor-poor tumors is negligible.[148]

Mechanisms for Tamoxifen Resistance

The results of randomized clinical trials have indicated that there is little reason to extend tamoxifen treatment of breast cancer patients beyond 5 years.[152, 162, 163] Indeed, the data suggested that survival and recurrence rates worsened with longer therapy, probably due to the emergence of tamoxifen-resistant tumors. There are several possible explanations for resistance, and whichever of these are operative, it is believed that a subpopulation resistant to tamoxifen is present from the beginning, and over time grows to be clinically apparent.[164]

1. Loss of estrogen receptors.

Generally it is believed that estrogen receptor expression is not a permanent phenotype of breast cancer cells; thus, tumors can change from receptor-positive to receptor-negative. But more than 50% of resistant tumors retain estrogen receptors.[165] The conventional wisdom has been that progression is associated with loss of cellular control and loss of estrogen receptor expression. However, the correlation between metastatic disease and estrogen receptor-negative state is not strong. Indeed, metastatic disease with estrogen receptor-positive cells despite an estrogen receptor-negative primary lesion has been reported. In addition, the rate of estrogen receptor expression is about the same in *in situ* disease and invasive disease. Most normal breast cells are estrogen receptor-negative, and in vitro, cell lines maintain their receptor status. Thus, there is little reason to believe that tamoxifen-resistant tumors lose receptor expression. *The importance of this is that resistance is not a wild, potentially uncontrollable dedifferentiation*.[166]

2. Variant and mutant estrogen receptors.

Mutations in resistant breast tumors are infrequent and are unlikely to account for resistance.[167] Studies of breast tumors from tamoxifen-resistant patients indicate that most express wild-type normal estrogen receptor; very few mutated estrogen receptors have been described.

3. Changes in coactivators.

If a breast cancer cell were to begin expressing these factors in a fashion similar to endometrium or bone, then agonistic actions would occur.

4. Cross talk between signaling pathways.

Because of the synergism between the estrogen receptor and protein kinase pathways, stimulation of protein kinase pathways can change an antagonist message to agonism.[168]

This mechanism operates through the phosphorylation of the estrogen receptor or proteins involved in estrogen receptor-mediated transcription. Stimulation of this protein kinase phosphorylation activates the agonist activity of tamoxifen-like antiestrogens. Furthermore, the lack of response of pure antiestrogens to this phosphorylation may be part of the reason for the response of resistant tumors to pure antiestrogens.

5. Binding to other proteins.

A remote possibility is the prevention of action by binding to other proteins, such as antiestrogen binding sites, microsomal proteins that bind to tamoxifen with high affinity but do not bind estrogen.[169]

6. Differential cellular transport.

Overexpression of the transmembrane efflux pump that excretes compounds from cells could diminish the intracellular amount of tamoxifen present.

7. Differential metabolism.

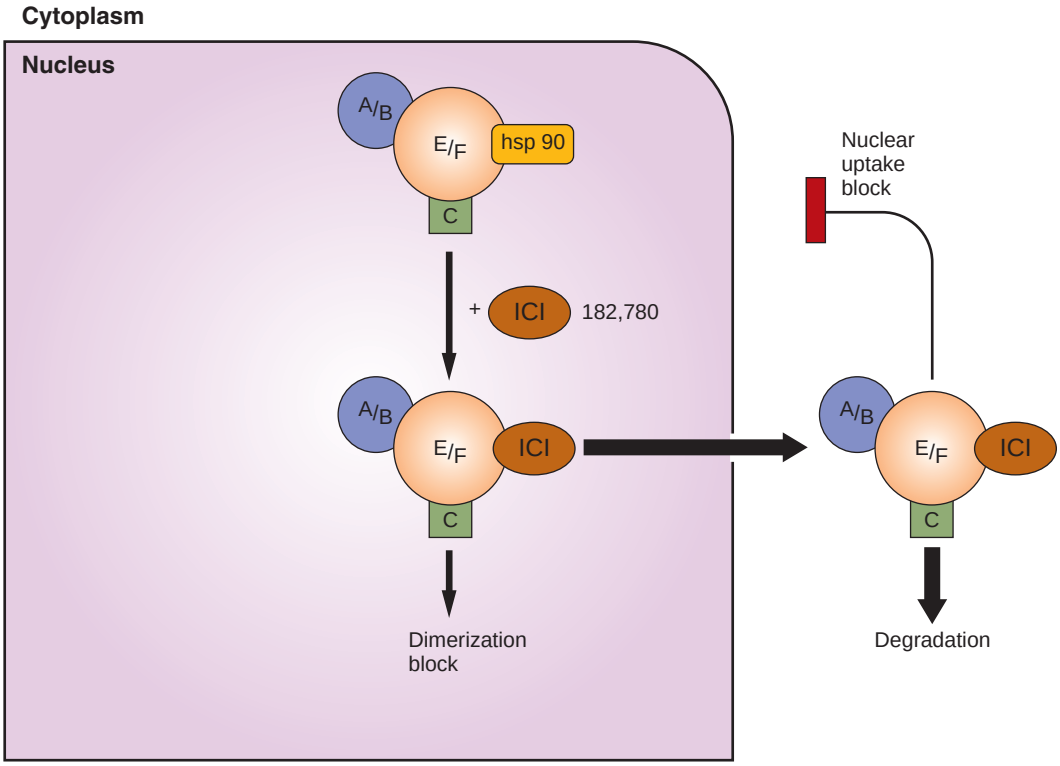
Changes in pharmacology and metabolism of tamoxifen might occur so that cells acquire the ability to metabolize the antagonist to greater agonist activity. Some breast cancer patients develop tumors that regress when tamoxifen is withdrawn. However, estrogenic metabolites of tamoxifen have not been identified.

Tamoxifen resistance occurs because essentially the estrogen receptor is not the dominant mechanism involved in the growth of these cells. Evidence supports growth factor stimulation and kinase phosphorylation as the predominant systems in tamoxifen-resistant cells. These cells are hypersensitive to estrogen and respond to tamoxifen as an agonist.[170]

Randomized trials have demonstrated the superiority of aromatase inhibitors compared with tamoxifen for the treatment of hormone-sensitive early breast cancer. This includes better disease-free survival, a reduction in new contralateral primary tumors, and an increased time to recurrence. Because of this superiority, the standard has shifted from tamoxifen to aromatase inhibitors (as discussed in Chapter 16). Nevertheless, tamoxifen made a major contribution to the molecular understanding of hormonal action.

The Pure Antiestrogens

The pure antiestrogens are derivatives of estradiol with long hydrophobic side chains at the 7 position. Binding with the pure antiestrogens prevents DNA binding. Because the site responsible for dimerization overlaps with the hormone-binding site, it is believed that pure antiestrogens sterically interfere with dimerization, and thus inhibit DNA binding. In addition, these compounds increase the cellular turnover of estrogen receptor, and this action contributes to its antiestrogen effectiveness. ICI 182,780, fulvestrant, is used to treat metastatic breast cancer that has failed to respond to the usual endocrine therapy.[171] The estrogen receptor combined with fulvestrant is immobilized in the cell and rapidly undergoes degradation.[172, 173] The half-life of the estrogen receptor when occupied with estradiol is about 5 hours; when occupied with a pure antiestrogen, it is less than 1 hour.



Estrogen Agonists/Antagonists (previously called Selective Estrogen Receptor Modulators, SERMS)

This class of synthetic compounds is characterized by a fundamental principle: the conformational shape produced after binding to the receptor results in modified action, influenced by the cellular context of adapter (regulating) proteins and which of these proteins are selected. Tamoxifen rightfully belongs to this family, and its use stimulated the pursuit of a related drug that would not stimulate the endometrium. Agents such as raloxifene and lasofoxifene have antiestrogenic activity in the uterus as well as in the breast, and at the same time exert agonistic effects in certain target tissues.[146] Raloxifene inhibits bone resorption and improves lipids (although there is no effect on HDL-cholesterol). By virtue of variations in conformational changes in the drug-receptor complex and the cellular context of specific tissues, drugs such as these can be developed to produce beneficial effects in certain target systems (such as bone) and to avoid unwanted actions (such as endometrial stimulation). The unique conformational shape raloxifene produces when it binds to the estrogen receptor prevents the involvement of a required coactivator protein at the TAF-2 site. In target tissues that respond principally to TAF-2 gene transcription, these agents will lack estrogenic activity; however in tissues with the appropriate cellular context of proteins, estrogenic gene transcription can occur through the TAF-1 mechanism. In tissues that respond principally to the estrogen receptor-β that lacks TAF-1 activity or when target tissues lack coactivating proteins that interact with TAF-1, these agents will be pure estrogen antagonists. Compounds are being developed that will interact with the progesterone and androgen receptors and produce selected target tissue responses.

Antiprogestins

Both progesterone and the antiprogestins, such as mifepristone (RU-486) and onapristone, form hormone-responsive element-receptor complexes that are similar, but the antiprogestin complex has a slightly different conformational change (in the hormone-binding domain) that prevents full gene activation.[174] RU 486 has some agonistic activity due to its ability to activate certain, but not all, of the transcription activation functions on the progesterone receptor; the final biologic response is modulated by the target tissue context of coactivators and corepressors.[175] New antiprogestins are in development that bind to the progesterone receptor and prevent the subsequent binding of the receptor to gene response elements.

The search for inhibitors of progesterone binding began many years ago, in the late 1960s, but it wasn't until the early 1980s that mifepristone, the first successful antiprogestin, was produced by scientists at Roussel Uclaf, a pharmaceutical company in Paris. Mifepristone is a 19-nortestosterone derivative. The dimethyl (dimethylaminophenyl) side chain at carbon 11 is the principal factor in its antiprogesterone action. Three major characteristics of its action are important: a long half-life, high affinity for the progesterone receptor, and active metabolites.

The affinity of RU 486 for the progesterone receptor is 5 times greater than that of the natural hormone. In the absence of progesterone, it can produce an agonistic (progesterone) effect. It does not bind to the estrogen receptor, but it can act as a weak antiandrogen because of its low-affinity binding to the androgen receptor. Mifepristone also binds to the glucocorticoid receptor, but higher doses are required to produce effects. The binding affinity of mifepristone and its metabolites for the glucocorticoid receptor is very, very high. The reason why it takes such a high dose to produce an effect is because the circulating level of cortisol is so high, 1,000-fold higher than progesterone. This allows titration of clinical effects by adjustments of dose.

Both progesterone and mifepristone induce conformational changes with the progesterone receptor, especially in the hormone-binding domain.[176, 177] Thus, the antiprogestin not only competes with progesterone for the progesterone receptor, but after binding to the hormone-binding domain, the receptor structure is altered in such a way that the transcription activity of the B progesterone receptor is inhibited. In cells where the A progesterone receptor is expressed, antiprogestin binding stimulates A receptor-induced inhibition of transcription activity for all steroid hormone receptors (this would explain the antiestrogen activity of mifepristone).

Mifepristone is most noted for its abortifacient activity and the political controversy surrounding it. However, the combination of its agonistic and antagonistic actions can be exploited for many uses, including contraception, therapy of endometriosis, induction of labor, treatment of Cushing's syndrome, and, potentially, treatment of various cancers. Hopefully, new antiprogestins will be free of political and emotional constraints, and the many potential applications will be pursued.

Androgen Antagonists

The two most commonly used androgen antagonists are cyproterone acetate and spironolactone. Cyproterone and spironolactone bind to the androgen receptor and exert mixed agonism–antagonism. In the presence of significant levels of androgens, the antagonism predominates, and these agents are effective for the treatment of hirsutism. Flutamide is a nonsteroidal pure antiandrogen, effectively blocking androgenic action at target sites by competitive inhibition.

Mechanism of Action for Tropic Hormones

Tropic hormones include the releasing hormones originating in the hypothalamus and a variety of peptides and glycoproteins secreted by the anterior pituitary gland and placenta. The specificity of the tropic hormone depends on the presence of a receptor in the cell membrane of the target tissue. Tropic hormones do not enter the cell to stimulate physiologic events but unite with a receptor on the surface of the cell.

The receptor protein in the cell membrane can act as the active agent and, after binding, operate as an ion channel or function as an enzyme. Alternatively, the receptor protein is coupled to an active agent, an intracellular messenger. The major intracellular messenger molecules are cyclic AMP, inositol 1,4,5-triphosphate (IP3), 1,2-diacylglycerol (1,2-DG), calcium ion, and cyclic guanosine 3′,5′–monophosphate (cyclic GMP).

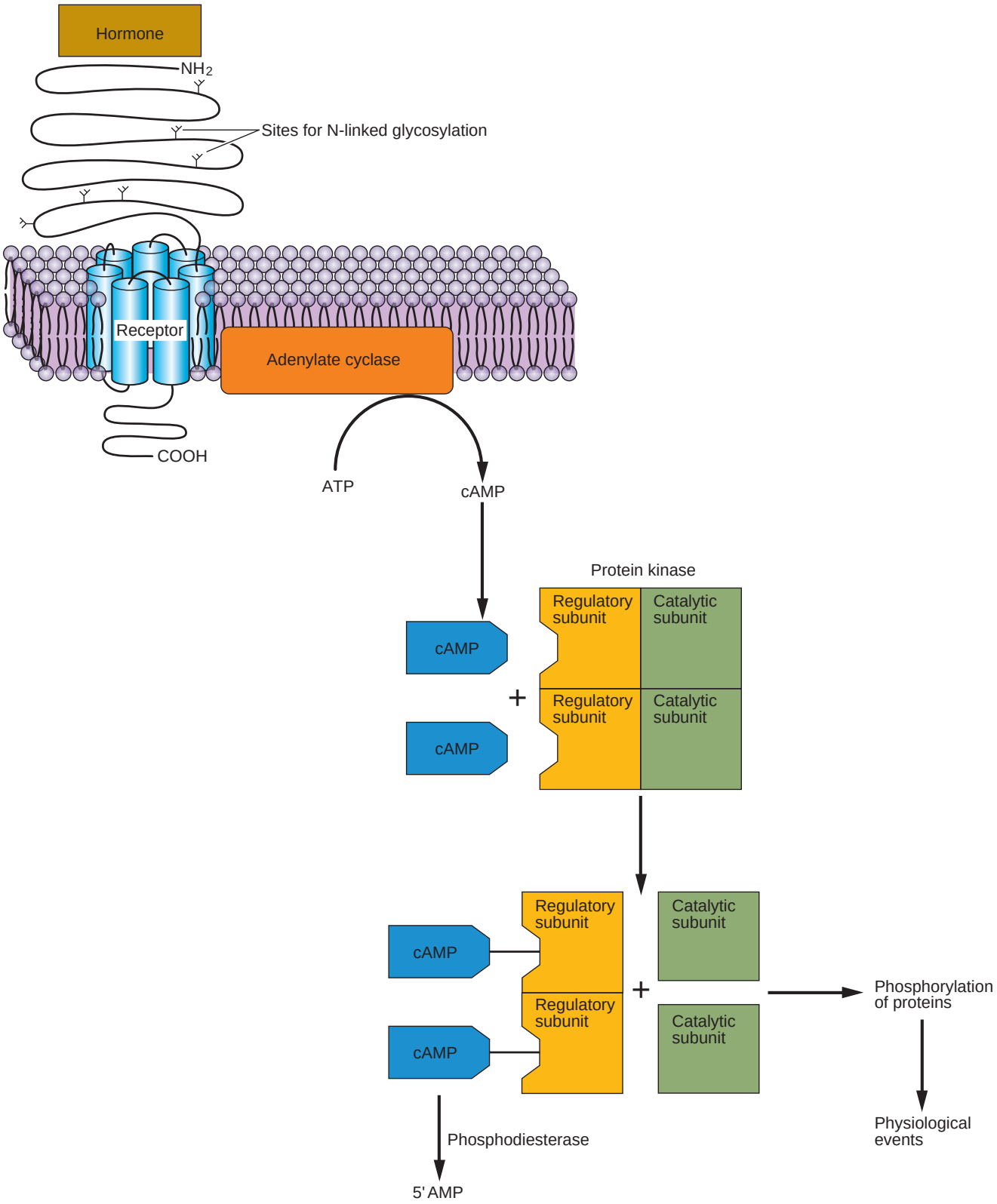
Receptors from this membrane family are also found in the membranes of lysosomes, endoplasmic reticulum, Golgi complex, and in nuclei. The regulation of these intracellular organelle receptors differs from those of the cell surface membranes.

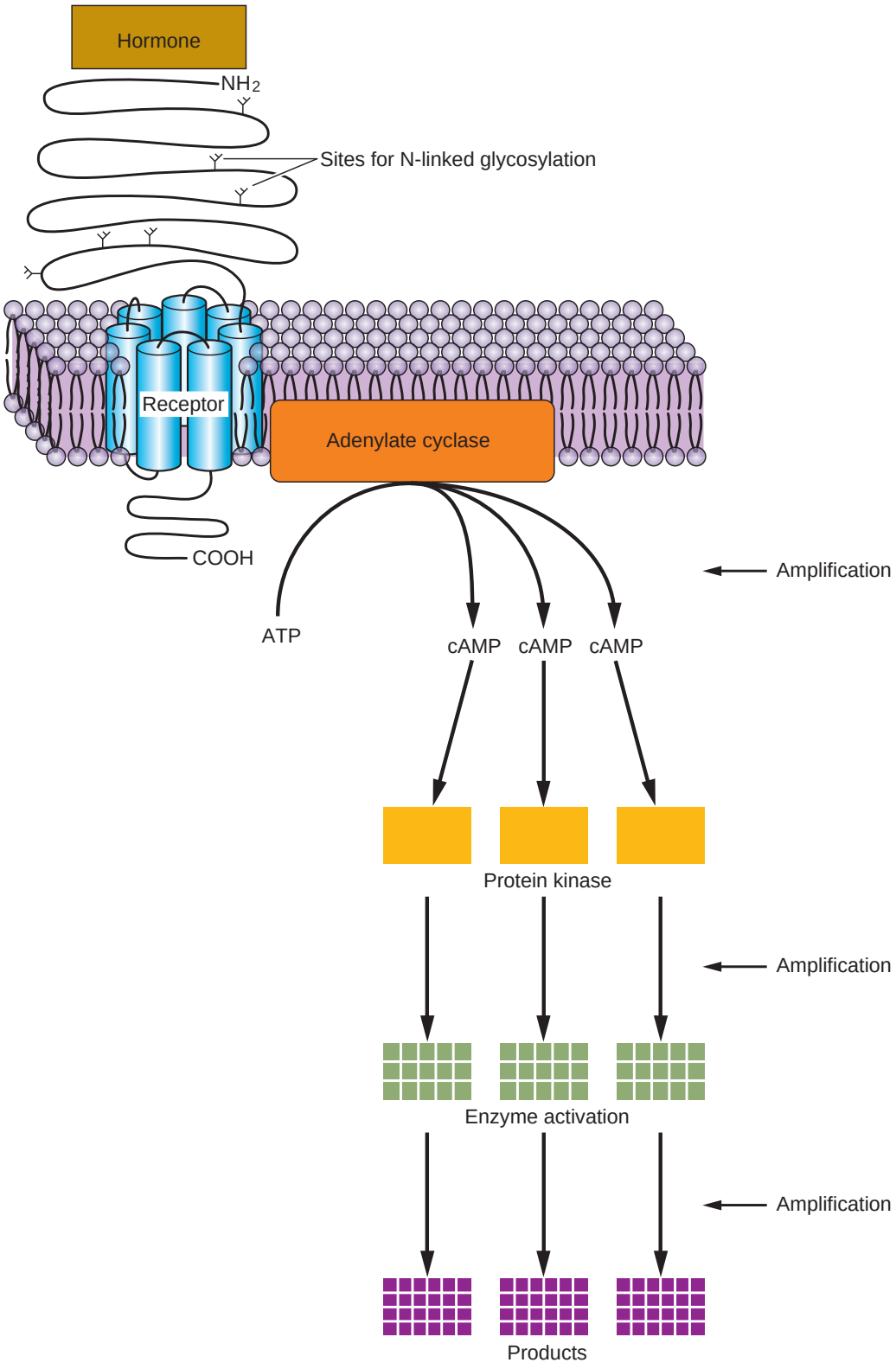
The Cyclic AMP Mechanism

Cyclic AMP is the intracellular messenger for FSH, LH, human chorionic gonadotropin (hCG), thyroid-stimulating hormone (TSH), and ACTH. Union of a tropic hormone with its cell membrane receptor activates the adenylate cyclase enzyme within the membrane wall leading to the conversion of adenosine 5'-triphosphate (ATP) within the cell to cyclic AMP. Specificity of action and/or intensity of stimulation can be altered by changes in the structure or concentration of the receptor at the cell wall binding site. In addition to changes in biologic activity due to target cell alterations, changes in the molecular structure of the tropic hormone can interfere with cellular binding and physiologic activity.

The cell's mechanism for sensing the low concentrations of circulating tropic hormone is to have an extremely large number of receptors but to require only a very small percentage (as little as 1%) to be occupied by the tropic hormone. The cyclic AMP released is specifically bound to a cytoplasm receptor protein, and this cyclic AMP-receptor protein complex activates a protein kinase. The protein kinase is present in an inactive form as a tetramer containing two regulatory subunits and two catalytic subunits. Binding of cyclic AMP to the regulatory units releases the catalytic units, with the regulatory units remaining as a dimer. The catalytic units catalyze the phosphorylation of serine and threonine residues of cellular proteins such as enzymes and mitochondrial, microsomal, and chromatin proteins. The physiologic event follows this cyclic AMP-mediated energy-producing event. Cyclic AMP is then degraded by the enzyme phosphodiesterase into the inactive compound, 5'-AMP.

Most noteworthy, DNA contains responsive elements that bind proteins phosphorylated by the catalytic units, thus leading to activation of gene transcription. The ***cyclic AMP responsive element (CRE)*** functions as an enhancer element upstream from the start of transcription.[178] A large family of transcription factors interact with the CRE, creating an important regulatory unit for gene transcription. Cyclic AMP activates a specific transcription factor, cyclic AMP regulatory element-binding protein (CREB); the binding of CREB to CRE activates many genes. This system can also involve DNA sequences upstream from the CRE site.


Hormone
NH2
Sites for N-linked glycosylation
Receptor
Adenylate cyclase
COOH
ATP
cAMP
Protein kinase
Regulatory subunit
Catalytic subunit
cAMP
cAMP
Regulatory subunit
Catalytic subunit
Regulatory subunit
Catalytic subunit
cAMP
cAMP
Phosphorylation of proteins
Physiological events
Phosphodiesterase
5' AMP


Hormone
NH2
Sites for N-linked glycosylation
Receptor
Adenylate cyclase
COOH
ATP
cAMP
cAMP
cAMP
Amplification
Protein kinase
Amplification
Enzyme activation
Amplification
Products

Because LH can stimulate steroidogenesis without apparent changes in cyclic AMP (at low hormone concentrations), it is possible that an independent pathway exists; i.e., a mechanism independent of cyclic AMP. Mechanisms independent of cyclic AMP could include ion flow, calcium distribution, and changes in phospholipid metabolism.

The cyclic AMP system can be regarded as an example of evolutionary conservation. Rather than developing new regulatory systems, certain critical regulators have been preserved from bacteria to mammals. How is it that a single intracellular mediator can regulate different events? This is accomplished by turning on different biochemical events governed by the different gene expression in individual cells. In addition, the adenylate cyclase enzyme exists in several isoforms, which respond either with stimulation or inhibition to various systems and agents.[179]

The cyclic AMP system provides a method for amplification of the faint hormonal signal swimming in the sea of the bloodstream. Each cyclase molecule produces a lot of cyclic AMP; the protein kinases activate a large number of molecules that in turn lead to an even greater number of products. This is an important part of the sensitivity of the endocrine system, a major reason why only a small percentage of the cell membrane receptors need be occupied in order to generate a response.

Prostaglandins stimulate adenylate cyclase activity and cyclic AMP accumulation. Despite the effect on adenylate cyclase, prostaglandins appear to be synthesized after the action of cyclic AMP. This implies that tropic hormone stimulation of cyclic AMP occurs first; cyclic AMP then activates prostaglandin synthesis and, finally, intracellular prostaglandin moves to the cell wall to facilitate the response to the tropic hormone. In addition to actions mediated by cyclic AMP, prostaglandins can also operate through changes in intracellular concentrations of calcium.

Prostaglandins and cyclic GMP can participate in an intracellular negative feedback mechanism governing the degree of, or direction of, cellular activity (e.g., the extent of steroidogenesis or shutting off of steroidogenesis after a peak of activity is reached). In other words, the level of cellular function can be determined by the interaction among prostaglandins, cyclic AMP, and cyclic GMP.

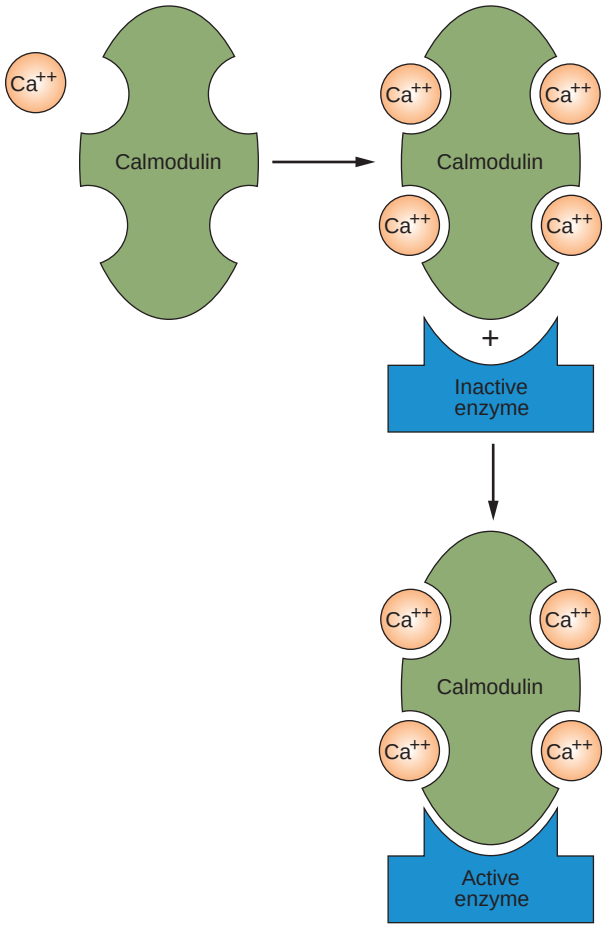
There are differences among the tropic hormones. Oxytocin, insulin, growth hormone, prolactin, and human placental lactogen (hPL) do not utilize the adenylate cyclase mechanism. Receptors for prolactin, growth hormone, and a number of cytokines (including erythropoietin and interleukins) belong to a single transmembrane domain receptor family.[180] Studies of this receptor family indicate that prolactin operates through various signal transduction mechanisms, including ion channels and nuclear kinase activation.

Gonadotropin-releasing hormone (GnRH) is calcium dependent in its mechanism of action and utilizes IP_3 and 1,2-DG as second messengers to stimulate protein kinase activity.[181] These responses require a G protein and are associated with cyclical release of calcium ions from intracellular stores and the opening of cell membrane channels to allow entry of extracellular calcium.

The Calcium Messenger System

The intracellular calcium concentration is a regulator of both cyclic AMP and cyclic GMP levels.[182] Activation of the surface receptor either opens a channel in the cell membrane that lets calcium ions into the cell, or calcium is released from internal stores (the latter is especially the case in muscle). This calcium flux is an important intracellular mediator of response to hormones, functioning itself as a second messenger in the nervous system and in muscle.

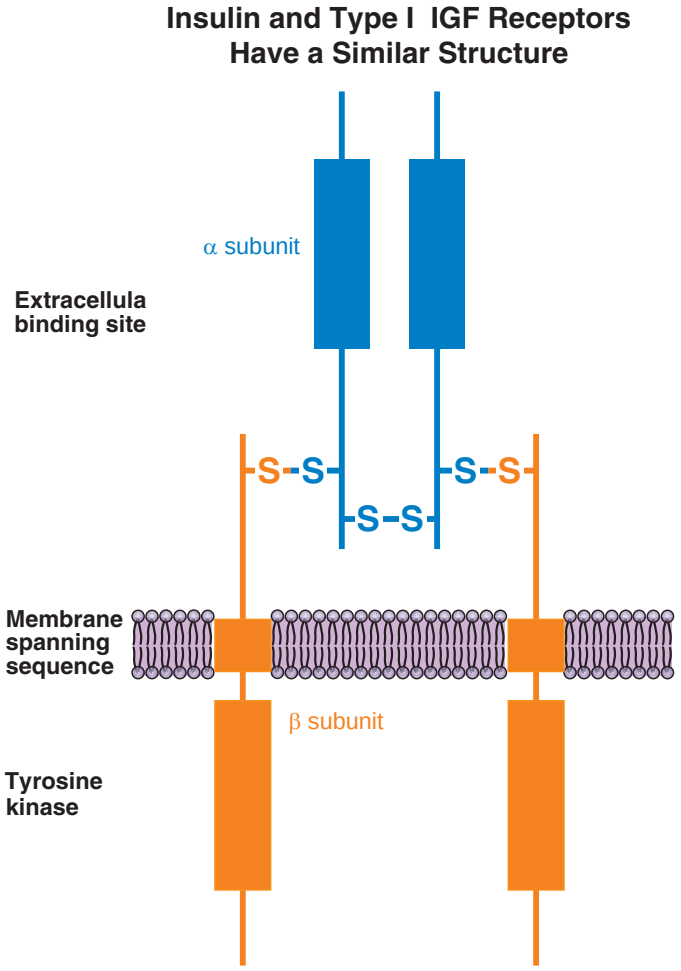
The calcium messenger system is linked to hormone-receptor function by means of a specific enzyme, phospholipase C, that catalyzes the hydrolysis of polyphosphatidylinositols, specific phospholipids in the cell membrane. Activation of this enzyme by hormone binding to its receptor leads to the generation of two intracellular messengers, inositol triphosphate (IP_3) and diacylglycerol (DAG), which initiate the function of the two parts of the calcium system. The first part is a calcium activated protein kinase, responsible for sustained cellular responses, and the second part involves a regulator called calmodulin, responsible for acute responses. These responses are secondary to alterations in enzyme activity, especially protein kinases and in transcription factors.



Calmodulin has been identified in all animal and plant cells that have been examined. Therefore, it is a very ancient protein. It is a single polypeptide chain of 148 amino acid residues whose sequence and structural and functional properties are similar to those of troponin C, the substance that binds calcium during muscle contractions, facilitating the interaction between actin and myosin. The calmodulin molecule has four calcium-binding sites, and binding with calcium gives a helical conformation, which is necessary for biologic activity. A typical animal cell contains more than 10 million molecules of calmodulin, constituting about 1% of the total cell protein. As a calcium regulatory protein, it serves as an intracellular calcium receptor and modifies calcium transport, enzyme activity, the calcium regulation of cyclic nucleotide and glycogen metabolism, and such processes as secretion and cell motility. Thus, calmodulin serves a role analogous to that of troponin C, mediating calcium's actions in noncontractile tissues, and cyclic AMP works together with calcium and calmodulin in the regulation of intracellular metabolic activity.

Kinase Receptors

The cell membrane receptors of insulin, insulin-like growth factor, epidermal growth factor, platelet-derived growth factor, and fibroblast growth factor are tyrosine kinases. All tyrosine kinase receptors have a similar structure: an extracellular domain for ligand binding, a single transmembrane domain, and a cytoplasmic domain. The unique amino acid sequences determine a three-dimensional conformation that provides ligand specificity. The transmembrane domains are not highly conserved (thus differing in makeup). The cytoplasmic domains respond to ligand binding by undergoing conformational changes and autophosphorylation. The structure of the receptors for insulin and insulin-like growth factor is more complicated, with two alpha- and two beta-subunits, forming two transmembrane domains connected extracellularly by disulfide bridges. The receptors for the important autocrine and paracrine factors, activin and inhibin, function as serine-specific protein kinases.



Kinase activation requires distinctive sequences; thus there is considerable homology among the kinase receptors in the cytoplasmic domain. Many of the substrates for these kinases are the enzymes and proteins in other messenger systems; e.g., the calcium messenger system. Thus, the kinase receptors can cross-talk with other receptor regulated systems that involve the G proteins.

Regulation of Tropic Hormones

Modulation of the peptide hormone mechanism is an important biologic system for enhancing or reducing target tissue response. The regulation of tropic hormone action can be divided into four major components.

1. Autocrine and paracrine regulation factors.
2. Heterogeneity of tropic hormones.
3. Up- and down-regulation of receptors.
4. Regulation of adenylate cyclase.

Autocrine and Paracrine Regulation Factors

Growth factors are polypeptides that modulate activity either in the cells in which they are produced or in nearby cells; hence, they are autocrine and paracrine regulators. Regulation factors of this type (yet another biologic family) are produced by local gene expression and protein translation, and they operate by binding to cell membrane receptors. The receptors usually contain an intracellular component with tyrosine kinase activity that is energized by a binding-induced conformational change that induces autophosphorylation. However, some factors work through other second messenger systems, such as cyclic AMP or IP_3. Growth factors are involved in a variety of tissue functions, including mitogenesis, tissue and cellular differentiation, chemotactic actions, and angiogenesis. The growth factors involved in reproductive physiology include activin, inhibin, insulin-like growth factor-I (IGF-I), insulin-like growth factor-II (IGF-II), transforming growth factor-β (TGF-β), fibroblast growth factor (FGF), and epidermal growth factor (EGF).

In addition to the growth factors, various immune factors, especially cytokines, modulate ovarian steroidogenesis. These factors, including interleukin-1, tumor necrosis factor, and interferon, are found in human follicular fluid and, in general, inhibit gonadotropin stimulation of steroidogenesis.

For mitogenesis to occur, cells may require exposure to a sequence of growth factors, with important limitations in duration and concentrations. Growth factors are important for the direction of embryonic and fetal growth and development. In cellular differentiation, growth factors can operate in a cooperative, competitive, or synergistic fashion with other hormones. For example, IGF-I plus FSH, but not IGF-I alone, increases the number of LH receptors, progesterone synthesis, and aromatase activity in granulosa cells.[183]

Activin and inhibin are disulfide-linked dimers composed of peptide subunits (one alpha subunit and two beta subunits) as follows:[184]

The major forms of activin:	Activin-A	$Beta_A$–$Beta_A$
	Activin-AB	$Beta_A$–$Beta_B$
	Activin-B	$Beta_B$–$Beta_B$
The major forms of inhibin:	Inhibin-A	Alpha–$Beta_A$
	Inhibin-B	Alpha-$Beta_B$

Each of the subunits is encoded by separate genes that produce precursor proteins that are cleaved to form the subunits. In addition, the free subunits and related monomeric products can be secreted. Despite the structural similarity between activin and inhibin, they function as antagonists in some systems (e.g., activin stimulates and inhibin inhibits FSH secretion). Activins, inhibins, and TGF-β come from the same gene family, which also includes antimüllerian hormone, and proteins active during insect and frog embryogenesis. The activity of activin is regulated by protein binding, specifically to follistatin. Follistatin is a single-chain glycosylated peptide, structurally unrelated to inhibin and activin, that regulates the activin-inhibin system. Signaling by this family of peptides is accomplished by several receptor isoforms that are transmembrane serine kinases.

TGF-β can either stimulate or inhibit growth and differentiation, depending on the target cell and the presence or absence of other growth factors. In the ovary, TGF-β promotes granulosa cell differentiation by enhancing the actions of FSH (especially in expression of FSH and LH receptors) and antagonizing the down-regulation of FSH receptors. TGF-β and the insulin-like growth factors are required for the maintenance of normal bone mass. EGF is a structural analog of TGF-α and is involved in mitogenesis. In the ovary, EGF, secreted by theca cells, is important for granulosa cell proliferation, an action opposed by TGF-β that is also secreted by the theca cells. The most potent mitogens are the two forms of fibroblast growth factor (FGF). Additional roles for FGF, secreted by the granulosa, include modulation of enzyme activity involved in the physical act of ovulation and angiogenic function during the development of the corpus luteum.

The Insulin-Like Growth Factors

The insulin-like growth factors (called somatomedins in the past) are single-chain polypeptides that resemble insulin in structure and function.[183] These factors are widespread and are involved in growth and differentiation in response to growth hormone, and as local regulators of cell metabolism. IGF-II is more prominent during embryogenesis, whereas IGF-I is more active postnatally. Only the liver produces more IGF-I than the ovary. According to animal studies, both IGF-I and IGF-II are secreted by granulosa cells. IGF-I amplifies the action of gonadotropins and coordinates the functions of theca and granulosa cells. IGF-I receptors on the granulosa are increased by FSH and LH and augmented by estrogen. In the theca, IGF-I increases steroidogenesis. In the granulosa, IGF-I is important for the formation and increase in numbers of FSH and LH receptors, steroidogenesis, the secretion of inhibin, and oocyte maturation. It should be noted that the endogenous insulin-like growth factor in the human ovarian follicle is IGF-II in both the granulosa and the theca cells.[28] Studies indicating activity of IGF-I with human ovarian tissue can be explained by the fact that both IGF-I and IGF-II activities can be mediated by the type I IGF receptor that is structurally similar to the insulin receptor.

Granulosa cells also contain receptors for insulin, and insulin can bind to the IGF-I receptor. The IGF-I receptor is a heterotetramer with two alpha- and two beta-subunits in a structure similar to that of the insulin receptor. Insulin can bind to the alpha-subunit ligand-binding domain and activate the beta-subunit, which is a protein kinase. Thus, insulin can modulate ovarian cellular functions either through its own receptor or through the IGF-I receptor.

The biologic potency and availability of the insulin-like growth factors are further modulated by a collection of IGF-binding proteins that bind circulating insulin-like growth factors and also alter cellular responsiveness. Six insulin-like growth factor binding proteins (IGFBP-1 through IGFBP-6) have been detected in serum and various tissues.[185] IGF-I and IGF-II circulate in the blood in a concentration 1,000 times greater than insulin; however, largely all of the circulating IGFs are bound to IGFBPs. The multiple IGFBPs and their proteases provide a mechanism for tissue-specific activities of IGFs. The various IGFBPs

differ in their actions and individual expression, depending on the specific cell type and tissue. The principal IGFBP that regulates IGF biologic availability can vary according to metabolic changes. There are many possible permutations because the IGFBPs are not simply transport proteins; there are inhibitory and stimulatory IGFBPs that inhibit or potentiate IGF actions. Tissue-specific regulation of IGFBP protease activity can change the bioavailability of IGFs at specific sites. In addition, the IGFBPs have been demonstrated to have direct effects of their own, independent of IGF. Therefore, this is a complex regulatory system that provides both endocrine signals and autocrine and paracrine functions.

Receptors Involved in Steroidogenesis

Steroidogenic factor-1 (SF-1) and DAX-1 (a name that represents: **D**osage-sensitive sex reversal-**A**drenal hypoplasia congenita critical region on the **X** chromosome) are nuclear receptors for which specific ligands have not been identified ("orphan receptors"). SF-1 influences the expression of genes that encode steroidogenic enzymes, and when genetic expression of SF-1 is disrupted in mice, gonads and adrenal glands fail to develop.[186, 187] In addition, SF-1 regulates transcription of the StAR gene.[188] Partial loss of SF-1 causes reduced ovarian activity and infertility.[189] Inactivating mutations in the *DAX1* gene result in X-linked adrenal hypoplasia, which is also associated with hypogonadotropic hypogonadism.[190] DAX-1 is believed to work with SF-1 in regulating development and function of steroid-producing tissues as well as the regulation of gonadotropins.[191] SF-1 regulates genes that encode the gonadotropin subunits, as well as the GnRH receptor.[187] Thus, SF-1 and DAX-1 are involved at all levels: the hypothalamus, the pituitary, and in the steroid-producing organs. These proteins function as transcription factors (as are the traditional nuclear hormone receptors such as the estrogen receptor) in the complex mechanisms being unraveled by molecular biologists.

Heterogeneity of Tropic Hormones

The glycoproteins, such as FSH and LH, are not single proteins but should be viewed as a family of heterogeneous forms of varying immunologic and biologic activity.[192] The various forms (isoforms) arise in various ways, including different DNA promoter actions, alterations in RNA splicing, point mutations, and posttranslational carbohydrate changes.[193] The impact of the variations is to alter structure and metabolic clearance, thus affecting binding and activity. The isoforms have different molecular weights, circulating half-lives, and biologic activities. Throughout the menstrual cycle, the amazing number of at least 20–30 isoforms of both FSH and LH are present in the bloodstream.[194] ***The overall activity of a glycoprotein, therefore, is due to the effects of the mixture of forms that reach and bind to the target tissue***.

The nonglycosylated subunit precursors of glycoprotein hormones are synthesized in the endoplasmic reticulum, followed by glycosylation. The glycosylated subunits combine and then are transported to the Golgi apparatus for further processing of the carbohydrate component. The units combine to form a compact heterodimer. The protein moiety binds to specific target tissue receptors, whereas the carbohydrate moiety plays a critical role in coupling the hormone-receptor complex to adenylate cyclase (perhaps by determining the necessary conformational structure).

The preciseness of the chemical makeup of the tropic hormones is an essential element in determining the ability of the hormone to mate with its receptor. The glycopeptides (FSH, LH, TSH, and hCG) are dimers composed of two glycosylated polypeptide subunits,

the α and β subunits. The α and β subunits are tightly bound in a noncovalent association. The three-dimensional structure and the active conformation of the subunits are maintained by internal disulfide bonds.[195] All of the glycopeptides of the human species (FSH, LH, TSH, and hCG) share a common α chain, an identical structure containing 92 amino acids. The β chains (or the β subunits) differ in both amino acid and carbohydrate content, conferring the specificity inherent in the relationship between hormones and their receptors. Therefore, the specific biologic activity of a glycopeptide hormone is determined by the β subunit; hypogonadism has been reported due to single amino acid substitution in the LH β subunit.[196]

β-hCG is the largest β subunit, containing a larger carbohydrate moiety and 145 amino acid residues, including a unique carboxyl-terminal tail piece of 29 amino acid groups. It is this unique part of the hCG structure that allows the production of highly specific antibodies and the utilization of very specific immunologic assays. The extended sequence in the carboxy-terminal region of β-hCG contains four sites for glycosylation, the reason why hCG is glycosylated to a greater extent than LH, a difference that is responsible for the longer circulating half-life for hCG.

These differences in structure are associated with a different promoter and transcriptional site that is located upstream in the hCG β subunit gene compared with the site in the LH β-subunit gene. The hCG β subunit gene transcription site does not contain a hormone response element, allowing hCG secretion to escape feedback regulation by the sex steroids, in contrast to FSH and LH.

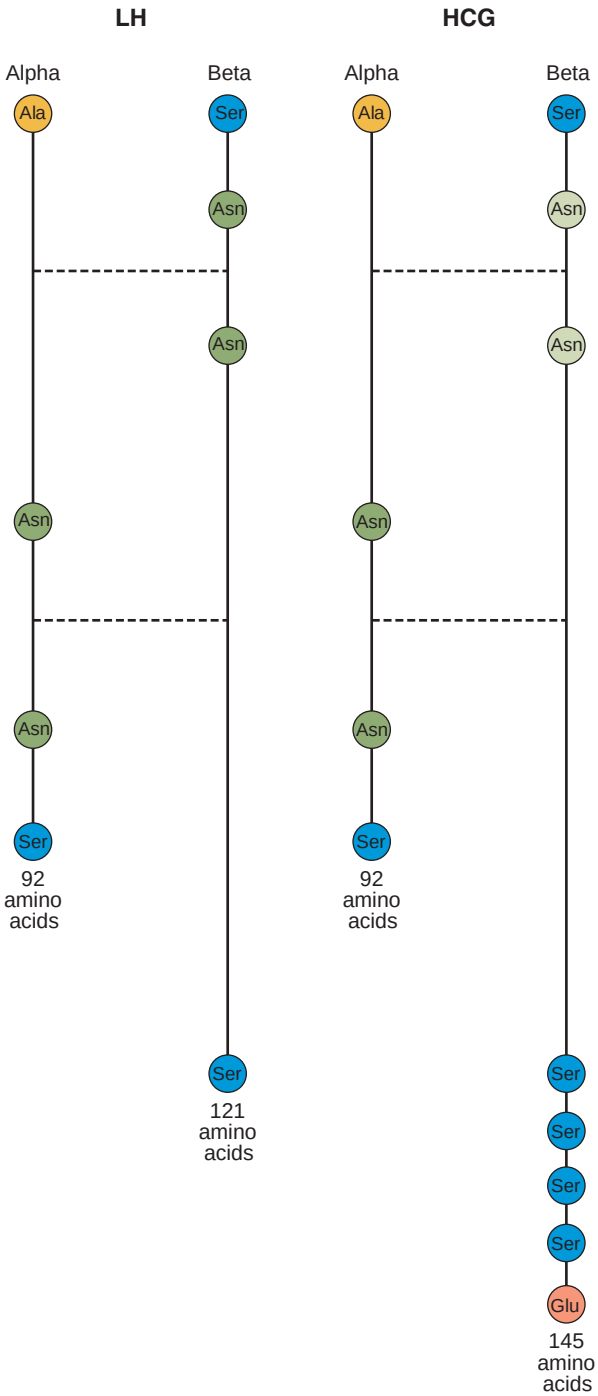
The rate-limiting step in the synthesis of gonadotropins and TSH is the availability of β subunits, because excess α subunits can be found in blood and in tissue. Furthermore, the three dimensional structure of the β subunit, accomplished by folding the subunit by the formation of the disulfide bonds, is an important conformational step that is essential for assembly with the α subunit.[197] This conformational change is not completed until the subunits are fully united to produce the final whole hormone.

The half-life of α-hCG is 6–8 minutes, and that of whole hCG from the placenta about 24 hours. All human tissues appear to make hCG as a whole molecule, but the placenta is different in having the ability to glycosylate the protein, thus reducing its rate of metabolism and giving it biologic activity through a long half-life. The carbohydrate components of the glycoproteins are composed of fructose, galactose, mannose, galactosamine, glucosamine, and sialic acid. Although the other sugars are necessary for hormonal function, sialic acid is the critical determinant of biologic half-life. Removal of sialic acid residues in hCG, FSH, and LH leads to very rapid elimination from the circulation.

FSH consists of the α subunit of 92 amino acids and a β subunit of 118 amino acids. It has four carbohydrate side chains, two on each subunit. The β subunit of LH consists of 121 amino acids. LH has three carbohydrate side chains with a single glycosylation site (with less than half of the sialic acid in FSH). The initial half-life of LH is approximately 20 minutes, compared with the initial half-life of FSH of 3–4 hours.

Genes for tropic hormones contain promoter and enhancer or inhibitor regions located in the 5′-flanking regions upstream from the transcription site. These sites respond to second messengers (cyclic AMP) as well as steroids and other yet unknown regulators. The protein cores of the two glycoprotein subunits are the products of distinct genes.[198] Using recombinant DNA technology, it has been demonstrated that there is a single human gene for the expression of the α subunit. The gene for the α subunit shared by FSH, LH, hCG, and TSH is located on chromosome 6q12.21. A single promoter site subject to multiple signals and hormones regulates transcription of the α gene in both placenta and pituitary. The α subunit gene is expressed in several different cell types, but the β subunit genes are restricted in

cell type. The TSH β gene is expressed only in thyrotropes regulated by thyroid hormone; the FSH β gene is expressed in gonadotropes regulated by GnRH, activin, inhibin, and gonadal steroids; the LH β gene, also expressed in gonadotropes, is regulated by GnRH and unaffected by activin and inhibin.[199]



The α subunit gene requires the activation of distinct regulatory elements in thyrotrope and gonadotrope cells, as well as in the placenta. It is the activation of these cell-specific elements that produces tissue specificity for α gene expression. In gonadotropes, the GnRH signaling pathway for α gene transcription utilizes phosphorylase stimulation of

diacylglycerol (DAG) and inositol triphosphate (IP_3) that leads to a release of intracellular calcium stores. GnRH also stimulates the influx of calcium at the cell membrane. DAG, IP_3, and calcium work together to stimulate protein kinase C activity. Protein kinase regulation of the α promoter is a principal part of the overall mechanism. This pituitary process is influenced by multiple factors, including growth factors and gonadal steroids. In the placenta, the mechanism also utilizes specific regulatory elements, but the primary signal is mediated by the cyclic AMP-protein kinase A pathway.

The gene for the FSH β subunit is on chromosome 11p13, and in the pituitary, it is markedly influenced by activin.[200, 201] Although FSH and LH both require GnRH stimulation, the FSH β gene is unique in that response to GnRH is dependent on activin.[202] With increasing GnRH stimulation, the role of activin is increasingly repressed by its binding protein, follistatin, the secretion of which is also stimulated by GnRH and activin. Activin is further antagonized by inhibin, the first of these factors recognized to suppress FSH secretion.[203]

The genes that encode for the β subunits of LH, hCG, and TSH are located in a cluster on chromosome 19q13.32. There are six genes for the β subunit of hCG, and only one for β-LH.[204] Transcription for the six hCG genes, each with different promoter activity, varies, and it is not certain why hCG requires multigenic expression (perhaps this is necessary to reach the extremely high level of production in early pregnancy). It is thought that β-hCG evolved relatively recently from β-LH, and the unique amino acid terminal extension of β-hCG arose by a read-through mutation of the translation stop codon in the β-LH gene; the DNA sequences of the β-hCG genes and the β-LH gene are 96% identical.[204] Only primates and the equine species (horse, donkey, and zebra) have been demonstrated to have genes for the β subunit of chorionic gonadotropin. In contrast to human chorionic gonadotropin, equine chorionic gonadotropin exerts both LH and FSH activities in many mammalian species because it contains peptide sequences in its β subunit that are homologous to those in the pituitary gonadotropins of other species. The equine β-chorionic gonadotropin gene is identical to the equine β-LH gene, and although the primate β-hCG gene evolved from the same ancestral β-LH gene, the horse chorionic gonadotropin gene evolved in a different way about 50 million years ago.[205] The β-LH gene is not expressed in the placenta.

A specific immunologic LH variant is relatively common. This variant is due to two point mutations in the LH β subunit gene and is more common in people of northern European descent, reaching a carrier frequency of 41.9% in Lapps of northern Finland.[206] The clinical significance of this mutation is not known; however, routine immunoassays can provide falsely low readings because this variant is not detected. Inherited disorders because of disruptions in the coding sequences of both LH and FSH are in fact quite rare.[207]

The placenta-specific expression of β-hCG is due to several differences in DNA sequences between the β-hCG and β-LH genes.[199] The cyclic AMP-mediated enhancement of the β-hCG promoter is influenced by several regulatory proteins. The study of the β subunit genes has been hampered by difficulties in maintaining glycoprotein-producing cell lines. The availability of choriocarcinoma cell lines, however, has allowed greater investigation of the β-hCG genes.

Although the β subunit specifies the biologic activity of an individual glycoprotein, the combination of the α and β subunits is necessary for full hormonal expression. Furthermore, the α subunit also plays an important role in accomplishing normal receptor binding and activation.[208, 209] Neither subunit alone can effectively bind to the receptor with high affinity or exert biologic effect. In other words, binding and activation occur only when the hormone is in the combined α-β form. In addition, the α subunit influences the overall bioactivity of the glycoprotein hormones.[210] Thus structural alterations in either the α subunit or the β subunit could alter target tissue responses.

Variations in Carbohydrate

The glycopeptide hormones can be found in the pituitary existing in a variety of forms, differing in their carbohydrate (oligosaccharides) makeup. The isoform mixture of gonadotropins is influenced both quantitatively and qualitatively by GnRH and the feedback of the steroid hormones, producing posttranslational carbohydrate modifications.[211, 212] This heterogeneity in structure (which is also associated with heterogeneity in charge) represents a mechanism under endocrine control that modulates half-lives and bioactivity.

Certain clinical conditions can be associated with alterations in the usual chemical structure of the glycopeptides, resulting in interference with the ability to bind to receptors and stimulate biologic activity. In addition to deglycosylation and the formation of antihormones, gonadotropins can be produced with an increased carbohydrate content. A low-estrogen environment in the pituitary gland, for example, favors the production of so-called big gonadotropins, gonadotropins with an increased carbohydrate component and, as a result, decreased biologic activity.[213] Immunoassay in these situations may not reveal the biologic situation; an immunoassay sees only a certain set of molecules but not all. Therefore, immunologic results do not always indicate the biologic situation.

Bioactive levels of FSH and LH are very low in women receiving oral contraceptives and during the luteal and late follicular phases. The highest values are during the midcycle surge and in postmenopausal women (including women with premature ovarian failure).[214] The levels of bioactive FSH parallel those of immunoactive FSH with a constant ratio throughout the cycle. The greater bioactivity of FSH at midcycle is associated with less sialyated, shorter-lived isoforms. These changes are effects of both GnRH and estrogen.

The carbohydrate component, therefore, affects target tissue response in two ways: (1) metabolic clearance and half-life, and (2) biologic activity. The latter action focuses on two functions for the hormone-receptor complex: binding and activation. One structural domain is important for binding and another for triggering the biologic response. Carbohydrate residues, especially the sialic acid residues, are less important in binding. Indeed, experimental data indicate that the carbohydrate chains have no role in the binding of gonadotropins to their receptors.[215] Nevertheless, removal of the carbohydrate moiety of either subunit diminishes gonadotropic activity. Therefore, the carbohydrate component affects the biologic activity of the hormone-receptor complex after binding. Specific studies indicate that the carbohydrate component plays a critical role in activation (coupling) of the adenylate cyclase system.[216]

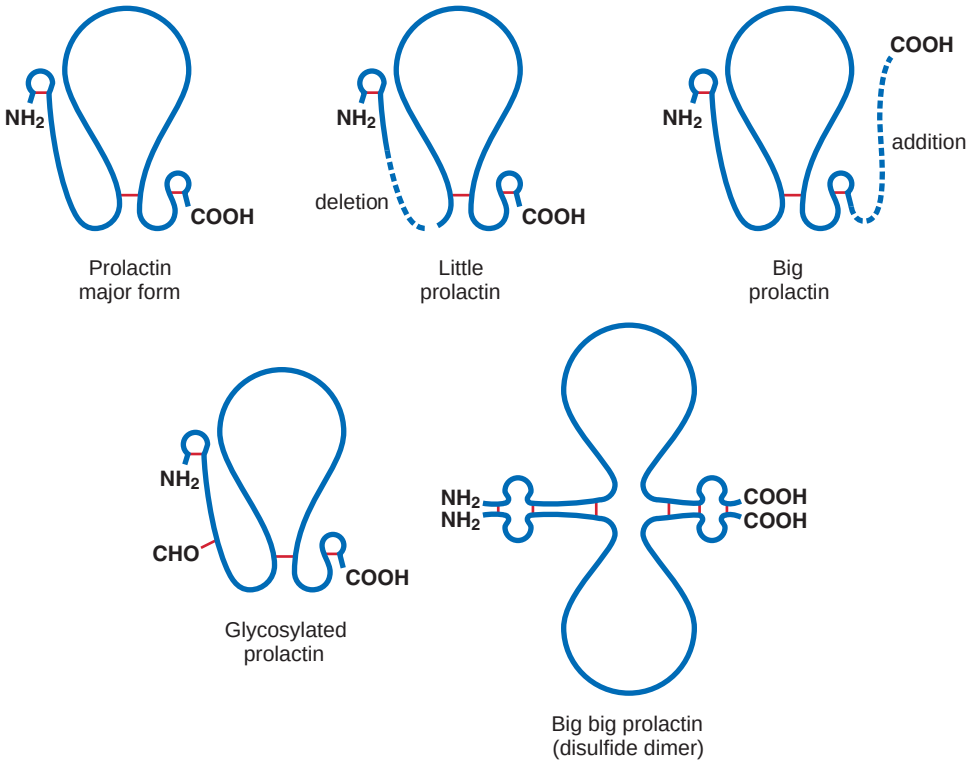
The circulating half-life of a gonadotropin is mainly proportional to the amount of sialic acid present.[217] The higher content of sialic acid in FSH compared with LH accounts for the more rapid clearance of LH from the circulation (the FSH half-life is several hours; the LH half-life is about 50 minutes). hCG is highly sialylated, and accordingly, has a half-life of 5–6 hours. However, clearance of gonadotropins as measured by half-lives is not explained totally by carbohydrate differences. Differences in amino acid sequences also contribute, and most importantly, the stability of the complete hormone (resisting dissociation into the rapidly cleared subunits) is a major factor.

Heterogeneity of Prolactin

In most mammalian species, prolactin is a single-chain polypeptide of 199 amino acids, 40% similar in structure to growth hormone and placental lactogen.[218] All three hormones are believed to have originated in a common ancestral protein about 400 million years ago. Many hormones, growth factors, and neurotransmitters affect the prolactin gene.

Simultaneous measurements of prolactin by both bioassay and immunoassay reveal discrepancies. At first, differences in prolactin were observed based on size, leading to the use of terms such as little, big, and the wonderfully sophisticated term, big big prolactin. Further chemical studies have revealed structural modifications that include glycosylation, phosphorylation, and variations in binding and charge. This heterogeneity is the result of many influences at many levels: transcription, translation, and peripheral metabolism.[219, 220]

Prolactin is encoded by a single gene on chromosome 6, producing a molecule that in its major form is maintained in three loops by disulfide bonds.[218] Most, if not all, variants of prolactin are the result of posttranslational modifications. Little prolactin probably represents a splicing variant resulting from the proteolytic deletion of amino acids. Big prolactin has little biologic activity and does not cross-react with antibodies to the major form of prolactin. The so-called big big variants of prolactin are due to separate molecules of prolactin binding to each other, either noncovalently or by interchain disulfide bonding. Some of the apparently larger forms of prolactin are prolactin molecules complexed to binding proteins. High levels of relatively inactive prolactin in the absence of a tumor can be due to the creation of macromolecules of prolactin by antiprolactin autoantibodies.[221, 222] Overall, big prolactins account for somewhere between 10% and 25% of the hyperprolactinemia reported by commercial assays.[223]



Other variations exist. Enzymatic cleavage of the prolactin molecule yields fragments that may be capable of biologic activity. Prolactin that has been glycosylated continues to exert activity; differences in the carbohydrate moities can produce differences in biologic activity and immunoreactivity. However, the nonglycosylated form of prolactin is the predominant form of prolactin secreted into the circulation.[224] Modification of prolactin also includes phosphorylation, deamidation, and sulfation.

The prolactin receptor is encoded by a gene on chromosome 5 that is near the gene for the growth hormone receptor. However, there is evidence for more than one receptor, depending on the site of action (e.g., decidua and placenta).[225] The prolactin receptor belongs to the receptor family that includes many cytokines and some growth factors, supporting

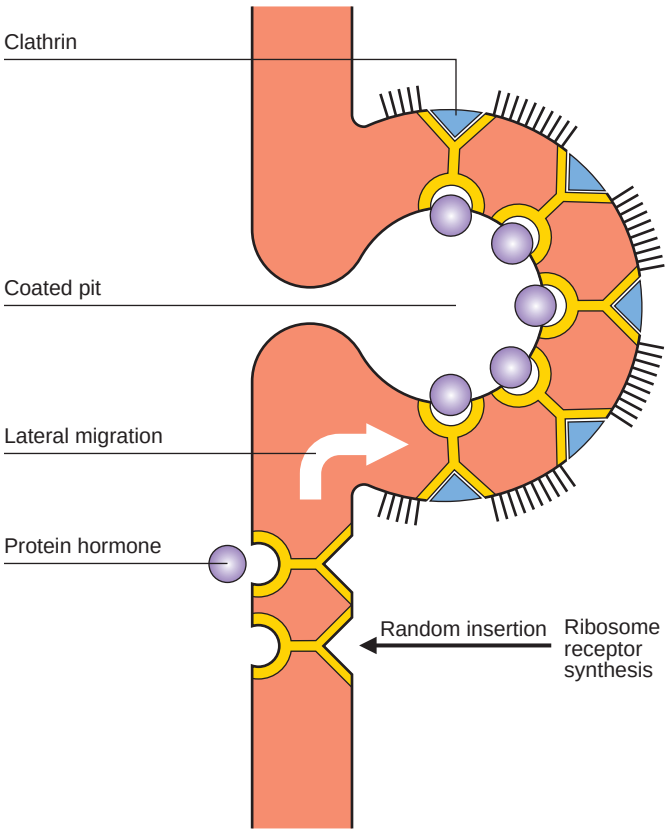
a dual role for prolactin as a classic hormone and as a cytokine. The prolactin signal is mediated through a cytoplasmic tyrosine kinase pathway.

At any one point of time, the bioactivity (e.g., galactorrhea) and the immunoreactivity (circulating level by immunoassay) of prolactin represent the cumulative effect of the family of structural variants. Remember, immunoassays do not always reflect the biologic situation (e.g., a normal prolactin level in a women with galactorrhea).

Up- and Down-Regulation of Receptors

Positive or negative modulation of receptors by homologous hormones is known as up- and down-regulation. Little is known regarding the mechanism of up-regulation; however, hormones such as prolactin and GnRH can increase the cell membrane concentrations of their own receptors.

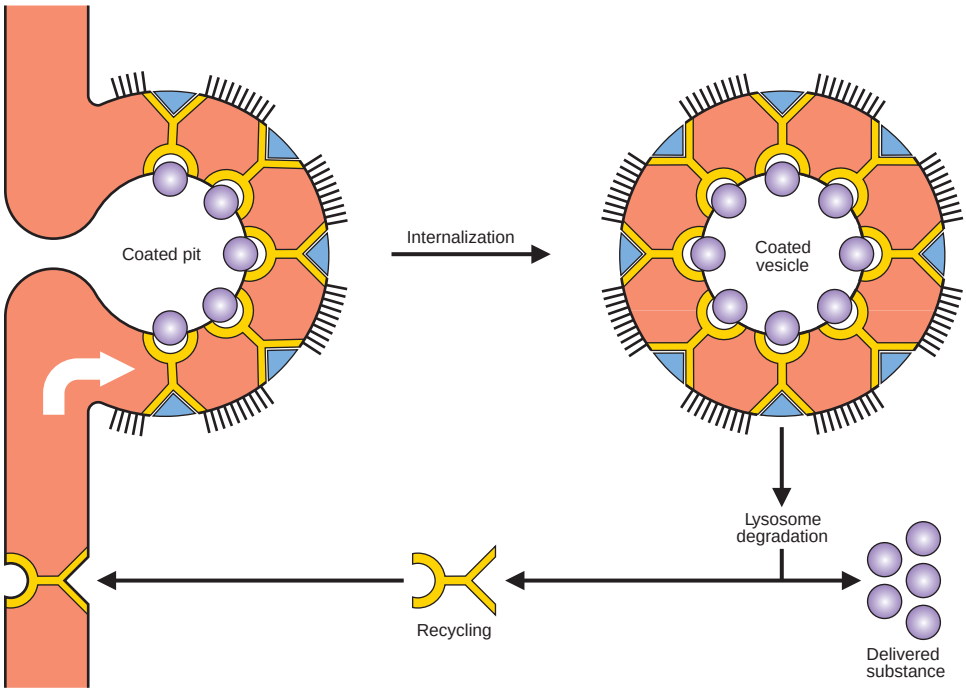
Theoretically, deactivation of the hormone-receptor complex could be accomplished by dissociation of the complex or loss of receptors from the cell, either by shedding (externally) or by internalization of the receptors into the cell. It is the process of ***internalization*** that is the major biologic mechanism by which polypeptide hormones down-regulate their own receptors and thus limit hormonal activity. As a general rule, an excess concentration of a tropic hormone, such as LH or GnRH, will stimulate the process of internalization, leading to a loss of receptors in the cell membrane and a decrease in biologic response. We now understand that the principal reason for the episodic (pulsatile) secretion of hormones is to avoid down-regulation and to maintain, if not up-regulate, its receptors. The pulse frequency is a key factor, therefore, in regulating receptor number.



It is believed that receptors are randomly inserted into the cell membrane after intracellular synthesis. The receptor may be viewed as having three important segments: an external binding site that is specific for a polypeptide hormone, the transmembrane region, and an internal site that plays a role in the process of internalization. When the receptor is bound to a polypeptide hormone and when high concentrations of the hormone are present in the circulation, the hormone-receptor complex moves through the cell membrane in a process called lateral migration. Lateral migration carries the complex to a specialized region of the cell membrane, ***the coated pit.*** Each cell in target tissues contains from 500 to 1,500 coated pits. Lateral migration, thus, concentrates hormone-receptor complexes in the coated pit (***clustering***), allowing increased internalization of the complex via the special mechanism of ***receptor-mediated endocytosis***.[226] The time course for this process (minutes rather than seconds) is too slow to explain the immediate hormone-induced responses, but other cellular events may be mediated by this mechanism that circumvents the intracellular messenger, cyclic AMP.

The coated pit is a lipid vesicle hanging on a basket of specific proteins, called ***clathrins*** (from the Latin "clathra" meaning "lattice"). The unit is a network of hexagons and pentagons, thus looking like a soccer ball. The internal margin of the pit has a brush border, hence the name coated pit. The clathrin protein network serves to localize the hormone-receptor complexes by binding to the internal binding site on the receptor.

When fully occupied, the coated pit invaginates, pinches off, and enters the cell as a coated vesicle also called a receptosome. The coated vesicle is delivered to the lysosomes in which the structure then undergoes degradation, releasing the substance (e.g., a polypeptide hormone) and the receptor. The receptor may be recycled; i.e., it may be reinserted into the cell membrane and used again. On the other hand, the receptor and the hormone may be metabolized, thus decreasing that hormone's biologic activity. The internalized hormones may also mediate biologic response by influencing cellular organelles such as the Golgi apparatus, the endoplasmic reticulum, and even the nucleus. For example, nuclear membranes from human ovaries bind hCG and LH, and there follows an enzyme response that is involved in the transfer of mRNA from nucleus to the cytoplasm.[227]



A similar process, called ***potocytosis***, utilizes cholesterol-rich membrane invaginations called ***caveolae*** (far fewer in number and smaller in structure than the clathrin coated pits) for the internalization of small molecules and ions.[228] This is another method of intracellular signaling in response to hormones, and many proteins involved in cell signaling have been detected in caveolae; e.g., G proteins, kinases, and growth factor receptors. Caveolin is the major protein structural component of caveolae. Nitric oxide, the important mediator of vascular events, resides in caveolae and is regulated by tyrosine phosphorylation and interaction with caveolin.[229, 230] Caveolae also facilitate endocytosis and exocytosis of substances, by the recycling of caveolin between the cell surface and the Golgi network.[231]

Besides down-regulation of polypeptide hormone receptors, the process of internalization can be utilized for other cellular metabolic events, including the transfer into the cell of vital substances such as iron or vitamins. Indeed, this is the basic mechanism for transporting large molecules across the cell wall into the interior.

Cell membrane receptors can be randomly distributed in the cell membrane and transmit information to modify cell behavior.[232] For these receptors, internalization is a method for down-regulation by degradation in lysosomes. Because of this degradation, recycling is usually not a feature of this class of receptors. Hormones that utilize this category of receptors include FSH, LH, hCG, GnRH, TSH, TRH, and insulin. For these hormones, the coated pit can be viewed as a trap to immobilize hormone-receptor complexes. The fate of the hormone, however, can vary from tissue to tissue. In some target tissues, hCG is internalized and the hCG-receptor complex is transferred intact from the coated vesicle into the lysosomes for dissociation and degradation. In other tissues, especially the placenta, it is thought that the hCG-receptor complex is recycled back to the cell surface as a means of transporting hCG across the placenta into both maternal and fetal circulations.[233]

Cell membrane receptors, located in the coated pits, when bound to ligands lead to internalization, thus providing the cell with required factors, the removal of noxious agents from the biologic fluid bathing the cell, or the transfer of substances through the cell (transendocytosis). These receptors are spared from degradation and can be recycled. Examples of this category include low-density lipoproteins (LDL), which supply cholesterol to steroid-producing cells; cobalamin and transferrin, which supply vitamin B12 and iron, respectively; and the transfer of immunoglobulins across the placenta to provide fetal immunity.

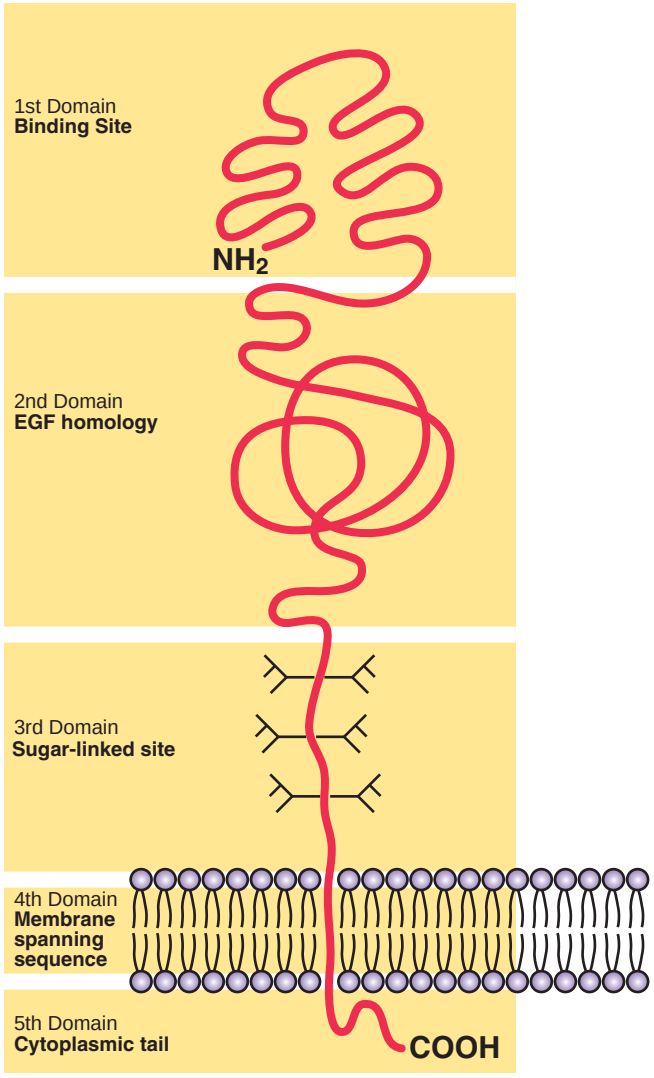
A closer look at LDL and its receptor is informative because it is the prototype for this system. The low-density lipoprotein particle is a sphere. It contains in its center about 1,500 molecules of cholesterol that are attached as esters to fatty acids. This core is contained by a bilayer lipid membrane. Protein-binding proteins (the apoproteins) project on the surface of this membrane, and it is these proteins that the receptor must recognize.

Remember, this is an important story, because all cells that produce steroids must use cholesterol as the basic building block. Such cells cannot synthesize enough cholesterol, and therefore must bring cholesterol into the cell from the bloodstream. LDL is the principal messenger delivering the cholesterol. Experimental evidence, however, indicates that HDL-cholesterol as well as LDL can provide cholesterol to steroid-producing cells.[234] Indeed, human ovarian granulosa cells use HDL-cholesterol in a system that differs from the LDL-cholesterol pathway: the lipoproteins are not internalized, but rather, the cholesteryl esters are extracted from the lipoproteins at the cell surface and then transferred into the cell.[235]

Different cell surface receptors and proteins contain similar structural parts.[236] For example, the receptor for LDL contains a region that is homologous to the precursor of epidermal growth factor and another region that is homologous to a component of complement. The LDL receptor is a “mosaic protein.” There are regions of proteins derived from the exons of different gene families. This is an example of a protein that evolved as a new combination of preexisting functional units of other proteins.

The LDL receptor is synthesized as a precursor of 860 amino acids. The precursor includes 21 amino acids that constitute a hydrophobic signal sequence that is cleaved prior to its insertion into the cell surface. This signal sequence presumably directs the protein where to go in the cell. This leaves an 839 amino acid protein that has five recognizable domains.

1. NH2-terminal of 292 amino acids, composed of a sequence of 40 amino acids repeated with some variation seven times. This domain is the binding site for LDL and is located on the external surface of the cell membrane.
2. Approximately 400 amino acids 35% homologous to epidermal growth factor precursor.
3. The sugar-linked site.
4. 22 Hydrophobic amino acids that cross the cell membrane. Deletion of the transmembrane signal sequence (found in a naturally occurring mutation) results in an LDL receptor that is secreted from the cell instead of being inserted into the membrane.
5. Cytoplasmic tail of 50 amino acids that is located internally and serves to cluster LDL receptors in coated pits.



When the coated pit is fully occupied with LDL, a coated vesicle is delivered into the cell in the process called endocytosis. The vesicle moves to the Golgi system and then is routed by an unknown mechanism (although a similar coated pit system in the Golgi appears to be involved) to the lysosomes in which the structure undergoes degradation, releasing cholesterol esters and the receptor. The receptor can be recycled or degraded. The intracellular level of free cholesterol influences the following important activities: the rate-limiting enzyme for cholesterol synthesis, the reesterification of excess cholesterol for storage as lipid droplets, and the synthesis of LDL receptors. The cholesterol derived from the LDL transport process can have any one of the following fates: utilization in the mitochondria for steroidogenesis, reesterification for storage, use in membrane structures, or excretion.[237] Excretion (release of free cholesterol into the circulation by means of the HDL mechanism) involves the cell surface caveolae.[228, 231] Thus, entry is via coated pits (endocytosis) and efflux from endoplasmic reticulum to the cell membrane is via caveolae (exocytosis).

Synthesis and insertion of new LDL receptors are a function of LH in the gonads and ACTH in the adrenal. This process is relatively fast. It has been calculated that the coated pit system turns over an amount of cell surface equivalent to the total amount of plasma membrane every 30–90 minutes.[237] The LDL receptor makes one round trip every 10 minutes during its 20-hour lifespan for a total of several hundred trips.[238] Genetic defects in receptors for LDL lead to a failure in internalization and hyperlipidemia.

Summary of Down-Regulation

Down-regulation is a decrease in response in the presence of continuous stimulation. It can involve any of the following mechanisms:

1. Densensitization by autophosphorylation of the cytoplasmic segment of the receptor.
2. Loss of receptors by internalization, a relatively slow mechanism.
3. Uncoupling of the regulatory and catalytic subunits of the adenylate cyclase enzyme.
4. Alterations in key intracellular regulatory proteins.

Regulation of Adenylate Cyclase

The biologic activity of polypeptide or glycoprotein hormones (such as FSH or LH) can be altered by autocrine and paracrine regulators, the heterogeneity of the molecules, up- and down-regulation of the receptors, and, finally, by modulation of the activity of the enzyme, adenylate cyclase.

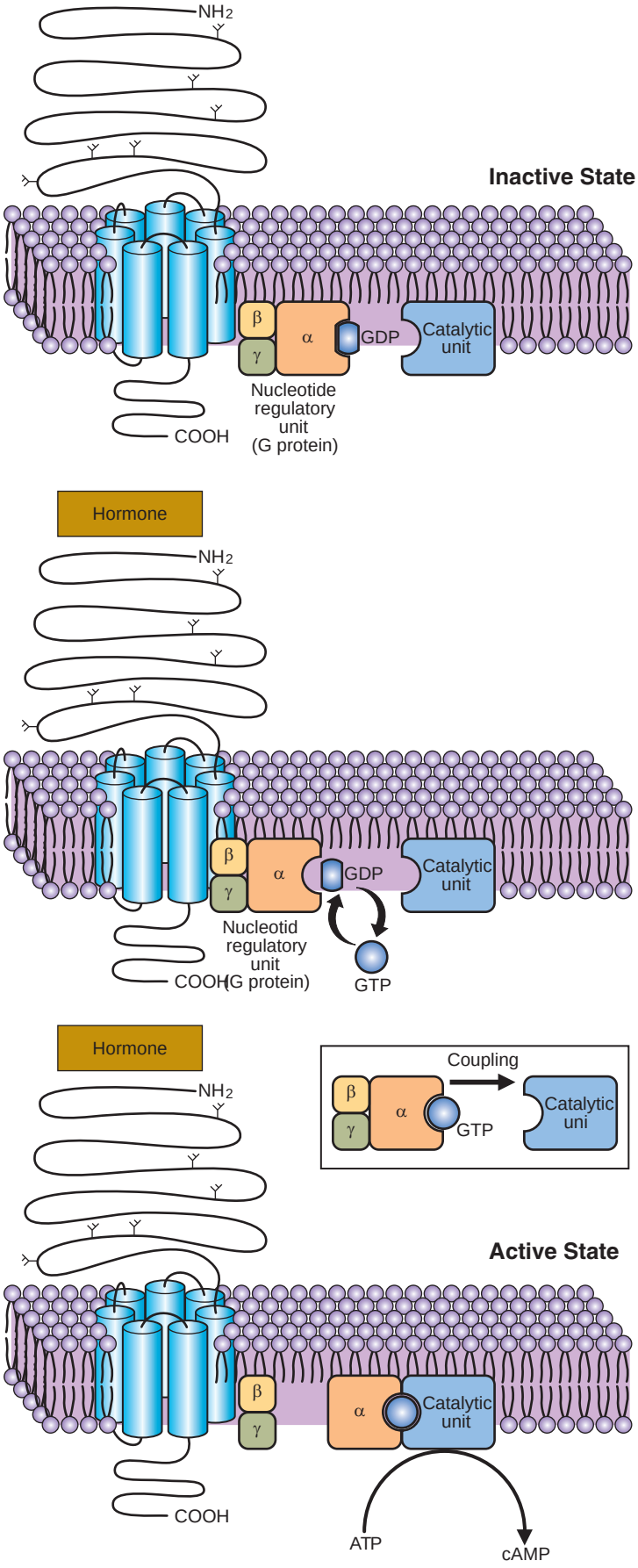
The G Protein System

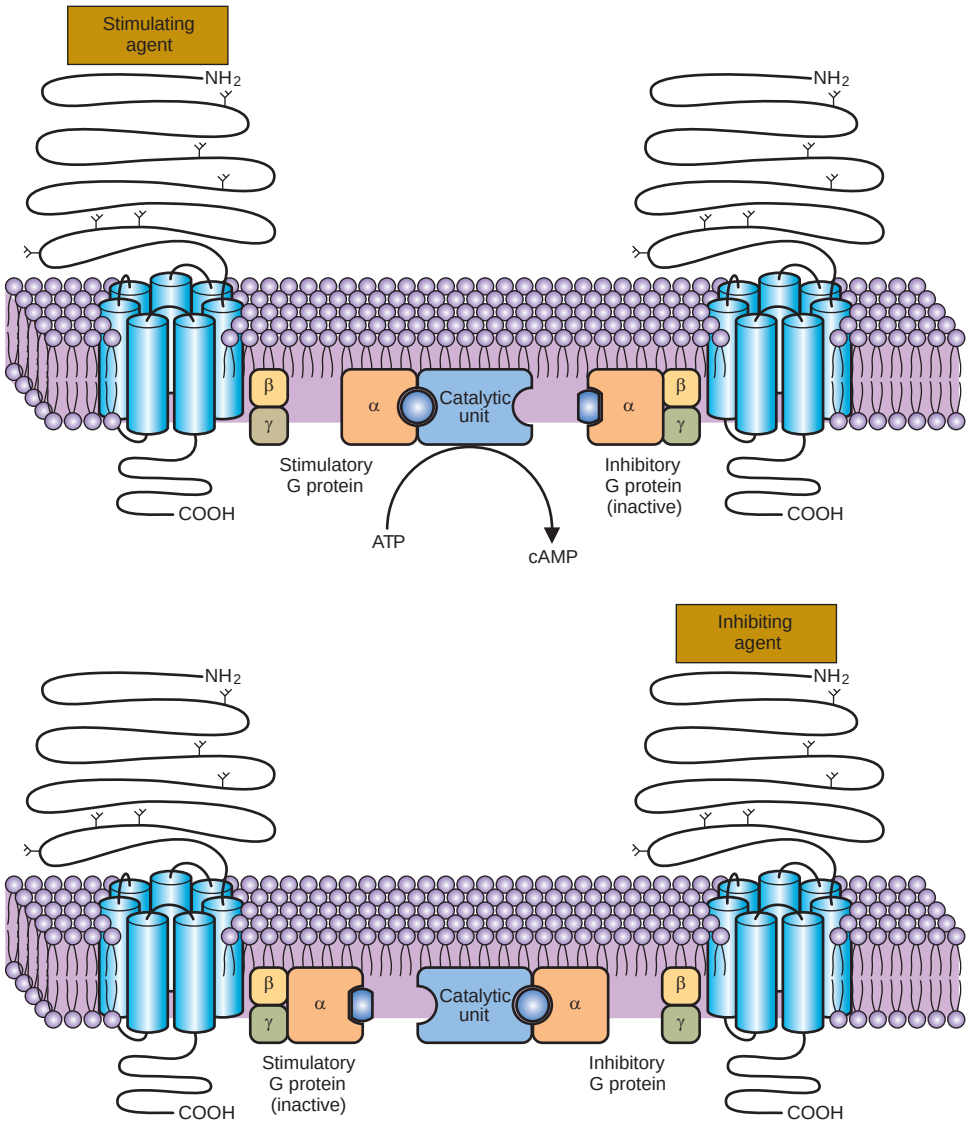
The 1994 Nobel Prize in Medicine and Physiology was awarded to Alfred G. Gilman and Martin Rodbell for the discovery and description of G proteins. Adenylate cyclase is composed of three protein units: the receptor, a guanyl nucleotide regulatory unit, and a catalytic unit.[239] The regulatory unit is a coupling protein, regulated by guanine nucleotides (specifically guanosine 5′–triphosphate, GTP), and therefore it is called GTP binding protein or G protein for short.[240, 241] The catalytic unit is the enzyme itself that converts ATP to cyclic AMP. The receptor and the nucleotide regulatory unit are structurally linked, but inactive until the hormone binds to the receptor. Upon binding, the complex of hormone, receptor, and nucleotide regulatory unit is activated leading to an uptake of GTP by the regulatory unit. The activation and uptake of GTP result in an active enzyme that can convert ATP to cyclic AMP. This result can be viewed as the outcome of the regulatory unit ***coupling*** with the catalytic unit, forming an intact complete enzyme. Enzyme activity is then terminated by hydrolysis of the GTP to guanosine 5′-diphosphate (GDP) returning the enzyme to its inactive state. Quick action and acute control of adenylate cyclase are assured because the G protein is a GTPase that self-activates upon binding of GTP.

The G protein has been purified. From the amino acid sequence, complementary DNA clones have been produced. These studies have indicated that a family of G proteins exists that couples receptors to active ligands, playing roles in signal transduction, intracellular transport, and exocytosis. The ability of the hormone-receptor complex to work through a common messenger (cyclic AMP) and produce contrasting actions (stimulation and inhibition) is thought to be due to the presence of both stimulatory nucleotide regulatory G proteins and inhibitory nucleotide regulatory G proteins.[242, 243] However, the G protein system is not limited to the cyclic AMP signal, but can activate other messenger-generating enzymes, as well as ion channels.

The G proteins are composed of α, β, and γ subunits, each the product of many distinct genes.[244] The β and γ subunits are not all alike, and they exhibit selectivity for specific ligands. Indeed, there are several hundred G protein receptors, with a common structure but sufficiently dissimilar to be activated by different ligands. Each G protein has a unique α subunit, and there are 16 mammalian α subunit genes. Based on amino acid similarities, they are grouped into four subfamilies: $G_s\alpha$, $G_q\alpha$, $G_i\alpha$, G_{i2}. G_s and G_q proteins mediate stimulatory events such as hormone secretion, whereas G_i proteins exert inhibition. The role of the G_{i2} group is not yet certain. These multiple subunits allow great variability in function to be expressed by many different combinations that produce conformational changes linked to message transmission.

In the inactive state GDP is bound to the α subunit. Hormone-receptor interaction and binding change the α subunit conformation. GTP replaces GDP on the α subunit, freeing the β and γ subunits, which allows the GTP α subunit to bind to the catalytic unit of adenylate cyclase, forming the active enzyme. The GTP α subunit can also activate other messengers, such as ion channels. Intrinsic GTPase activity quickly hydrolyzes the GTP-α to GDP-α, which leads to reassociation with the β and γ subunits, reforming the G protein complex for further activation. The functional specificity is due to the α subunit, which differs for each G protein, and therefore there are many different α subunits encoded by different genes.


NH2
Inactive State
β
γ
α
GDP
Catalytic unit
Nucleotide regulatory unit (G protein)
COOH
Hormone
NH2
β
γ
α
GDP
Catalytic unit
Nucleotid regulatory unit (G protein)
COOH
GTP
Hormone
Coupling
β
γ
α
GTP
Catalytic uni
NH2
Active State
β
γ
α
Catalytic unit
COOH
ATP
cAMP



The G Protein Receptors

The more than 200 receptors linked to G proteins are derived from a supergene family, the 800 genes presumably originated from a common ancestral gene.[245] The gonadotropin receptor contains a transmembrane region that has the structural features of a receptor that couples with G protein and a large extracellular domain.[246] Receptors that utilize the G proteins are inserted in membranes and consist of a long polypeptide chain that folds into seven helices, the amino acid loops that connect the helices extend either into the cytoplasm or into the extracellular space. The amino end extends outside the cell, and the carboxyl end extends into the cell. The large extracellular segment is the site for specific gonadotropin recognition and binding. Binding changes the conformation (which is associated with phosphorylation), leading to interaction with the G proteins, which in turn activate second messengers, either enzymes or ion channels. These are ancient proteins; e.g., they are used by yeast to detect mating pheromones (perhaps this is why this protein is the basic structure for sight and smell in higher organisms; rhodopsin is a G protein located in the light-sensitive rod of the retina). Thus, the G receptors can be activated by hormones, neurotransmitters, growth factors, odorants, and photons of light.

LH and hCG bind to a common receptor, encoded by genes on chromosome 2p21. The LH/hCG receptor is highly conserved in mammals; the human receptor is very similar to that of rat and bovine receptors.[247] It is likely that expression of the LH/hCG receptor is regulated by many factors, including endocrine, paracrine, and autocrine mechanisms, but the primary requirement is FSH. In addition to the G protein pathway, activation of the LH/hCG receptor stimulates the calcium messenger system.

The receptor for FSH is very similar to the LH/hCG receptor, but it is structurally distinct.[248, 249] Appropriately (for specificity), the extracellular segment contains the major sequence divergence. The FSH receptor gene is located on chromosome 2p21–16, near the LH/hCG receptor gene. The FSH receptor is also regulated by its hormone environment, especially by FSH and estradiol. Other members of this family include receptors for TSH, catecholamines, vasopressin, angiotensin II, and dopamine.

Mutations in the G Protein System

Rare mutations that alter the structure and activity of G proteins can result in disease.[241, 244, 250–253] Loss of function mutations of a G protein or a given receptor will result in hormone deficiency syndromes; e.g., the TSH receptor and hypothyroidism, the LH receptor and male pseudohermaphroditism, and pseudohypoparathyroidism due to a $G_s\alpha$ mutation. Activation of gonadotropin-releasing hormone (GnRH) at puberty is mediated by kisspeptin 1, a ligand that binds to its G protein receptor on GnRH neurons; an inactivating mutation in this receptor causes hypogonadism and delayed or absent puberty. Mutations in the G protein receptor for GnRH have been identified that are responsible for familial hypogonadism. The McCune-Albright syndrome (sexual precocity, polyostotic fibrous dysplasia, café-au-lait skin pigmentation, and autonomous functioning of various endocrine glands) is due to unregulated activity (gain in function) of the adenylate cyclase system because of a mutation in the $G_s\alpha$ gene. $G_s\alpha$ protein mutations have also been found in adrenal and ovarian tumors, growth hormone-secreting pituitary adenomas, and thyroid adenomas. It is possible that alterations in the G protein system may ultimately explain abnormalities in endocrine-metabolic functions, as well as oncogenic mutations.[242, 243, 254]

Besides mutations, altered G protein receptor function can be due to polymorphisms, slight genetic changes that can be associated with altered physiology or disease. Single nucleotide polymorphisms (SNPs) associated with these receptors are being identified, and correlations are being made with risks and outcomes for diseases.[255] This is an avenue of genetic research that is bound to reveal explanations for some instances of infertility or problems of sexual differentiation.

Some Genetic Diseases due to Specific G Protein System Mutations

Mutation	*Disorder*
Activating LH receptor	Precocious puberty in boys
Inactivating LH receptor	Male pseudohermaphroditism
Activating FSH receptor	Ovarian hyperstimulation
Inactivating FSH recaptor	Premature ovarian failure
Inactivating TRH receptor	Hypothyroidism
$G_s\alpha$ (stimulatory)	McCune-Albright syndrome
$G_i\alpha$ (inhibitory)	Hypothyroidism
Rhodopsin	Retinitis pigmentosa
Activating Vasopressin receptor	Diabetes insipidus

Coupling and Uncoupling, Desensitization, and Alterations in Regulatory Proteins

Another way to explain stimulating and inhibiting actions at the adenylate cyclase level focuses on the mechanism of coupling. LH stimulates steroidogenesis in the corpus luteum and works through the coupling of stimulatory regulatory units to the catalytic units of adenylate cyclase. Prostaglandin $F_{2\alpha}$ is directly luteolytic, inhibiting luteal steroidogenesis through a mechanism that follows binding to specific receptors. This luteolytic action may be exerted via an inhibitory regulatory unit that leads to uncoupling with the catalytic unit, thus interfering with gonadotropin action.

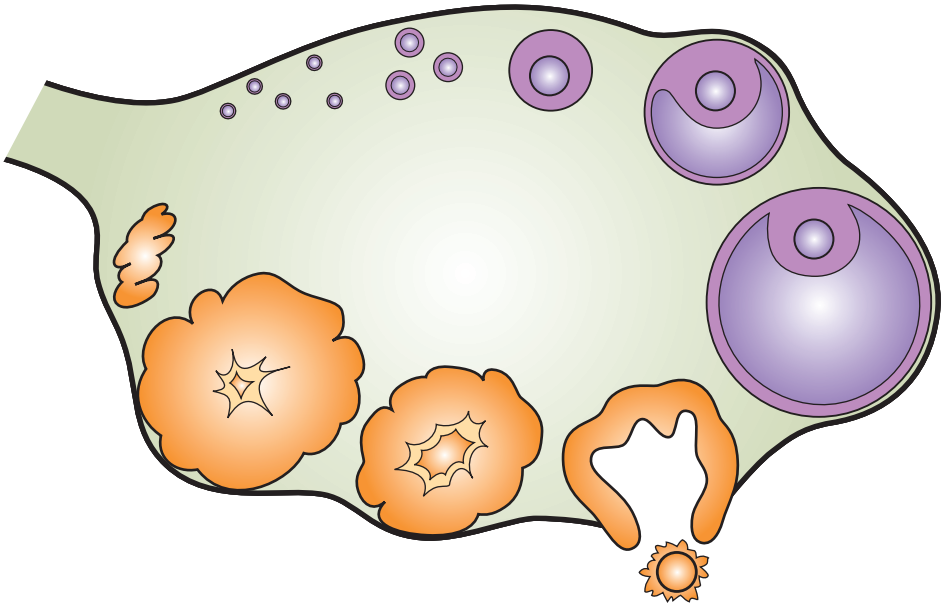
Increasing concentrations of tropic hormones, such as gonadotropins, are directly associated with desensitization of adenylate cyclase independent of the internalization of receptors. Desensitization is a rapid, acute change without loss of receptors in contrast to the slower process of internalization and true receptor loss. The desensitization process after prolonged agonist exposure involves receptor phosphorylation (which uncouples the receptor from the G protein). The LH/hCG receptor, a member of the G protein family, undergoes desensitization/uncoupling in response to LH or hCG in a process that involves phosphorylation of the C-terminal cytoplasmic tail of the receptor.[256, 257] Decreased gonadotropin secretion in the presence of prolonged continuous GnRH stimulation is a desensitization response that can occur followed by recovery within the time frame of a normal endogenous GnRH secretory pulse.[258]

Desensitization can also follow enzymatic alterations that affect the key intracellular proteins that regulate steroidogenesis. Activation of the mitogen-activated protein kinase (MAPK) system increases levels of SF-1, which in turn attenuates StAR expression, which is essential for the transport of cholesterol in gonadal cells producing steroids.[259]

All references are available online at: http://www.clinicalgynendoandinfertility.com

3

The Ovary—Embryology and Development



The great names of early Western medicine were Hippocrates, Soranus, and Galen. Although Aristotle (384–322 B.C.) referred to castration as a common agricultural practice, it was Soranus who provided the first anatomic description of the ovaries. Soranus of Ephesus (a city founded by Greeks on the coast of what is now Turkey) lived from 98 to 138 A.D. and has often been referred to as the greatest gynecologist of antiquity.[1] He studied in Alexandria and practiced in Rome. His great text was lost for centuries and was not published until 1838.

Galen was born in 130 A.D. in Pergamum, a Greek city in eastern Turkey, studied in Alexandria and became a famous practitioner and teacher of medicine also in Rome. He lived 70 years and wrote about 400 treatises, 83 of which are still in existence. Galen preserved in his own writings (in Greek) Aristotle's descriptions of reproduction. He was a true scholar and was regarded as the ultimate authority on anatomy and physiology until the sixteenth century.[2] It was Galen who established bleeding as the appropriate treatment for almost every disorder. Although in retrospect Galen's conclusions and teachings contained many errors, how many other individuals have been able to satisfy the needs of scholars and physicians for hundreds of years?

After Galen, no further thoughts or advances were recorded for well over 1,000 years as the dark weight of the medieval ages descended on Western civilization. During the medieval

years, it was safe to copy Galen's works but literally dangerous to contribute anything original. Medieval scholars believed it was impossible to progress in knowledge beyond Galen. The doctrine according to Galen was not challenged until the introduction of printing made Galen's works available to scholars.

Although Leonardo da Vinci (1452–1519) drew accurately the anatomy of the uterus and the ovaries, the major advances in anatomic knowledge can be traced to the University of Padua, the famed Italian university where a succession of anatomists made important contributions.[3] It was Andreas Vesalius (1514–1564), who while still in his 20s, because of his own human dissections, realized that Galen described only animals. Appointed Professor of Surgery and Anatomy at the University of Padua at the age of 23, he published *De Humani Corporis Fabrica*, his authoritative, illustrated book on human anatomy, in 1543, at the age of 29. Vesalius was harshly attacked by the medical establishment, and one year after the publication of his book, he left Padua to become the court physician in Spain.

Vesalius was the first to describe ovarian follicles and probably the corpus luteum. Fallopius (1523–1562), remembered for his description of the fallopian tubes, was a pupil of Vesalius, and then a successful and popular teacher of anatomy at Padua. Fabricius (Girolamo Fabrici d'Aquapendente, 1533–1619), a pupil of Fallopius, succeeded Fallopius as chair of anatomy at Padua and made major contributions to embryology. Studying an organ in birds and observing that it contained eggs, Fabricius called it the "ovary." During this period of time, the ovaries came to be recognized as structures, but their function remained a mystery.

William Harvey published the first original English book on reproductive anatomy and physiology in 1651, at the age of 69, 35 years after his discovery of the circulation of blood. He obtained his medical education at the University of Padua, where he learned to describe accurately his own observations, a practice he was to continue and that culminated in his writings. Unfortunately, Harvey promoted and maintained the Aristotelian belief that the egg was a product of conception, a result of an interaction between semen and menstrual blood. This view was corrected by Bishop Niels Stensen of Denmark in 1667, and in 1672, at the age of 31, the Dutch physician Regnier de Graaf published his great work on the female reproductive organs, *De Mulierum Organis Generationi Inservientibus Tractatus Novus* (A New Treatise on the Female Reproductive Organs), that established the ovary as the source of the ovum.

Ovarian follicles had been described by Vesalius and Fallopius, but the impact of his publication earned de Graaf eternal recognition as the ovarian follicle became known as the graafian follicle, even though de Graaf believed that the whole follicle was the egg. de Graaf was the first to accurately describe the corpus luteum, although Marcello Malpighi, whose works were published posthumously in 1697, invented the name "corpus luteum."

With the discovery of mammalian spermatozoa by van Leeuwenhoek in 1677, it became possible to speculate that fertilization resulted from the combination of a spermatozoon and the graafian follicle. It would be another 150 years before it was appreciated that the oocyte resides within the follicle (described in 1827 by Carl Ernst von Baer), and that there is a relationship between the ovaries and menstruation. The process of fertilization was described by Newport in 1853–1854, and Oscar Hertwig, studying sea urchins, reported in 1876 the penetration of a spermatozoon into an egg and chromosome reduction during meiosis, bringing to a close the era of descriptive anatomy of the ovary and marking the beginning of scientific explorations into physiology and endocrinology.

The Human Ovary

The physiologic responsibilities of the ovary are the periodic release of gametes (eggs, oocytes) and the production of the steroid hormones estradiol and progesterone. Both activities are integrated in the continuous repetitive process of follicle maturation, ovulation, and corpus luteum formation and regression. The ovary, therefore, cannot be viewed as a relatively static endocrine organ whose size and function expand and contract, depending on the vigor of stimulating tropic hormones. Rather, the female gonad is a heterogeneous ever-changing tissue whose cyclicity is measured in weeks, rather than hours.

The ovary consists of three major portions: the outer cortex, the central medulla, and the rete ovarii (the hilum). The hilum is the point of attachment of the ovary to the mesovarium. It contains nerves, blood vessels, and hilus cells, which have the potential to become active in steroidogenesis or to form tumors. These cells are very similar to the testosterone-producing Leydig cells of the testes. The outermost portion of the cortex is called the tunica albuginea, topped on its surface by a single layer of cuboidal epithelium, referred to as the ovarian surface epithelium or the ovarian mesothelium (epithelial ovarian carcinomas account for 90% of human ovarian cancers). The oocytes, enclosed in complexes called follicles, are in the inner part of the cortex, embedded in stromal tissue. The stromal tissue is composed of connective tissue and interstitial cells, which are derived from mesenchymal cells, and have the ability to respond to luteinizing hormone (LH) or human chorionic gonadotropin (hCG) with androgen production. The central medullary area of the ovary is derived largely from mesonephric cells.

The Fetal Ovary

During fetal life, the development of the human ovary can be traced through four stages:[4, 5] (1) the indifferent gonad stage, (2) the stage of differentiation, (3) the period of oogonal multiplication and oocyte formation, and finally, (4) the stage of follicle formation.

The Indifferent Gonad Stage

At approximately 5 weeks of gestation, the paired gonads are structurally consolidated coelomic prominences overlying the mesonephros, forming the gonadal ridges. At this point, the gonad is morphologically indistinguishable as a primordial testis or ovary. The gonad is composed of primitive germ cells intermingled with coelomic surface epithelial cells and an inner core of medullary mesenchymal tissue. Just below this ridge lies the mesonephric duct. This indifferent stage lasts about 7–10 days. Together, the mesonephros and the genital ridge are called the urogenital ridge, indicating the close association of the urinary and reproductive systems.

The origin of the gonadal somatic cells is still not certain. The earliest recognizable gonad contains, besides the germ cells, somatic cells derived from at least three different tissues: coelomic epithelium, mesenchyme, and mesonephric tissue. In one early model, the gonad was believed to be formed by the invasion of the “germinal epithelium” into the underlying mesenchyme. The germinal epithelium is simply that part of the coelomic epithelium

that gives rise to gonadal tissue. The invading cells were thought to form the primary sex cords that contain the germ cells surrounded by somatic cells (the cells destined to form the tissue that holds the germ cells). In a newer model, well-supported by experimental data, the somatic cells of the gonad are believed to arise from the mesonephros and not the coelomic epithelium.[6–8] Ultrastructural studies have even suggested that both coelomic epithelial and underlying mesonephric cells provide the somatic cells that are destined to become follicular cells.[9]

The contribution of mesonephric cells requires migration into the gonad. This movement is regulated by fibroblast growth factor 9 in the male, and it may be repressed by a gene encoding a protein known as Sprouty2.[10] Thus the Y chromosome directs sexual development by influencing gene expression via specific regulating and signaling proteins.

The primordial germ cells originate within the primitive ectoderm, but the specific cells of origin cannot be distinguished.[11] The germ cells are first identified at the end of the third week after fertilization in the primitive endoderm at the caudal end in the dorsal wall of the adjacent yolk sac, and, soon, they also appear in the splanchnic mesoderm of the hindgut.[12, 13] The gonadal ridge is the one and only site where the germ cells can survive. By displacement because of growth of the embryo and also by active ameboid movement along the dorsal mesentery of the hindgut toward the genital ridges, the germ cells "migrate" from the yolk sac around the hindgut to their gonadal sites between weeks 4 and 6 of gestation.[11, 14]

The factors that initiate and guide the migration of the germ cells are not known, although chemotactic and adhesive peptides, such as fibronectin and laminin, are involved. In rodents, germ cell proliferation and migration involve stem cell factor (kit ligand) and the expression of its receptor (c-Kit), a transmembrane tyrosine kinase receptor encoded by the c-kit oncogene.[15] In gonads obtained from individuals with intersex disorders that have a high risk of testicular tumors, the expression of kit protein was detected at a later gestational age than in normal controls, consistent with both later germ cell migration and a change in the oncogene expression.[16] The kit gene is located on chromosome 4, and mutations in this gene have not been discovered in women with premature ovarian failure.[17]

The germ cells begin their proliferation during their migration.[9] The germ cells are the direct precursors of sperm and ova, and by the sixth gestational week, on completion of the indifferent state, these primordial germ cells have multiplied by mitosis to a total of 10,000. By the sixth week of gestation, the indifferent gonads contain the germ cells and supporting cells derived from the coelomic epithelium and the mesenchyme of the gonadal ridge.

Male or female differentiation of the gonad is directed by the sex chromosomes. But the decision to be male or female must be communicated to the cells of the indifferent gonad.[18] This communication, too, is genetic in origin, involving signaling proteins and their receptors, programmed by the fundamental impact that depends on whether a Y chromosome is present. A genetic program involving hundreds of genes is initiated in the indifferent gonad at the urogenital ridge that leads to differentiation into a testis or an ovary.[19]

The Stage of Differentiation

If the indifferent gonad is destined to become a testis, differentiation along this line will take place at 6–9 gestational weeks. The absence of testicular evolution (formation of medullary primary sex cords, primitive tubules, and incorporation of germ cells) gives implicit evidence of the existence of a primitive, albeit momentarily quiescent, ovary. In contrast to the male, female internal and external genitalia differentiation precedes gonadal maturation. These events are related to the genetic constitution and the territorial receptivity of the

mesenchyme. If either factor is deficient or defective, improper development occurs. As has been noted, primitive germ cells are unable to survive in locations other than the gonadal ridge. If partial or imperfect gonadal tissue is formed, the resulting abnormal nonsteroidal and steroidal events have wide ranging morphologic, reproductive, and behavioral effects.

The Testes

The factor that determines whether the indifferent gonad will become a testis is called, appropriately, the testes-determining factor (TDF), a product of a single gene located on the Y chromosome.[20, 21] The testicular-determining factor gene is located within a region named SRY, the sex-determining region on the Y chromosome.[22] The protein product of the *SRY* gene contains a DNA-binding domain to activate gene transcription that diverts the development of gonadal somatic cells from the pathway to follicle cells to Sertoli cells.[23] Rare cases have been reported of phenotypic, infertile males with a 46,XX karyotype, the male differentiation is due to a translocation of a Y chromosome fragment containing *SRY* to an autosome or an X chromosome.

Normal testis development requires not only the presence of the *SRY* gene but its interaction with autosomal genes.[24, 25] Genes similar to *SRY* have been named *SOX* genes (the similarity is with the *SRY* box region that contains the DNA-binding sequence).[23] *SRY* expression precedes *SOX* expression; indeed, *SRY* cooperatively with steroidogenic factor-1, SF1, and Fgf9, a member of the fibroblast growth factor family, upregulates the *SOX9* gene.[21] Thus, *SRY* along with partner proteins targets male-specific genes that are essential for testicular development. This process is discussed in detail in Chapter 9. In mice, the developmental consequences of activating and inactivating mutations in *Sox9* resemble those of similar mutations in *Sry*, implying not only that *Sox9* is required for testis differentiation, but also that *Sry* activation of *Sox9* may be all that is necessary to activate other genes important to testis development, such as *Fgf9* (fibroblast growth factor 9), and to repress genes that induce ovary development, such as *Wnt4* (a member of the wingless family of genes), *Rspo1* (R-spondin 1), *Dax1* (dosage-sensitive sex reversal, adrenal hypoplasia critical region, on chromosome X, gene 1), and *Foxl2* (forkhead box L2).[26]

Both testis and ovary differentiation require dominantly acting genes, with *SRY* inducing testis development via up-regulation of *SOX9*, and with other genes, primarily *WNT4* and *RSPO1*, teaming to promote ovary development via repression of *SOX9*. This concept views the fate of the bipotential gonad as balanced between opposing forces and *SRY* as the key factor. In XY gonads, *SRY* induces *SOX9* and tips differentiation toward testis development, and in XX gonads lacking *SRY*, other genes combine to repress *SOX9* and promote ovary development.[27]

The expression of the *SRY* gene is confined to the genital ridge during fetal life, but the gene is also active in the germ cells of the adult, perhaps playing a role in spermatogenesis.[20] The traditional view assigns ***active*** gene control and expression for testicular differentiation and a ***passive***, "default" mode of development for the ovary. However, recent evidence has challenged that concept, revealing that ovarian development requires the actions of genes such as *WNT4*, *RSPO1*, and *DAX1*, which team to repress the expression of genes in the testis pathway (e.g. *SOX9*), and other genes that promote ovarian development.[28, 29] "Passive" is not an accurate description because even before the gonad differentiates into an ovary, robust gene activity associated with the female genotype occurs in the urogenital ridge.[19]

When the Y chromosome containing *SRY* is present, the gonads develop into testes. The male phenotype is dependent on the products (antimüllerian hormone and testosterone) of the fetal testes, whereas the female phenotype is the result of an absence of these fetal gonadal products.[30] Antimüllerian hormone (AMH), which inhibits the formation of the müllerian ducts, is secreted at the time of Sertoli cell differentiation, beginning at 7 weeks.

AMH expression is altered only by mutations in the *AMH* gene, located on chromosome 19p13.3.[31] Regression of the müllerian ducts is dependent on the presence of an adequate number of Sertoli cells, and the regulation of AMH receptor.[32] Mutations in the AMH receptor gene results in the presence of the uterus, fallopian tubes, and the upper vagina in 46,XY men with normal external virilization.

After involution of the müllerian system, AMH continues to be secreted, but there is no known function. However, evidence in the mouse suggests a role in early germ cell transformation during spermatogenesis.[33] In the ovary, very small amounts of AMH mRNA are present early in life, and although there may be no role in female development, its production later in life by the granulosa cells leads to autocrine and paracrine actions in oocyte maturation and follicular development.[34] Serum AMH levels in adult women correlate with the number of ovarian follicles present and predict the response to stimulation with ovulation-inducing therapy.[35]

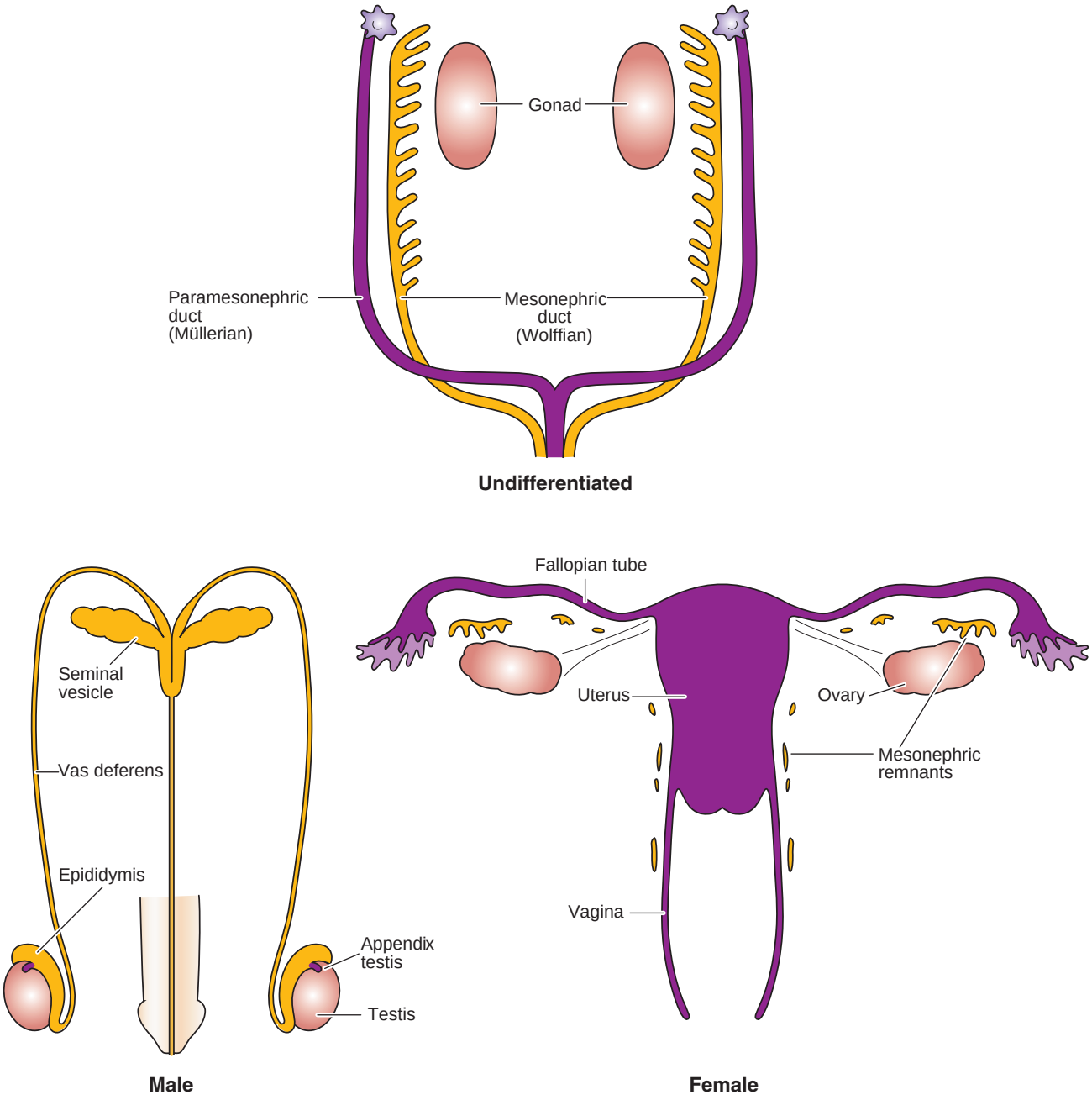
The testis begins its differentiation in week 6–7 of gestation by the appearance of Sertoli cells that aggregate to form the testicular cords. The primordial germ cells are embedded in the testicular cords that will form the Sertoli cells and spermatogonia. The mature Sertoli cells are the site of production of ABP (androgen-binding protein, important in maintaining the high local androgen environment necessary for spermatogenesis) and inhibin.

The Leydig cells differentiate (beginning week 8) from mesenchymal cells of the interstitial component surrounding the testicular cords. Thus, secretion of AMH precedes steroidogenesis in Leydig cells. Shortly after the appearance of the Leydig cells, secretion of testosterone begins. Androgen secretion increases in conjunction with increasing Leydig cell numbers until a peak is reached at 15–18 weeks. At this time, Leydig cell regression begins, and at birth only a few Leydig cells are present.

The cycle of fetal Leydig cells follows the rise and decline of fetal human chorionic gonadotropin (hCG) levels during pregnancy. This relationship and the presence of hCG receptors in the fetal testes indicate a regulatory role for hCG.[4] The pattern of hCG in the fetus parallels that of the mother, peaking at about 10 weeks and declining to a nadir at 20 weeks of gestation, but the concentrations are only 5% of maternal concentrations. Human chorionic gonadotropin stimulation produces Leydig cell hypertrophy, and peak fetal testosterone levels are seen at 15–18 weeks.[36] However, normal masculine differentiation occurs in mouse models lacking LH receptors, and molecular evidence indicates that fetal Leydig cells (but not adult cells) respond to adrenocorticotropic hormone (ACTH) as well as hCG. [37] A primary role for ACTH is supported by the report of a male with an inactivating mutation of the gene for the hCG/LH receptor who developed female genitalia along with a vas deferens and epididymis.[38]

Testosterone synthesis in human fetal testes begins at the 8th week of gestation, reaches a peak between 15–18 weeks, and then declines. Testicular function in the fetus can be correlated with the fetal hormonal patterns. Although the initial testosterone production and sexual differentiation are in response to the fetal levels of ACTH and hCG, further testosterone production and masculine differentiation are maintained by the fetal pituitary gonadotropins. Decreased testosterone levels in late gestation probably reflect the decrease in gonadotropin levels. The fetal Leydig cells, by an unknown mechanism, avoid down-regulation and respond to high levels of hCG and LH by increased steroidogenesis and cell multiplication. This generation of cells is replaced by the adult generation of Leydig cells that becomes functional at puberty and responds to high levels of hCG and LH with down-regulation and decreased steroidogenesis. Leydig cells, therefore, are composed of two distinct populations, one active during fetal life and one active during adult life.

The fetal spermatogonia, derived from the primordial germ cells, are in the testicular cords, surrounded by the Sertoli cells. In contrast to the female, male germ cells do not start meiotic division before puberty.



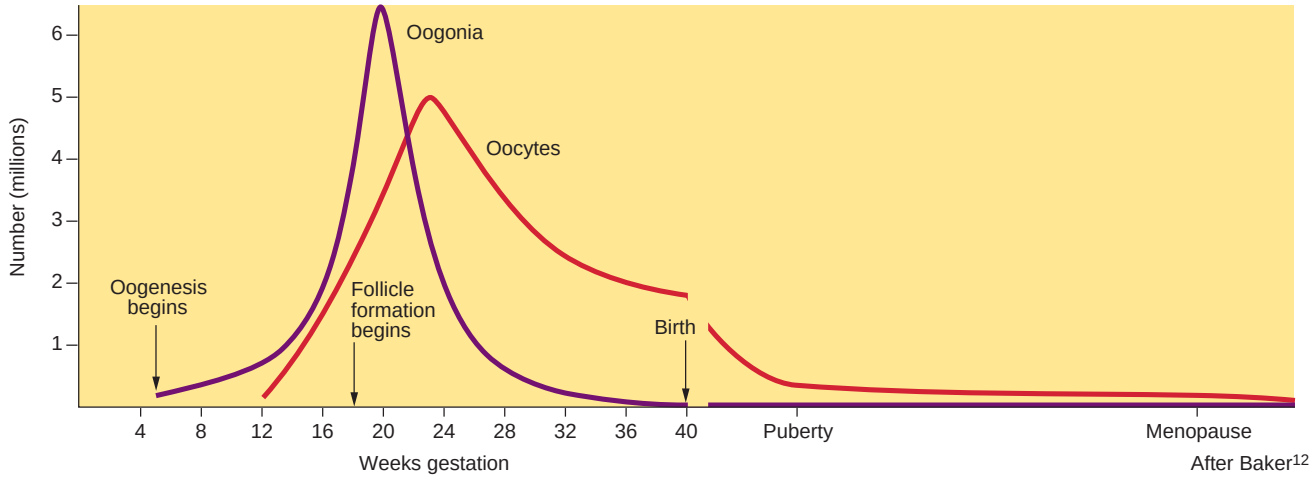
The differentiation of the wolffian system begins with the increase in testicular testosterone production. The classic experiments by Jost indicate that this effect of testosterone is due to local action, explaining why male internal genitalia in true hermaphrodites are only on the side of the testis.[30] Not all androgen-sensitive tissues require the prior conversion of testosterone to dihydrotestosterone (DHT). In the process of masculine differentiation, the development of the wolffian duct structures (epididymis, the vas deferens, and the seminal vesicle) is dependent on testosterone as the intracellular mediator, whereas development of the urogenital sinus and urogenital tubercle into the male external genitalia, urethra, and prostate requires the conversion of testosterone to DHT.[39] In the female, the loss of the wolffian system is due to the lack of locally produced testosterone.

The Stage of Oogonal Multiplication and Oocyte Formation

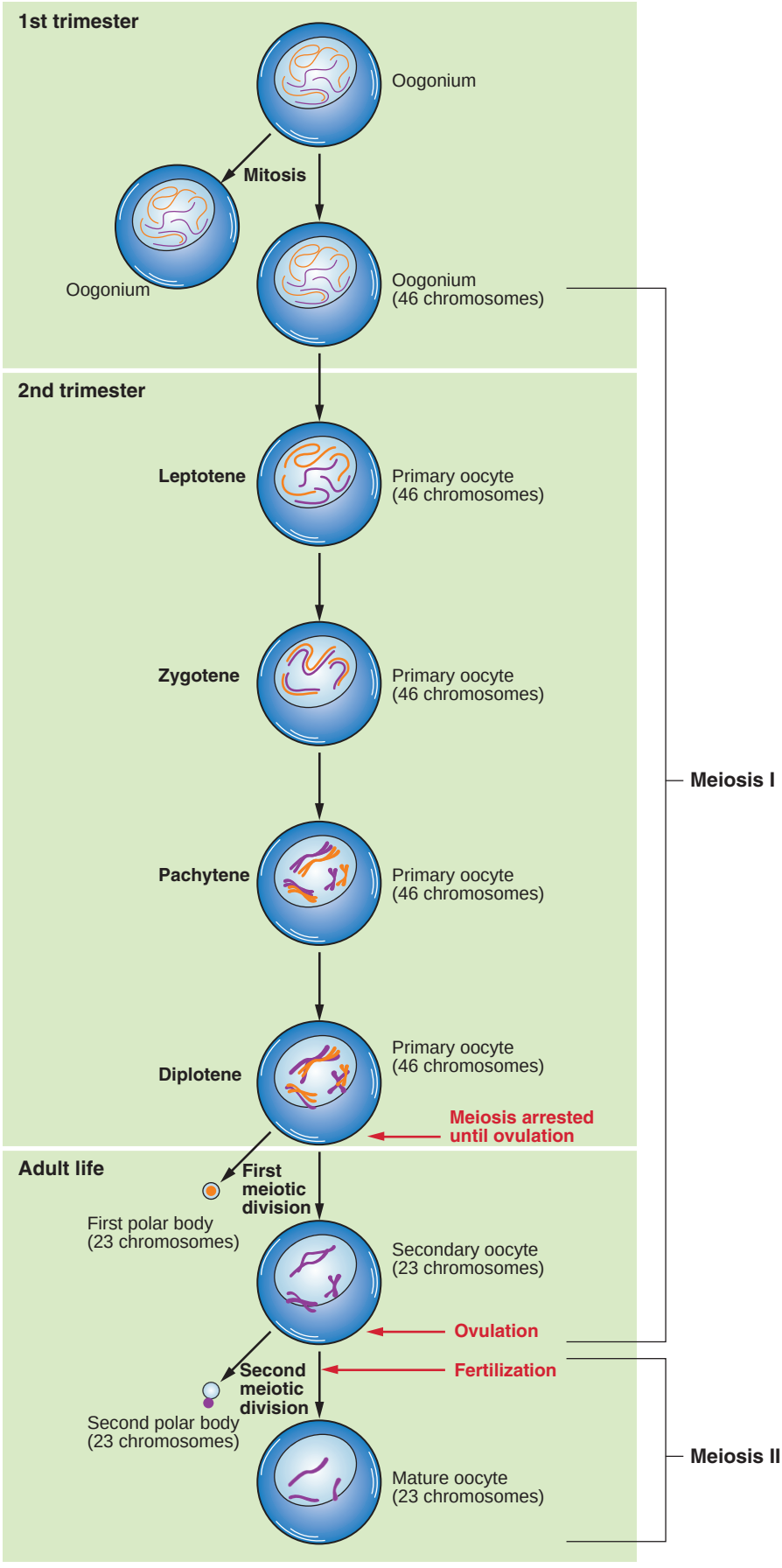
Just as testicular development requires *SRY* and *SOX9* expression, specific molecular signaling pathways are necessary for ovarian differentiation and oocyte survival. In the developing ovary, the most important genes are *WNT4* and *RSPO1*. Both activate the beta-catenin signaling pathway in somatic cells, which results in the loss of cell-cell adhesion between female germ cells that is required for their entry into meiosis.[40] *WNT4* and *RSPO1* also combine to suppress somatic cell *SOX9* expression. Thus, ovarian differentiation is dependent on essential genetic molecular signaling that is active in the absence of *SRY*. Evidence suggests that *RSPO1* also may act directly to suppress male differentiation, even in the absence of *SRY*.[41] The ultimate fate of the gonad depends on which molecular signaling pathway dominates.[42]

At 6–8 weeks, the first signs of ovarian differentiation are reflected in the rapid mitotic multiplication of germ cells, reaching 6–7 million oogonia by 16–20 weeks.[12, 43] This represents the maximal oogonal content of the gonad. From this point in time, germ cell content will irretrievably decrease until, some 50 years later, the store of oocytes will be finally exhausted.

By mitosis, the germ cells give rise during week 9 to the oogonia. The oogonia are transformed to oocytes as they enter the first meiotic division and arrest in prophase. This process begins at 11–12 weeks, perhaps in response to a factor or factors produced by the rete ovarii,[44] that may act directly on the germ cells, or indirectly via actions on the somatic cells. Studies in mice suggest that retinoic acid derived from the mesonephros may act as a functional meiosis-inducing factor in the female germ cells.[45] Progression of meiosis to the diplotene stage is accomplished throughout the rest of pregnancy and completed by birth. Arrest of meiosis at the end of the first stage is probably maintained by inhibiting substances produced by granulosa cells. A single ovum is formed from the two meiotic divisions of the oocyte, one just before ovulation and the second (forming the haploid ovum) at the time of sperm penetration. The excess genetic material is extruded as one polar body at each meiotic division. Gonadotropins and various growth factors (but not sex steroids) can induce resumption of meiosis in vitro, but only in oocytes enclosed by cumulus-granulosa cells. A family of sterols is present in follicular fluid, presumably secreted by the cumulus cells in response to gonadotropins, that activates oocyte meiosis and oocyte maturation.[46, 47] Follicle-stimulating hormone (FSH) induces the resumption of meiosis, a reaction that requires the presence of the gap junction network that allows communication between the cumulus cells and the oocyte.



After Baker[12]


1st trimester
Oogonium
Mitosis
Oogonium
Oogonium
(46 chromosomes)
2nd trimester
Leptotene
Primary oocyte
(46 chromosomes)
Zygotene
Primary oocyte
(46 chromosomes)
Pachytene
Primary oocyte
(46 chromosomes)
Diplotene
Primary oocyte
(46 chromosomes)
Meiosis arrested
until ovulation
Adult life
First
meiotic
division
First polar body
(23 chromosomes)
Secondary oocyte
(23 chromosomes)
Ovulation
Second
meiotic
division
Fertilization
Second polar body
(23 chromosomes)
Mature oocyte
(23 chromosomes)
Meiosis I
Meiosis II

Loss of germ cells takes place throughout all of these events: during mitosis of germ cells, during the various stages of meiosis, and finally, after follicle formation. The massive loss of oocytes during the second half of pregnancy is the consequence of several mechanisms. Besides follicular growth and atresia, substantial numbers of oocytes regress during meiosis, and those oogonia that fail to be enveloped by granulosa cells undergo degeneration. This process is influenced by genes that actively repress germ cell death.[48] In addition, germ cells (in the cortical area) migrate to the surface of the gonad and become incorporated into the surface epithelium or are eliminated into the peritoneal cavity.[49, 50] In contrast, once all oocytes are encased in follicles (shortly after birth), the loss of oocytes will be only through the process of follicular growth and atresia.

Chromosomal anomalies can accelerate germ cell loss. Individuals with Turner syndrome (45,X) experience normal migration and mitosis of germ cells, but the oogonia do not undergo meiosis, and rapid loss of oocytes leaves the gonad without follicles early in life, and it appears as a fibrous streak. The rate of loss varies, and 10–20% experience spontaneous menstruation; rare pregnancies have been reported in those who have had spontaneous menstruation.[51, 52] However, the presence of menstrual function and reproduction in a patient with Turner phenotype may be due to an undetected mosaic complement, such as a 46,XX line in addition to 45,X.

The Stage of Follicle Formation

At 18–20 weeks, the highly cellular cortex is gradually perforated by vascular channels originating in the deeper medullary areas, and this marks the beginning of follicle formation.[53] As the finger-like vascular projections enter the cortex, it takes on the appearance of secondary sex cords. As blood vessels invade and penetrate, they divide the previously solid cortical cell mass into smaller and smaller segments. Drawn in with the blood vessels are perivascular cells that originate in the mesonephros or in the coelomic epithelium. Some believe that the coelomic epithelium is the origin of all ovarian somatic cells; others favor a mesenchymal or dual origin.[5, 11] These cells give rise to the pregranulosa cells that surround the oocytes, which have completed the first stage of meiosis. The resulting unit is the ***primordial follicle—an oocyte arrested in prophase of meiosis, enveloped by a single layer of spindle-shaped pregranulosa cells, surrounded by a basement membrane.*** Eventually all oocytes are covered in this fashion. Residual mesenchyme not utilized in primordial follicle formation is noted in the interstices between follicles, forming the primitive ovarian stroma. This process of primordial follicular development continues until all oocytes in the diplotene stage can be found in follicles, some time shortly after birth.

As soon as the oocyte is surrounded by the rosette of pregranulosa cells, the entire follicle can undergo variable degrees of maturation before arresting and becoming atretic. The formation of a ***primary follicle*** is marked by a change of the pregranulosa layer to a cuboidal layer of granulosa cells. This change is associated with proliferation. In the human it is estimated that about 13 pregranulosa cells surround the oocyte and with the change to a primary follicle, the number increases to about 76 granulosa cells.[54] A later and perhaps more accurate study concluded that the primary follicle contains about 105 granulosa cells, associated with an increase in average diameter from 40 to 54 μm.[55]

Further differentiation into a ***preantral follicle*** is marked by more complete granulosa proliferation. Call–Exner body formation (coalescence to form an antrum) and occasionally a minor theca layer system that differentiates from surrounding mesenchymal cells can be seen. Preantral follicles can be found in the sixth month of gestation, and ***antral follicles*** (the graafian follicle, characterized by a fluid-filled space) are present by the end

of pregnancy, but not in large numbers. It is only during the last third of gestation that theca cells can be found surrounding follicles.[43]

Even in fetal life, the cycle of follicle formation, variable ripening, and atresia occurs. Although these steps are precisely those typical of adult reproductive life, full maturity, as expressed in ovulation, does not occur. Estrogen production does not occur until late in pregnancy when follicular development takes place, and even then steroidogenesis is not significant. Unlike the male, gonadal steroid production is not required for development of a normal phenotype. The development of the müllerian duct into the fallopian tubes, the uterus, and the upper third of the vagina is totally independent of the ovary.

The ovary at birth and in the first year of life can contain cystic follicles of varying size, undoubtedly stimulated by the reactive gonadotropin surge accompanying the withdrawal of the neonatal hypothalamus and pituitary from the negative feedback of fetoplacental steroids.[56] Ovarian cysts can also be occasionally detected in fetuses by ultrasonography.

The anterior pituitary begins development between 4 and 5 weeks of fetal life. The median eminence is apparent by week 9 of gestation, and the hypothalamic-pituitary portal circulation is functional by the 12th week. Pituitary levels of follicle-stimulating hormone (FSH) peak at 20–23 weeks, and circulating levels peak at 28 weeks.[57] Levels are higher in female fetuses than in males until the last 6 weeks of gestation. Ovaries in anencephalic fetuses, which lack gonadotropin-releasing hormone (GnRH) and gonadotropin secretion, lack antral follicles and are smaller at term, but progression through meiosis and development of primordial follicles occurs, obviously not dependent on gonadotropins.[4] The ovary develops receptors for gonadotropins only in the second half of pregnancy. Thus, the loss of oocytes during fetal life cannot be solely explained by the decline in gonadotropins. The follicular growth and development observed in the second half of pregnancy, however, is gonadotropin dependent.[58] Hypophysectomy of a fetal monkey is followed by an increase in oocyte loss by atresia.[59]

The Neonatal Ovary

The total cortical content of germ cells falls to 500,000–2 million by birth as a result of prenatal oocyte depletion.[60–62] This huge depletion of germ cell mass (close to 4–5 million) has occurred over as short a time as 20 weeks. No similar rate of depletion will be seen again. Because of the fixed initial endowment of germ cells, the newborn female enters life, still far from reproductive potential, having lost 80% of her oocytes.

The ovary is approximately 1 cm in diameter and weighs about 250–350 mg at birth, although sizable cystic follicles can enlarge the total dimensions. Intriguingly, the gonad on the right side of the body in both males and females is larger, heavier, and greater in protein and DNA content than the gonad on the left side.[63] Compartmentalization of the gonad into cortex and a small residual medulla has been achieved. In the cortex, almost all the oocytes are involved in primordial follicle units. Each ovary contains a similar total number of follicles.[64] Varying degrees of maturation in some units can be seen as in the fetal state.

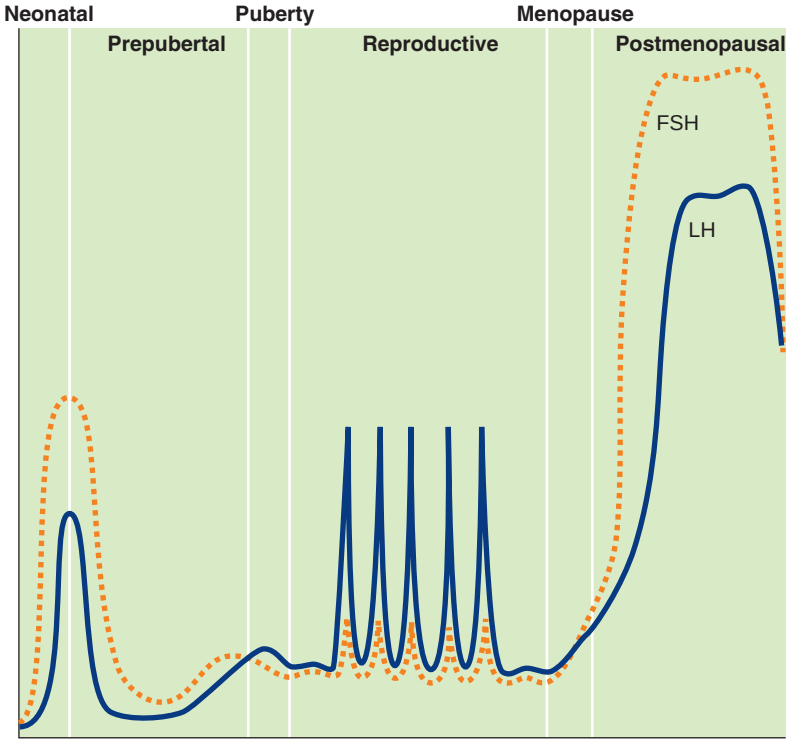
There is a sex difference in fetal gonadotropin levels. There are higher pituitary and circulating FSH and pituitary LH levels in female fetuses. The lower male levels are undoubtedly due to testicular testosterone and inhibin production. In infancy, the postnatal FSH rise is more marked and more sustained in females, whereas LH values are not as high. The FSH levels are greater than the levels reached during a normal adult menstrual cycle, decreasing to low levels usually by 1 year of age, but sometimes later.[65] LH levels are in the range of lower adult levels. This early activity is accompanied by inhibin levels comparable to the

low range observed during the follicular phase of the menstrual cycle. Follicular response to the antral stage is relatively common in the first 6 months of life in response to these elevated gonadotropin levels. The most common cause of abdominal masses in fetuses and newborns is ovarian cysts, a consequence of gonadotropin stimulation.[66]

Interference with the postnatal rise in gonadotropins in monkeys is associated with disturbances in normal hypothalamic-pituitary function at puberty.[67] Indeed, in male monkeys, the administration of a GnRH analogue in the neonatal period has an adverse impact on subsequent immunologic and behavioral functions as well as normal reproduction.[67] After the postnatal rise, gonadotropin levels reach a nadir during early childhood (by about 6 months of age in males and 1–2 years in females) and then rise slightly between ages 4 and 10 years.

The Ovary in Childhood

The childhood period is characterized by low levels of gonadotropins in the pituitary and in the blood, little response of the pituitary to GnRH, and maximal hypothalamic suppression. The ovary, however, is not quiescent during childhood. Follicles begin to grow at all times and frequently reach the antral stage. Ultrasonography can commonly demonstrate ovarian follicular cysts during childhood, ranging in size from 2 to 15 mm.[68] These small unilocular ovarian cysts are not clinically significant.[69] The process of atresia with an increasing contribution of follicular remnants to the stromal compartment yields progressive ovarian enlargement during childhood, about a 10-fold increase in weight.[70] Of course, the lack of gonadotropin support prevents full follicular development and function. There is no evidence that ovarian function is necessary until puberty. However, the oocytes during this time period are active, synthesizing messenger RNAs and protein. Furthermore, ovariectomy in prepubertal monkeys indicates that the prepubertal suppression of GnRH and gonadotropins is partially dependent on the presence of ovaries, suggesting some functional activity of the ovary in childhood.[71]



The Adult Ovary

At the onset of puberty, the germ cell mass has been reduced to 300,000 to 500,000 units.[12, 72] During the next 35–40 years of reproductive life, 400 to 500 will be selected to ovulate, and the primary follicles will eventually be depleted to a point at menopause when only a few hundred to a thousand remain.[73, 74,75] A gradually increasing rate of follicular depletion occurs throughout life accounting for a progressively declining number of follicles. In the last 10–15 years before menopause, follicular depletion correlates with a subtle but real increase in FSH and decrease in inhibin-B as well as insulin-like growth factor-I (IGF-I).[76–80] Fewer follicles grow per cycle as a woman ages, and cycles are at first shorter because follicular growth begins sooner during the late luteal phase, a consequence of a greater rise in FSH between cycles, and then longer as anovulation becomes more common.[81–85] These changes, including the increase in FSH (which is probably due to the decrease in inhibin-B), may partly reflect the reduced quality and capability of aging follicles. Beginning in their late 30s, women have smaller oocytes and smaller follicles, perhaps of lesser quality.[55] However, the rise in FSH due to a decrease in inhibin-B is also believed to be the consequence of a decreasing number of follicles in each cohort of active follicles.[86]

The acceleration of follicular loss in the later reproductive years has been questioned by a mathematical analysis of normal human ovaries.[87] This model indicates a constantly accelerating loss of follicles from birth to menopause. The rise in FSH and decrease in inhibin-B in the late 30s could still reflect both lesser quality in the remaining follicles and fewer follicles participating in each menstrual cycle.

This classic concept that the mammalian ovary cannot produce new oocytes (and follicles) after fetal life may need revision. Experimental work that is not without controversy identified stem cells in blood and bone marrow, which after transplantation to a chemically- or genetically-sterile recipient could generate oocytes within follicles.[88] This raises the intriguing possibility of a new approach to treatment for men and women with infertility or reproductive diseases.[89] Putative stem cells have even been isolated from the ovarian surface endothelium, obtained from postmenopausal women and women with premature ovarian failure.[90] Is it possible that the spontaneous resumption of ovarian function occasionally experienced by women with premature ovarian failure is due to stem cell-generated postnatal oogenesis? This is work in progress, including new studies that support the conventional wisdom that each female mammal is born with a finite number of oocytes that were formed in fetal life.[91]

The loss of oocytes (and follicles) through atresia is a response to changes in many factors. Certainly gonadotropin stimulation and withdrawal are important, but ovarian steroids and autocrine and paracrine factors are also involved. The consequence of these unfavorable changes, atresia, is a process called ***apoptosis***, programmed cell death. This process is heralded by alterations in mRNAs required for cell proteins that maintain follicle integrity.[92] Indeed, the process is a consequence of an orderly expression of key gene products that either promote or repress the apoptotic events.

Human ovaries and nonhuman primate ovaries are innervated by sympathetic and sensory neurons.[93] This neuronal network innervates the ovarian vasculature, interstitial tissue, and developing follicles. The neurons are connected synaptically to the paraventricular nucleus of the hypothalamus.[94] These neuronal cells produce catecholamines and nerve growth factor. The precise function of this unique ovarian nervous system is not known. Vasoactive intestinal peptide derived from these nerve fibers suppresses follicular atresia (apoptosis) in a mechanism that also involves IGF-I.[95] It has been suggested that sympathetic innervation of the ovary continues to develop at puberty, and that neurotransmitters are involved in the process in which follicles acquire FSH receptors and respond to FSH.[96, 97]

During the reproductive years, the typical cycle of follicle maturation, including ovulation and corpus luteum formation, is realized. This results from the complex but well-defined sequence of hypothalamic-pituitary-gonadal interactions in which follicle and corpus luteum steroid hormones, pituitary gonadotropins, and autocrine and paracrine factors are integrated to yield ovulation. These important events are described in detail in Chapters 5 and 6. For the moment, our attention is exclusively directed to a description of the events as the gonad is driven inexorably to final and complete exhaustion of its germ cell supply. The major feature of this reproductive period in the ovary's existence is the full maturational expression of some follicle units in ovulation and corpus luteum formation and the accompaniment of varying steroid output of estradiol and progesterone. For every follicle that ovulates, close to 1,000 will pursue abortive growth periods of variable length.

Follicular Growth

In the adult ovary, the stages of follicle development observed even in the prenatal period are repeated but to a more complete degree. Initially, the oocyte enlarges and the granulosa cells proliferate markedly. A solid sphere of cells encasing the oocyte is formed. At this point, the theca interna is noted in initial stages of formation. The zona pellucida begins to form.

The time that elapses in progressing from a primary follicle to ovulation is approximately 85 days.[98, 99] The majority of this time passes in development that is independent of gonadotropins, achieving a state of readiness that will yield further growth in response to FSH stimulation. If gonadotropin increments are available, as can be seen early in a menstrual cycle, a further FSH-dependent stage of follicle maturation is seen. The number of follicles that mature is dependent on the amount of FSH available to the gonad and the sensitivity of the follicles to the gonadotropins. FSH receptor expression is greatest in granulosa cells, but significant expression can be detected in ovarian surface epithelium and fallopian tube epithelium, where the function is uncertain, but a role in epithelium-derived tumors is possible.[100]

The antrum first appears as a coalescence of numerous intragranulosa cavities called Call–Exner bodies, which were described by Emma Call and Siegmund Exner in Vienna, in 1875. Emma Call was one of the first woman physicians in the United States.[101] After receiving her medical degree from the University of Michigan in 1873, she went to Vienna as Exner's postgraduate student. She returned to Boston and practiced as an obstetrician for more than 40 years. Emma Call was the first woman elected to the Massachusetts Medical Society (in 1884). Her description of the Call-Exner bodies was her only publication.

Whether Call–Exner bodies represent liquefaction or granulosa cell secretion is uncertain. At first, the cavity is filled with a coagulum of cellular debris. Soon a liquor accumulates, which is essentially a transudation of blood filtered through the avascular granulosa from the theca vessels. With antral formation, the theca interna develops more fully, expressed by increased cell mass, increased vascularity, and the formation of lipid-rich cytoplasmic vacuoles within the theca cells. As the follicle expands, the surrounding stroma is compressed and is called the theca externa.

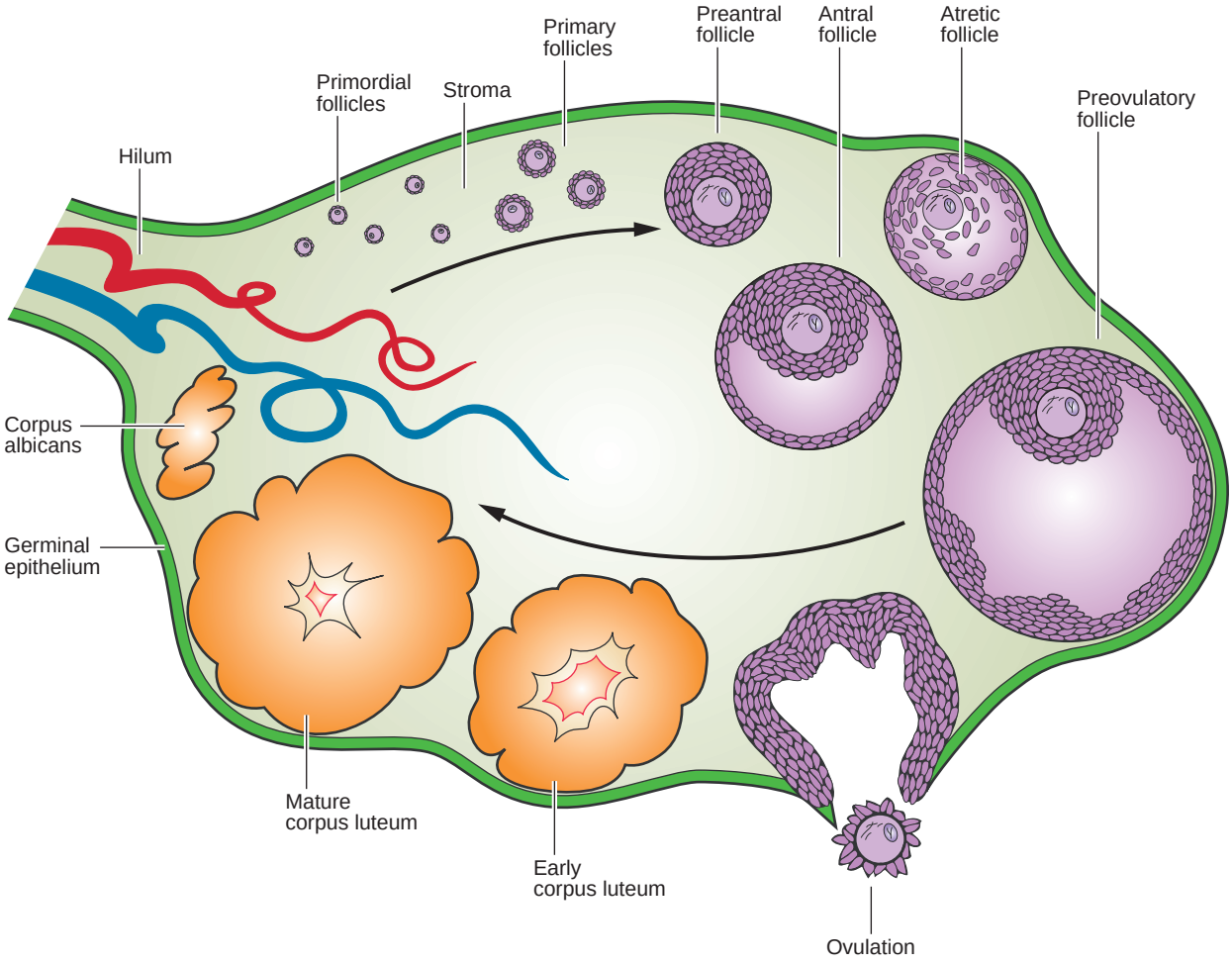
The granulosa cells that surround the oocyte are avascular and separated from the surrounding stroma by a basement membrane. Deprived of a vascular supply until after ovulation, the granulosa cells depend on specialized gap junctions that connect cells and communicate with the oocyte for the purpose of metabolic exchange and the transport of signaling molecules. It is this structure that allows repression and stimulation for the correct timing of meiosis. The granulosa cells differ in function and activity; e.g., LH receptor concentrations are highest in those cells closest to the basement membrane and lowest in those that surround the oocyte.[102]

At any point in this development, individual follicles become arrested and eventually regress in the apoptotic process known as atresia. At first the granulosa component

begins to disrupt. The antral cavity constituents are resorbed, and the cavity collapses and obliterates. The oocyte degenerates in situ. Finally, a ribbon-like scarred streak surrounded by theca is seen. Eventually this theca mass loses its lipid and becomes indistinguishable from the growing mass of stroma. Thus, the process of apoptosis is extensive in the granulosa, and the theca layer is largely spared to be incorporated into the interstitial tissue. Prior to regression, cystic follicles can be retained in the cortex for variable periods of time.

Ovulation

If gonadotropin stimulation is adequate, one of the several follicle units propelled to varying degrees of maturity will advance to ovulation. Morphologically, these events include distention of the antrum by increments of antral fluid and compression of the granulosa against the limiting membrane separating the avascular granulosa and the luteinized, vascularized theca interna. In addition, the antral fluid increment gradually pinches off the cumulus oophorous, the mound of granulosa enveloping the oocyte. The mechanisms of the thinning of the theca over the surface of the now protruding, distended follicle, the creation of an avascular area weakening the ovarian capsule, and the final acute distention of the antrum with rupture and extrusion of the oocyte in its cumulus are multiple and complex (discussed in Chapter 6). Repeated evaluation of intrafollicular pressures has failed to indict an explosive factor in this crucial event.



As demonstrated in a variety of animal experiments, the physical expulsion of the oocyte is dependent on a preovulatory surge in prostaglandin synthesis within the follicle. Inhibition of this prostaglandin synthesis produces a corpus luteum with an entrapped oocyte. Both prostaglandins and the midcycle surge of gonadotropins are thought to increase the concentration and activity of local proteases, such as plasminogen conversion to plasmin. As a result of generalized tissue weakening (loss of intercellular gap junction integrity and disruption of elastic fibers), there is swift accumulation of antral fluid followed by rupture of the weakened tissue envelope surrounding the follicle.

Corpus Luteum

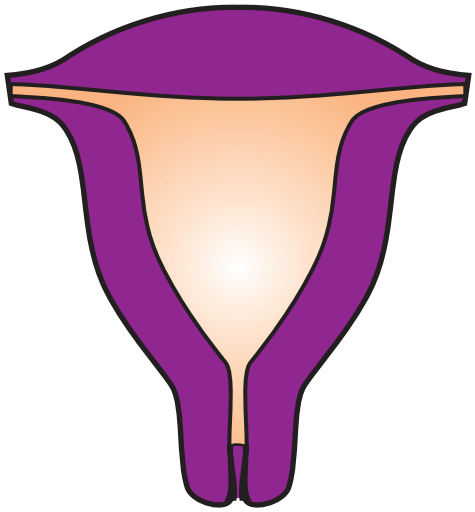
Shortly after ovulation, profound alterations in cellular organization occur in the ruptured follicle that go well beyond simple repair. After tissue integrity and continuity are retrieved, the granulosa cells hypertrophy markedly, gradually filling in the cystic, sometimes hemorrhagic, cavity of the early corpus luteum. In addition, for the first time, the granulosa becomes markedly luteinized by incorporation of lipid-rich vacuoles within its cytoplasm. Both these properties had been the exclusive features of the theca prior to ovulation. For its part, the theca of the corpus luteum becomes less prominent, vestiges being noted eventually only in the interstices of the typical scalloping of the mature corpus luteum. As a result, a new yellow body is formed, now dominated by the enlarged, lipid-rich, fully vascularized granulosa. In the 14 days of its life, dependent on the low but important quantities of LH available in the luteal phase, this unit produces estradiol and progesterone. Unless rescued by rising levels of human chorionic gonadotropin (hCG) from a successful implantation, the corpus luteum rapidly ages. Its vascularity and lipid content wane, and the sequence of scarification (albicantia) ensues.

Modulators of Function

The complex events that yield an ovum for fertilization and ovarian structures that provide hormonal secretion are the products of essentially every regulating mechanism in human biology. This includes classic endocrine signals, autocrine and paracrine/intracrine regulation, neuronal input, and immune system contributions. Representatives of the white blood cell series constitute a major component of the ovarian stromal (interstitial) compartment. Macrophages present in permanent, noncyclic numbers may influence ovarian function through the secretion of regulatory cytokines.[103] During the adult ovarian cycle, there is an infiltration of white blood cells in a pattern characterized by increasing numbers of mast cells culminating in degranulation and release of histamine that is associated with hyperemia at ovulation.[104] The corpus luteum attracts eosinophils and T lymphocytes, which signal and activate monocytes and macrophages involved in luteolysis. However, this immune mechanism should be viewed not just as a healing and resolving response, but also as an important regulatory system (involving the secretion of cytokines and growth factors) for ovarian function.[103]

All references are available online at: http://www.clinicalgynendoandinfertility.com

4 The Uterus



Anatomic knowledge of the uterus was slow to accumulate.[1, 2] Papyrus writings from 2500 B.C. indicate that the ancient Egyptians made a distinction between the vagina and uterus. Because the dead had to be embalmed, dissection was precluded, but prolapse was recognized because it was important to return the uterus into its proper place prior to mummification. Next to the Egyptian papyri in antiquity were Hindu writings in which descriptions of the uterus, tubes, and vagina indicate knowledge gained from dissections. This was probably the earliest description of the fallopian tubes.

There is little information in Greek writings about female anatomy; however, Herophilus (fourth century B.C.), the great anatomist in Alexandria and the originator of scholarly dissection, recorded the different positions of the uterus. Soranus of Ephesus (98–138 A.D.) accurately described the uterus (probably the first to do so), obviously from multiple dissections of cadavers. He recognized that the uterus is not essential for life, acknowledged the presence of leiomyomas, and treated prolapse with pessaries.

Herophilus and Soranus were uncertain about the function of the fallopian tubes, but Galen, Rufus, and Aetisu correctly guessed their function. Galen promoted the practice of bleeding for the treatment of almost every disorder. In his argument that nature prevented disease by discharging excess blood, Galen maintained that women were healthier because their superfluous blood was eliminated by menstruation.[3] The writings of Galen (130–200 A.D.) represented the knowledge of medicine for over 1,000 years until the end of the medieval Dark Ages. Galen's description of the uterus and tubes indicates that he had only seen the horned uteri of animals.

In the 16th century, Berengarius, Vesalius, Eustachius, and Fallopius made significant contributions to the anatomic study of the female genitalia. Berengarius (Giacomo Berengario da Carpi) was the first anatomist to work with an artist. His anatomic text, published in 1514, depicted dissected subjects as if they were still alive.

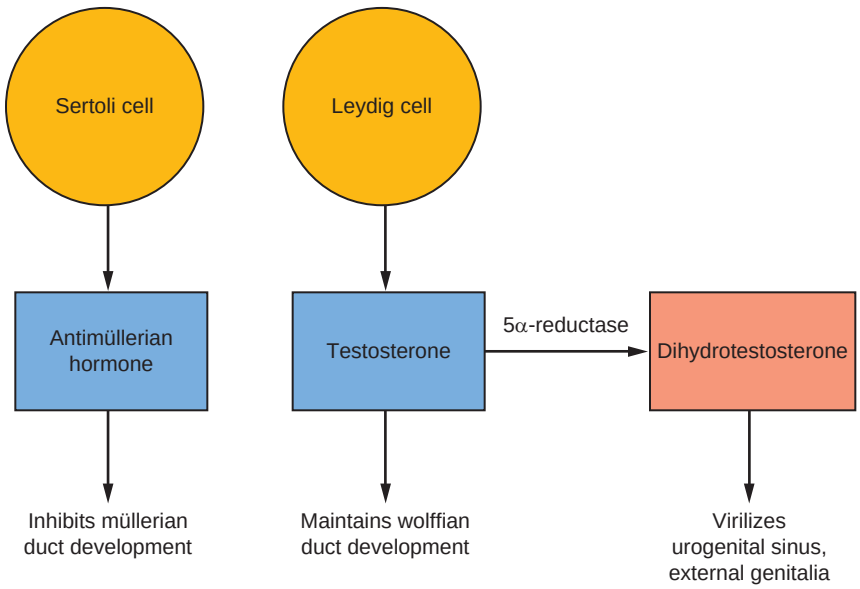
Gabriele Fallopio (or Fallopius) published his work, *Observationes Anatomicae,* in Venice in 1561, 1 year before his death from pleurisy at age 40. He provided the first descriptions of the clitoris and the hymen and the first exact descriptions of the ovaries and the tubes. He named the vagina and the placenta and called the tubes the uteri tuba (the trumpet of the uterus), but soon they were known universally as the fallopian tubes. It was his professor and mentor at the University of Padua, however, Andreas Vesalius, who was the first to accurately reveal the presence of the endometrial cavity.

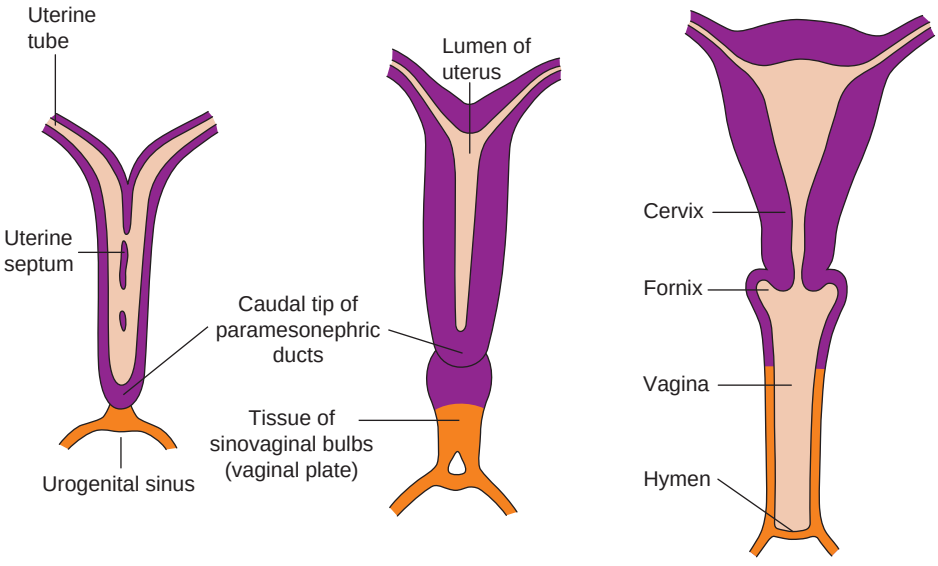
Development of the Müllerian System

The wolffian (mesonephric) and müllerian (paramesonephric) ducts are discrete primordia that temporarily coexist in all embryos during the ambisexual period of development (up to 8 weeks). Thereafter, one type of duct system persists normally and gives rise to special ducts and glands, whereas the other disappears during the third fetal month, except for nonfunctional vestiges.

Hormonal control of mammalian somatic sex differentiation was established by the classic experiments of Alfred Jost.[4] In Jost's landmark studies, the active role of male-determining factors, as opposed to the constitutive nature of female differentiation, was defined as the directing feature of sex differentiation. This principle applies not only to the internal ducts but to the gonad, external genitalia, and even the brain. The critical factors in determining which of the duct structures stabilize or regress are the secretions from the testes: AMH (antimüllerian hormone, also known as müllerian-inhibiting substance or müllerian-inhibiting factor) and testosterone.

AMH is a member of the transforming growth factor-β family of glycoprotein differentiation factors that include inhibin and activin. The gene for AMH has been mapped to chromosome 19. AMH is synthesized by Sertoli cells soon after testicular differentiation and is responsible for the ipsilateral regression of the müllerian ducts by 8 weeks. Despite its presence in serum up to puberty, lack of regression of the uterus and tubes is the only consistent expression of AMH gene mutations. In the absence of AMH, the fetus will develop fallopian tubes, uterus, and upper vagina from the paramesonephric ducts (the müllerian ducts). This development requires the prior appearance of the mesonephric ducts, and for this reason,





After Sadler [6]

abnormalities in development of the tubes, uterus, and upper vagina are associated with abnormalities in the renal system.

The internal genitalia possess the intrinsic tendency to feminize. In the absence of a Y chromosome and a functional testis, the lack of AMH allows retention of the müllerian system and development of fallopian tubes, uterus, and upper vagina. In the absence of testosterone, the wolffian system regresses. In the presence of a normal ovary or the absence of any gonad, müllerian duct development takes place. This process is discussed in greater detail in Chapter 9.

The paramesonephric ducts come into contact in the midline to form a Y-shaped structure, the primordium for the uterus, tubes, and the upper one-third of the vagina.[5] The fallopian tubes, uterus, and the upper portion of the vagina are created by the fusion of the müllerian ducts by the tenth week of gestation. Canalization to create the uterine cavity, the cervical canal, and the vagina is complete by the 22nd week of gestation. Under the epithelium lies mesenchymal tissue that will be the origin of the uterine stroma and smooth muscle cells. By the 20th week of pregnancy, the uterine mucosa is fully differentiated into the endometrium.

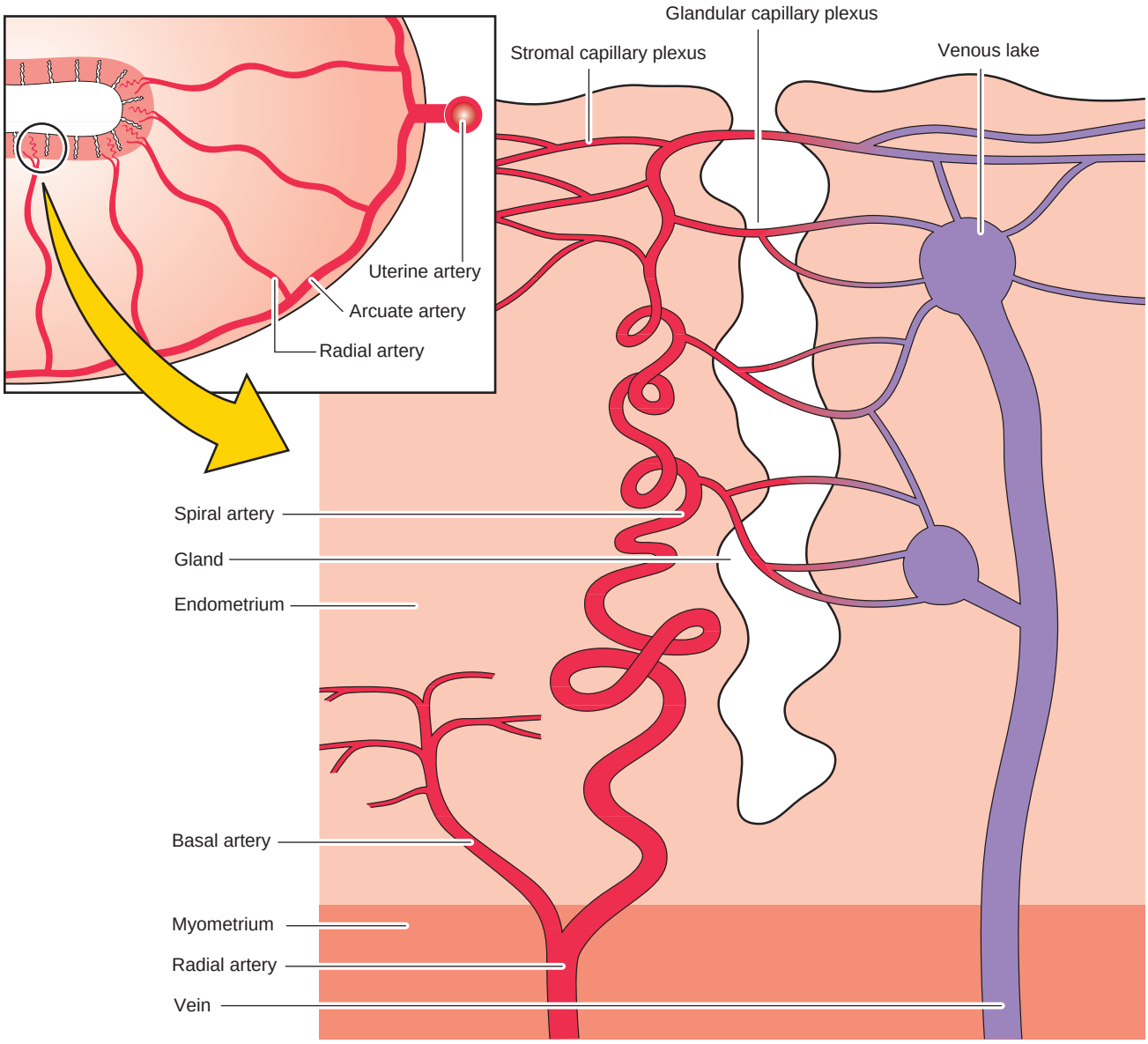
The endometrium, derived from the mucosal lining of the fused müllerian ducts, is essential for reproduction and may be one of the most complex tissues in the human body. It is always changing, responding to the cyclic patterns of estrogen and progesterone of the ovarian menstrual cycle and to a complex interplay among its own autocrine and paracrine factors.

The Histologic Changes in Endometrium During an Ovulatory Cycle

The sequence of endometrial changes associated with an ovulatory cycle has been carefully studied by Noyes in the human and Bartlemez and Markee in the subhuman primate.[7–11]

From these data a description of menstrual physiology has been developed based on specific anatomic and functional changes within glandular, vascular, and stromal components of the endometrium.[12–14] These changes will be discussed in five phases: (1) the menstrual endometrium, (2) the proliferative phase, (3) the secretory phase, (4) preparation for implantation, and finally, (5) the phase of endometrial breakdown. Although these distinctions are not entirely arbitrary, it must be recalled that the entire process is an integrated evolutionary cycle of endometrial growth and regression, which is repeated some 400 times during the adult life of the human female.

The endometrium can be divided morphologically into an upper two-thirds "functionalis" layer and a lower one-third "basalis" layer. The purpose of the functionalis layer is to prepare for the implantation of the blastocyst; therefore, it is the site of proliferation, secretion, and degeneration. The purpose of the basalis layer is to provide the regenerative endometrium following menstrual loss of the functionalis.[15]



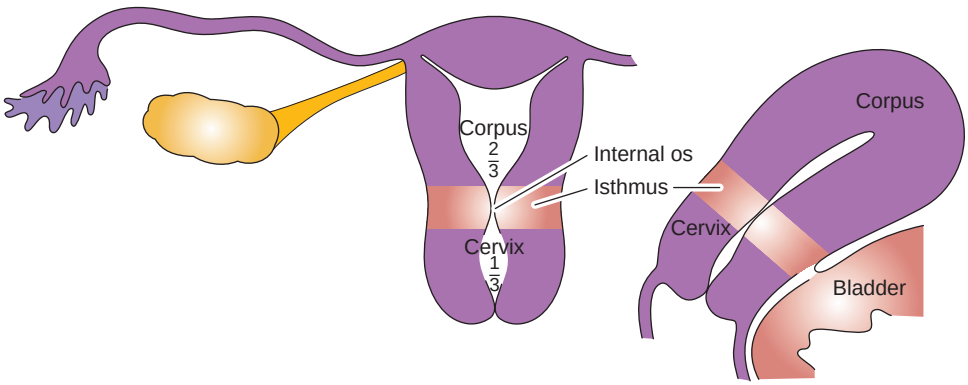
The Uterine Vasculature

The two uterine arteries that supply the uterus are branches of the internal iliac arteries. At the lower part of the uterus, the uterine artery separates into the vaginal artery and an ascending branch that divides into the arcuate arteries. The arcuate arteries run parallel to the uterine cavity and anastomose with each other, forming a vascular ring around the cavity. Small centrifugal branches (the radial arteries) leave the arcuate vessels, perpendicular to the endometrial cavity, to supply the myometrium. When these arteries enter the endometrium, small branches (the basal arteries) extend laterally to supply the basalis layer. These basal arteries do not demonstrate a response to hormonal changes. The radial arteries continue in the direction of the endometrial surface, now assuming a corkscrew appearance (and now called the spiral arteries), to supply the functionalis layer of the endometrium. It is the spiral artery (an end artery) segment that is very sensitive to hormonal changes. One reason that the functionalis layer is more vulnerable to vascular ischemia is that there are no anastomoses among the spiral arteries. The endometrial glands and the stromal tissue are supplied by capillaries that emerge from the spiral arteries at all levels of the endometrium. The capillaries drain into a venous plexus and eventually into the myometrial arcuate veins and into the uterine veins. This unique vascular architecture is important in allowing a repeated sequence of endometrial growth and desquamation.

The Menstrual Endometrium

The menstrual endometrium is a relatively thin but dense tissue. It is composed of the stable, nonfunctioning basalis component and a variable, but small, amount of residual stratum spongiosum. At menstruation, this latter tissue displays a variety of functional states including disarray and breakage of glands, fragmentation of vessels and stroma with persisting evidence of necrosis, white cell infiltration, and red cell interstitial diapedesis. Even as the remnants of menstrual shedding dominate the overall appearance of this tissue, evidence of repair in all tissue components can be detected. Endometrial regeneration originates in epithelial and stromal stem cells.[16] Endometrial epithelial stem cells are located in the base of the endometrial glands and stromal stem cells around blood vessels in the basalis layer.

The menstrual endometrium is a transitional state bridging the more dramatic proliferative and exfoliative phases of the cycle. Its density implies that the shortness of height is not entirely due to desquamation. Collapse of the supporting matrix also contributes significantly to the shallowness. Reticular stains in Rhesus endometrium confirm this



"deflated" state. Nevertheless, as much as two-thirds of the functioning endometrium is lost during menstruation. The more rapid the tissue loss, the shorter the duration of flow. Delayed or incomplete shedding is associated with heavier flow and greater blood loss.

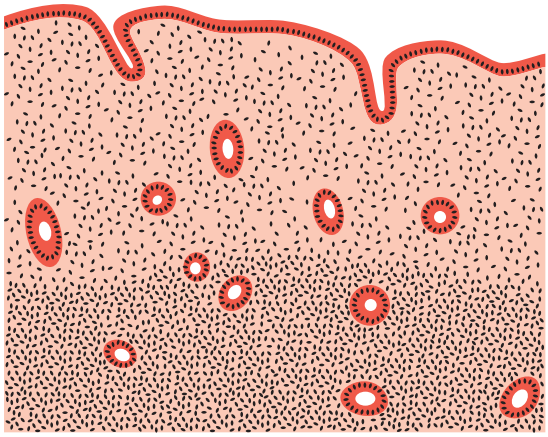
DNA synthesis is occurring in those areas of the basalis that have been completely denuded by day 2–3 of the menstrual cycle (the endometrium in the isthmic area, the narrow area between the cervix and the corpus, and the endometrium in the cornual recesses at the ostia of the tubes remain intact). The new surface epithelium emanates from the flanks of stumps of glands in the basalis layer left standing after menstrual desquamation.[17] Rapid re-epithelialization follows the proliferation of the cells in the basalis layer and the surface epithelium in the isthmic and tubal ostial endometrium. This epithelial repair is supported by underlying fibroblasts. The stromal fibroblast layer forms a compact mass over which the resurfacing epithelium can "migrate." In addition, it is likely that the stromal layer contributes important autocrine and paracrine factors for growth and migration. Because hormone levels are at their nadir during this repair phase, the response may be due to injury rather than hormone mediated. However, the basalis layer is rich in its content of estrogen receptors. This "repair" is fast; by day 4 of the cycle, more than two-thirds of the cavity is covered with new epithelium.[17] By day 5–6, the entire cavity is re-epithelialized, and stromal growth begins.

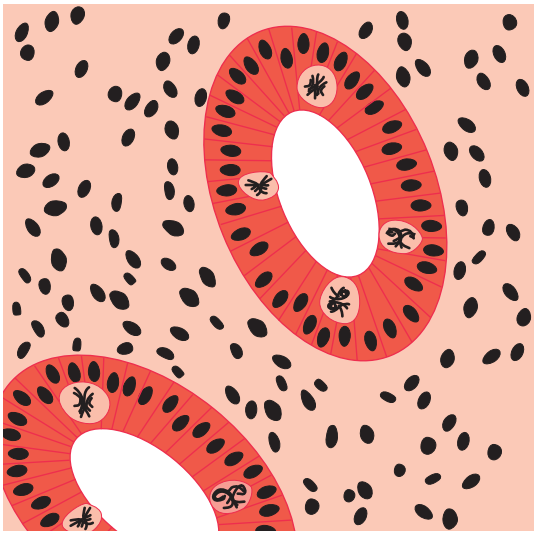
The Proliferative Phase

The proliferative phase is associated with ovarian follicle growth and increased estrogen secretion. Undoubtedly as a result of this steroidal action, reconstruction and growth of the endometrium are achieved. The glands are most notable in this response. At first they are narrow and tubular, lined by low columnar epithelium cells. Mitoses become prominent and pseudostratification is observed. As a result, the glandular epithelium extends peripherally and links one gland segment with its immediate neighbor. A continuous epithelial lining facing the endometrial cavity is formed. The stromal component evolves from its dense cellular menstrual condition through a brief period of edema to a final loose syncytial-like status. Coursing through the stroma, spiral vessels extend (unbranched and uncoiled in the early proliferative phase) to a point immediately below the epithelial binding membrane. Here they form a loose capillary network. All of the tissue components (glands, stromal cells, and endothelial cells) demonstrate proliferation, which peaks on days 8–10 of the cycle, reflecting rising estradiol levels in the circulation and maximal estrogen receptor concentration in the endometrium.[18] This proliferation is marked by increased mitotic activity and increased nuclear DNA and cytoplasmic RNA synthesis, which is most intense in the functionalis layer in the upper two-thirds of the uterus, the usual site of blastocyst implantation.

During proliferation, the endometrium grows from approximately 0.5 mm to 3.5–5.0 mm in height of a singular layer. Restoration of tissue constituents has been achieved by estrogen-induced new growth as well as incorporation of ions, water, and amino acids. The stromal ground substance has re-expanded from its menstrual collapse. Although true tissue growth has occurred, a major element in achievement of endometrial height is "reinflation" of the stroma.

An important feature of this estrogen-dominant phase of endometrial growth is the increase in ciliated and microvillous cells. Ciliogenesis begins on days 7–8 of the cycle.[17] This response to estrogen is exaggerated in hyperplastic endometrium that is the result of hyperestrogenism. The concentration of these ciliated cells around gland openings and the ciliary beat pattern influence the mobilization and distribution of endometrial secretions during

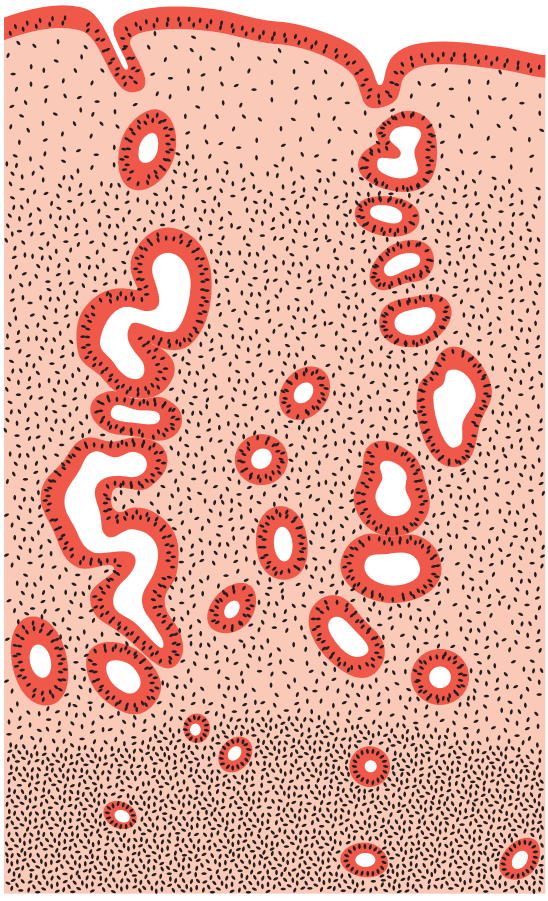


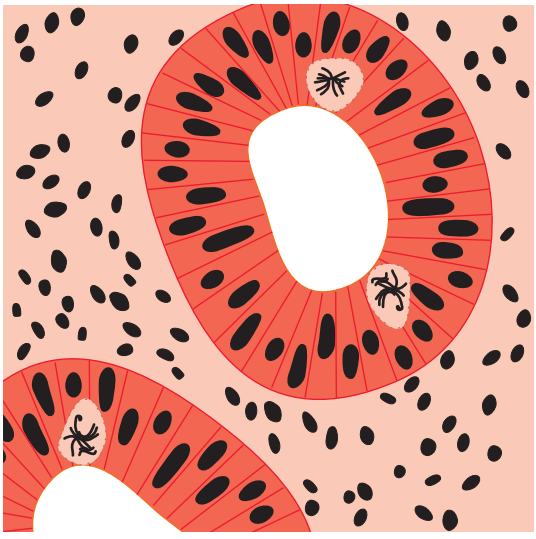


Early Proliferative

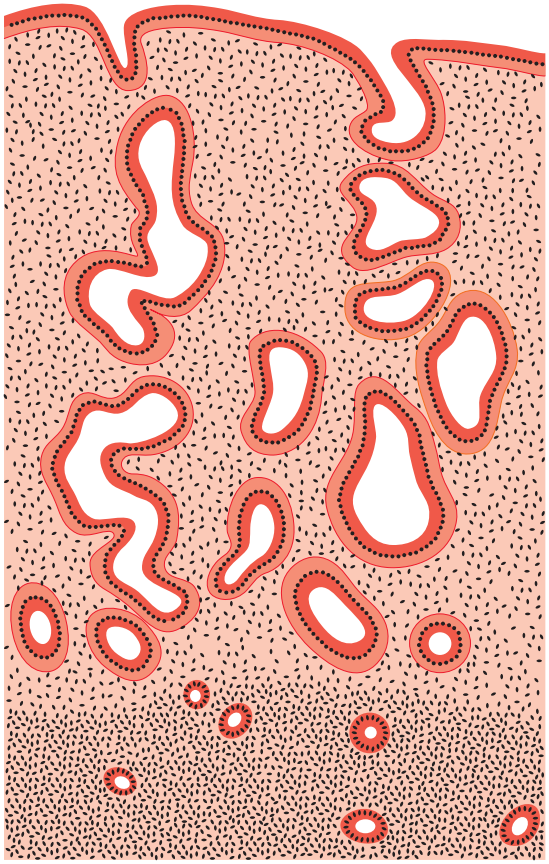
the secretory phase. Cell surface microvilli, also a response to estradiol, are cytoplasmic extensions and serve to increase the active surface of cells.

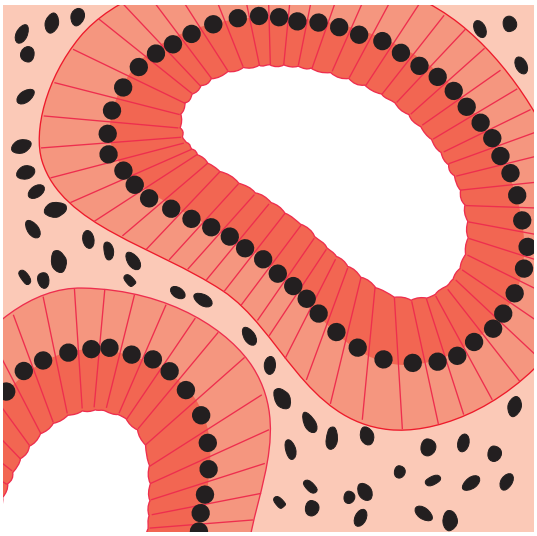
At all times, a large number of cells derived from bone marrow are present in the endometrium. These include lymphocytes and macrophages, diffusely distributed in the stroma.





Late Proliferative





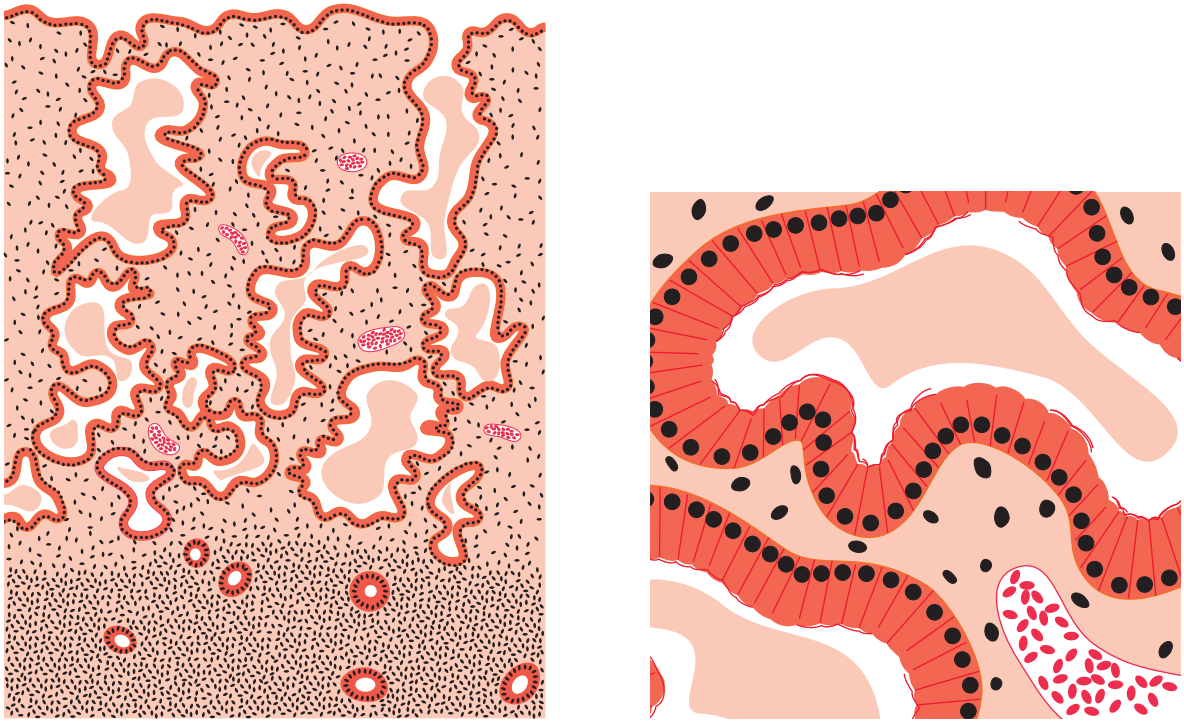
Early Secretory

The Secretory Phase

After ovulation, the endometrium now demonstrates a combined reaction to estrogen and progesterone activity. Most impressive is that total endometrial height is fixed at roughly its preovulatory extent (5–6 mm) despite continued availability of estrogen. Epithelial proliferation ceases 3 days after ovulation.[19] This restraint or inhibition is believed to be induced by progesterone. This limitation of growth is associated with a decline in mitosis and DNA synthesis, significantly due to progesterone interference with estrogen receptor expression and progesterone stimulation of 17β-hydroxysteroid dehydrogenase and sulfotransferase, which convert estradiol to estrone sulfate (which is rapidly excreted from the cell).[20, 21] In addition, estrogen stimulates many oncogenes that probably mediate estrogen-induced growth. Progesterone antagonizes this action by suppressing the estrogen-mediated transcription of oncogene mRNA.[22]

Individual components of the tissue continue to display growth, but confinement in a fixed structure leads to progressive tortuosity of glands and intensified coiling of the spiral vessels. The secretory events within the glandular cells, with progression of vacuoles from intracellular to intraluminal appearance, are well known and take place over a 7-day postovulatory interval. At the conclusion of these events, the glands appear exhausted, the tortuous lumina variably distended, and individual cell surfaces fragmented in a sawtooth appearance. Stroma is increasingly edematous, and spiral vessels are prominent and densely coiled.

The first histologic sign that ovulation has occurred is the appearance of subnuclear intracytoplasmic glycogen vacuoles in the glandular epithelium on cycle days 17–18. Giant



Late Secretory

mitochondria and the "nucleolar channel system" appear in the gland cells. The nucleolar channel system has a unique appearance due to progesterone, an infolding of the nuclear membranes. Individual components of the tissue continue to display growth, but confinement in a fixed structure leads to progressive tortuosity of glands and intensified coiling of the spiral vessels. These structural alterations are soon followed by active secretion of glycoproteins and peptides into the endometrial cavity. Transudation of plasma also contributes to the endometrial secretions. Important immunoglobulins are obtained from the circulation and delivered to the endometrial cavity by binding proteins produced by the epithelial cells. The peak secretory level is reached 7 days after the midcycle gonadotropin surge, coinciding with the time of blastocyst implantation.

The Implantation Phase

Significant changes occur within the endometrium from the 7th to the 13th day postovulation (days 21–27 of the cycle). At the onset of this period, the distended tortuous secretory glands have been most prominent with little intervening stroma. By 13 days postovulation, the endometrium has differentiated into three distinct zones. Something less than one-fourth of the tissue is the unchanged basalis fed by its straight vessels and surrounded by indifferent spindle-shaped stroma. The midportion of the endometrium (approximately 50% of the total) is the lace-like ***stratum spongiosum***, composed of loose edematous stroma with tightly coiled but ubiquitous spiral vessels and exhausted dilated glandular ribbons. Overlying the spongiosum is the superficial layer of the endometrium (about 25% of the height) called the ***stratum compactum***. Here the prominent histologic feature is the stromal cell, which has become large and polyhedral. In its cytoplasmic expansion one cell abuts the other, forming a compact, structurally sturdy layer. The necks of the glands traversing this segment are compressed and less prominent. The subepithelial capillaries and spiral vessels are engorged.

At the time of implantation, on days 21–22 of the cycle, the predominant morphologic feature is edema of the endometrial stroma. This change may be secondary to the estrogen- and progesterone-mediated increase in prostaglandin and vascular endothelial growth factor (VEGF) production by the endometrium that cause an increase in capillary permeability. Receptors for the sex steroids are present in the muscular walls of the endometrial blood vessels, and the enzyme system for prostaglandin synthesis is present in both the muscular walls and the endothelium of the endometrial arterioles. Mitoses are first seen in endothelial cells on cycle day 22. Vascular proliferation leads to the coiling of the spiral vessels, a response to the sex steroids, the prostaglandins, and the autocrine and paracrine factors produced in response to estrogen and progesterone.

During the secretory phase, so-called K (Körnchenzellen) cells appear, reaching a peak concentration in the first trimester of pregnancy. These are granulocytes that have an immunoprotective role in implantation and placentation. They are located perivascularly and are believed to be derived from the blood. By day 26–27, the endometrial stroma is infiltrated by extravasated polymorphonuclear leukocytes. The majority of the leukocytes are killer cells and macrophages, believed to be involved in the process of endometrial breakdown and menstruation. The appearance and function of these cells are regulated by the complex array of peptides and cytokines in the endometrium in response to hormonal signaling.

The gene expression pattern in the endometrium throughout the menstrual cycle is being established, with a focus on the implantation window.[23–25] As expected, microarray analyses reveal a changing pattern of gene expression that correlates with each hormonal and morphological stage in the endometrial menstrual cycle.[26] Ultimately this will yield a comprehensive picture, with the gene signature of each event in the estrogen and progesterone regulation of the endometrium. The regulating growth factors, cytokines, and peptide hormones that are essential for implantation will be identified.

The stromal cells of the endometrium respond to hormonal signals, synthesize prostaglandins, and, when transformed into decidual cells, produce an impressive array of substances, some of which are prolactin, relaxin, renin, insulin-like growth factors (IGFs), and insulin-like growth factor binding proteins (IGFBPs). The endometrial stromal cells, the progenitors of decidual cells, were originally believed to be derived from the bone marrow (from cells invading the endometrium), but they are now considered to emanate from the primitive uterine mesenchymal stem cells.[27]

The decidualization process begins in the luteal phase under the influence of progesterone and mediated by autocrine and paracrine factors. On cycle days 22–23, predecidual cells can be identified, initially surrounding blood vessels, characterized by cytonuclear enlargement, increased mitotic activity, and the formation of a basement membrane. The decidua, derived from stromal cells, becomes an important structural and biochemical tissue of pregnancy. Decidual cells control the invasive nature of the trophoblast, and the products of the decidua play important autocrine and paracrine roles in fetal and maternal tissues.

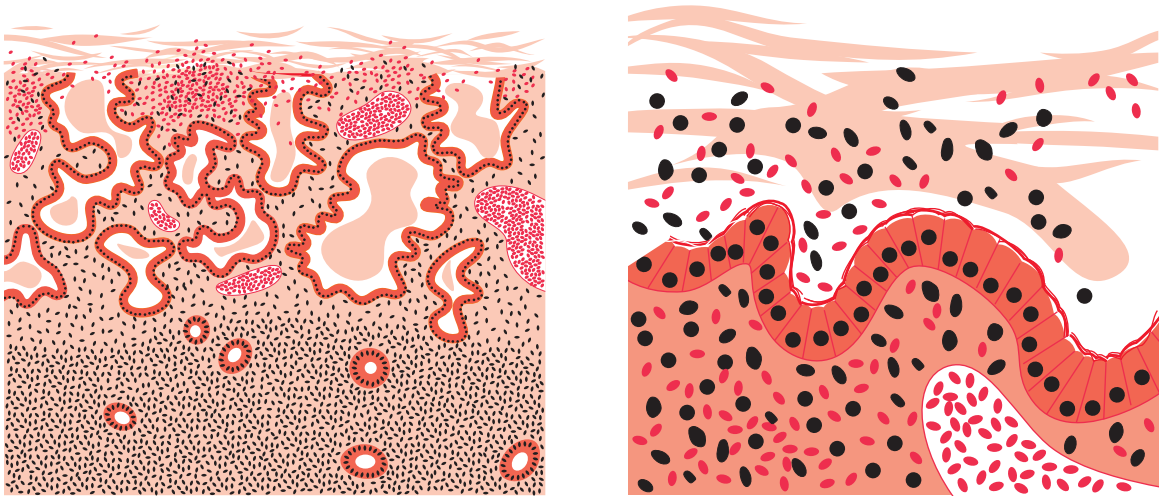
Lockwood and his colleagues assign a key role to decidual cells in both the process of endometrial bleeding (menstruation) and the process of endometrial hemostasis (implantation and placentation).[28–30] Implantation requires endometrial hemostasis and the maternal uterus requires resistance to invasion. Inhibition of endometrial hemorrhage can be attributed, to a significant degree, to appropriate changes in critical factors as a consequence of decidualization; e.g., lower plasminogen activator levels, reduced expression of the enzymes that degrade the stromal extracellular matrix (such as the metalloproteinases), and increased levels of plasminogen activator inhibitor-1. Withdrawal of estrogen and progesterone support, however, leads to changes in the opposite directions, consistent with endometrial breakdown.

The Phase of Endometrial Breakdown

Predecidual transformation has formed the "compacta" layer in the upper part of the functionalis layer by day 25 (3 days before menstruation). In the absence of fertilization, implantation, and the consequent lack of sustaining quantities of human chorionic gonadotropin from the trophoblast, the otherwise fixed lifespan of the corpus luteum is completed, and estrogen and progesterone levels wane.

The withdrawal of estrogen and progesterone initiates important endometrial events: vasomotor reactions, the process of apoptosis, tissue loss, and, finally, menstruation. The most prominent immediate effect of this hormone withdrawal is a modest shrinking of the tissue height and remarkable spiral arteriole vasomotor responses. The classic concept of the vascular sequence was constructed from direct observations of Rhesus endometrium transplanted to the anterior chamber of the eye.[7, 8] With shrinkage of height, blood flow within the spiral vessels diminished, venous drainage was decreased, and vasodilation ensued. Thereafter, the spiral arterioles underwent rhythmic vasoconstriction and relaxation. Each successive spasm was more prolonged and profound, leading eventually to endometrial blanching. Thus these reactions were proposed to lead to menstruation because of endometrial ischemia and stasis caused by vasoconstriction of the spiral arterioles. A new model of menstruation, as discussed in Chapter 15, emphasizes enzymatic autodigestion of the functional layer of the endometrium and its capillary plexus.

In the first half of the secretory phase, acid phosphatase and potent lytic enzymes are confined to lysosomes. Their release is inhibited by progesterone stabilization of the lysosomal membranes. With the waning of estrogen and progesterone levels, the lysosomal membranes are not maintained, and the enzymes are released into the cytoplasm of epithelial, stromal, and endothelial cells and eventually into the intercellular space. These active enzymes will digest their cellular constraints, leading to the release of prostaglandins, extravasation of red blood cells, tissue necrosis, and vascular thrombosis. This process is one of ***apoptosis***, (programmed cell death, characterized by a specific morphologic pattern that involves cell shrinkage and chromatin condensation culminating in cell fragmentation) mediated by cytokines.[31] An important step in this breakdown is the dissolution of cell-to-cell adhesion by key proteins. Binding of endometrial epithelial cells utilizes transmembrane proteins, ***cadherins***, that link intercellularly with each other and intracellularly with catenins that are bound to actin filaments.[32]



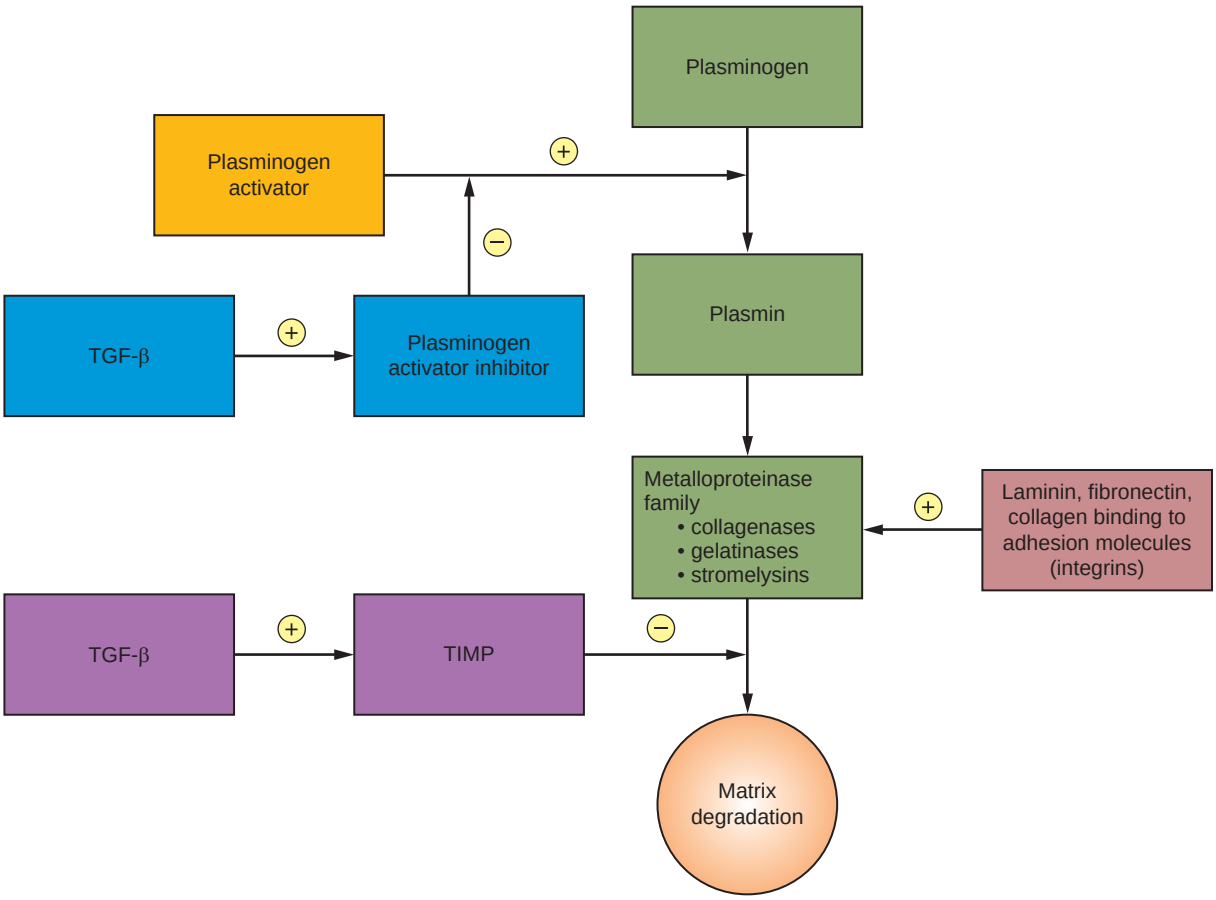
Menstruation

Endometrial tissue breakdown also involves a family of enzymes, matrix metalloproteinases, that degrade components (including collagens, gelatins, fibronectin, and laminin) of the extracellular matrix and basement membrane.[33, 34] The metalloproteinases include collagenases that degrade interstitial and basement membrane collagens; gelatinases that further degrade collagens; and stromelysins that degrade fibronectin, laminin, and glycoproteins. The expression of metalloproteinases in human endometrial stromal cells follows a pattern correlated with the menstrual cycle, indicating a sex steroid response as part of the growth and remodeling of the endometrium with a marked increase in late secretory and early menstrual endometrium.[35] Progesterone withdrawal from endometrial cells increases VEGF production and induces matrix metalloproteinase secretion, probably from both endometrial stromal cells and leukocytes, which is followed by the irreversible breakdown of cellular membranes and the dissolution of extracellular matrix.[36–38] Appropriately, this enzyme expression increases in the decidualized endometrium of the late secretory phase, during the time of declining progesterone levels. With the continuing progesterone secretion of early pregnancy, the decidua is maintained and metalloproteinase expression is suppressed, in a mechanism mediated by transforming growth factor-beta (TGF-β).[39] In a nonpregnant cycle, metalloproteinase expression is suppressed after menses, presumably by increasing estrogen levels.

Metalloproteinase activity is restrained by specific tissue inhibitors designated as TIMP.[40] The balance of metalloproteinase and TIMP activity is an important event in successful implantation. Thus, progesterone withdrawal can lead to endometrial breakdown through a mechanism that is independent of vascular events (specifically ischemia), a mechanism that involves cytokines.[31] During bleeding, both normal and abnormal, there is evidence indicating that specific genes are activated in the endometrium; one such gene has the structural features of the TGF-β family.[41]

There is considerable evidence to support a major role for a cytokine, tumor necrosis factor-α (TNF-α), in menstruation.[31] TNF-α is a transmembrane protein whose receptor belongs to the nerve growth factor/TNF family for inducing apoptotic signals. The key change is an increase in secretion because TNF-α secretion by endometrial cells reaches a peak at menstruation, but there is no cycle change in receptor content. TNF-α inhibits endometrial proliferation and induces apoptosis; this cytokine causes a loss of adhesion proteins (the cadherin-catenin-actin complex) and induces cell-to-cell dissolution. In addition to endometrial cells, TNF-α also causes damage to vascular endothelium.

Progesterone withdrawal is also associated with an increase in vascular endothelial growth factor receptor concentrations in the stromal cells of the layers of endometrium destined to be sloughed.[42] Although the vascular endothelial growth factor system is usually involved with angiogenesis, in this case these factors are involved in the preparation for menstrual bleeding, perhaps influencing the expression of matrix metalloproteinases (MMPs). Endometrial genes without classic steroid response elements can respond to the sex steroids by utilizing a family of proteins (the Sp family) that mediate steroid activity at the level of transcription (acting in a fashion similar to the steroid receptors). These proteins, induced by progesterone in stromal (decidual) and epithelial cells, can activate tissue factor, plasminogen activator inhibitor-1, IGF binding protein-1, uteroglobin, and uteroferrin. Tissue factor is involved in the clotting mechanism to sustain hemostasis. Uteroglobin is a small protein expressed in endometrial epithelial cells.[43] The physiologic function of uteroglobin is uncertain. Uteroglobin, with high affinity, binds progestins and may play a role in immunosuppression. Uteroglobin gene expression is stimulated by estrogen, and this response is enhanced by progesterone. Human endometrium can secrete β-endorphin, yet another candidate for involvement in endometrial immunologic events, and its release is inhibited by both estrogens and glucocorticoids.[44]



Eventually, considerable leakage occurs as a result of diapedesis, and finally, interstitial hemorrhage occurs due to breaks in superficial arterioles and capillaries. White cells migrate through capillary walls, at first remaining adjacent to vessels but then extending

throughout the stroma. The leukocytes add important regulatory cytokines, chemokines, and enzymes that are involved in the degradation of the extracellular matrix. During arteriolar vasomotor changes, red blood cells escape into the interstitial space. Thrombin-platelet plugs also appear in superficial vessels. The prostaglandin content ($PGF_{2\alpha}$ and PGE_2) in the secretory endometrium reaches its highest levels at the time of menstruation. The vasoconstriction and myometrial contractions associated with the menstrual events are believed to be significantly mediated by prostaglandins from perivascular cells and the potent vasoconstrictor endothelin-1, derived from stromal decidual cells.

As ischemia and weakening progress, the continuous binding membrane is fragmented, and intercellular blood is extruded into the endometrial cavity. New thrombin-platelet plugs form intravascularly upstream at the shedding surface, limiting blood loss. Increased blood loss is a consequence of reduced platelet numbers and inadequate hemostatic plug formation. Menstrual bleeding is influenced by activation of clotting and fibrinolysis. Fibrinolysis is principally the consequence of the potent enzyme plasmin, formed from its inactive precursor plasminogen. Endometrial stromal cell tissue factor (TF) and plasminogen activators and inhibitors are involved in achieving a balance in this process. TF stimulates coagulation, initially binding to factor VII. TF and plasminogen activator inhibitor-1 (PAI-1) expression accompanies decidualization, and the levels of these factors may govern

the amount of bleeding.[30, 45] PAI-1, in particular, exerts an important restraining action on fibrinolysis and proteolytic activity.[46] Blood loss is also controlled by constriction of the spiral arteries, mediated by the perivascular cells, myofibroblasts that surround the spiral arteries.[47] These cells respond to progesterone withdrawal by expressing prostaglandins, cytokines, and MMPs, causing not only cycling vasoconstriction and vasodilation but also modulating leukocyte entry (an important additional source of metalloproteinases) into the endometrium. Disordered growth and function of the perivascular cells are likely contributing factors in menstrual bleeding problems.

High Progesterone Levels

↓

Perivascular Growth and Decidualization

↓

Suppression of Prostaglandin, Cytokine, and MMP Expression

Progesterone Withdrawal

↓

Prostaglandin, Cytokine, and VEGF Expression

↓

Vasoconstriction, Vasodilation, Leukocyte Infiltration, and Increase in MMPs

With progressive enzymatic degradation of the endometrium, the subsurface capillary and venous vascular system is disrupted, causing hemorrhage and escape of blood into the endometrial cavity. Additional ischemic breakdown ensues with necrosis of cells and defects in vessels adding to the menstrual effluvium. Degeneration extends to the deepest extent of the functional layer where rupture of the basal arterioles contributes to bleeding. A natural cleavage point exists between basalis and spongiosum, and, once breached, the loose, vascular, edematous stroma of the spongiosum desquamates and collapses. The process is initiated in the fundus and inexorably extends throughout the uterus. In the end, the typical deflated, shallow, dense, menstrual endometrium results. Within 13 hours, the endometrial height shrinks from 4 to 1.25 mm.[13] Menstrual flow stops as a result of the combined effects of prolonged vasoconstriction of the radial arteries and the spiral arteries in the basalis, tissue collapse, vascular stasis, and estrogen-induced "healing." In contrast to postpartum bleeding, myometrial contractions are not important for control of menstrual bleeding. Thrombin generation in the basal endometrium in response to extravasation of blood is essential for hemostasis. Thrombin promotes the generation of fibrin, the activation of platelets and clotting cofactors, and angiogenesis.

The basalis endometrium remains during menses, and repair takes place from this layer. This endometrium is protected from the lytic enzymes in the menstrual fluid by a mucinous layer of carbohydrate products that are discharged from the glandular and stromal cells.[48]

Normal Menses

Approximately 50% of the menstrual detritus is expelled in the first 24 hours of menstrual flow. The menstrual fluid is composed of the autolysed functionalis, inflammatory exudate, red blood cells, and proteolytic enzymes (at least one of which, plasmin, lyses fibrin clots as they form). The high fibrinolytic activity advances emptying of the uterus by liquefaction of tissue and fibrin. If the rate of flow is great, clotting can and does occur.

Most women (90%) have menstrual cycles with an interval of 24 to 35 days (Chapter 6).[49, 50] Menarche is followed by approximately 5–7 years of increasing regularity as cycles shorten to reach the usual reproductive-age pattern. In the 40s, cycles begin to lengthen again. The usual duration of flow is 4–6 days, but many women flow as little as 2 days and as much as 8 days. The normal volume of menstrual blood loss is 30 mL; greater than 80 mL is abnormal. Normal and abnormal characteristics and definitions of menstrual flow are discussed in detail in Chapter 15.

A Teleologic Theory of Endometrial-Menstrual Events

Menstruation is a very recent phenomenon in the evolutionary time line. It occurs in very few species, even among viviparous animals. An unabashedly teleologic view of menstrual events was offered many years ago by Rock et al.[51] The basic premise of this thesis is that every endometrial cycle has, as its only goal, nourishing support of an early embryo. Failure to accomplish this objective is followed by orderly elimination of unutilized tissue and prompt renewal to achieve a more successful cycle.

The ovum must be fertilized within 12–24 hours of ovulation. Over the next 4 days, it remains unattached within the tubal lumen utilizing tubal fluids and residual cumulus cells to sustain nutrition and energy for early cellular cleavage. After this stay, the solid ball of cells (morula), which is the embryo, leaves the tube and enters the uterine cavity. Here the embryo undergoes another 2–3 days of unattached but active existence. Fortunately, by this time endometrial gland secretions have filled the cavity and they bathe the embryo in nutrients. This is the first of many neatly synchronized events that mark the conceptus-endometrial relationship. By 6 days after ovulation, the embryo (now a blastocyst) is ready to attach and implant. At this time, it finds an endometrial lining of sufficient depth, vascularity, and nutritional richness to sustain the important events of early placentation to follow. Just below the epithelial lining, a rich capillary plexus has been formed and is available for creation of the trophoblast-maternal blood interface. Later, the surrounding zona compactum, occupying more and more of the endometrium, will provide a sturdy splint to retain endometrial architecture despite the invasive inroads of the burgeoning trophoblast.

Failure of the appearance of human chorionic gonadotropin, despite otherwise appropriate tissue reactions, leads to the vasomotor changes associated with estrogen-progesterone withdrawal and menstrual desquamation. However, not all the tissue is lost, and, in any event, a residual basalis is always available, making resumption of growth with estrogen a relatively rapid process. Indeed, even as menses persists, early regeneration can be seen. As soon as follicle maturation occurs (in as short a time as 10 days), the endometrium is ready once again to perform its reproductive function.

The Uterus Is an Endocrine Organ

The uterus is dynamic. It not only responds and changes in a sensitive fashion to classic hormonal signals (the endocrine events of the menstrual cycle) but it is also composed of complex tissues, with important autocrine and paracrine functions that serve not only the uterus but also the contiguous tissues of the fetoplacental unit during pregnancy. The most dynamic component of the uterus is the endometrium.

Endometrial Products

The endometrium secretes many substances, the functions of which (and their interrelationships) represent a major investigative challenge.[52] In addition to producing a nourishing, supportive environment for the early embryo, the endometrium plays an important role in suppressing the immune response within the pregnant uterus. The mechanisms controlling the immune response in decidual cells are not understood, but hormonal influence is undoubtedly important.

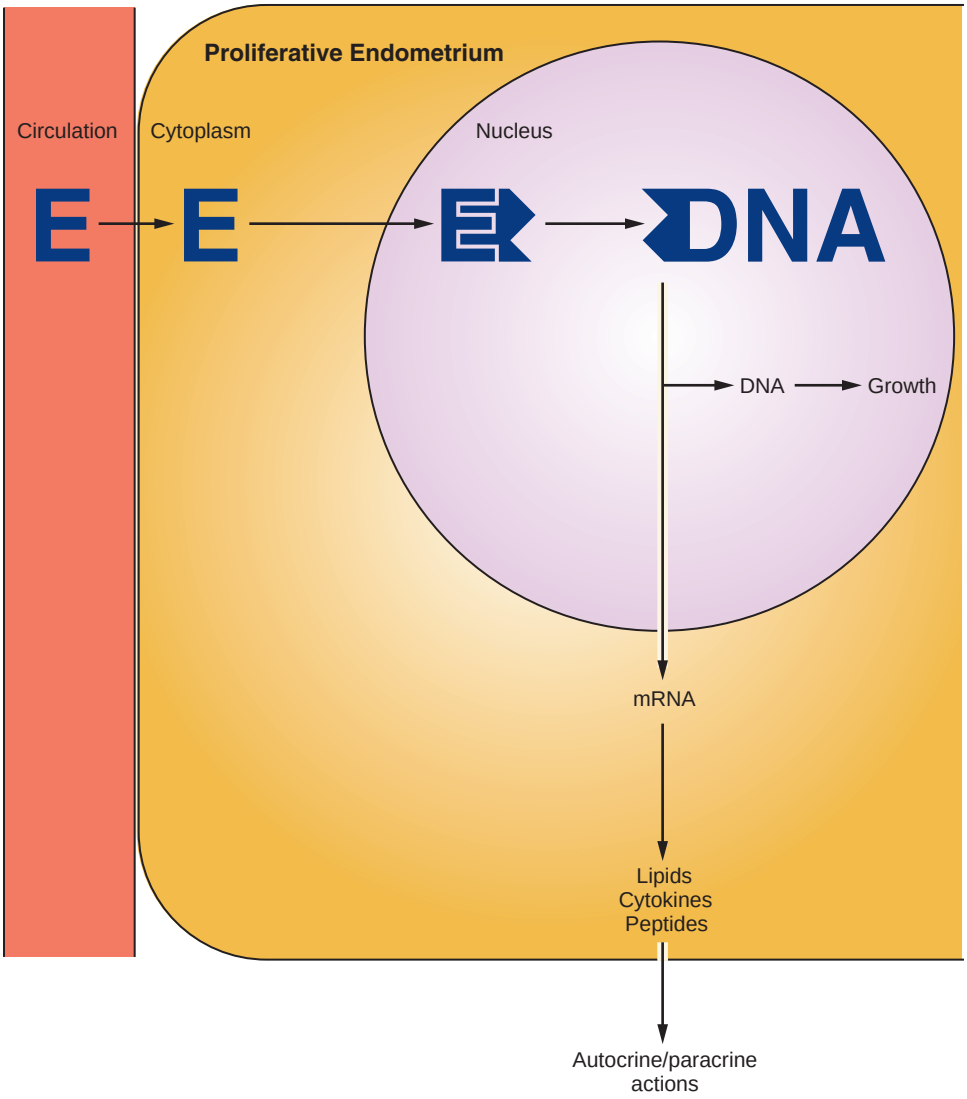
The presence of the cytokine family, involved in inflammation and immune responses, is not surprising in a tissue that undergoes cyclic degeneration. The interleukins stimulate the production of prostaglandins as well as other cytokines.[53] Colony-stimulating factor-1 is a cytokine that influences cellular proliferation and the presence of macrophages. Interferon-γ is produced by activated T lymphocytes and inhibits endometrial epithelial proliferation. Leukemia-inhibiting factor (LIF) is expressed in response to a variety of other cytokines

A Partial List of Endometrial Regulating Molecules

Lipids	*Cytokines*	*Peptides*
Prostaglandins	Interleukin-1α	Prolactin
Thromboxanes	Interleukin-1β	Relaxin
Leukotrienes	Interleukin-6	Prorenin and Renin
	Interferon-γ	Endorphin
	Colony-stimulating factor-1	Endothelin-1
	Tumor necrosis factor-α	Corticotropin-releasing hormone
	Leukemia-inhibiting factor	Fibronectin
		Uteroglobin
		Lipocortin-1
		Parathyroid hormone-like protein
		Integrins
		Epidermal growth factor family
		EGF
		Heparin-binding EGF
		TGF-α
		Insulin-like growth factor family
		IGF-I
		IGF-II
		IGFBPs 1-6
		Transforming growth factor-β family
		Activins
		Inhibins
		Follistatin
		Platelet-derived growth factor
		Fibroblast growth factor
		Vascular endothelial growth factor
		Gonadotropin-releasing hormone (GnRH)

and growth factors. Like the interleukins, LIF is most abundant during the progesterone-dominated secretory phase and early decidua and may have a role in embryo implantation.[54, 55] Tumor necrosis factor-α (TNF-α) gene expression is present in endometrium, and its activity is increased during the proliferative phase, decreased early in the secretory phase, and increased again in the midsecretory phase.[56] TNF-α exerts multiple influences on cellular growth.

Growth factors are peptides that bind to specific cell membrane receptors and initiate intracellular signaling pathways. Because the growth factors are potent mitogens, it is also not surprising that the follicular phase of the cycle, associated with proliferative activity of the endometrium, is marked by dramatic alterations in growth factors. Estrogen stimulates gene expression for epidermal growth factor (EGF) (and its receptor) and insulin-like growth factor (IGF) production. In turn, EGF elicits estrogen-like actions by interacting with the estrogen receptor mechanism.[57] EGF, a potent mitogen, is present in endometrial stromal and epithelial cells during the follicular phase of the cycle and in the stromal cells during the luteal phase.[58] Transforming growth factor-α (TGF-α) and EGF work through the same receptor and are important mediators of estrogen-induced growth of the endometrium. TGF-α levels peak at midcycle, in contrast to EGF levels, which are relatively stable and noncyclic.[59–61] Platelet-derived growth factor is a potent mitogen localized to stromal cells.



The insulin-like growth factors promote cellular mitosis and differentiation. They are expressed in a pattern controlled by estrogen and progesterone. IGF-I is predominant in proliferative and early secretory endometrium, and IGF-II appears in the mid to late secretory phase and persists in early pregnancy decidua.[62] Endometrial IGF-I expression is correlated with the circulating estrogen levels during the menstrual cycle.[63] This suggests that IGF-I synthesis is regulated by estrogen and mediates estrogen-induced growth of the endometrium, and IGF-II is involved in differentiation in response to progesterone. Evidence in the monkey indicates that IGF-I is the primary regulator of myometrial growth in response to estrogen as well as to estrogen plus progesterone.[64]

As elsewhere in the body, the myometrial IGF activity is modulated by the IGF binding proteins, which respond to the sex steroids in a differential manner; IGFBP-2 parallels IGF-I response, whereas IGFBP-3 is decreased in muscle but increased in vascular endothelium by estrogen.[65] IGFBP-4 and IGFBP-5 respond to estrogen but are unaffected by the addition of progesterone. IGFBP-1, as discussed later, is a major product of decidualized endometrium.

Gonadotropin-releasing hormone (GnRH) is present in endometrium and in increased amounts in secretory endometrium and decidua.[66] In human decidual cells, GnRH increases the expression of matrix metalloproteinases, suggesting a role for GnRH in the regulation of the enzymes involved in implantation.[67] Like all of these molecules, GnRH is involved in signaling pathways associated with cell proliferation and breakdown, interacting with adhesion factors such as integrins, enzymes, and angiogenic substances.[68]

Human myometrial smooth muscle and endometrial stromal cells express mRNA for parathyroid hormone-like protein, the function of which is unknown.[69] Transforming growth factor-β (TGF-β) stimulates the production of the parathyroid hormone-like protein. TGF-β production is greatest in the secretory phase and may inhibit cellular proliferation by increasing IGFBP-3 synthesis.

Prostaglandins are produced by both epithelial and stromal cells, and the prostaglandin content in the endometrium reaches a peak level in late secretory endometrium. The predominant prostaglandin produced by endometrium is prostaglandin $F_{2\alpha}$, a potent stimulus for myometrial contractions.[70] Endometrial prostaglandin production decreases dramatically after implantation, suggesting the presence of an active mechanism for suppression.[71] The production of prostaglandins requires estrogen support, but the increased production by secretory endometrium indicates progesterone enhancement, and acute withdrawal of progesterone promotes a further increase.[70] Endometrial stromal cells produce prostacyclin and thromboxane in response to estrogen, a response that can be blocked by progestins.[72] The myometrium principally produces prostacyclin, utilizing precursors derived from the endometrium. However, receptors for all members of the prostaglandin family are present on human myometrial cells, and contraction of the myometrium is a major consequence of prostaglandin $F_{2\alpha}$.[73]

Thromboxane is synthesized by uterine tissues. Gene expression for the thromboxane synthase and for the thromboxane receptor can be identified in endometrial glands, stromal cells, myometrial smooth muscle, and uterine blood vessels.[74] Thromboxane A_2 is a potent vasoconstrictor and stimulator of smooth muscle cells. Because of its rapid metabolism, it is limited to autocrine and paracrine activity.

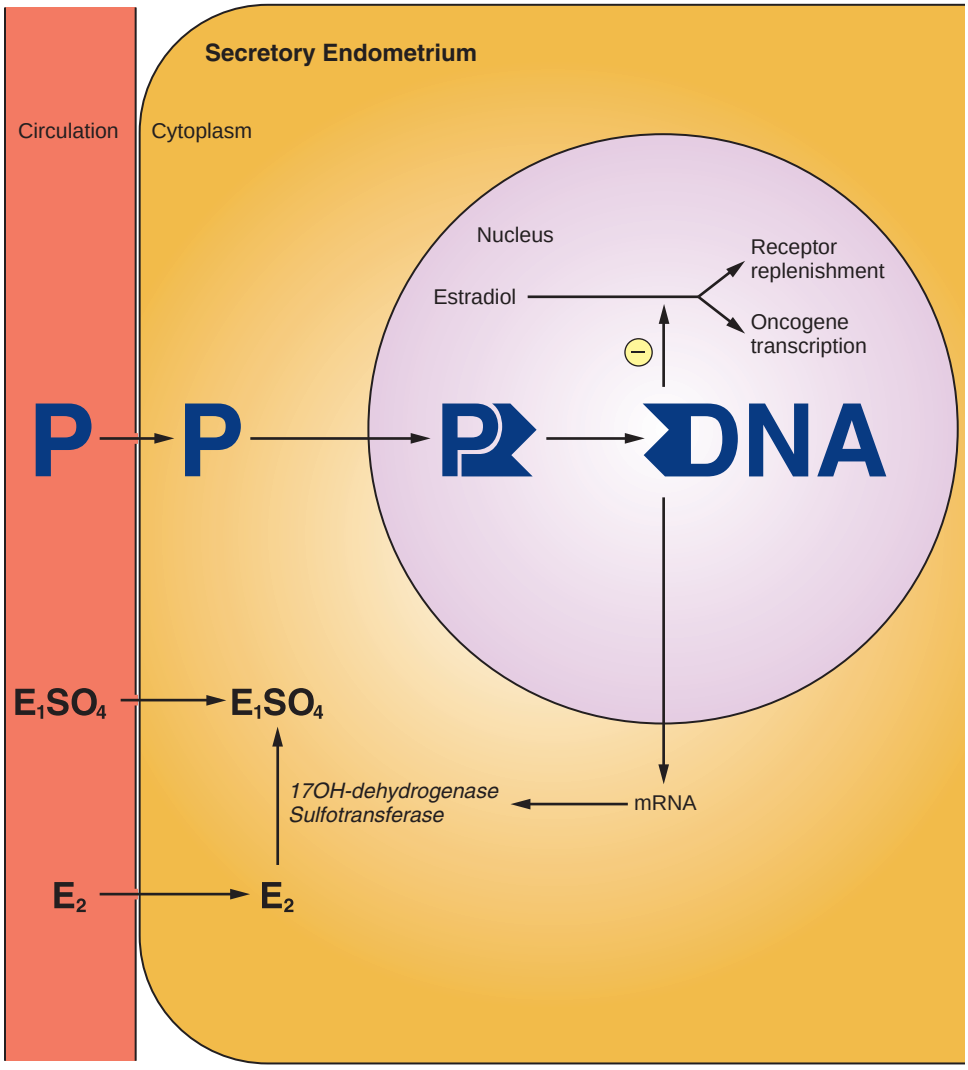
Women with excessive menstrual bleeding have alterations in the normal rates of prostaglandin production. For this reason, effective reductions in menstrual blood loss can be achieved with treatment utilizing one of the nonsteroidal anti-inflammatory agents that inhibit prostaglandin synthesis. These agents are also effective treatment for prostaglandin-mediated dysmenorrhea.

Fibronectin and laminin are extracellular matrix substances that are secreted by stromal cells of the endometrium in response to progesterone.[75] These proteins are important

adhesion molecules during implantation. Integrins are a family of glycoproteins that function as receptors for proteins such as collagen, fibronectin, and laminin. The integrins are highly expressed in endometrium and are important for cell-to-cell and cell-to-matrix interactions.[76] The expression of integrins is regulated by cytokines and growth factors, not estrogen and progesterone.[77]

Endothelins are potent vasoconstrictors produced in the vascular endothelial cells. The vasoconstrictor activity of endothelin-1, present in the endometrium, is balanced by the fact that it promotes the synthesis of the vasodilators nitric oxide and prostacyclin. Endothelin-1 is synthesized in endometrial stromal cells and the glandular epithelium, stimulated by both TGF-β and interleukin-1α.[78] Endothelin-1 is at least one agent responsible for the vasoconstriction that shuts off menstrual bleeding. It is also a potent stimulator of myometrial contractions and can contribute to dysmenorrhea. Finally, endothelin-1 is a mitogen and can promote the healing re-epithelialization of the endometrium. Human decidual cells also synthesize and secrete endothelin-1, from where it may be transported into the amniotic fluid.[79]

Angiogenesis, the formation of new blood vessels, is an essential process in tissue growth and development. Angiogenesis is necessary for tumor growth, and, in normal tissues, it is usually kept in check by regulating factors. The female reproductive tissues (specifically ovarian follicles, the trophoblast, and the endometrium), however, must experience periodic



and rapid growth and regression. In these tissues, angiogenesis is part of normal events. The endometrium is a major source for angiogenic factors during the menstrual cycle and during pregnancy.[80] Vascular endothelial growth factors (VEGFs), a collection of specific mitogens for endothelial cells, are abundantly expressed in human endometrium, reaching a peak that correlates with the maximal angiogenesis reached during the secretory phase.[81, 82] The VEGF family contains six growth factors and utilizes three different receptors. During the proliferative phase, estrogen stimulates VEGF synthesis. VEGF expression is also stimulated by hypoxia, specifically the hypoxia associated with endometrial breakdown, and the new blood vessel growth as well as the re-epithelialization of the endometrium in the new proliferative phase are dependent on these growth factors in response to estrogen.[83, 84]

Angiogenesis is also influenced by many other growth factors and other substances such as fibronectin and prostaglandins. The fibroblast growth factor family, in particular, is highly mitogenic for endothelial cells and endometrial stromal cells. Angiopoietins sustain the endometrium by preventing apoptosis and stabilizing blood vessels. The endometrium also produces inhibitory proteins, and the final growth of blood vessels reflects the balance between the inhibitory and stimulatory factors.

In all types of endometrial and myometrial cells, estrogen receptor expression reaches a maximum in the late proligerative phase.[85, 86] The concentration is greatest in the glandular epithelium. During the early secretory phase, estrogen receptor expression declines, followed by an increase in the mid and late secretory phases. These changes reflect the cyclic changes in estradiol (which increases estrogen receptor expression) and progesterone (which decreases estrogen receptor expression). Although estrogen receptor-beta is present in human endometrium, it is less prominent than estrogen receptor-alpha and exhibits less change during the cycle, except when it becomes the predominant estrogen receptor in the endometrial vasculature in the late secretory period.[87] Estrogen stimulation of proliferation is largely, if not totally, mediated by estrogen receptor-alpha.

Progesterone receptor expression in endometrial glandular epithelium reaches a maximum in the late proliferative and early secretory phases (reflecting induction of progesterone receptor by estrogen) and then declines to nearly undetectable levels by the midpoint of the secretory phase. Stromal cells in the endometrium show only minor fluctuations in progesterone receptors during the menstrual cycle. Decidualizing stromal cells exhibit strong progesterone receptor expression, although progesterone receptors are absent from decidual epithelial cells. Smooth muscle cells of the uterus demonstrate strong progesterone receptor expression throughout the menstrual cycle. Many of the events in uterine growth and function are regulated by the interplay between estrogen and progesterone. In general, progesterone antagonizes estrogen stimulation of proliferation and metabolism. This antagonism can be explained by the effects of progestins on the estrogen receptor (a decrease in levels), on the enzymes that lead to excretion of estrogen from cells, and by progesterone suppression of estrogen-mediated transcription of oncogenes.

Androgen receptor is present in endometrium at all stages of the menstrual cycle, in postmenopausal endometrium, and in the decidua of pregnancy.[88] Surprisingly, the androgen receptor concentration is constant throughout the cycle. Androgens suppress the proliferative effects of estrogen on the endometrium, and experimental evidence suggests that the suppressive effects on the endometrium induced by antiprogestational agents are mediated by the androgen receptor.[89]

The complexity of the endometrium can be appreciated by viewing the results of complementary DNA microarray studies. In one effort directed just to the endometrial breakdown associated with menstruation, 571 transcripts were identified that were involved in 131 biochemical pathways, including thyroid hormone synthesis and metabolism![90] Gene expression studies are just beginning to profile the patterns associated with specific hormones and pharmacologic agents.[91]

The Decidua

The decidua is the specialized endometrium of pregnancy. The biochemical dialogue between the fetoplacental unit and the mother must pass back and forth through the decidua. The classic view of the decidua conformed to its designation as a thin line in anatomic diagrams, a minor, inactive structural component. We now know that the decidua is a vigorous, active tissue.

Decidual cells are derived from the stromal cells of the endometrium, under the stimulation of progesterone. This transformation is regulated by members of the transforming growth factor beta family, including activin A.[92, 93] In addition, ghrelin acting via the growth hormone receptor is involved in this process.[94]

Decidual cells appear during the secretory phase and continue to proliferate during early pregnancy, eventually lining the entire uterus including the implantation site. The decidual cell is characterized by the accumulation of glycogen and lipid droplets and the new expression of a host of substances, including prolactin, relaxin, renin, insulin-like growth factors (IGFs), and insulin-like growth factor binding proteins (IGFBPs). There is no evidence that these proteins are secreted into the circulation; therefore, they serve as autocrine and paracrine agents.[95, 96]

Riddick was the first to detect prolactin in the decidualizing endometrium of the late luteal phase.[97] The amino acid sequence and the chemical and biologic properties of decidual prolactin are identical to those of pituitary prolactin. Decidual prolactin synthesis and release are controlled by the placenta, fetal membranes, and decidual factors. Dopamine, bromocriptine, and thyrotropin-releasing hormone (TRH), in contrast to their action in the pituitary, have no effect on decidual synthesis and release of prolactin. A protein named decidual prolactin-releasing factor has been purified from the placenta, and an inhibiting protein, which blocks the stimulatory activity of the releasing factor, has been purified from decidua.[96] IGF-1, relaxin, and insulin all stimulate decidual prolactin synthesis and release, each through its own receptor. The same decidual cells produce both prolactin and relaxin. Prolactin exerts an overall inhibitory effect on the process of decidualization, perhaps an autocrine mechanism to limit the extent of decidualization.[98]

Lipocortin-1 is a calcium- and phospholipid-binding protein, present in the placenta and decidua, that inhibits phospholipase A_2 and responds to glucocorticoids. Lipocortin-1 inhibits decidual prolactin release but in a mechanism independent of phospholipase action and independent of glucocorticoids. The prostaglandin system is not involved in decidual prolactin production, and corticosteroids do not affect decidual prolactin release.[99]

There is good reason to believe that amniotic fluid prolactin is derived from the decidua. In vitro experiments indicate that the passage of prolactin across the fetal membranes is in the direction of the amniotic cavity. The amniotic fluid concentration correlates with the decidual content, not with maternal circulating levels. Amniotic fluid prolactin reaches peak levels in the first half of gestation (about 4,000 ng/mL) when maternal plasma levels are approximately 50 ng/mL and fetal levels about 10 ng/mL. Maternal circulating prolactin reaches maximal levels near term. Finally, amniotic fluid prolactin is unaffected by bromocriptine treatment (which reduces both fetal and maternal circulating levels to baseline levels).

It is believed that decidual prolactin regulates amniotic fluid volume and electrolyte concentrations. Prolactin regulates water and ion transport in lower animals, and prolactin binds to amniotic membranes. Disorders in human pregnancy associated with abnormal amniotic fluid volumes may be explained by this mechanism, especially idiopathic polyhydramnios, which is associated with a decrease in the number of prolactin receptors in the

membranes. Prolactin may be involved in the regulation of surfactant synthesis in the fetus, and prolactin may inhibit uterine muscle contractility. Prolactin suppresses the immune response and helps to prevent immunologic rejection of the conceptus. Prolactin can also function as an autocrine and paracrine growth factor in the uterus.[100]

Fibroblast growth factor, derived from decidua, stimulates blood vessel growth in early pregnancy. Another factor, endothelial-cell-stimulating angiogenesis factor (a nonprotein mitogen), is also derived from decidua and contributes to the vascularization of the decidua during the first trimester of pregnancy.[101] The expression of corticotropin-releasing hormone (CRH) has been demonstrated in human decidua, and many actions for decidual CRH are possible: activation of prostaglandins, stimulation of myometrial contractions, and a contribution to both maternal and fetal stress responses during pregnancy and labor.[102]

Prorenin (the inactive precursor of renin) is produced in decidua in response to IGF-1, insulin, endothelin, and relaxin.[103–105] A uterine role for renin has not been determined.

The insulin-like growth factor-binding proteins, IGFBP-1, -2, -3, and -4, are produced by endometrial stromal cells.[106] Large amounts of IGFBP-1 are present in amniotic fluid. The IGFBPs appear to be regulated by insulin, the IGFs, and relaxin.[107] Relaxin is related structurally to insulin and the IGFs, and it stimulates IGFBP-1 production in endometrial stromal cells.[108] IGFBP-1 is considered to be a marker for decidualization. Because it binds growth-promoting IGFs, the appearance of IGFBP-1 contributes to differentiation rather than proliferation of the endometrial stromal cells.

IGFBP-1 begins to appear in midsecretory phase endometrium and reaches a level of major production in decidua by late in the first trimester of pregnancy. IGFBP-1, when first identified, was known as placental protein 12 and then as pregnancy-associated α-globulin. By the second trimester of pregnancy, high levels of IGFBP-1 are present in the amniotic fluid and the circulation, and then fall significantly during the third trimester. The decidual production of IGFBP-1 is correlated with the morphologic and histologic changes induced by progesterone and regulated by progesterone, relaxin, insulin, IGF-I, and IGF-II. In fact, IGFBP-1 is a mediator of progesterone-induced decidualization of endometrial stromal cells.[109] Binding of the insulin-like growth factors to the IGFBPs would limit further mitogenic activity in the endometrium in the secretory phase and during pregnancy. In addition, decidual IGFBP-1 may contribute to the limitation of trophoblast invasion.

The continuous stimulation of IGFBP-1 production by human endometrium can be maintained in women as long as they retain an intrauterine device that releases a progestin into the endometrial cavity.[110] In endometrial samples from these women, areas of endometrial atrophy correlate with intense staining for IGFBP-1. This makes a strong argument for the importance of insulin-like growth factors for endometrial growth and the potential for prevention of endometrial growth by providing IGFBP-1.

The glycoprotein α subunit, common to follicle-stimulating hormone (FSH), luteinizing hormone (LH), thyroid-stimulating hormone (TSH), and hCG, is secreted into the circulation by the pituitary and placenta. A specific role for the α subunit has not been apparent; however, gonadotropin receptors are present in the endometrium, and α subunit acts synergistically with progesterone to induce decidualization of endometrial cells in vitro.[111] In addition, the α subunit stimulates decidual prolactin secretion.[112]

The chorion laeve, villous trophoblast, and decidua are all sites of TGF-β production.[113] TGF-β can signal its own production; thus, TGF-β can be a messenger from fetal tissues to decidua. TGF-β is also believed to play a role in limiting trophoblastic invasion.[114] This may be accomplished by stimulating the production of plasminogen activator inhibitor and the factor that causes tissue inhibition of metalloproteinases.[115]

SUMMARY: The Uterus Is an Endocrine Organ

One cannot dispute the fact that the uterus is an endocrine organ, but the vast array of active substances with their complicated names can be bewildering and overwhelming. It is helpful to keep in mind a fundamental and relatively simple description: the endometrium is necessary for reproduction, and the synchronous, complex cycle of events is dependent on the endocrine guidance of estradiol and progesterone, modulated and mediated by the plethora of locally produced biochemical agents. Each and every signaling substance utilizes one of the pathways discussed in Chapter 2 and makes a contribution to the dynamic sequence of morphological and biochemical events repeatedly dedicated to nourishing and supporting an early embryo.

Anatomical Abnormalities of the Uterus

Congenital abnormalities of the müllerian ducts are relatively common, occurring in 7% to 10% of all women, and contributing to the problems of infertility, recurrent pregnancy loss, and poor pregnancy outcomes that occur in approximately 25% of women with uterine anomalies.[116–121] Major anomalies are about 3 times more common in women with recurrent miscarriages.[122] The problems encountered in pregnancy include preterm labor, breech presentations, and complications that lead to interventions and greater perinatal mortality. Cervical cerclage is often indicated for prevention of preterm labor due to these anomalies. In addition, these abnormalities can produce the symptoms of dysmenorrhea and dyspareunia and even amenorrhea. Endometriosis in young women, especially adolescents, should raise clinical suspicion of genital tract malformations. Because the embryologic origin of the ovaries is separate and distinct from that of the müllerian structures, patients with müllerian anomalies have normal ovaries and ovarian function. Conception and implantation are not prevented. Surgical correction is recommended for pain, endometriosis due to obstruction, and poor obstetrical outcomes.

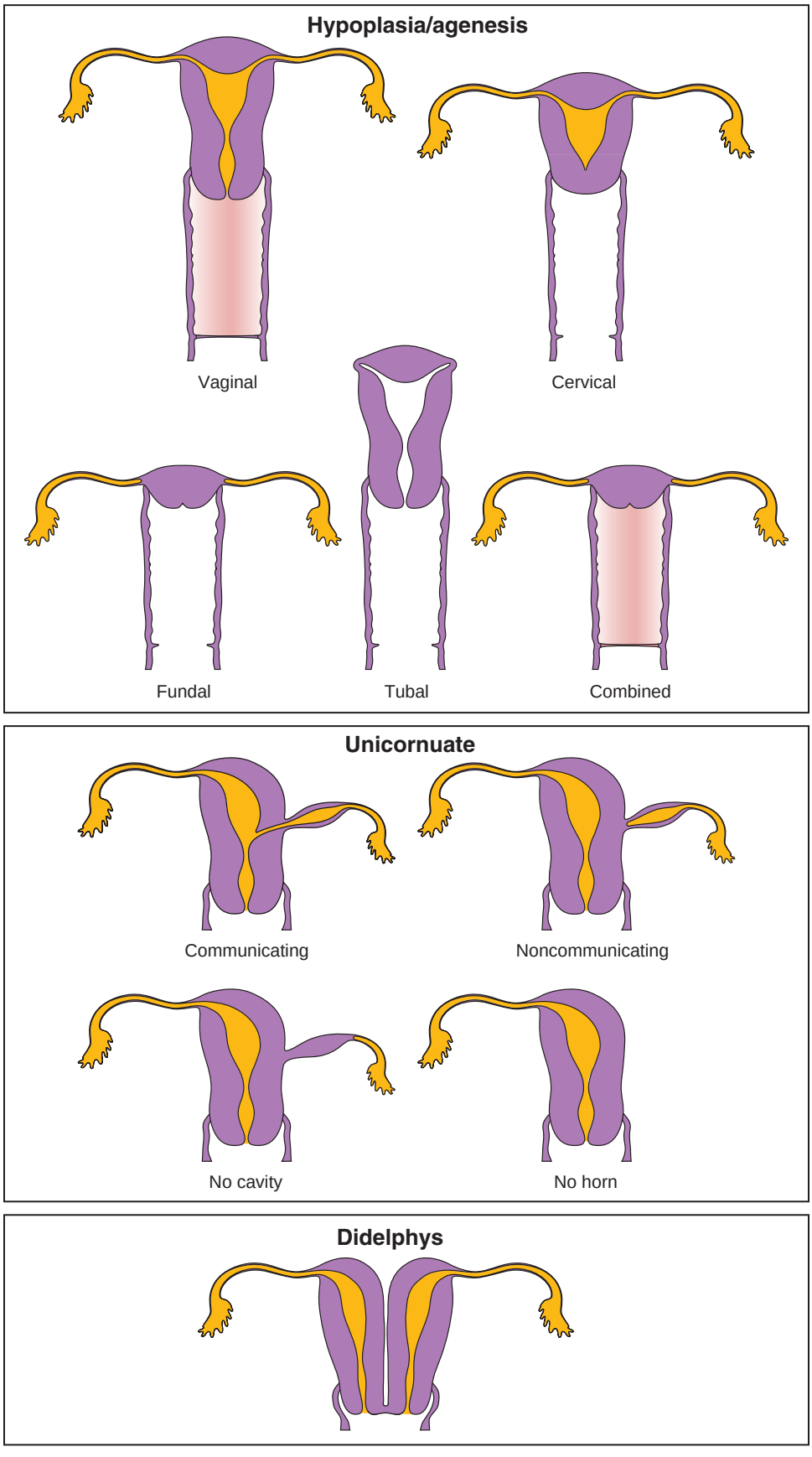
Incidence of Müllerian Defects

Fertile and infertile women	3–4%[123]
Women with recurrent miscarriages	5–10%[120]
Women with late miscarriages and preterm deliveries	>25%[120]

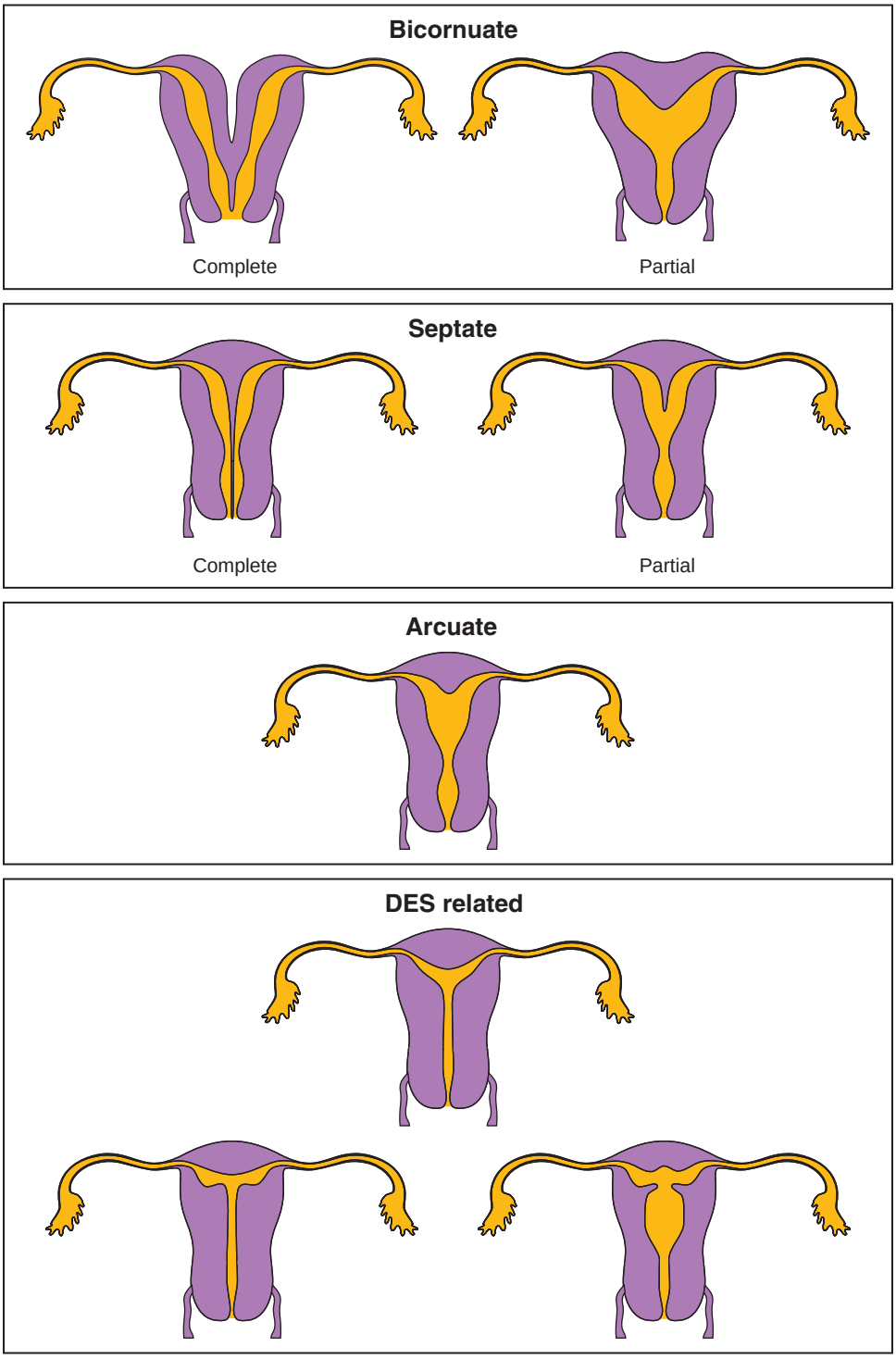
Anomalies can result from the failure of the müllerian ducts to fuse in the midline, to connect with the urogenital sinus, or to create the appropriate lumen in the upper vagina and uterus by resorption of the central vaginal cells and the septum between the fused müllerian ducts. Because fusion begins in the midline and extends caudally and cephalad, abnormal results can exist at either end. Formation of the uterine cavity begins at the lower pole and extends cephalad with dissolution of midline tissue; hence, incomplete resorption of tissue commonly yields persistence of the midline uterine wall intruding into the cavity. The molecular pathophysiology of these abnormalities has been insufficiently studied; however, the association with other somatic anomalies and occasional reports of familial transmission suggest genetic linkages.

Vaginal outflow tract obstruction can be minimal with a transverse septum or complete due to agenesis. A septum is the result of a defect in the connection of the fused müllerian

Classification of Müllerian Anomalies[124]



Classification of Müllerian Anomalies[124]



ducts to the urogenital sinus or a failure of canalization of the vagina. The location of the septum varies, although it is usually in the upper or middle third of the vagina. Vaginal agenesis is the result of a complete failure in canalization; these patients present with amenorrhea or pain due to accumulated menstrual effluvium. Surgical correction is frequently necessary to relieve the relative constriction (and obstruction) of the vaginal canal. An absent vagina is usually accompanied by an absent uterus and tubes, the

classic müllerian agenesis of the Mayer-Rokitansky-Kuster-Hauser syndrome (discussed in Chapter 11).

Distribution of Specific Anomalies[123]

Septate uterus	35%
Bicornuate uterus	26%
Arcuate uterus	18%
Unicornuate uterus	10%
Uterus didelphys	8%

Uterine anomalies can be organized into the following categories.[124] Each of these can be associated with obstructions that present during adolescence with amenorrhea and cyclic pain.[125]

Uterus Didelphus (Double Uterus)

Lack of fusion of the two müllerian ducts results in duplication of corpus and cervix. These patients usually have no difficulties with menstruation and coitus, except when there is also a midline longitudinal vaginal septum. Occasionally, one side is obstructed and symptomatic. In addition, a double uterus is occasionally associated with an obstructed hemivagina (often with ipsilateral renal agenesis); early diagnosis and excision of the obstructing vaginal septum will preserve fertility. Pregnancy is associated with an increased risk of miscarriage, malpresentations, and premature labor, although many patients will have no reproductive difficulties.[123, 126, 127] Unification of the two endometrial cavities by metroplasty is indicated in patients with repeated poor obstetrical outcomes.

Unicornuate Uterus

An abnormality that is unilateral obviously is due to a failure of development in one müllerian duct (probably a failure of one duct to migrate to the proper location). The altered uterine configuration is associated with an increase in endometriosis and in obstetrical complications (early spontaneous miscarriage, ectopic pregnancy, abnormal presentations, intrauterine growth retardation, and premature labor).[126, 128–131] There may be a rudimentary horn present, and implantation in this horn is followed by a very high rate of pregnancy wastage or tubal pregnancies. A rudimentary horn can also be a cause of chronic pain, and surgical excision is worthwhile. However, most rudimentary horns are asymptomatic because they are noncommunicating, and the endometrium is not functional. Because of the potential for problems, prophylactic removal of the rudimentary horn is recommended when it is encountered during a surgical procedure. Approximately 40% of patients with a unicornuate uterus will have a urinary tract anomaly (usually of the kidney).[132] Surgical reconstructive procedures do not improve obstetrical outcomes; however, cervical cerclage may be beneficial when indicated.

The Bicornuate Uterus

Partial lack of fusion of the two müllerian ducts produces a single cervix with a varying degree of separation in the two uterine horns. This anomaly is relatively common, and

pregnancy outcome has usually been reported to be near normal. Some, however, find a high rate of early miscarriage, preterm labor, and breech presentations.[119, 126] With a history of repeated poor pregnancy outcome, surgical metroplasty is worth consideration.

The Septate Uterus and the Arcuate Uterus

Partial lack of resorption of the midline septum between the two müllerian ducts results in fibromuscular defects that range from a slight midline septum (the arcuate, heart-shaped cavity) to a significant midline division of the endometrial cavity. A total failure in resorption can leave a longitudinal vaginal septum (a double vagina). This defect is not a cause of infertility, but once pregnant, the greater the septum the greater the risk of recurrent spontaneous miscarriage, especially in the second trimester. The complete septate uterus is associated with a high risk of spontaneous miscarriage, preterm labor, intrauterine growth retardation, and breech presentation.[119, 133, 134] Even a small septum is associated with these poor obstetrical outcomes.[135] Outcomes are excellent with treatment by hysteroscopy.[123, 134, 136–140] Post treatment miscarriage rates are approximately 10% in contrast to the 90% pretreatment rates with a complete septum. A longitudinal vaginal septum usually does not have to be excised (unless dyspareunia is a problem). In some reports, the arcuate uterus has no adverse impact on reproductive outcome.[126] Prophylactic surgery is considered appropriate for a septate uterus in older women and in women being treated with in vitro fertilization. A surgical procedure is not indicated for the arcuate uterus.

Very Rare Anomalies

Isolated agenesis of the cervix or the endometrium is incredibly rare. Absence of the cervix can lead to so much pain and obstruction that hysterectomy is the best solution. Attempts to preserve fertility by creating a fistulous communication between uterus and vagina have achieved some success, but repeat surgery due to reappearance of obstruction is common.[141, 142] In asymptomatic patients, consideration should be given to preservation of structures for the possibility of pregnancy that can be achieved by means of one of the techniques of assisted reproduction (Chapter 32).

The Diethylstilbestrol-Associated Anomaly

Mothers who were treated in 1938 to 1975 with high doses of estrogen early in their pregnancies had children who developed a variety of anomalies, ranging from the hypoplastic T-shaped uterus to irregular cavities with adhesions.[143] Women with uterine abnormalities usually also have cervical defects. In these individuals, the chance of term pregnancy is decreased because of higher risks of ectopic pregnancy, spontaneous miscarriage, and premature labor.[144] An incompetent cervix is common. Poor outcome is correlated with an abnormal uterus on hysterosalpingography. No treatment is available beyond cervical cerclage.

Accurate Diagnosis of Anomalies

In the past, full diagnosis required surgical intervention, first laparotomy and then, more recently, laparoscopy. Today, vaginal ultrasonography, especially three-dimensional ultra-

sound, sonohysterography, and magnetic resonance imaging are highly accurate, and surgical intervention is usually not necessary.[145–147] Hysterosalpingography alone can yield inaccurate results due to a failure to perfuse both uterine horns on either side of a midline division, and cannot reliably distinguish bicornuate and septate uteri. Decisions should not be based on hysterosalpingography alone. Congenital anomalies of the müllerian ducts are frequently accompanied by abnormalities in the urinary tract, such as a horseshoe or pelvic kidney. Renal agenesis can be present on the same side as a müllerian defect.

Pedro Acién at the San Juan University Hospital in Alicante, Spain, is an acknowledged expert on the many and varied malformations of the female genital tract. He advocates a more complete classification, that includes müllerian anomalies with anomalies of the urogenital ridge, the mesonephric structures, and the cloaca.[148] The embryologic origins of the various anomalies and an understanding of unusual cases can be obtained through Acién's publications.[5, 148]

Leiomyomas (Uterine Fibroids)

Uterine leiomyomas are benign neoplasms that arise from uterine smooth muscle and cause abnormal uterine bleeding and symptoms secondary to a large pelvic mass. It is hypothesized that leiomyomas originate from somatic mutations in myometrial cells, resulting in progressive loss of growth regulation.[149, 150] The tumor grows as genetically abnormal clones of cells derived from a single progenitor cell (in which the original mutation took place). Studies indicate that leiomyomas are monoclonal.[151] Different rates of growth can reflect the different chromosomal abnormalities present in individual tumors. Multiple myomas within the same uterus are not clonally related; each myoma arises independently.

The presence of multiple myomas (which have a higher recurrence rate than single myomas) argues in favor of a genetic predisposition for myoma formation. There is about a 2.5-fold increased risk of developing myomas in first-degree relatives of women with these tumors.[152] Hereditary leiomyomatosis and renal cell carcinoma is an autosomal dominant syndrome with both cutaneous and uterine leiomyomas. The risk of renal cell carcinoma and that of leiomyosarcoma are increased in this syndrome.[153, 154] The gene involved is *fumarate hydratase*, coding for an enzyme involved in the Kreb's cycle. A family history of cutaneous leiomyomata should trigger screening for this gene mutation. Renal cell cancer occurs in 10–16% of women with this syndrome. Studies of DNA polymorphisms will undoubtedly yield patterns identifying women at high risk for uterine leiomyomata, and perhaps risk for recurrence following ablation treatments and for malignant progression to leiomyosarcoma. Thus far, chromosomal abnormalities have been described in about 40% of myomas.[155] Another approach is to identify the microRNA pattern associated with leiomyoma size, growth rates, and ethnic prevalence.[156]

It is not certain whether leiomyosarcomas arise independently or from leiomyomas. However, the incidence of leiomyosarcomas in patients with leiomyomas is very low (less than 1%).[157] Gene profiling has not discovered shared abnormalities or a common molecular pathway comparing myomas with leiomyosarcomas.[158]

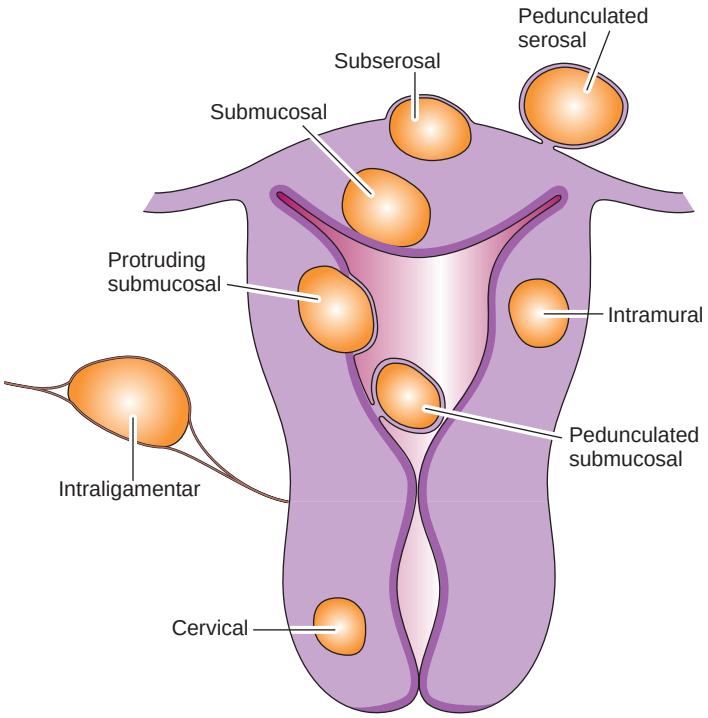
If surgical specimens are serially sectioned, about 77% of women who come to hysterectomy will have myomas, many of which are occult.[159] By the age of menopause, ultrasound can identify myomas in about 80% of black American women and 70% of white American women.[160] In the United States, about 40% of abdominal hysterectomies and 17% of vaginal hysterectomies are performed for leiomyomas.[161] The peak incidence for myomas requiring surgery occurs around age 45, approximately 8 cases per 1,000 women

each year.[162] In the United States, approximately 10–15% of women require hysterectomy for myomas. For unknown reasons, uterine leiomyomas are 2–3 times more prevalent in black women compared with white, Hispanic, and Asian women and account for 75% of hysterectomies among black women.[160, 163, 164]

Myomas are present (diagnosed by ultrasonography) in about 30% of women, and in about 1–2% of pregnancies.[165, 166] The risk of myoma is decreased with increasing parity and with increasing age at last term birth.[166, 167] Women with at least two full-term pregnancies have half the risk for myomas. Smoking decreases the risk (presumably by decreasing estrogen levels), and obesity increases the risk (presumably by increasing estrogen levels). Although a lower risk for myomas is associated with factors that decrease estrogen levels, including leanness, smoking, and exercise, the use of oral contraceptives is not associated with an increased risk of uterine myomas, although the Nurses' Health Study reported a slightly increased risk when oral contraceptives were first used in early teenage years.[167–169]

The hormone sensitivity of leiomyomas is further indicated by the following clinical observations. Leiomyomas develop during the reproductive (hormonally active) years and regress after menopause. Occasionally, leiomyomas grow during pregnancy, and the hypogonadal state induced by treatment with gonadotropin-releasing hormone (GnRH) agonists often causes shrinkage of myomas.

The environment within the leiomyoma is hyperestrogenic. The estradiol concentration is increased, and leiomyomas contain more estrogen and progesterone receptors.[170–173] Aromatase gene and enzyme expression are present in significant levels in leiomyomas.[174] Indeed, leiomyoma tissue is hypersensitive to estrogen and appears to have lost a regulatory influence that limits estrogen response.[175] Endometrial hyperplasia can be observed at the margins of submucosal myomas.[176] In the myometrium and in leiomyomas, peak mitotic activity occurs during the luteal phase, and mitotic activity is increased by the administration of high doses of progestational agents.[177, 178] These facts indicate that progesterone stimulates mitotic activity in leiomyomas, although animal studies indicate both stimulation and inhibition of myometrial growth. Similarly, clinicians have reported both regression



and growth with progestational treatment. Nevertheless, most of the evidence supports a growth-promoting role for progestins. The association with estrogen can be explained by the estrogen enhancement of progesterone receptor expression.[179, 180] Treatment with mifepristone, the progesterone antagonist, or with asoprisnil, a selective progesterone receptor agonist/antagonist, is associated with a reduction in leiomyoma size.[181, 182]

At least one pathway for the growth-promoting effect of progestins is the induction of *BCL2* gene expression increasing the production of the Bcl-2 protein that inhibits apoptosis and promotes cell replication.[183] Bcl-2 protein expression is increased in leiomyoma cells and markedly increases in response to progesterone.[184] In contrast, normal myometrial cells do not respond to estradiol or progesterone with Bcl-2 protein expression, and there is no cyclic change throughout the menstrual cycle.

As in the normal uterus, the effects of estrogen and progestins on leiomyomas are mediated by growth factors.[185] EGF is overexpressed in myomas, EGF receptors are present in leiomyomas, and GnRH agonist treatment (and hypogonadism) decreases EGF concentration in myomas (but not in normal myometrium).[186, 187] IGF-I and IGF-II and their receptors are abundant in myometrium and actively overexpressed in leiomyomas.[188, 189] Leiomyomas express more IGF-II and less IGFBP-3 than myometrium, a situation that would enhance growth factor availability and activity in the tumor.[190] Leiomyoma cells express more parathyroid hormone-related protein (another growth factor) than normal myometrium.[191] Like the endometrium and myometrium, leiomyomas secrete prolactin, and prolactin functions in the uterus as a growth factor.[100] Even hematopoiesis is possible in a leiomyoma.[192]

One of the consequences of altered growth factor expression in myomas is an abnormal vasculature, characterized by a dilated venous plexus.[193] This morphologic feature may be the result of specific vascular regulators of angiogenesis, such as fibroblast growth factor and vascular endothelial growth factor. These changes probably contribute to the heavy menstrual bleeding associated with submucosal myomas.

Uterine growth and signaling molecules are highly expressed in leiomyomas.[194] As with all tumors, these pathways in leiomyomas may one day be targeted by gene therapy. For example, specific adenoviruses can deliver altered genes to myoma cells that can interfere with the gene expression required for tumor growth and cellular functions.

Reproductive Function and Leiomyomas

Leiomyomas are an infrequent cause of infertility, either by mechanical obstruction or distortion (and interference with implantation).[195, 196] When a mechanical obstruction of fallopian tubes, cervical canal, or endometrial cavity is present and no other cause of infertility or recurrent miscarriage can be identified, myomectomy is usually followed by a prompt achievement of pregnancy in a high percentage of patients (usually within the first year).[196, 197] Small submucosal myomas are best treated by hysteroscopic resection. Preoperative visualization is important, and mapping of myomas by sonohysterography or magnetic resonance imaging (MRI) is superior to standard ultrasonography (which is relatively inaccurate).[198] It is difficult to distinguish between submucosal myomas and endometrial polyps with ultrasonography.[199] Very large myomas (greater than 4–5 cm) and myomas that do not have greater than 50% protrusion into the cavity are not good candidates for hysteroscopic removal.

The 5-year recurrence rate after abdominal myomectomy for a single myoma is about 10%, and 25% with multiple myomas, with subsequent hysterectomy necessary in one-third of patients with recurrence.[200] In a series with long-term follow-up, the recurrence rate over 10 years after single myomectomy reached 27%.[201] Women who gave birth after

myomectomy had a recurrence rate (over 10 years) of 16%, compared to a rate of 28% in those who did not give birth. In an Italian study of recurrence, the rate at 5 years reached 55% in those who did give birth after surgery and 42% in those with no childbirth.[202] These differences may reflect the diligence and sensitivity of the ultrasonographic assessments.

An increased incidence of spontaneous miscarriage because of myomas has not been definitively documented in the literature. Myomectomy for infertility or recurrent miscarriage requires a deliberate and careful decision after all factors have been considered. Intracavitary myomas, however, usually require surgery. Submucosal myomas are associated with general cavitary alterations in the expression of proteins involved with implantation, not just an effect confined to the endometrium over the myoma.[203] Intramural myomas that do not affect the endometrial cavity do not affect implantation or increase the risk of miscarriage.[204, 205] Because of the rapid regrowth of myomas following cessation of GnRH agonist therapy, medical therapy for infertility is not recommended.

Most myomas do not grow during pregnancy.[206] When they do, most of the growth is in the first trimester, and most myomas regress in size after the pregnancy. The size of a myoma will not predict its course; large myomas will not necessarily grow more than small ones. Most pregnancies, in the presence of myomas, will, therefore, be uncomplicated (although a higher incidence of cesarean section has been observed).[165, 207] Nevertheless, the risks of malpresentations, preterm delivery, and spontaneous miscarriage are increased.[208] So-called red degeneration of myomas is occasionally observed during late pregnancy, a condition due to central hemorrhagic infarction of the myoma. Pain is the hallmark of this condition, occasionally associated with rebound tenderness, mild fever, leukocytosis, nausea, and vomiting. Usually pain is the only symptom and resolution follows rest and analgesic treatment.[209] Surgery should be a last resort. The larger the myoma, the greater the risk of premature labor.[210]

Medical Therapy of Leiomyomas

The goals of medical therapy for leiomyomas are to ***temporarily*** reduce symptoms and to reduce myoma size, and the therapy of choice is treatment with a GnRH agonist.[211] Any treatment that lowers endogenous estrogen levels should be effective, and therefore, the use of aromatase inhibitors is another option.[212] Prolonged medical regimens are expensive and complicated. With few exceptions, surgical treatment is preferred for symptomatic uterine leiomyomas. Medical therapy is provided preoperatively to improve anemia and reduce surgical complexity and recovery times.[213]

The short half-life of GnRH is due to rapid cleavage of the bonds between amino acids 5–6, 6–7, and 9–10. By altering amino acids at these positions, analogues of GnRH can be synthesized with different properties. Substitution of amino acids at the 6 position or replacement of the C-terminal glycine-amide (inhibiting degradation) produces agonists. An initial agonistic action (the so-called flare effect) is associated with an increase in the circulating levels of follicle-stimulating hormone (FSH) and luteinizing hormone (LH). This response is greatest in the early follicular phase when GnRH and estradiol have combined to create a large reserve pool of gonadotropins. After 1–3 weeks, desensitization and down-regulation of the pituitary produce a hypogonadotropic, hypogonadal state. The initial response is due to desensitization, the uncoupling of the receptor from its effector system, whereas the sustained response is due to a loss of receptors by down-regulation and internalization. Furthermore, postreceptor mechanisms lead to secretion of biologically inactive gonadotropins, which, however, can still be detected by immunoassay.

The GnRH analogues cannot escape destruction if administered orally. Higher doses administered subcutaneously can achieve nearly equal effects as those observed with intravenous

treatment; however, the smaller blood peaks are slower to develop and take longer to return to baseline. Other forms of administration include nasal spray, sustained-release implants, and intramuscular injections of biodegradable microspheres.

Treatment with GnRH Agonists

Summarizing the experience with GnRH agonist treatment of leiomyomas, the mean uterine size decreases 30–64% after 3–6 months of treatment.[211] Maximal response is usually achieved by 3 months. The reduction in size correlates with the estradiol level and with body weight. Menorrhagia, anemia, pelvic pressure, and urinary frequency all respond favorably to GnRH agonist treatment.[214, 215] A decrease in operative blood loss can be achieved when the pretreatment uterus is as large as a 16-week pregnancy or larger. However, some studies find no benefit in terms of surgical blood loss or length of hospital stay, and surgical dissection may be more difficult because of softening of the myoma.

Why is there a variation in response? When one considers the many factors involved in myoma growth (estrogen, progesterone, growth factors, and receptors), it makes sense that not every myoma is the same. After cessation of GnRH agonist therapy, menses return in 4–10 weeks, and myoma and uterine size return to pretreatment levels in 3–4 months. The rapid regrowth is consistent with the fact that reduction in size is not due to a cytotoxic effect.

Preoperative GnRH agonist therapy offers several advantages for hysteroscopic removal of submucosal tumors. In addition to a decrease in myoma size, endometrial atrophy will improve visualization, and decreased vascularity will reduce blood loss.

Leiomyomatosis Peritonealis Disseminata is a condition in which multiple small nodules of benign smooth muscle are found throughout the abdominal cavity and occasionally in the pulmonary cavity. This condition appears to be sensitive to estrogen because it has been aggravated by postmenopausal estrogen treatment, and regression has been achieved with GnRH agonist treatment.[216]

Adenomyosis is the ectopic presence of endometrial glands within the myometrium. This diagnosis can be made by magnetic resonance imaging, and successful treatment with a GnRH agonist has been reported.[217, 218]

Side Effects with GnRH Agonists

Hot flushes are experienced by more than 75% of patients, usually in 3–4 weeks after beginning treatment. Approximately 5–15% of patients will complain of headache, mood changes, vaginal dryness, joint and muscle stiffness, and depression. About 30% of patients will continue to have irregular (although light) vaginal bleeding. It is useful to measure the circulating estradiol level. If the level is greater than 30 pg/mL, suppression is inadequate. A small number (10%) of patients will experience a localized allergic reaction at the site of injection of depot forms of GnRH analogues. More serious reaction is rare, but immediate and delayed anaphylaxis can occur, requiring intense support and management.[219]

Bone loss occurs with GnRH therapy, but not in everyone, and it is reversible (although it is not certain if it is totally reversible in all patients). A significant vaginal hemorrhage 5–10 weeks after beginning treatment is encountered in about 2% of treated women, due to degeneration and necrosis of submucosal myomas.[220] A disadvantage of agonist treatment is a delay in diagnosis of a leiomyosarcoma. Keep in mind that almost all leiomyosarcomas

present as the largest or only uterine mass. Close monitoring is necessary and surgery has been the usual recommendation when either enlargement or no shrinkage of myomas occurs during GnRH agonist treatment.[221] The use of Doppler ultrasonography or magnetic resonance imaging offers greater accuracy of evaluation. However, the incidence of leiomyosarcoma, even in patients with "rapidly growing leiomyomas," is very low (less than 0.5%) and almost unheard of in premenopausal women.[157] In premenopausal women, a conservative approach is warranted.

Escape of suppression can result in an unexpected pregnancy. No adverse effects of fetal exposure to GnRH agonists have been reported, even when exposure has persisted throughout the early weeks of pregnancy.[222]

GnRH Agonists and Steroid Add-Back

Treatment with a GnRH agonist with steroid add-back has been explored to permit long-term therapy without bone loss.[211] Two strategies have been employed: simultaneous agonist and steroid add-back treatment or a sequential regimen in which the agonist is used alone for 3 months, followed by the combination of the agonist and steroid add-back. This long-term treatment is attractive for women who are perimenopausal, perhaps avoiding surgery. In addition, long-term treatment would be useful for women with coagulopathies, and in women with medical problems who need to postpone surgery.

Simultaneous treatment with agonist and medroxyprogesterone acetate (20 mg daily) or norethindrone (10 mg daily) effectively reduced hot flushing but was less effective (consistent with a major supportive role for progestins in myomas) in reducing uterine volume.[211, 223] A sequential program, adding a traditional postmenopausal hormone regimen (0.625 mg conjugated estrogens on days 1–25 and 10 mg medroxyprogesterone acetate on days 16–25) effectively reduced uterine volume and maintained the reduced volume for 2 years (and avoided any loss in bone density)[211] A daily 2.5 mg dose of tibolone also prevents bone loss and inhibits vasomotor symptoms without reducing the therapeutic efficacy of GnRH agonist treatment.[224] The addition of raloxifene to GnRH agonist treatment appeared to produce a greater reduction in leiomyoma size,[225] but the effect was not sufficiently different to be of clinical significance. Raloxifene treatment by itself, even in a large dose, failed to reduce leiomyoma size in premenopausal women, although in postmenopausal women, raloxifene produced a 30% to 40% reduction in size after 1 year.[226, 227] Treatment with raloxifene, alendronate, or tibolone prevents the bone loss associated with agonist therapy, but only tibolone also prevents hot flushing.[224, 228–230]

We recommend 1 month of GnRH agonist treatment followed by agonist treatment combined with a daily, continuous add-back of estrogen and progestin using one of the available postmenopausal daily regimens. In view of the sensitivity of leiomyoma tissue to progestational agents, it makes sense to keep the dose of progestin relatively low. Preoperative GnRH agonist treatment is not indicated in every patient with uterine leiomyomas. The best candidates for treatment are women with bleeding and anemia to allow time for a response to iron supplementation and when the surgeon's clinical judgment suggests that a reduction in size may influence the choice of technique (e.g., laparoscopic or vaginal hysterectomy instead of laparotomy).

Treatment with a GnRH Antagonist

GnRH antagonist treatment can suppress pituitary-gonadal function without the initial stimulatory (flare) response observed with GnRH agonists. Results with depot Cetrorelix

preoperative treatment of uterine fibroids are similar to those with GnRH agonist treatment; however, the response is faster (a maximal reduction in size within 14 days), probably because there is no initial flare response.[231, 232]

Treatment with Mifepristone

Mifepristone, the progestin antagonist, effectively reduces the size of uterine leiomyomas and produces amenorrhea in most patients. The initial study was relatively short-term (12 weeks), and fibroid shrinkage was observed with doses of 25 and 50 mg daily.[181] A lower dosage is effective without the high rate of hot flushing observed at higher doses. In a 6-month study, a dose of 5 mg mifepristone daily was associated with a 48% reduction in uterine volume, a decrease in pressure and pain, an increase in hemoglobin levels, and a nonsignificant increase in hot flushing.[233] A similar reduction in uterine volume was observed in a 3-month study with the 5 mg dose, also with improvements in pain and bleeding.[234] However, long-term mifepristone treatment can result in endometrial hyperplasia, a consequence of the antiprogestin action of the drug. This endometrial effect makes mifepristone an unacceptable choice for on-going treatment of leiomyomas until large clinical trials are performed to establish its safety. Short-term treatment prior to surgery is appropriate. Asoprisnil, a progesterone receptor modulator, has also successfully reduced uterine volume and bleeding.[182] It is necessary to be cautious regarding the use of progesterone receptor modulators, as with progesterone antagonists, until endometrial safety is established.

Treatment with the Levonorgestrel-releasing Intrauterine System

When uterine enlargement because of leiomyomas is no greater than the size of a 12-week pregnancy, the insertion of a levonorgestrel-releasing intrauterine system is followed by a decrease in uterine size in many but not all patients and a dramatic reduction in menstrual blood loss, with 40% of patients achieving amenorrhea.[235–237] The contraceptive efficacy is not diminished, but expulsion rates are higher. This method of treatment is not recommended when distortion of the uterine cavity is evident on examination with ultrasonography. The beneficial effect of locally applied levonorgestrel is unexplained, contrasting with the studies that indicate growth promotion of myomas by progestins.

Treatment with Uterine Artery Embolization

Uterine artery embolization effectively reduces bleeding, pain, and fibroid size.[238–241] In a procedure under local anesthesia that takes about one hour, a catheter is advanced from the femoral artery to the uterine arteries to allow direct injection of polyvinyl particles or gelatin microspheres that occlude the blood flow. Myomas undergo necrosis in response to the transient ischemia, but normal tissue generates fibrinolysis and survives. The procedure is not recommended for large fibroids. After 5 years, recurrence of symptoms is about 10% to 25%. Most patients experience pain, nausea, and low-grade fever with a very high white blood count for 1 to 2 days following the procedure. In addition, serious complications occur, including complication-related hysterectomy, amenorrhea, premature menopause, septicemia from uterine infection, bowel obstruction, and pulmonary embolus. Several deaths have been reported, giving a rate comparable to that with hysterectomy. A significant number of patients with larger myomas acquire intra-abdominal adhesions after the procedure.[242] The general recommendation is that embolization should not be performed

in women who desire to retain their fertility. However, a substantial number of completed pregnancies have been reported after the procedure[243, 244]; nevertheless, the fertility rates and the complication rates after pregnancy is achieved are not known with certainty. A randomized comparison with myomectomy indicated a higher rate of infertility and miscarriages after embolization.[245]

Treatment with Ultrasound

A magnetic resonance mapping system for heat can be used to visualize myomas and direct high-energy ultrasound to destroy myomas.[246, 247] The temperature achieved produces instant necrosis within a limited volume of tissue, and, therefore, the method requires multiple treatments over several hours. Thermal injury to skin and normal tissues are potential side effects. Overall safety and long-term efficacy remain to be established; the early pregnancy experience in 51 women documented a 41% live birth rate and a 28% miscarriage rate.[248]

Transient uterine ischemia can be produced by placing vaginal clamps in the vaginal fornices, guided by ultrasonography, to compress the uterine arteries against the cervix for about 6 hours. Short-term studies have demonstrated efficacy comparable to embolization, but long-term follow-up data are not yet available.[249]

All references are available online at: http://www.clinicalgynendoandinfertility.com

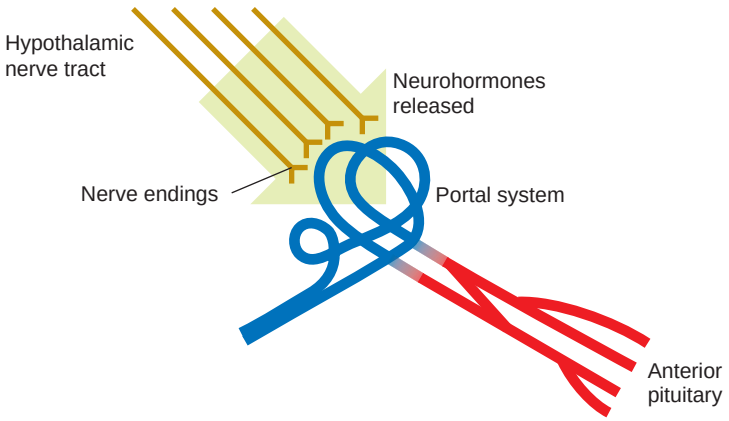
5 Neuroendocrinology

There are two major sites of action within the brain that are important in the regulation of reproductive function—the hypothalamus and the pituitary gland. In the past, the pituitary gland was viewed as the master gland. Then a new concept emerged in which the pituitary was relegated to a subordinate role as part of an orchestra, with the hypothalamus as the conductor, responding to both peripheral and central nervous system messages and exerting its influence by means of neurotransmitters transported to the pituitary by a portal vessel network. Regardless of which site was dominant, the conventional thesis was that the central nervous system-pituitary complex determined and directed the chronology of developmental events within a responsive ovary. However, there is now a new concept—the complex sequence of events known as the menstrual cycle is controlled by the sex steroids and peptides produced within the very follicle destined to ovulate. The hypothalamus and its direction, and the pituitary, are essential for the operation of the entire mechanism, but the endocrine function that leads to ovulation is brought about by endocrine feedback on the anterior pituitary.

A full understanding of this feature of reproductive biology will benefit the clinician who faces problems in gynecologic endocrinology. With this understanding, the clinician can comprehend the hitherto mysterious but significant effects of stress, diet, exercise, and other diverse influences on the pituitary-gonadal axis. Furthermore, we will be prepared to make advantageous use of the numerous neuropharmacologic agents that are the dividends of neuroendocrine research. To these ends, this chapter offers a clinically oriented review of the current status of reproductive neuroendocrinology.

Hypothalamic-Hypophyseal Portal Circulation

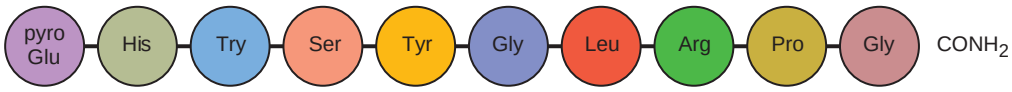
The hypothalamus is at the base of the brain just above the junction of the optic nerves. In order to influence the anterior pituitary gland, the brain requires a means of transmission or connection. A direct nervous connection does not exist. The blood supply of the anterior pituitary, however, originates in the capillaries that richly lace the median eminence area of the hypothalamus. The superior hypophyseal arteries form a dense network of capillaries within the median eminence, which then drain into the portal vessels that descend along the pituitary stalk to the anterior pituitary. The direction of the blood flow in this hypophyseal portal circulation is from the brain to the pituitary. Section of the neural stalk, which interrupts this portal circulation, leads to inactivity and atrophy of the gonads, along with a decrease in adrenal and thyroid activity to basal levels. With regeneration of the portal vessels, anterior pituitary function is restored. Thus, the anterior pituitary gland is under the influence of the hypothalamus by means of neurohormones released into this portal circulation. There also exists retrograde flow so that pituitary hormones can be delivered directly to the hypothalamus, creating the opportunity for pituitary feedback on the hypothalamus. An additional blood supply is provided by short vessels that originate in the posterior pituitary that in turn receives its arterial supply from the inferior hypophyseal arteries.



The Neurohormone Concept

The influence of the pituitary by the hypothalamus is achieved by materials secreted in the cells of the hypothalamus and transported to the pituitary by the portal vessel system. Indeed, pituitary cell proliferation and gene expression are controlled by hypothalamic peptides and their receptors. In addition to the stalk section experiments cited previously, transplantation of the pituitary to ectopic sites (e.g., under the kidney capsule) results in failure of gonadal function. With retransplantation to an anatomic site under the median eminence, followed by regeneration of the portal system, normal pituitary function is regained. This retrieval of gonadotropic function is not accomplished if the pituitary is transplanted to other sites in the brain. Hence, there is something very special about the blood draining the basal hypothalamus. An exception to this overall pattern of positive influence is the control of prolactin secretion. Stalk secretion and transplantation cause release of prolactin from the anterior pituitary, implying a negative, inhibitory hypothalamic control. Furthermore, cultures of anterior pituitary tissue release prolactin in the absence of hypothalamic tissue or extracts.

Gonadotropin releasing hormone—
a decapeptide



Neuroendocrine agents originating in the hypothalamus have positive stimulatory effects on growth hormone, thyroid-stimulating hormone (TSH), adrenocorticotropic hormone (ACTH), as well as the gonadotropins, and represent the individual neurohormones of the hypothalamus. The neurohormone that controls gonadotropins is called gonadotropin-releasing hormone (GnRH). The neurohormone that controls prolactin is called prolactin-inhibiting hormone and is dopamine. Human corticotropin-releasing hormone (CRH) is a 41 amino acid peptide that is the principal regulator of ACTH secretion, and that also activates the sympathetic nervous system. As we shall see, CRH can suppress gonadotropin secretion, an action partly mediated by endorphin inhibition of GnRH.

In addition to their effects on the pituitary, behavioral effects within the brain have been demonstrated for several of the releasing hormones. Thyrotropin-releasing hormone (TRH) antagonizes the sedative action of a number of drugs and also has a direct antidepressant effect in humans. GnRH evokes mating behavior in male and female animals.[1]

Initially, it was believed that there were two separate releasing hormones, one for follicle-stimulating hormone (FSH) and another for luteinizing hormone (LH). It is now accepted that there is a single neurohormone (GnRH) for both gonadotropins. GnRH is a small peptide with 10 amino acids with some variation in the amino acid sequence among various mammals. Purified or synthesized GnRH stimulates both FSH and LH secretion. The divergent patterns of FSH and LH in response to a single GnRH are due to the modulating influences of the endocrine environment, specifically the feedback effects of steroids on the anterior pituitary gland.

The classic neurotransmitters are secreted at the nerve terminal. Brain peptides require gene transcription, translation, and posttranslational processing, all within the neuronal cell body. The final product is transported down the axon to the terminal for secretion. Small neuroendocrine peptides share common large precursor polypeptides, called polyproteins or polyfunctional peptides. These proteins can serve as precursors for more than one biologically active peptide.

The gene that encodes for the 92 amino acid precursor protein for GnRH, pro-GnRH, is located on the short arm of chromosome 8.[2] The precursor protein for GnRH contains (in the following order) a 23 amino acid signal sequence, the GnRH decapeptide, a 3 amino acid proteolytic processing site, and a 56 amino acid sequence called GAP (GnRH-associated peptide).[3] GAP is a potent inhibitor of prolactin secretion as well as a stimulator of gonadotropins; however, a physiologic role for GAP has not been established.[4] Its primary role may be to provide appropriate conformational support for GnRH.

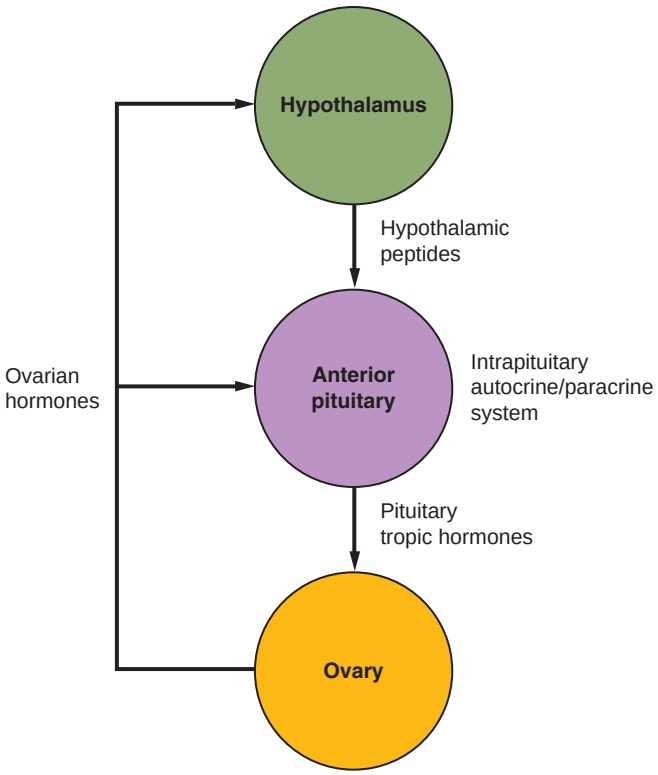
It is now apparent that GnRH has autocrine-paracrine functions throughout the body. It is present in both neural and nonneural tissues, and receptors are present in many extrapituitary tissues (e.g., the ovarian follicle and the placenta). Although GnRH is identical in all mammals, other nonmammalian forms exist, indicating that the GnRH molecule has existed for at least 500 million years.[5, 6] The central sequence, Tyr-Gly-Leu-Arg, is the nonconserved segment of GnRH, the segment with the most variability in other species. Accordingly, substitutions in this segment are well tolerated.

A second form of GnRH, known as GnRH-II, has been known to exist in many other species. GnRH-II consists of the following sequence: pGln-His-Trp-Ser-His-Gly-Trp

-Tyr-Pro-Gly. Prompted by its existence in other species, a search for its presence in humans was ultimately successful. A gene encoding GnRH-II is located on the human chromosome 20p13, obviously distinct from the GnRH-I gene on 8p11.2-p21.[7] Both genes produce a peptide with a signal sequence, a GnRH decapeptide, a proteolytic site, and a GAP. Human GnRH-II expression is highest outside the brain. An analysis of the evolution of GnRH indicates three major forms: GnRH localized to the hypothalamus (GnRH-I), forms in midbrain nuclei and outside the brain (GnRH-II), and forms in several fish species (GnRH-III), thus indicating appearance of the various forms before the emergence of vertebrates.[7] A total of 24 forms of GnRH have been identified in multiple species, but GnRH-I and GnRH-II are the primary GnRHs in mammals.[8] GnRH-I is the main form found in the brain, whereas GnRH-II is widely distributed in other organs.

A hypothalamic 12-amino acid peptide that inhibits pituitary secretion of gonadotropins was isolated from the brain of the Japanese quail.[9] This peptide was identified in mammals as well and labeled gonadotropin-inhibitory hormone (GnIH), but its presence and possible role in primate reproduction remain to be determined.[10]

Perhaps the notion that the pituitary is a master gland should not be discarded. Although it is highly regulated by input from other sites, its function is essential for sustaining life. The hormones from the pituitary gland regulate puberty, growth, reproduction, metabolism, osmotic balance, and responses to stress. Pituitary development and activity are under the control of the hypothalamus (with input from other central nervous system sites), and pituitary response is finely tuned by hormonal messages from tissues that are the targets of the pituitary trophic hormones. In addition, the pituitary has its own autocrine-paracrine system for enhancement and suppression of growth and function. But the pituitary gland is the focus for all of this activity, and this central, coordinating role is critical for normal life.



Prolactin Secretion

Prolactin gene expression occurs in the lactotropes of the anterior pituitary gland, in decidualized endometrium, and in the myometrium. The prolactin secreted in these various sites is identical, but there are differences in mRNA indicating differences in prolactin gene regulation.

Transcription of the prolactin gene is regulated by a transcription factor (a protein named Pit-1) that binds to the 5' promoter region and that is also necessary for growth hormone and TSH secretion.[11, 12] The expression of Pit-1 is in turn regulated by a transcription factor, Prop1 (Prophet of Pit-1); mutations in the Prop1 binding site in the Pit-1 gene result in deficiencies in multiple pituitary hormones.[12, 13] These mutations account for the majority of inherited cases of combined pituitary hormone deficiency. The remainder arise from mutations involving other transcription factors, specifically HESX1, LHX3, LHX4, TBX19, SOX2, AND SOX3.[14]

Prolactin gene transcription is regulated by the interaction of estrogen and glucocorticoid receptors with 5' flanking sequences. Mutations in the sequences of these flanking regions or in the gene for the Pit-1 protein can result in the failure to secrete prolactin. The Pit-1 gene is also involved in differentiation and growth of anterior pituitary cells; therefore, mutations in this gene can lead not only to absent secretion of growth hormone, prolactin, and TSH but also to an absence of their trophic cells in the pituitary—the result is significant hypopituitarism.[15] Molecular studies indicate that Pit-1 participates in mediating both stimulatory and inhibitory hormone signals for prolactin gene transcription. However, alterations in Pit-1 gene expression are not involved in pituitary tumor formation.[16]

The main function of prolactin in mammals is lactogenesis, whereas, in fish, prolactin is important for osmoregulation. The prolactin gene from the Chinook salmon contains coding sequences that are similar to those in mammals, and it is regulated similarly in the pituitary.[17] Pit-1, the pituitary specific transcription factor, therefore, appears to be highly conserved among species. Sexual arousal and orgasm in men and women produce marked elevations of circulating levels of prolactin that persist for at least 1 hour, perhaps contributing to the suppression of sexuality immediately after arousal and orgasm.[18, 19]

Prolactin gene expression is further regulated by other species-specific factors. Prolactin gene transcription is stimulated by estrogen and mediated by estrogen receptor binding to estrogen-responsive elements. This activation by estrogen requires interaction with Pit-1, in a manner not yet determined. Proximal promoter sequences are also activated by peptide hormones binding to cell surface receptors, e.g., TRH and growth factors. In addition, various agents that control cyclic AMP and calcium channels can stimulate or inhibit prolactin promoter activity.

Pituitary secretion of prolactin is chiefly, if not totally, under the inhibitory control of hypothalamic dopamine released into the portal circulation, a tonic inhibition that requires a high output of dopamine.[20] The action of dopamine in the pituitary is mediated by receptors that are coupled to the inhibition of adenylate cyclase activity. There are 5 forms of the dopamine receptor divided into 2 functional groups, D_1 and D_2, encoded by a single gene on chromosome 5 near the growth hormone receptor gene.[21, 22] The D_2 type is the predominant receptor in the anterior pituitary gland. The structure and function of the dopamine receptors are of the G protein system described in Chapter 2. Binding of dopamine to the receptor leads to suppression of adenylate cyclase and cyclic AMP maintenance of prolactin gene transcription and prolactin secretion. Other mechanisms are also activated, including suppression of intracellular calcium levels. Pit-1 binding sites are involved in this

dopamine response. In addition to direct inhibition of prolactin gene expression, dopamine binding to the D_2 receptor also inhibits lactotroph development and growth. These multiple effects of dopamine explain the ability of dopamine agonists to suppress prolactin secretion and the growth of prolactin-secreting pituitary adenomas. No activating or inactivating mutations of the dopamine receptors have been reported.

Several factors exert a stimulatory effect on prolactin secretion (prolactin-releasing factors), especially TRH, vasoactive intestinal peptide (VIP), epidermal growth factor, and perhaps GnRH. These factors interact with each other, affecting the overall lactotroph responsiveness; however, the only important clinical manifestation is the association of hyperprolactinemia with the elevation in TRH secretion that occurs with hypothyroidism. ***Prolactin homeostasis is regulated mainly by prolactin itself, feeding back on the dopamine-releasing neurons.***

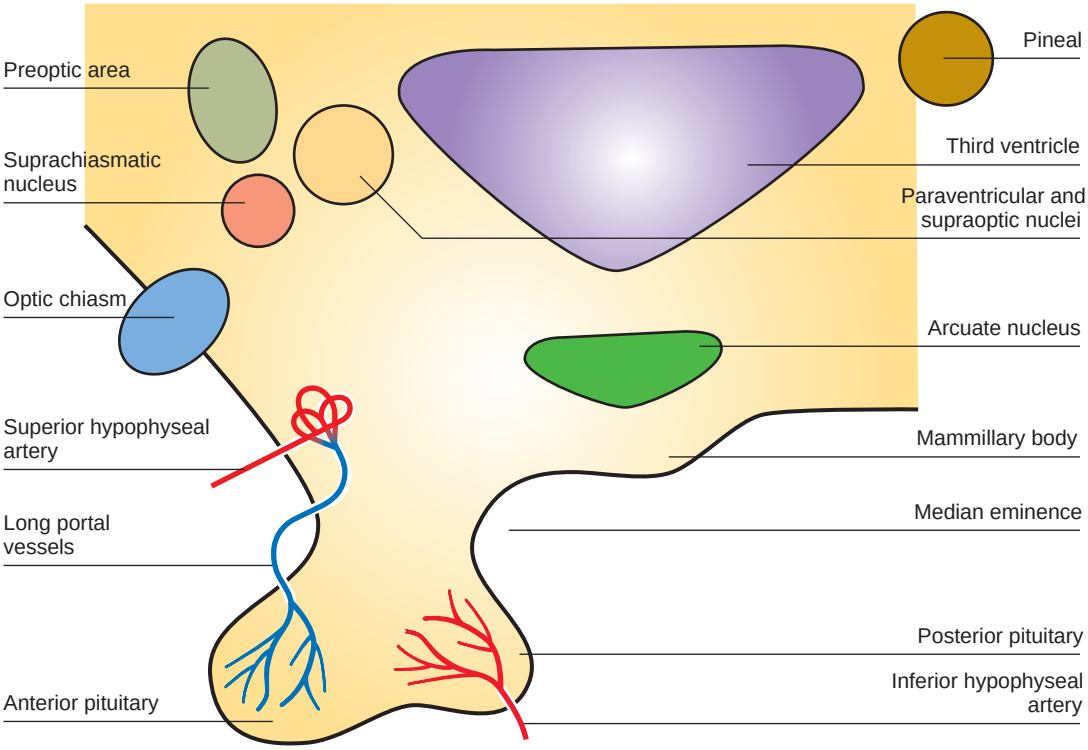
This dopaminergic mechanism is highly influenced by estrogen, either directly or through other neurotransmitters. Prolactin secretion can be understood by viewing dopamine received through the portal system as responsible for tonic inhibition. The dopaminergic system is stimulated by prolactin (decreasing secretion) and inhibited by estrogen (increasing secretion). Modulating influences include the inhibitory activity of endogenous opioids on dopamine release and stimulation by many substances, including serotonin and neuropeptide Y. Prolactin levels are highest during sleep.

Drugs used to treat psychological disorders are dopamine receptor antagonists. The inhibitory activity of dopamine on pituitary secretion of prolactin is blocked by these drugs, and prolactin levels increase in the circulation. Some of the newer drugs in this class do not affect prolactin secretion; these include clozapine, olanzapine, quetiapine, aripiprazole, and ziprasidone. Risperidone and amisulpride, however, act like the older drugs and increase prolactin secretion. The differences reflect the ability of each drug to cross the blood-brain barrier and the variations in affinity for the dopamine receptor.

The Hypothalamus and GnRH Secretion

The hypothalamus is the part of the diencephalon at the base of the brain that forms the floor of the third ventricle and part of its lateral walls. Within the hypothalamus are peptidergic neural cells that secrete the releasing and inhibiting hormones. These cells share the characteristics of both neurons and endocrine gland cells. They respond to signals in the bloodstream, as well as to neurotransmitters within the brain in a process known as neurosecretion. In neurosecretion, a neurohormone or neurotransmitter is synthesized on the ribosomes in the cytoplasm of the neuron, packaged into a granule in the Golgi apparatus, and then transported by active axonal flow to the neuronal terminal for secretion into a blood vessel or across a synapse.

The cells that produce GnRH originate from the olfactory area. By migration during embryogenesis, the cells move along cranial nerves connecting the nose and the forebrain to their primary location, where eventually 1,000–3,000 GnRH-producing cells can be found in the arcuate nucleus of the hypothalamus extending their axons to the median eminence.[23] The GnRH neurons appear in the medial olfactory placode (a thickened plate of ectoderm from which a sense organ develops) and enter the brain with the nervus terminalis, a cranial nerve that projects from the nose to the septal-preoptic nuclei in the brain.[24] This amazing journey accounts for Kallmann's syndrome, an association between an absence of GnRH and a defect in smell (a failure of both olfactory axonal and GnRH neuronal migration from the olfactory placode). Three modes of transmission have been documented: X-linked,

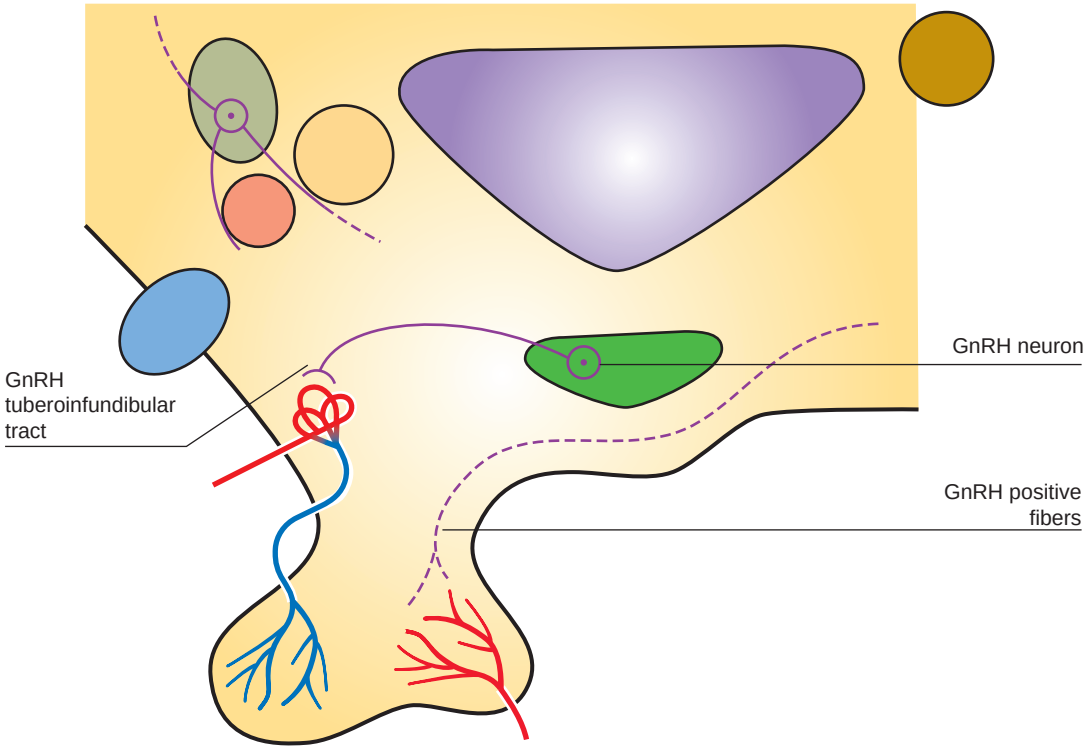


autosomal-dominant, and autosomal-recessive.[25] The 5–7-fold increased frequency in males indicates that X-linked transmission is the most common. The mutations responsible for this syndrome result in the failure to produce proteins with functions that are necessary for neuronal migration, anosmin-1 in X-linked forms, a protein that is homologous to members of the fibronectin family and responsible for cell adhesion and protease inhibition, and the receptors for fibroblast growth factor and prokineticin in the autosomal forms.[26–28] Normal GnRH neuron development and migration also depend on receptors that are tyrosine kinases, and abnormalites in these receptors might explain some clinical cases of GnRH deficiency.[29] Like olfactory epithelial cells in the nasal cavity, GnRH neurons have cilia.[30] The olfactory origin and the structural similarity of GnRH neurons and nasal epithelial cells suggest an evolution from reproduction controlled by pheromones.

Pheromones are airborne chemicals released by one individual that can affect other members of the same species. Odorless compounds obtained from the axillae of women in the late follicular phase of their cycles accelerated the LH surge and shortened the cycles of recipient women, and compounds from the luteal phase had the opposite effects.[31] This may be one mechanism by which women who are together much of the time have been reported to exhibit a synchrony in menstrual cycle timing.[32–36] However, the work on menstrual synchrony has been criticized, emphasizing that methodological problems have led to incorrect conclusions.[37]

In primates, GnRH cell bodies are primarily located within the medial basal hypothalamus.[38–40] Most of these cell bodies can be seen within the arcuate nucleus where GnRH is synthesized in GnRH neurons. The GnRH neurons exist in a complex network and are connected to each other and to many other neurons. This physical arrangement allows multiple interactions with neurotransmitters, hormones, and growth factors to modulate GnRH release. The delivery of GnRH to the portal circulation is via an axonal pathway, the GnRH tuberoinfundibular tract.

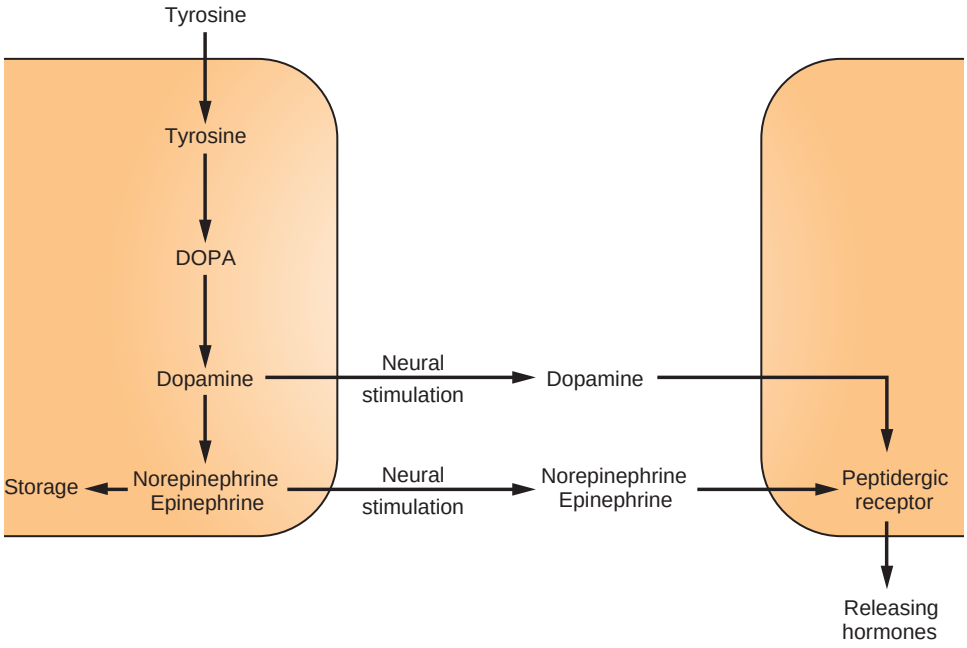
Fibers, identified with immunocytochemical techniques using antibodies to GnRH, can also be visualized in the posterior hypothalamus, descending into the posterior pituitary,



and in the anterior hypothalamic area projecting to sites within the limbic system.[38] Using hybridization techniques, messenger RNA for GnRH has been localized to the same sites previously identified by immunoreactivity. However, lesions that interrupt GnRH neurons projecting to regions other than the median eminence do not affect gonadotropin release. Only lesions of the arcuate nucleus in the monkey lead to gonadal atrophy and amenorrhea.[41] Therefore, the arcuate nucleus can be viewed with the median eminence as a unit, the key locus within the hypothalamus for GnRH secretion into the portal circulation. The other GnRH neurons may be important for a variety of behavioral responses.

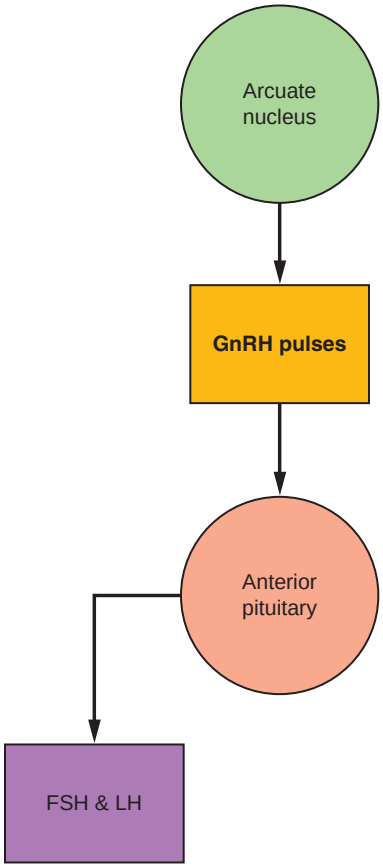
GnRH Secretion

The half-life of GnRH is only 2–4 minutes. Because of this rapid degradation, combined with the enormous dilution on entry into the peripheral circulation, biologically effective amounts of GnRH do not escape the portal system. Therefore, control of the reproductive cycle depends on constant release of GnRH. This function, in turn, depends on the complex and coordinated interrelationships among this releasing hormone, other neurohormones, the pituitary gonadotropins, and the gonadal steroids. The interplay among these substances is governed by feedback effects, both positive stimulatory and negative inhibitory. ***The long feedback loop*** refers to the feedback effects of circulating levels of target gland hormones, and this occurs both in the hypothalamus and the pituitary. ***The short feedback loop*** indicates a negative feedback of pituitary hormones on their own secretion, presumably via inhibitory effects on releasing hormones in the hypothalamus. ***Ultrashort feedback*** refers to inhibition by the releasing hormone on its own synthesis. These signals as well as signals from higher centers in the central nervous system may modify GnRH secretion through an array of neurotransmitters, primarily dopamine, norepinephrine, and endorphin but also serotonin and melatonin. Dopamine and norepinephrine are synthesized in the nerve terminals by decarboxylation of dihydroxyphenylalanine (DOPA), which in turn is synthesized by hydroxylation of tyrosine. Dopamine is the immediate precursor of



norepinephrine and epinephrine, but dopamine itself functions as a key neurotransmitter in the hypothalamus and the pituitary.[20]

A most useful concept is to view the arcuate nucleus as the central site of action, releasing GnRH into the portal circulation in pulsatile fashion. In a classic series of experiments, it was demonstrated that normal gonadotropin secretion requires pulsatile GnRH



discharge within a critical range in frequency and amplitude.[42] Even pituitary hormone gene transcription is sensitive to the pulsatile nature of GnRH release.[43]

Experimental manipulations have indicated that the critical range of GnRH pulsatile secretion is rather narrow. The administration (to monkeys) of 1 μg GnRH per minute for 6 minutes every hour (one pulse per hour) produces a portal blood concentration about equal to the peak concentration of GnRH in human portal blood, about 2 ng/mL. Increasing the frequency to 2 and 5 pulses per hour extinguishes gonadotropin secretion. A similar decline in gonadotropin secretion is obtained by increasing the dose of GnRH. Decreasing the pulse frequency decreases LH secretion but increases FSH secretion.

Like GnRH, gonadotropins are also secreted in pulsatile fashion, and, indeed, the pulsatile pattern of gonadotropin release reflects the pulsatile GnRH pattern.[44, 45] GnRH and gonadotropin secretion are always pulsatile in nature, but an augmentation of the pulsatile pattern of gonadotropin secretion occurs just before puberty with nighttime increases in LH. After puberty, enhanced pulsatile secretion is maintained throughout the 24-hour period, but it varies in both amplitude and frequency. In puberty, arcuate activity begins with a low frequency of GnRH release and proceeds through a cycle of acceleration of frequency, characterized by passage from relative inactivity, to nocturnal activation, to the full adult pattern. The progressive changes in FSH and LH reflect this activation of GnRH pulsatile secretion. Ovarian steroid release is also pulsatile, coordinated with LH pulses, the major stimulator of ovarian steroidogenesis.[46] In the absence of ovarian regulation, the GnRH pulse frequency is approximately one pulse per hour.[47]

Timing of GnRH Pulses

The measurement of LH pulses is used as an indication of GnRH pulsatile secretion (the long half-life of FSH precludes its use for this purpose).[48] The characteristics of LH pulses (and presumably of GnRH pulses) during the menstrual cycle are as follows:[46, 49, 50]

LH Pulse Mean Amplitude:

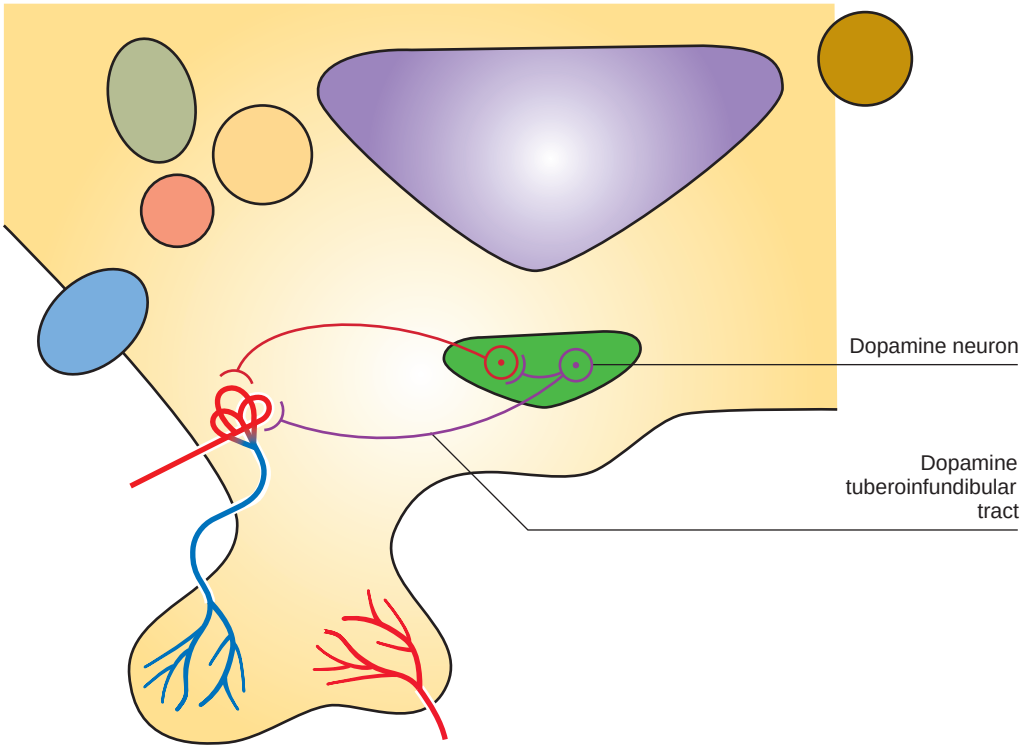
Early follicular phase	6.5 IU/L.
Midfollicular phase	5.0 IU/L.
Late follicular phase	7.2 IU/L.
Early luteal phase	15.0 IU/L.
Midluteal phase	12.2 IU/L.
Late luteal phase	8.0 IU/L.

LH Pulse Mean Frequency:

Early follicular phase	90 minutes.
Late follicular phase	60–70 minutes.
Early luteal phase	100 minutes.
Late luteal phase	200 minutes

Pulsatile secretion is more frequent but lower in amplitude during the follicular phase compared with the luteal phase. The slowing of GnRH pulse frequencies in the late luteal phase is an important change, favoring FSH synthesis and secretion; therefore, allowing the rise in FSH essential for the next cycle.[51] An increase in frequency and amplitude in mid-cycle GnRH pulsatile secretion favors the surge of LH necessary for ovulation and the beginning of the luteal phase.

It should be emphasized that these numbers are not inviolate. There is considerable variability between and within individuals, and a wide normal range exists.[52] Despite the



handicap of the long half-life, it has been ascertained that FSH secretion is correlated with LH secretion. The changes in amplitude are relatively small; therefore, increasing and decreasing circulating levels of the gonadotropins are affected mostly by changes in pulse frequency. During the luteal-follicular transition, pulse frequency increases approximately 4.5-fold.[50] In rodents, the frequency of GnRH pulses determines which gonadotropin subunit gene will be preferentially expressed.[53]

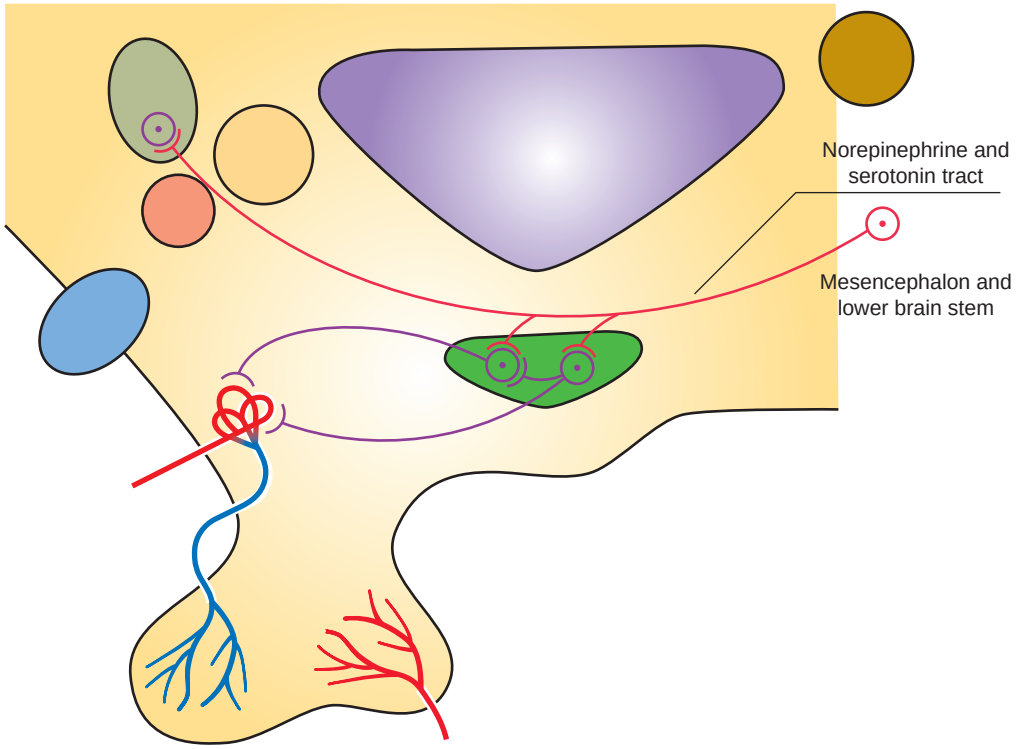
The anterior pituitary gland and GnRH neurons have an intrinsic pulsatile pattern. Although pulses of significant amplitude are linked to GnRH, small-amplitude pulses of high frequency represent spontaneous secretion from the anterior pituitary (at least as demonstrated in isolated pituitary glands in vitro).[54] It is not known whether this has any importance physiologically, and, at the present time, the pituitary secretory pattern is thought to reflect GnRH. The pulsatile secretion of GnRH correlates with episodic *GnRH-I* gene expression in the hypothalamus.[8] A promoter site in the *GnRH-I* gene is responsible for the pulsatile nature of the secretions, regulated by the usual array of transcription factors.[55]

Control of GnRH Pulses

Normal menstrual cycles require the maintenance of the pulsatile release of GnRH within a critical range of frequency and amplitude. Pulsatile, rhythmic activity is an intrinsic property of GnRH neurons, and the effect of various hormones and neurotransmitters must be viewed as modulating actions.[56]

The Dopamine Tract

Cell bodies for dopamine synthesis can be found in the arcuate and periventricular nuclei. The dopamine tuberoinfundibular tract arises within the medial basal hypothalamus and its



short axons terminate in the median eminence; it provides the major dopaminergic effect on the pituitary.

The administration of dopamine by intravenous infusion to men and women is associated with a suppression of circulating prolactin and gonadotropin levels.[57] Dopamine does not exert a direct effect on gonadotropin secretion by the anterior pituitary; thus, this effect is mediated through GnRH release in the hypothalamus. Dopamine is directly secreted into the portal blood, thus behaving like a neurohormone. Therefore, dopamine may directly suppress arcuate GnRH activity and also be transported via the portal system to directly and specifically suppress pituitary prolactin secretion. The hypothalamic tuberoinfundibular dopamine pathway is not the only dopamine pathway in the CNS, and it is only one of two major dopamine pathways in the hypothalamus. But it is this pathway that directly participates in the regulation of prolactin secretion. In addition, prolactin delivered to the intermediate lobe of the pituitary suppresses melanocyte-stimulating hormone release.

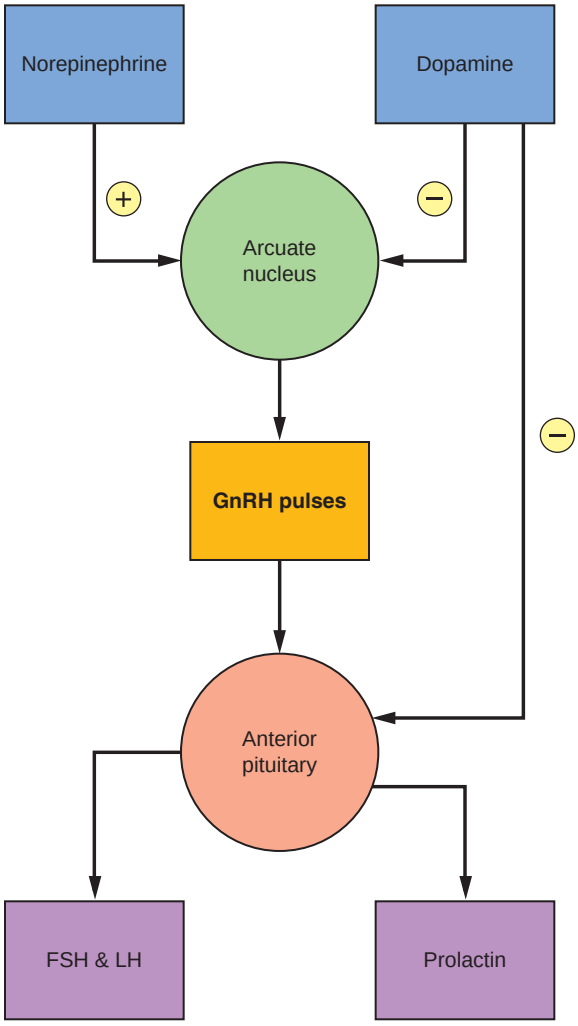
The Norepinephrine Tract

Most of the cell bodies that synthesize norepinephrine are located in the mesencephalon and lower brainstem. These cells also synthesize serotonin. Axons for amine transport ascend into the medial forebrain bundle to terminate in various brain structures including the hypothalamus.

The biogenic catecholamines modulate GnRH pulsatile release.[58] Norepinephrine is thought to exert stimulatory effects on GnRH, whereas dopamine and serotonin exert inhibitory effects. For an understanding of clinical problems, it is best to view dopamine as an inhibitor of both GnRH and prolactin. Little is known, however, about the role of serotonin. The probable mode of action of catecholamines is to influence the frequency (and perhaps the amplitude) of GnRH discharge. Thus, pharmacologic or psychologic factors that affect pituitary function probably do so by altering catecholamine synthesis or metabolism and, thus, the pulsatile release of GnRH.

Neuropeptide Y

Neuropeptide Y is an important peptide in the mechanism by which leptin and insulin inform the hypothalamus about the nutritional state of an individual. The secretion and gene expression of neuropeptide Y in hypothalamic neurons are regulated by gonadal steroids.[59] Neuropeptide Y stimulates appetite and the pulsatile release of GnRH; in the pituitary it potentiates gonadotropin response to GnRH.[60] It, thus, may facilitate pulsatile secretion of GnRH and gonadotropins. In the absence of estrogen, neuropeptide Y inhibits gonadotropin secretion. Because undernutrition is associated with an increase in neuropeptide Y (see Chapter 19) and increased amounts have been measured in cerebrospinal fluid of women with anorexia and bulimia nervosa, neuropeptide Y is viewed as at least one link between nutrition and reproductive function.[61, 62] The level of the neuroendocrine activity involved in reproduction is sensitive to an individual's energy state, or more simply, to the availability of sufficient body fuel to support reproduction.



Kisspeptins

Kisspeptins are peptides encoded by the gene *Kiss1* that is expressed in the hypothalamus. Kisspeptin and its G-protein receptor, GPR54, are essential for the normal development of puberty.[63] This signaling pathway is involved in the activation of GnRH neurons and GnRH secretion.[64, 65] The kisspeptin neurons contain estrogen and progesterone receptors, and are

involved not only in puberty but in the changes associated with ovulatory cycles, in at least one pathway by which the sex steroids affect GnRH secretion.

Pituitary Gonadotropin Secretion

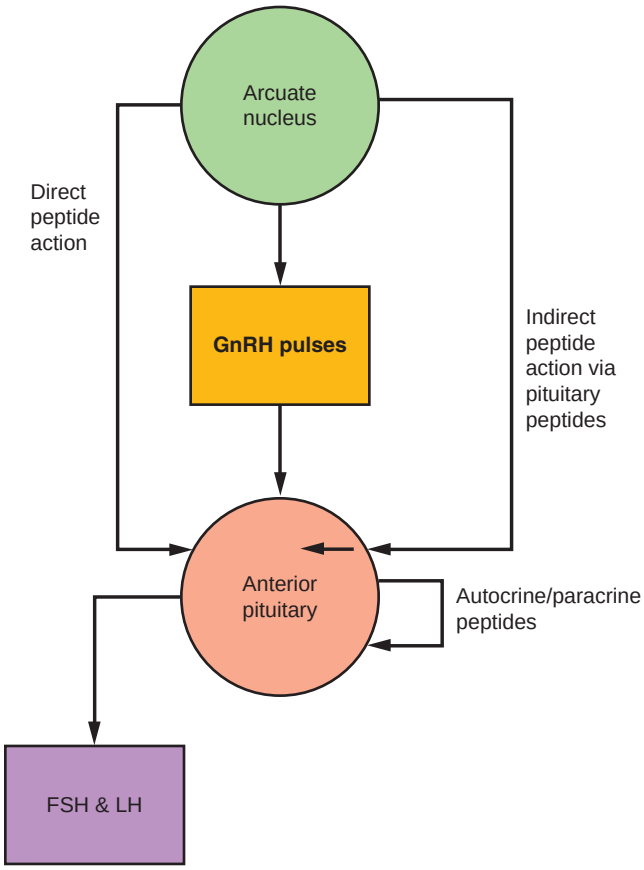
The gene for the α subunit of the gonadotropins is expressed in both the pituitary and placenta. The β subunit for human chorionic gonadotropin (hCG) is expressed in the placenta but only minimally (and with alterations in structure) in the pituitary, whereas the LH β subunit, as expected, is expressed in the pituitary but not significantly in the placenta.[66, 67] Studies of gonadotropin gene expression confirm the relationships established by earlier studies. The sex steroids decrease and castration increases the rate of gonadotropin gene transcription as reflected by the levels of specific messenger RNAs. In addition, the sex steroids can act at the membrane level, affecting the interaction of GnRH with its receptor.[68]

Both LH and FSH are secreted by the same cell, the gonadotrope, localized primarily in the lateral portions of the pituitary gland and responsive to the pulsatile stimulation by GnRH. GnRH is calcium dependent in its mechanism of action and utilizes inositol 1,4,5-triphosphate (IP_3) and 1,2-diacylglycerol (1,2-DG) as second messengers to stimulate protein kinase and cyclic AMP activity (Chapter 2).[69] These responses require a G protein receptor and are associated with cyclical release of calcium ions from intracellular stores and the opening of cell membrane channels to allow entry of extracellular calcium. Thus, calmodulin, protein kinase, and cyclic AMP are mediators of GnRH action. Gonadotrope gene transcription is mediated by several transcription factors, providing a mechanism by which the various subunits of FSH and LH can be synthesized and secreted by the same cell.[70] The rate-limiting step for the secretion of FSH and LH is the synthesis of the beta-subunits of each gonadotropin.

The GnRH type I receptor, a member of the G protein family, is encoded by a gene on chromosome 14q13.1-q21.1.[71, 72] The location of the type II receptor is uncertain; in the marmoset, it is on chromosome 1q.[73] The exact roles of the two GnRHs and GnRH receptors in the human are yet to be determined. GnRH receptors are regulated by many agents, including GnRH itself, inhibin, activin, and the sex steroids.[74] The number of GnRH receptors available is significantly regulated by the frequency of GnRH pulses. The signaling pathways include the induction and modification of stimulatory and inhibitory transcription factor proteins. No mutations in the GnRH gene have been described in patients with hypogonadotropic hypogonadism; however, multiple mutations in the GnRH receptor gene have been reported.

Synthesis of gonadotropins takes place on the rough endoplasmic reticulum. The hormones are packaged into secretory granules by the Golgi cisternae of the Golgi apparatus and then stored as secretory granules. Secretion requires migration (activation) of the mature secretory granules to the cell membrane where an alteration in membrane permeability results in extrusion of the secretory granules in response to GnRH. The rate-limiting step in gonadotropin synthesis is the GnRH-dependent availability of the beta subunits.

Binding of GnRH to its receptor in the pituitary activates multiple messengers and responses. The immediate event is a secretory release of gonadotropins, whereas delayed responses prepare for the next secretory release. One of these delayed responses is the self-priming action of GnRH that leads to even greater responses to subsequent GnRH pulses due to a complex series of biochemical and biophysical intracellular events. This self-priming action is important to achieve the large surge in secretion at midcycle; it requires estrogen exposure, and it can be augmented by progesterone. This important action of progesterone depends on estrogen exposure (for an increase in progesterone receptors) and activation of the progesterone receptor by GnRH-stimulated phosphorylation. This latter action is an example of cross-talk between peptide and steroid hormone receptors.



Five different types of secretory cells coexist within the anterior pituitary gland: gonadotropes, lactotropes, thyrotropes, somatotropes, and corticotropes. Autocrine and paracrine interactions combine to make anterior pituitary secretion subject to more complicated control than simply reaction to hypothalamic-releasing factors and modulation by feedback signals. Substantial experimental evidence exists to indicate stimulatory and inhibitory influences of various substances on the pituitary secretory cells. Although the GnRH system is a primary mechanism, other hypothalamic peptides can influence GnRH secretion. Peptides can interact with GnRH at the pituitary; peptides can be transported to the pituitary gland where they can directly affect the gonadotropes (e.g., oxytocin, CRH, and neuropeptide Y) or indirectly have an effect on FSH and LH secretion by stimulating the release of active substances within the pituitary (e.g., glalanin, interleukins). Autocrine-paracrine activities involve peptides synthesized by pituitary cells.

The Intrapituitary Autocrine-Paracrine System

Intrapituitary cytokines and growth factors provide an autocrine-paracrine system for regulating pituitary cell development and replication as well as pituitary hormone synthesis and secretion. The pituitary contains the familiar cast of substances encountered in organs throughout the body, including the interleukins, epidermal growth factor, fibroblast growth factors, the insulin-like growth factors, nerve growth factor, activin, inhibin, endothelin, and many others.[75–77] As in most tissues, the interaction among these substances is a complex story, but the activin-inhibin mechanism deserves emphasis.

Activin, Inhibin, and Follistatin

Activin and inhibin are peptide members of the transforming growth factor-β family.[78] Inhibin consists of two dissimilar peptides (known as alpha and beta subunits) linked by disulfide bonds. Two forms of inhibin (inhibin-A and inhibin-B) have been purified, each containing an identical alpha subunit and distinct but related beta subunits. Thus, there are three subunits for inhibins: alpha, beta-A, and beta-B. Each subunit is a product of different messenger RNA; therefore, each is derived from its own large precursor molecule.

Inhibin is secreted by granulosa cells, but messenger RNA for the alpha and beta chains has also been found in pituitary gonadotropes.[79] Inhibin selectively inhibits FSH but not LH secretion. Indeed, while suppressing FSH synthesis, inhibin may enhance LH activity.[80, 81] Cells actively synthesizing LH respond to inhibin by increasing GnRH receptor number; FSH dominant cells are suppressed by inhibin.[80] Inhibin has little or no effect on growth hormone, ACTH, and prolactin production. The mechanism of FSH inhibition may be secondary to inhibin competing with activin for the activin receptor.

Activin, also derived from granulosa cells, but present as well in the pituitary gonadotropes, contains two subunits that are identical to the beta subunits of inhibins A and B. In addition, activins have been identified with variants of the beta subunits, designated as beta-C, beta-D, and beta-E.[82] The activin beta-C and beta-E genes have been demonstrated to be nonessential in mouse knockout models.[83] Activin augments the secretion of FSH and inhibits prolactin, ACTH, and growth hormone responses.[81, 84–86] Activin increases pituitary response to GnRH by enhancing GnRH receptor formation.[74, 87] The effects of activin are blocked by inhibin and follistatin.[88] The roles for inhibin and activin in regulating the events of the menstrual cycle are discussed in Chapter 6. In contrast to inhibin's major role in suppressing FSH secretion, activin has a broad range of activities, involving bone, neurons, wound healing, and autocrine-paracrine functions in many organs.

The Forms of Inhibin

Inhibin-A:	Alpha-BetaA
Inhibin-B:	Alpha-BetaB

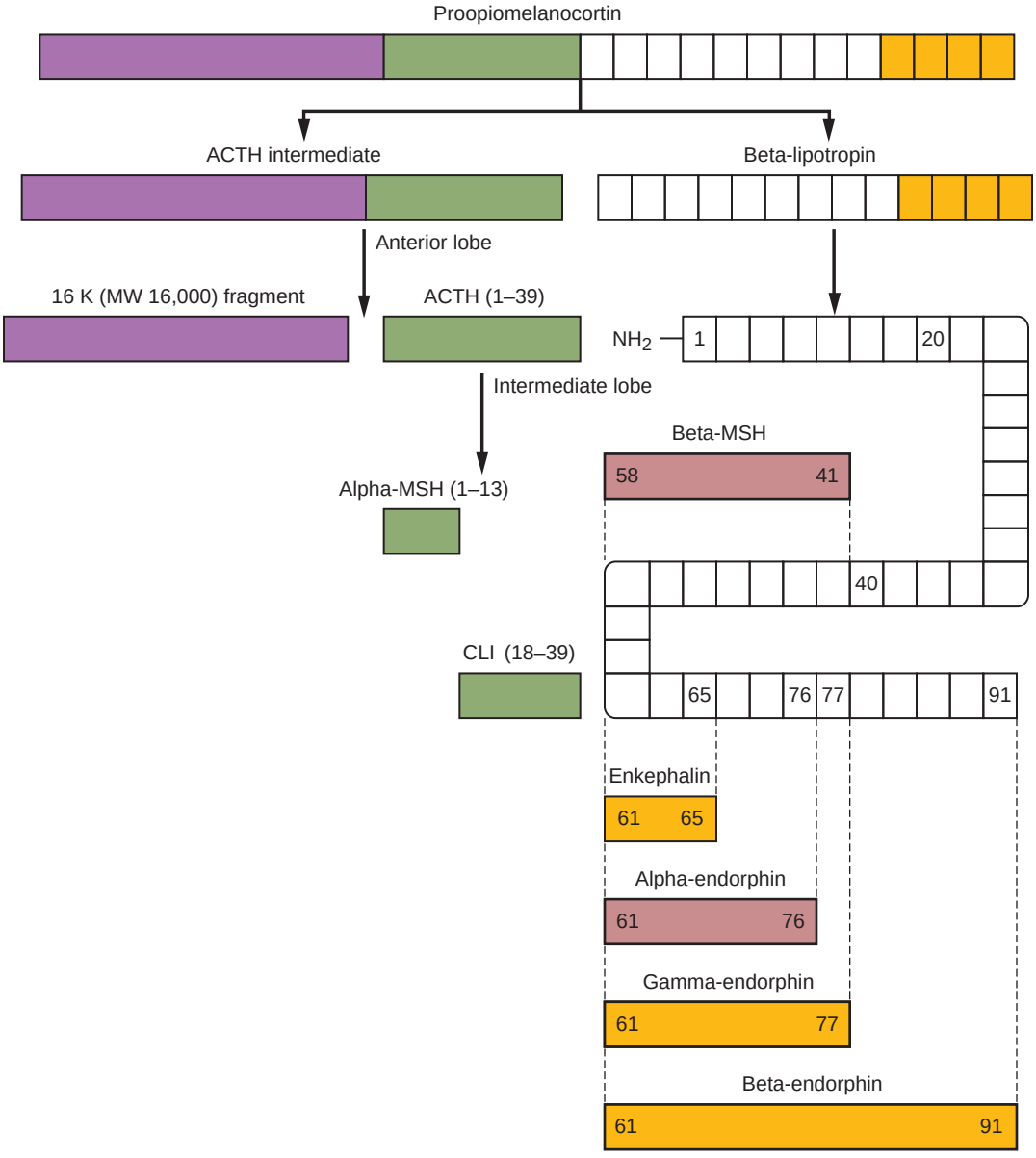
The Forms of Activin

Activin-A:	BetaA-BetaA
Activin-AB:	BetaA-BetaB
Activin-B:	BetaB-BetaB
Activin-C:	$Beta_C$-$Beta_C$
Activin-AC:	$Beta_A$-$Beta_C$
Activin-E:	$Beta_E$-$Beta_E$

Follistatin is a peptide secreted by a variety of pituitary cells, including the gonadotropes.[89] This peptide has also been called FSH-suppressing protein because of its main action: inhibition of FSH synthesis and secretion and the FSH response to GnRH, by binding to activin and in that fashion decreasing the activity of activin.[90, 91] Activin stimulates follistatin production, and inhibin prevents this response.

In summary, GnRH stimulates gonadotropin synthesis and secretion, as well as activin, inhibin, and follistatin. Activin enhances and follistatin suppresses GnRH activity. Evidence in vivo and in vitro indicates that gonadotropin response to GnRH requires activin activity, and gonadotropin response can be blocked by follistatin.[92–94] This relationship contributes to the down-regulation of pituitary gonadotropin secretion by

prolonged GnRH stimulation. Increasing GnRH pulsatile frequency first increases FSH production, and, then with high frequency or continuous GnRH stimulation, follistatin production is increased.[92] Selective synthesis and secretion of FSH can be explained by the decrease in the inhibiting factors, inhibin and follistatin, allowing activin to enhance the actions of GnRH, involving the transcription factors that promote FSH beta subunit expression.[95] LH secretion is primarily regulated by GnRH, without involvement of the inhibin-activin-follistatin system.



The Endogenous Opiates

The most fascinating peptide group is the endogenous opioid peptide family.[96] β-Lipotropin is a 91 amino acid molecule that was first isolated from the pituitary in 1964. Its function

remained a mystery for more than 10 years until receptors for opioid compounds were identified, and, by virtue of their existence, it was postulated that endogenous opioid compounds must exist and serve important physiologic roles. Endorphin was a word coined to denote morphine-like action and endogenous origin in the brain.

Opiate production is regulated by gene transcription and the synthesis of precursor peptides and at a posttranslational level where the precursors are processed into the various bioactive smaller peptides.[97] All opiates derive from one of three precursor peptides.

Proopiomelanocortin (POMC)—the source of endorphins.
Proenkephalin A and B—the source of several enkephalins.
Prodynorphin—yields dynorphins.

POMC was the first precursor peptide to be identified. It is made in the anterior and intermediate lobes of the pituitary; in the hypothalamus and other areas of the brain; in the sympathetic nervous system; and in other tissues including the gonads, the placenta, the gastrointestinal tract, and the lungs. The highest concentration is in the pituitary gland.

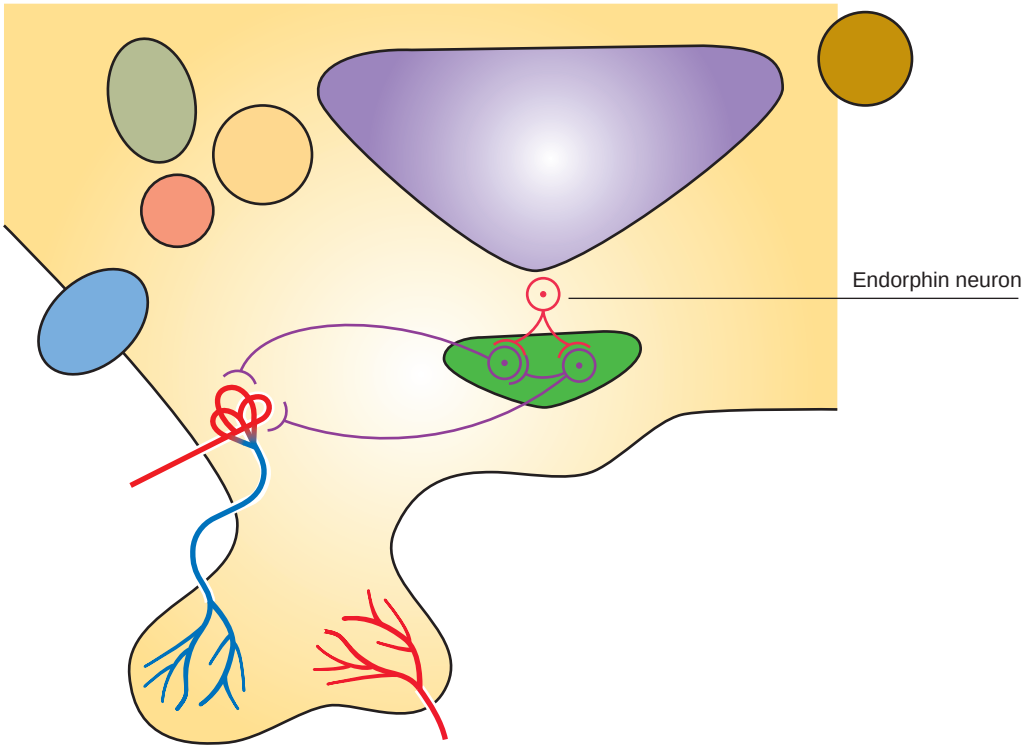
Proopiomelanocortin is split into two fragments, an ACTH intermediate fragment and β-lipotropin. β-Lipotropin has no opioid activity but is broken down in a series of steps to β-melanocyte-stimulating hormone (β-MSH), enkephalin; and α-, γ-, and β-endorphins. Melanocyte-stimulating hormone acts in lower animals to stimulate melanin granules within cells, causing darkening of the skin. In humans, there is no known function.

Enkephalin and the α- and γ-endorphins are as active as morphine on a molar basis, whereas β-endorphin is 5–10 times more potent. In the adult pituitary gland, the major products are ACTH and β-lipotropin, with only small amounts of endorphin. Thus, ACTH and β-lipotropin blood levels show similar courses, and they are major secretion products of the anterior pituitary in response to stress. In the intermediate lobe of the pituitary (which is prominent only during fetal life), ACTH is cleaved to CLIP (corticotropin-like intermediate lobe peptide) and β-MSH. In the placenta and adrenal medulla, POMC processing yields α-MSH-like and β-endorphin peptides. β-Endorphin has also been detected in the ovaries and in the testes.

In the brain, the major products are the opiates, with little ACTH. In the hypothalamus the major products are β-endorphin and α-MSH in the region of the arcuate nucleus and the ventromedial nucleus. ***The pituitary opiate system is a system for secretion into the circulation whereas the hypothalamic opiate system allows for distribution via axons to regulate other brain regions and the pituitary gland***.

β-Endorphin is appropriately considered a neurotransmitter, a neurohormone, and a neuromodulator. β-Endorphin influences a variety of hypothalamic functions, including regulation of reproduction, temperature, and cardiovascular and respiratory function, as well as extrahypothalamic functions such as pain perception and mood. POMC gene expression in the anterior pituitary is controlled mainly by corticotropin-releasing hormone and influenced by the feedback effects of glucocorticoids. In the hypothalamus, regulation of POMC gene expression is via the sex steroids.[98] In the absence of sex steroids, little, if any, secretion occurs.

Proenkephalin A is produced in the adrenal medulla, the brain, the posterior pituitary, the spinal cord, and the gastrointestinal tract. It yields several enkephalins: methionine-enkephalin, leucine-enkephalin, and other variants. The enkephalins are the most widely distributed endogenous opioid peptides in the brain and are probably mainly involved



as inhibitory neurotransmitters in the modulation of the autonomic nervous system. Prodynorphin, found in the brain (concentrated in the hypothalamus) and the gastrointestinal tract, yields dynorphin, an opioid peptide with high analgesic potency and behavioral effects, as well as α-neoendorphin, β-neoendorphin, and leumorphin. The last 13 amino acids of leumorphin constitute another opioid peptide, rimorphin. The prodynorphin products probably function in a fashion similar to endorphin.

It is simpler to say that there are three classes of opiates: enkephalin, endorphin, and dynorphin.

Opioid peptides are able to act through different receptors, although specific opiates bind predominantly to one of the various receptor types. Naloxone, used in most human studies, does not bind exclusively to any one receptor type, and, thus, results with this antagonist are not totally specific. Localization of opioid receptors explains many of the pharmacologic actions of the opiates. Opioid receptors are found in the nerve endings of sensory neurons, in the limbic system (site of euphoric emotions), in brainstem centers for reflexes such as respiration, and widely distributed in the brain and the spinal cord.

Opioid Peptides and the Menstrual Cycle

The opioid tone is an important part of menstrual function and cyclicity.[99] Although estradiol alone increases endorphin secretion, the highest levels of endorphin occur with sequential therapy of both estradiol and progesterone (in ovariectomized monkeys). Endogenous endorphin levels, therefore, increase throughout the cycle from nadir levels during menses to highest levels during the luteal phase. Normal cyclicity, thus, requires sequential periods of high (luteal phase) and low (during menses) hypothalamic opioid activity.

A reduction in LH pulse frequency is linked to increased endorphin release.[100] Naloxone increases both the frequency and the amplitude of LH pulses. ***Thus, the endogenous opiates inhibit gonadotropin secretion by suppressing the hypothalamic release of GnRH.*** Opiates have no effect on the pituitary response to GnRH. The gonadal steroids modify endogenous opioid activity, and the negative feedback of steroids on gonadotropins appears to be mediated by endogenous opiates. Because the fluctuating levels of endogenous opiates in the menstrual cycle are related to the changing levels of estradiol and progesterone, it is attractive to speculate that the sex steroids directly stimulate endogenous opioid receptor activity. There is an absence of opioid effect on postmenopausal and oophorectomized levels of gonadotropins, and the response to opiates is restored with the administration of estrogen, progesterone, or both.[101] Both estrogen and progesterone alone increase endogenous opiates, but estrogen enhances the action of progesterone, which could explain the maximal suppression of GnRH and gonadotropin pulse frequency during the luteal phase.[102, 103] In pubertal boys and girls, however, naloxone could not prevent the suppression of LH by estradiol administration, indicating that in this circumstance estradiol may directly inhibit GnRH secretion.[104, 105] Nevertheless, the evidence overall indicates that endogenous opiates exert an inhibiting influence over GnRH secretion.[106] The negative feedback of progesterone on GnRH secretion (the major mechanism for the inhibition of ovulation associated with progestin contraception) is definitely partly mediated by endogenous opiates but also by other neural mechanisms not yet determined.[107]

The inhibiting tone of endogenous opiates is reduced at the time of the ovulatory surge, allowing a release from suppression.[108] This is probably a response to estrogen, specifically an estrogen-induced decrease in opioid receptor binding and opioid release.[109, 110]

Experiments with naloxone administration suggest that the suppression of gonadotropins during pregnancy and the recovery during the postpartum period reflect steroid-induced opioid inhibition, followed by a release from central opioid suppression.

The principal endogenous opiates affecting GnRH release are β-endorphin and dynorphin, and it is probable that the major effect is modulation of the catecholamine pathway, principally norepinephrine. The action does not involve dopamine receptors, acetylcholine receptors, or alpha-adrenergic receptors. On the other hand, endorphin may affect GnRH release directly, without the involvement of any intermediary neuroamine.

Because α-MSH counteracts the effects of β-endorphin, posttranslational processing of POMC can affect hypothalamic-pituitary function by altering the amounts of α-MSH and β-endorphin.[111] This introduces another potential site for neuroendocrine regulation of reproductive function. Gonadal hormones likely have multiple sites for feedback signals.

Clinical Implications

A change in opioid inhibitory tone is not important in the changes of puberty because the responsiveness to naloxone does not develop until after puberty. A change in opioid tone does seem to mediate the hypogonadotropic state seen with elevated prolactin levels, exercise, and other conditions of hypothalamic amenorrhea, whereas endogenous opioid inhibition does not seem to play a causal role in delayed puberty or hereditary problems such as Kallmann's syndrome.[112, 113] Treatment of patients with hypothalamic amenorrhea (suppressed GnRH pulsatile secretion) with a drug (naltrexone) that blocks opioid receptors restores normal function (ovulation and pregnancy).[114] Thus, the reduced GnRH secretion associated with hypothalamic amenorrhea is mediated by an increase in endogenous opioid inhibitory tone.

Experimental evidence indicates that corticotropin-releasing hormone (CRH) inhibits hypothalamic GnRH secretion, both directly and by augmenting endogenous opioid secretion. Women with hypothalamic amenorrhea demonstrate hypercortisolism, suggesting that this is the pathway by which stress interrupts reproductive function.[115] Mathematical analysis of the associations among FSH, LH, β-endorphin and cortisol pulses support the existence of significant functional coupling between the neuroregulatory systems that control the gonadal and adrenal axes.[116] The CRH gene contains two segments that are similar to estrogen response elements, allowing estrogen enhancement of CRH activity, perhaps explaining the greater vulnerability of the reproductive axis to stress in females.[117] Besides the CRH-induced inhibition of GnRH release, the increase in cortisol generated by CRH stimulation of pituitary ACTH secretion also contributes to the suppression of reproduction; cortisol directly inhibits pituitary responsiveness to GnRH.[118]

Cumming concluded that most studies indicate an exercise-induced increase in endogenous opiates, but a significant impact on mood remains to be substantiated.[119] He noted that *runners' high* is more common in California than in Canada (euphoria is hard to come by when running in below freezing temperatures!).

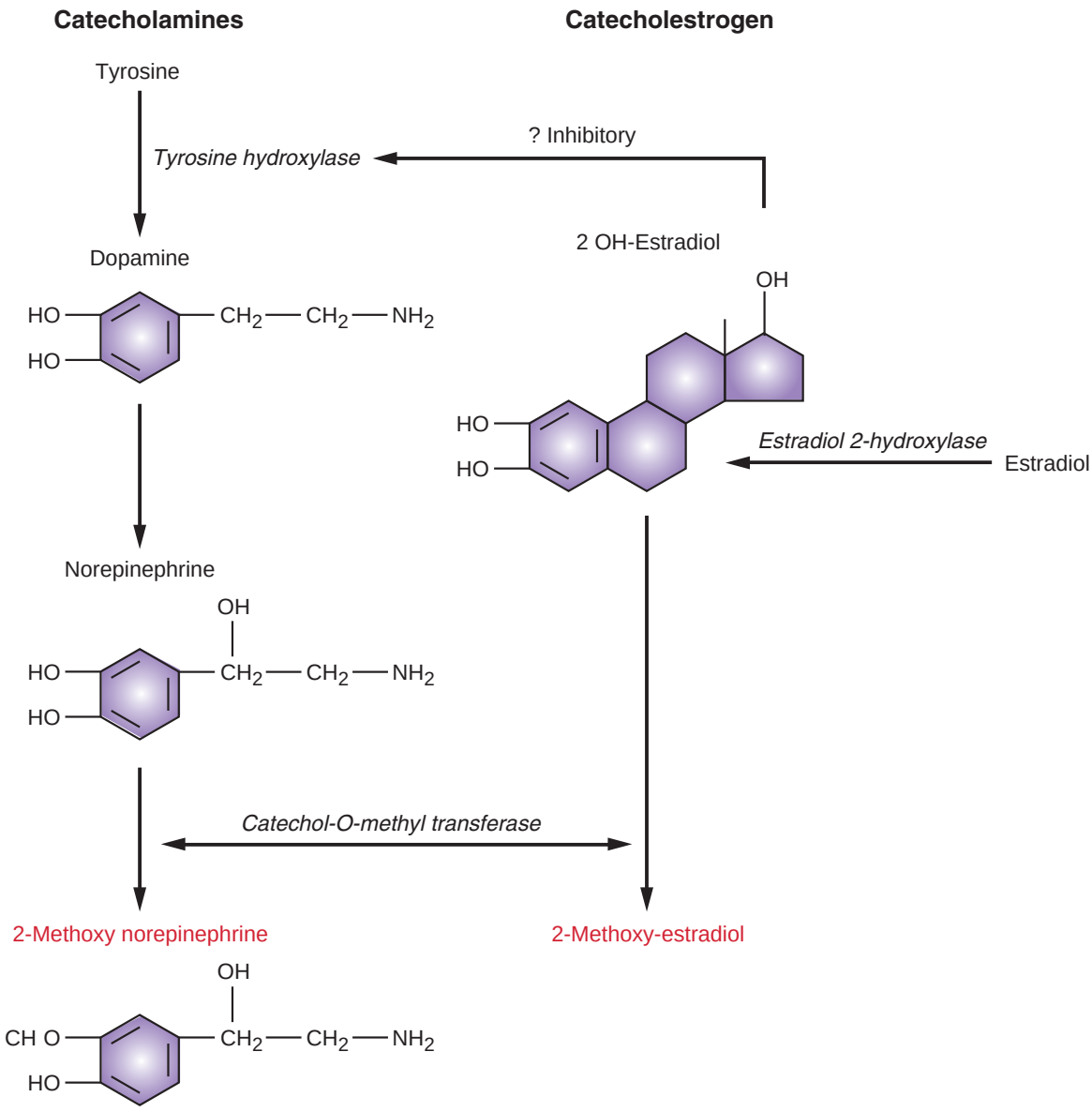
Administration of morphine, enkephalin analogs, and β-endorphin causes release of prolactin. The effect is mediated by inhibition of dopamine secretion in the tuberoinfundibular neurons in the median eminence. Most studies have reported no effect of naloxone on basal, stress-induced, or pregnant levels of prolactin nor on secretion by prolactinomas. Thus, a physiologic role for endogenous opioid regulation of prolactin does not appear to exist in men and women. However, suppression of GnRH secretion associated with hyperprolactinemia does appear to be mediated by endogenous opiates.[120]

Every pituitary hormone appears to be modulated by opiates. Physiologic effects are important with ACTH, gonadotropins, and possibly vasopressin. Opioid compounds have no direct action on the pituitary nor do they alter the action of releasing hormones on the pituitary.

POMC-like mRNA is present in the ovary and the placenta.[121] Expression is regulated by gonadotropins in the ovary but not in the placenta. Reasons for endorphin presence in these tissues are not yet apparent. High concentrations of all of the members of the POMC family are found in human ovarian follicular fluid, but only β-endorphin shows significant changes during the menstrual cycle, reaching highest levels just before ovulation.[122]

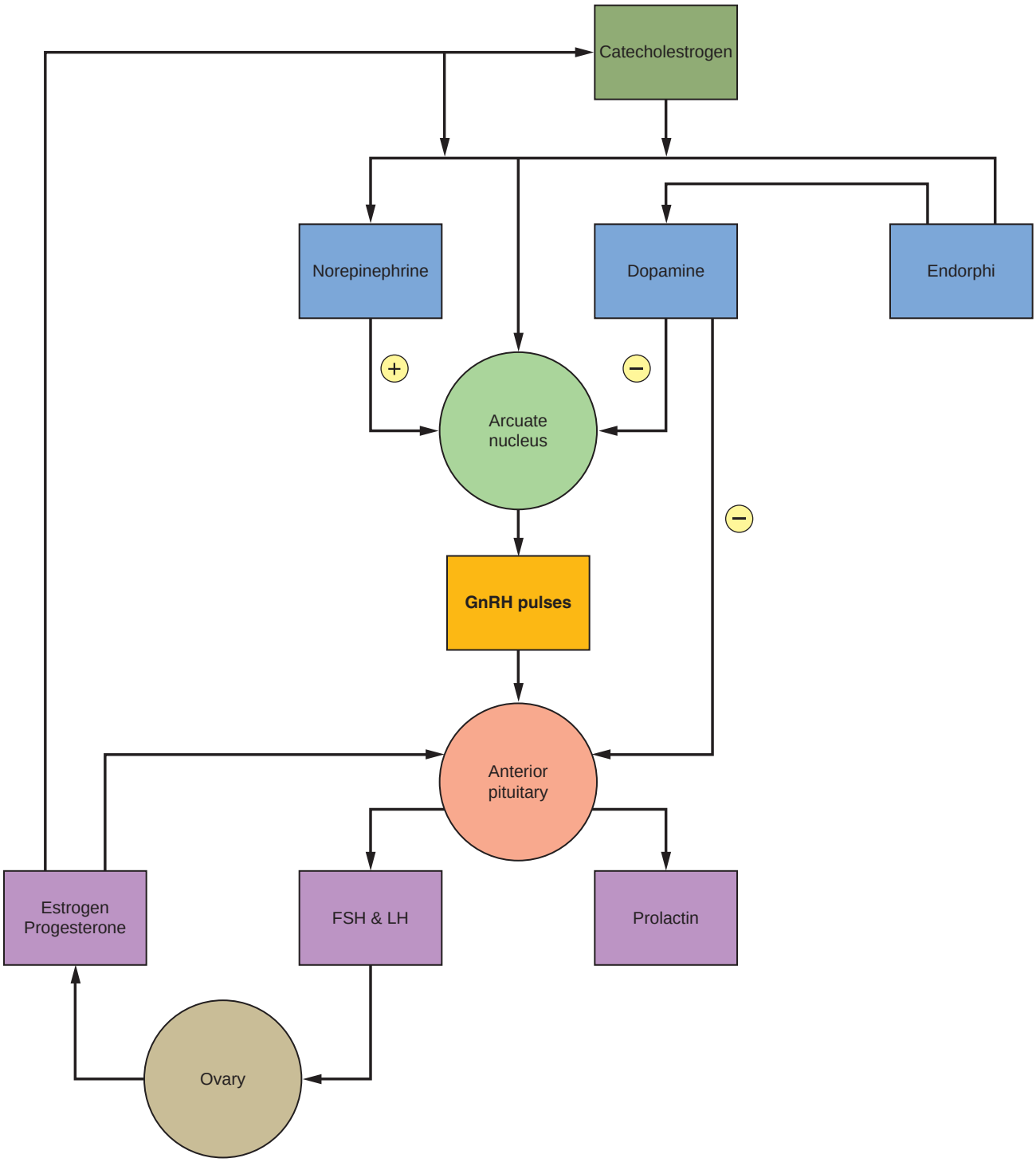
Catecholestrogens

The enzyme that converts estrogens to catecholestrogens (2-hydroxylase) is richly concentrated in the hypothalamus; hence there are higher concentrations of catecholestrogens than estrone and estradiol in the hypothalamus and pituitary gland. Catecholestrogens have two faces, a catechol side and an estrogen side. Because catecholestrogens have two faces, they have the potential for interacting with both catecholamine and estrogen-mediated systems.[123] To be specific, catecholestrogens can inhibit tyrosine hydroxylase (which would decrease catecholamines) and compete for catechol-*o*-methyltransferase (which would increase catecholamines). Because GnRH, estrogens, and catecholestrogens are located in similar sites, it is possible that catecholestrogens may serve to interact between catecholamines and GnRH secretion. However, these functions remain speculative because a definite role for catecholsteroids has not been established.



SUMMARY: Control of GnRH Pulses

The key concept is that normal menstrual function requires GnRH pulsatile secretion in a critical range of frequency and amplitude.[45, 48, 124] The normal physiology and pathophysiology of the menstrual cycle, at least in terms of central control, can be explained by mechanisms that affect the pulsatile secretion of GnRH. The pulses of GnRH are directly under the influence of a dual catecholaminergic system: norepinephrine facilitatory and dopamine inhibitory. In turn, the catecholamine system can be influenced by endogenous opioid activity. The feedback effects of steroids can be mediated through this system via catecholsteroid messengers or directly by influencing the various neurotransmitters.



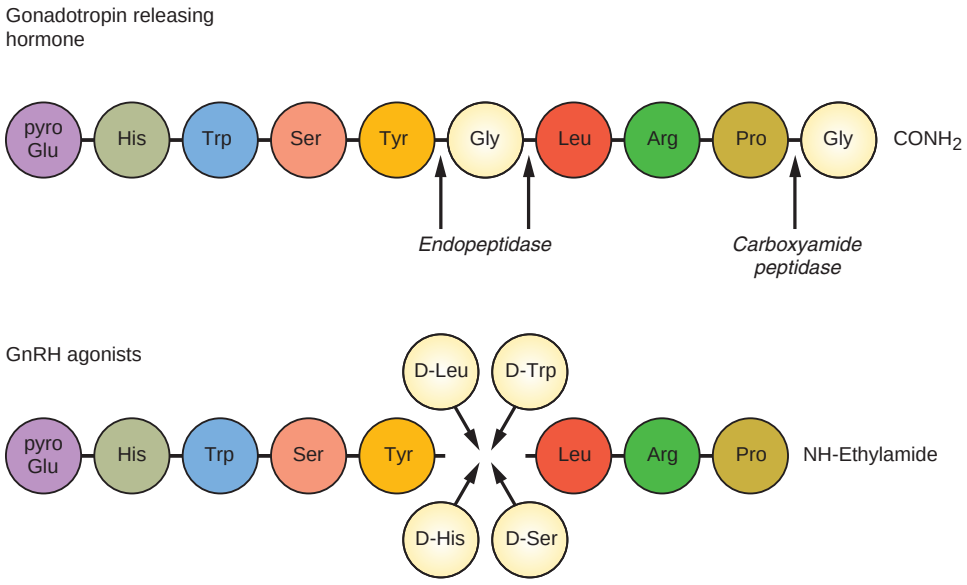
GnRH Agonists and Antagonists

The short half-life of GnRH is due to rapid cleavage of the bonds between amino acids 5–6, 6–7, and 9–10. By altering amino acids at these positions, analogues of GnRH can be synthesized with different properties. Substitution of amino acids at the 6 position or

replacement of the C-terminal glycine-amide (inhibiting degradation) produces agonists. The GnRH agonists are administered either intramuscularly or subcutaneously or by intranasal absorption. An initial agonistic action (the so-called flare effect) is associated with an increase in the circulating levels of FSH and LH. This response is greatest in the early follicular phase when GnRH and estradiol have combined to create a large reserve pool of gonadotropins. After 1–3 weeks, desensitization and down-regulation of the pituitary produce a hypogonadotropic, hypogonad state. The initial response is due to desensitization, whereas the sustained response is due to loss of receptors and the uncoupling of the receptor from its effector system. Furthermore, postreceptor mechanisms lead to secretion of biologically inactive gonadotropins, which, however, can still be detected by immunoassay.

Suppression of pituitary secretion of gonadotropins by a GnRH agonist can be used for the treatment of endometriosis, uterine leiomyomas, precocious puberty, or the prevention of menstrual bleeding in special clinical situations (e.g., in thrombocytopenic patients).

GnRH antagonists are synthesized with multiple amino acid substitutions. GnRH antagonists bind to the GnRH receptor and provide competitive inhibition of the naturally occurring GnRH. Thus, GnRH antagonists produce an immediate decline in gonadotropin levels with an immediate therapeutic effect within 24–72 hours. The early products either lacked potency or were associated with undesirable side effects due to histamine release. Products are now available for use in the treatment of endometriosis, prostate cancer, precocious puberty, and female infertility.

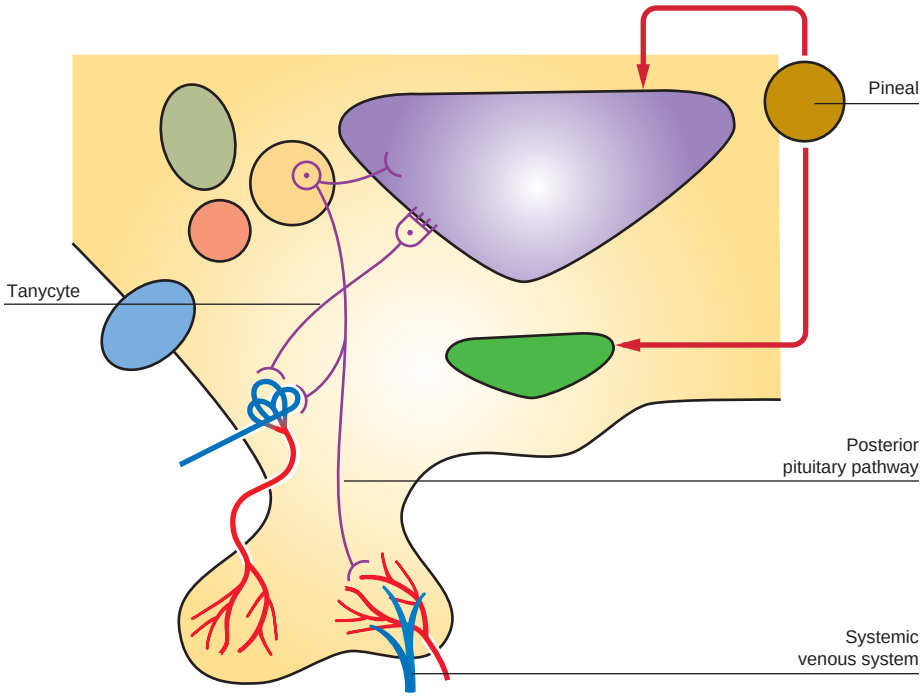


The GnRH analogues cannot escape destruction if administered orally. Higher doses administered subcutaneously can achieve nearly equal effects as observed with intravenous and intramuscular treatment; however, the smaller blood peaks are slower to develop and take longer to return to baseline. Other forms of administration include nasal spray, sustained-release implants, and injections of biodegradable microspheres. With the nasal route, absorption enhancers have to be added to increase bioavailability; these agents produce considerable nasal irritation. The depot formulations of GnRH agonists are administered intramuscularly and monthly.

GnRH Agonists and Antagonists in Clinical Use										
Position	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Native GnRH-I	pGlu	His	Trp	Ser	Tyr	Gly	Leu	Arg	Pro	Gly-NH_2
Native GnRH-II	pGlu	His	Trp	Ser	His	Gly	Trp	Tyr	Pro	Gly-NH_2
Leuprolide						D-Leu				NH-Ethylamide
Buserelin						D-Ser (tertiary butanol)				NH-Ethylamide
Nafarelin						D-Naphthylalanine (2)				
Histrelin						D-His (tertiary benzyl)				NH-Ethylamide
Goserelin						D-Ser (tertiary butanol)				Aza-Gly
Deslorelin						D-Trp				NH-Ethylamide
Triptorelin						D-Trp				
Abarelix	D-Ala	D-Phe	D-Ala			D-Asp		Lys-(iPr)		D-Ala
Cetrorelix	D-Nal	D-Phe	D-Pal			D-Cit				D-Ala
Ganirelix	D-Nal	D-Phe	D-Pal			D-hArg		hArg		D-Ala

Tanycytes

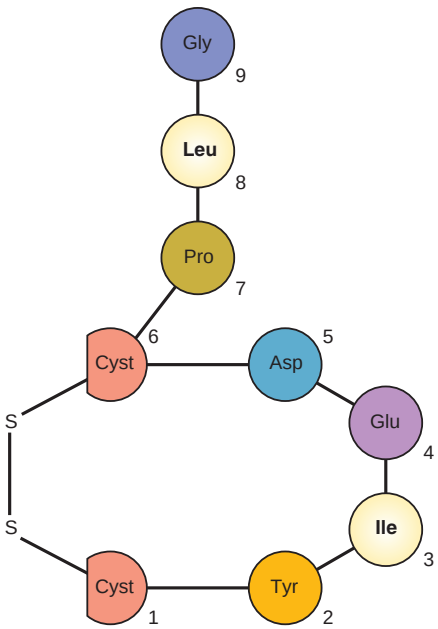
A significant pathway for hypothalamic influence may be via the cerebrospinal fluid (CSF). Tanycytes are specialized ependymal cells whose ciliated cell bodies line the third ventricle over the median eminence. The cells terminate on portal vessels, and they can transport materials from ventricular CSF to the portal system, e.g., substances from the pineal gland, or vasopressin, or oxytocin. Tanycytes change morphologically in response to steroids and exhibit morphologic changes during the ovarian cycle.



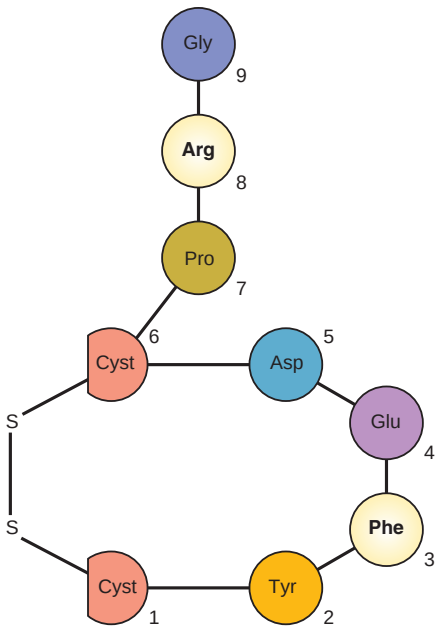
The Posterior Pituitary Pathway

The posterior pituitary is a direct prolongation of the hypothalamus via the pituitary stalk, whereas the anterior pituitary arises from pharyngeal epithelium that migrates into position with the posterior pituitary. Separate neurosecretory cells in both the supraoptic and paraventricular nuclei make vasopressin and oxytocin as parts of large precursor molecules that also contain the transport peptide, neurophysin.[125] Both oxytocin and vasopressin consist of nine amino acid residues, two of which are half cystines forming a bridge between positions 1 and 6. In the human, vasopressin contains arginine, unlike animals that have lysine vasopressin. The neurophysins are polypeptides with a molecular weight of about 10,000. There are two distinct neurophysins, estrogen-stimulated neurophysin known as neurophysin I and nicotine-stimulated neurophysin, known as neurophysin II.

The genes for oxytocin and vasopressin are closely linked on chromosome 20, derived from a common ancestor about 400 million years ago.[126] The transcriptional activity of these genes is regulated by endocrine factors, such as the sex steroids and thyroid hormone, through hormone-response elements located upstream. The neurons secrete two large protein molecules, a precursor called pro-pressophysin, which contains vasopressin and its neurophysin, and a precursor called pro-oxyphysin, which contains oxytocin and its neurophysin.[125] Neurophysin I is specifically related to oxytocin, and neurophysin II accompanies vasopressin. Because of this unique packaging, the hormones and their neurophysins are stored together and released at the same time into the circulation. The neurophysins are cleaved from their associated neurohormones during axonal transport from the neuronal cell bodies in the supraoptic and paraventricular nuclei to the posterior pituitary. The only known function for the neurophysins is axonal transport for oxytocin and vasopressin. Mutations in the gene encoding the precursor prohormone protein produce an alteration in the neurophysin, preventing the conformational shape necessary for transport of vasopressin to the pituitary gland and resulting in diabetes insipidus.[127, 128]



Oxytocin



Arginine-Vasopressin

The posterior pathway is complex and not limited to the transmission of vasopressin and oxytocin to the posterior pituitary. The transportation of vasopressin and oxytocin to the posterior pituitary occurs via nerve tracts that emanate from the supraoptic and paraventricular nuclei and descend through the median eminence to terminate in the posterior pituitary. However, these hormones are also secreted into the cerebrospinal fluid and directly into the portal system. Therefore, vasopressin and oxytocin can reach the anterior pituitary and influence ACTH secretion (in the case of vasopressin), and gonadotropin secretion (in the case of oxytoxin). Vasopressin cooperates with corticotropin-releasing hormone to cause an increased yield of ACTH. Vasopressin and oxytocin-like materials are also found in the ovary, the oviduct, the testis, and the adrenal gland, suggesting that these neurohypophyseal peptides have roles as paracrine or autocrine hormones.[129] The concentrations of these substances in the cerebrospinal fluid exhibit a circadian rhythm (with peak levels occurring during the day), suggesting a different mechanism for CSF secretion compared with posterior pituitary release.[130]

Neurophysin II is called nicotine neurophysin because the administration of nicotine or hemorrhage increases the circulating levels. Neurophysin I is called estrogen neurophysin because estrogen administration increases the levels in the peripheral blood, and peak levels of both neurophysin I and oxytocin are found at the time of the LH surge.[131] Oxytocin neurons and vasopressin neurons have been demonstrated in the rat to contain the estrogen receptor-β.[132] The rise in estrogen neurophysin begins 10 hours after the rise in estrogen and precedes that of the LH surge, and the elevation of neurophysin lasts longer than the LH surge. Because GnRH and oxytocin are competing substrates for hypothalamic degradation enzymes, it has been hypothesized that oxytocin in the portal blood at the midcycle can inhibit the metabolism of GnRH, thus increasing the amount of GnRH available. Furthermore, oxytocin may have direct actions on the pituitary, ovary, uterus, and fallopian tube during ovulation.

Neurophysin-containing pathways have been traced from the hypothalamic nuclei to various centers in the brainstem and the spinal cord. In addition, behavioral studies suggest a role for vasopressin in learning and memory. Administration of vasopressin has been associated with improvement in memory in brain-damaged human subjects and enhanced cognitive responses (learning and memory) in both young, normal individuals and depressed patients.

Both oxytocin and vasopressin circulate as the free peptides with a rapid half-life (initial component less than 1 minute, second component of 2–3 minutes). Three major stimuli for vasopressin secretion are changes in osmolality of the blood, alterations in blood volume, and psychogenic stimuli such as pain and fear. The osmoreceptors are located in the hypothalamus; the volume receptors are in the left atrium, aortic arch, and carotid sinus. Angiotensin II also produces a release of vasopressin, suggesting another mechanism for the link between fluid balance and vasopressin. Cortisol may modify the osmotic threshold for the release of vasopressin.

The major functions of vasopressin mainly involve the regulation of osmolality and blood volume, but also the release of insulin and ACTH, and influences on behavioral responses such as memory.[133] Vasopressin is a powerful vasoconstrictor and antidiuretic hormone. Vasopressin release increases when plasma osmolality rises and is inhibited by water loading (resulting in diuresis). Diabetes insipidus is a condition marked by loss of water because of a lack of vasopressin action in the tubules of the kidney, secondary to a defect in synthesis or secretion of vasopressin. The opposite condition is the continuous and autonomous secretion of vasopressin, the syndrome of inappropriate ADH (antidiuretic hormone) secretion. This syndrome, with its resultant retention of water, is associated with a variety of brain disorders and with the production of vasopressin and its precursor by malignant tumors.

Oxytocin stimulates muscular contractions in the uterus and myoepithelial contractions in the breast. Thus, it is involved in parturition and the letdown of milk. The release of oxytocin is so episodic that it is described as spurts. Ordinarily, there are about three spurts every 10 minutes. Oxytocin is released during coitus, probably by the Ferguson reflex (vaginal and cervical stimulation) but also by olfactory, visual, and auditory pathways. Perhaps oxytocin has some role in muscle contractions during orgasm.[134] In the male, release of oxytocin during coitus may contribute to sperm transport during ejaculation. Oxytocin is also released in specific areas of the brain, especially the ventromedial hypothalamus where it can act to inhibit appetite and stimulate sexual behavior.[135]

Using sensitive assays, an increase in maternal levels of oxytocin can be detected prior to parturition, occurring at first only at night.[136, 137] Once labor has begun, oxytocin levels rise significantly, especially during the second stage. Thus, oxytocin may be important for developing the more intense uterine contractions. Extremely high concentrations of oxytocin can be measured in the cord blood at delivery, and release of oxytocin from the fetal pituitary may also be involved in labor. However, this is controversial, and studies in monkeys fail to indicate a role for fetal oxytocin in parturition.[137] Part of the contribution of oxytocin to parturition is the stimulation of prostaglandin synthesis in decidua and myometrium.[138] Cervical dilation appears to be dependent on oxytocin stimulation of prostaglandin production, probably in the decidua. The greater frequency of labor and delivery at night may be due to greater nocturnal oxytocin secretion.[136] In addition, oxytocin is synthesized in the amnion, chorion, and, significantly, in the decidua.[136] This locally produced oxytocin may be a significant stimulus for myometrial and membrane production of prostaglandins.

It is likely that oxytocin action during the initial stages of labor depends on myometrial sensitivity to oxytocin in addition to the levels of oxytocin in the blood. The concentration of oxytocin receptors in the myometrium is low in the nonpregnant state and increases steadily throughout gestation (an 80-fold increase), and, during labor, the concentration doubles. This receptor concentration correlates with the uterine sensitivity to oxytocin.[139] The mechanism for the increase is unknown, but it likely is due to a change in the prostaglandin and hormonal milieu of the uterus. The local production and effects of oxytocin, estrogen, and progesterone combine in a complicated process of autocrine, paracrine, and endocrine actions to result in parturition.

Oxytocin is released in response to suckling, mediated through impulses generated at the nipple and transmitted via the third, fourth, and fifth thoracic nerves to the spinal cord and to the hypothalamus. In addition to causing milk ejection, the reflex is responsible for the uterine contractions associated with breastfeeding. Opioid peptides inhibit oxytocin release, and this may be the means by which stress, fear, and anger inhibit milk output in lactating women. Oxytocin is also expressed in many tissues where it exerts autocrine-paracrine actions.

The Brain and Ovulation

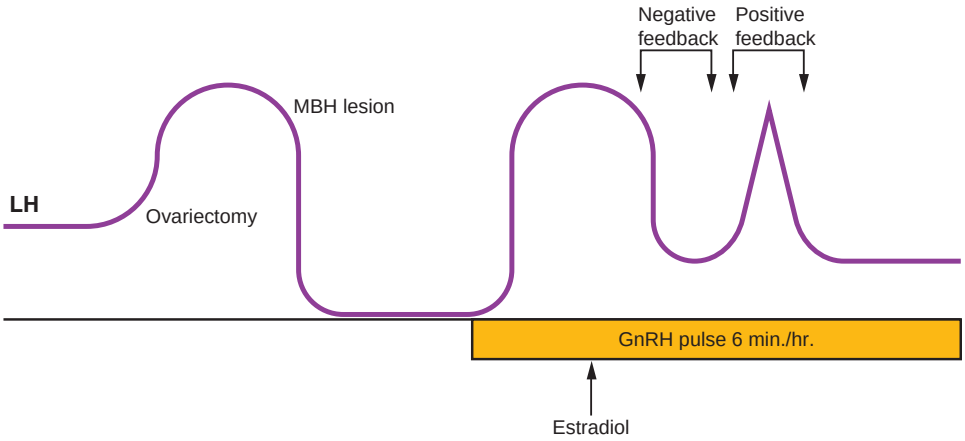
Classic studies in a variety of rodents indicated the presence of feedback centers in the hypothalamus that responded to steroids with the release of GnRH. The release of GnRH was the result of the complex, but coordinated, relationships among the neurohormones, the pituitary gonadotropins, and the gonadal steroids designated by the time-honored terms positive and negative feedback.

FSH levels were thought to be largely regulated by a negative inhibitory feedback relationship with estradiol. For LH, both a negative inhibitory feedback relationship with estradiol and a positive stimulatory feedback with high levels of estradiol were demonstrated. The feedback centers were located in the hypothalamus and were called the tonic and cyclic centers. The tonic center controlled the day-to-day basal level of gonadotropins and was responsive to the negative feedback effects of sex steroids. The cyclic center in the female brain was responsible for the midcycle surge of gonadotropins, a response mediated by the positive feedback of estrogen. Specifically, the midcycle surge of gonadotropins was thought to be due to an outpouring of GnRH in response to the positive feedback action of estradiol on the cyclic center of the hypothalamus.

This classic concept was not inaccurate. The problem was that the concept accurately described events in the rodent, but the mechanism is different in the primate. Neuroendocrine genetic expression studies based on rodents duplicate the earlier studies, and once again support a hypothalamic "center" as the primary locus for sex steroid actions, but the primate operates differently.

In the primate, the "center" for the midcycle surge of gonadotropins moved from the hypothalamus to the pituitary. Experiments in the monkey demonstrated that GnRH, originating in the hypothalamus, plays a permissive and supportive role. Its pulsatile secretion is an important prerequisite for normal pituitary function,[124] but the feedback responses regulating gonadotropin levels are controlled by ovarian steroid feedback on the anterior pituitary cells.

The present concept is derived from experiments in which the medial basal hypothalamus was either destroyed[42] or the hypothalamus was surgically separated from the pituitary.[140] In a typical (and now classic) experiment, lesion of the medial basal hypothalamus by radiofrequency waves was followed by loss of LH levels as the source of GnRH was eliminated.[41] Administration of GnRH in a pulsatile fashion by an intravenous pump restored LH secretion. The administration of estradiol was then able to produce both negative and positive feedback responses, actions that certainly must be directly on the anterior pituitary because the hypothalamus was absent and GnRH was being administered in a steady and unchanging frequency and dose.

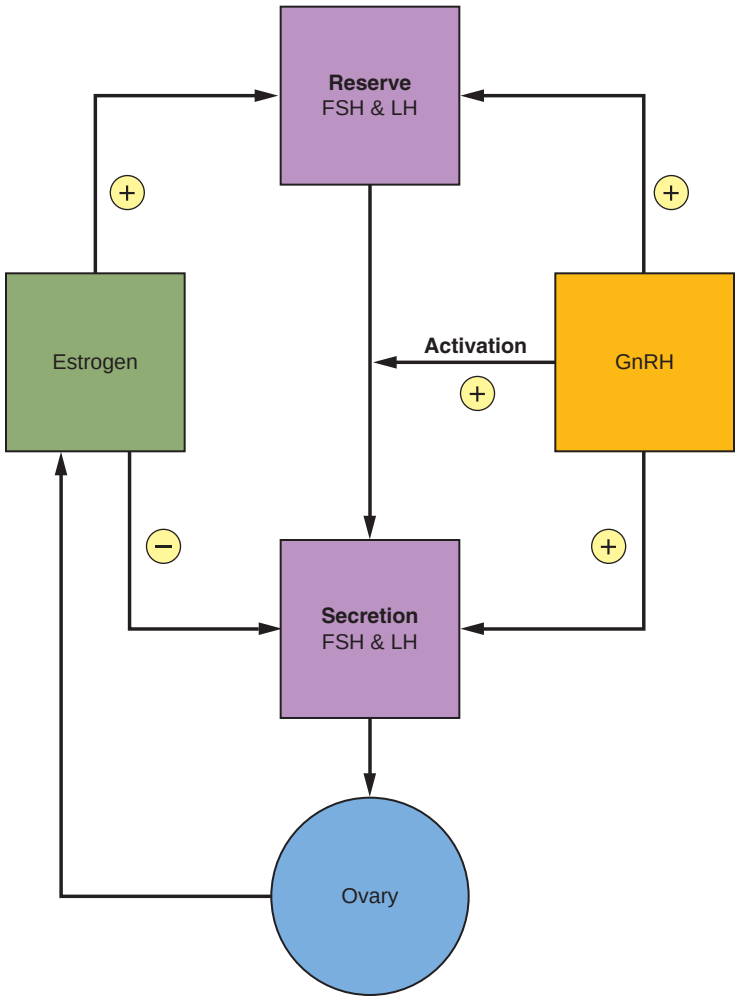


Administration of GnRH intravenously as a bolus produces an increase in blood levels of LH and FSH within 5 minutes, reaching a peak in about 20–25 minutes for LH and 45 minutes for FSH. Levels return to pretreatment values after several hours. When administered by constant infusion at submaximal doses, there is first a rapid rise with a peak at 30 minutes, followed by

a plateau or fall between 45 and 90 minutes, and then a second and sustained increase at 225–240 minutes. This biphasic response suggests the presence of two functional pools of pituitary gonadotropins.[141] The readily releasable pool (secretion) produces the initial response, and the later response is dependent on a second, reserve pool of stored gonadotropins.

There are three principal positive actions of GnRH on gonadotropin elaboration:

1. Synthesis and storage (the reserve pool) of gonadotropins.
2. Activation—movement of gonadotropins from the reserve pool to a pool ready for direct secretion, a self-priming action.
3. Immediate release (direct secretion) of gonadotropins.



Secretion, synthesis, and storage change during the cycle. At the beginning of the cycle, when estrogen levels are low, both secretion and storage levels are low. With increasing levels of estradiol, a greater increase occurs in storage, with little change in secretion. Thus, in the early follicular phase, estrogen has a positive effect on the synthesis and storage response, building up a supply of gonadotropins to meet the requirements of the midcycle surge. Premature release of gonadotropins is prevented by a negative (inhibitory) action of estradiol on the pituitary secretory response to GnRH.

As the midcycle approaches, subsequent responses to GnRH are greater than initial responses, indicating that each response not only induces release of gonadotropins but also activates the storage pool for the next response. This sensitizing or priming action of GnRH also involves an increase in the number of its own receptors and requires the presence of estrogen.[142, 143] Estrogen itself is capable of increasing the number of GnRH receptors.[80, 144] The rise in estrogen at midcycle prepares the gonadotrope to further respond to GnRH.

Because the midcycle surge of LH can be produced in the experimental monkey in the absence of a hypothalamus and in the face of unchanging GnRH, the ovulatory surge of LH is believed to be a response to positive feedback action of estradiol on the anterior pituitary. When the estradiol level in the circulation reaches a critical concentration and this concentration is maintained for a critical time period, the inhibitory action on LH secretion changes to a stimulatory action. The mechanism of this steroid action is not known with certainty, but experimental evidence suggests that the positive feedback action involves many mechanisms, including an increase in GnRH receptor concentration and an increase in pituitary sensitivity to GnRH. This action of estrogen is classically referred to as positive feedback, but it can also be viewed as an acute decrease in inhibitory influences.

The negative feedback of estrogen operates through different systems; at the pituitary level, estrogen inhibition of FSH secretion is associated with decreasing pituitary expression of activin.[145–148] In addition, estradiol directly inhibits the FSH beta subunit gene by influencing corepressor proteins (adapter proteins) to bind to the gene and suppress transcription.[149]

What a logical mechanism! The midcycle surge must occur at the right time of the cycle to ovulate a ready and waiting mature follicle. What better way to achieve this extreme degree of coordination and timing than by the follicle itself, through the feedback effects of the sex steroids originating in the follicle destined to ovulate.

The presence of GnRH is certainly essential; the administration of a GnRH antagonist to women at midcycle prevents the LH surge.[150] GnRH is increased in the peripheral blood of women and the portal blood of monkeys at midcycle.[151] Although this increase may not be absolutely necessary (as demonstrated in the monkey experiments), studies do indicate that increased activity is occurring in both the hypothalamus and the pituitary.[146, 152–154] Therefore, although the system can operate with only an unwavering, permissive action of GnRH, fine-tuning probably takes place by means of simultaneous effects on GnRH pulsatile secretion and pituitary response to GnRH. This is supported by gonadotropin gene expression studies, indicating steroid effects at both the hypothalamus and the pituitary.[155] The upstream region of the LH β-subunit gene binds the estrogen receptor, providing a means for direct steroid hormone modulation in the pituitary.[156]

The effect of estrogen on the hypothalamic release of GnRH was confusing at first because initial studies failed to detect the presence of estrogen receptors in GnRH neurons.[157–159] Later studies identified the more recently discovered estrogen receptor-β, and experiments in knockout mice indicated that estrogen receptor-β mediates estrogen effects on GnRH neurons.[160–162] Finally, more sophisticated methods and the knockout model demonstrated that both alpha and beta estrogen receptors are present in GnRH neurons and are involved in cyclic changes in GnRH release.[163–165] The human GnRH gene contains a hormone-responsive element that binds estrogen and its receptor.[166] In molecular studies, estrogen decreased messenger RNA for GnRH-II but increased messenger RNA for GnRH-I.[167] In vivo studies in the sheep have demonstrated that estradiol has both negative and positive feedback effects on hypothalamic GnRH secretion and that a GnRH surge is involved in the preovulatory LH surge.[168–170] In monkeys, estrogen treatment decreases GnRH gene expression.[171] There no longer is any question that estrogen can regulate the activity of the hypothalamic GnRH neurons.[172]

Influencing the hypothalamic frequency of GnRH secretion can in turn influence pituitary response to GnRH. Faster or slower frequencies of GnRH pulses result in lower GnRH receptor numbers in the pituitary.[173] Thus, a critical peak frequency is necessary for peak numbers of GnRH receptors and the peak midcycle response. Here is a method for the fine-tuning at both the hypothalamus (pulse frequency) and the pituitary (receptor number). Indeed, turning off the surge may involve down-regulation because of excessive GnRH. Studies in sheep indicate that a surge of GnRH at the time of the LH surge is associated with a switch from episodic secretion to continuous secretion into the portal circulation, producing the high exposure known to result in down-regulation.[174]

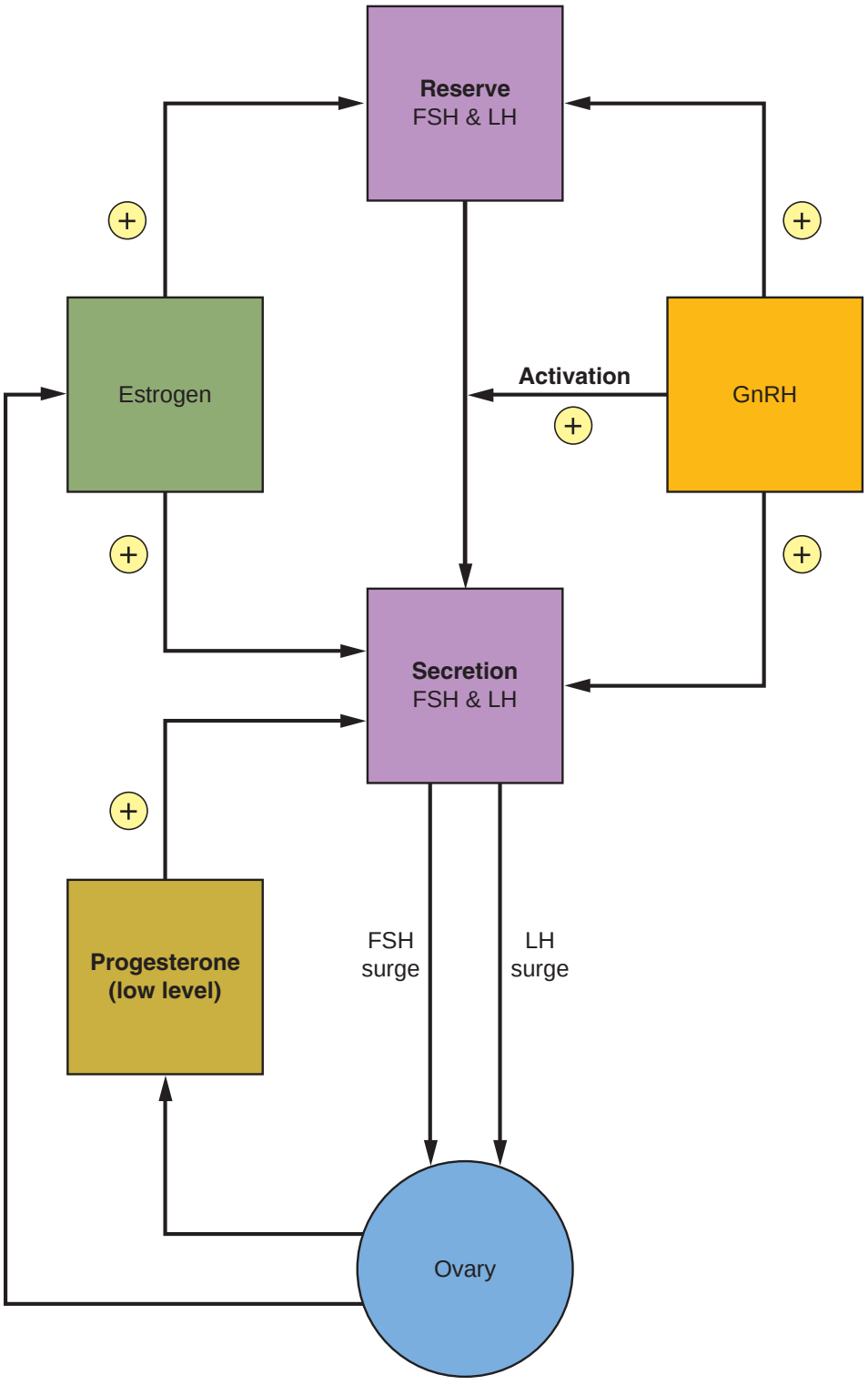
The differential response of FSH and LH, arising from the same pituitary cell, to the frequency of GnRH pulses is a consequence of the interplay among the regulating factors: the classic endocrine effects of estrogen, progesterone, and inhibin and the autocrine/paracrine actions of activin and follistatin.[175] The biphasic response of FSH, in particular, is modulated by the balance of activin and follistatin, created by the effects of estrogen and inhibin.

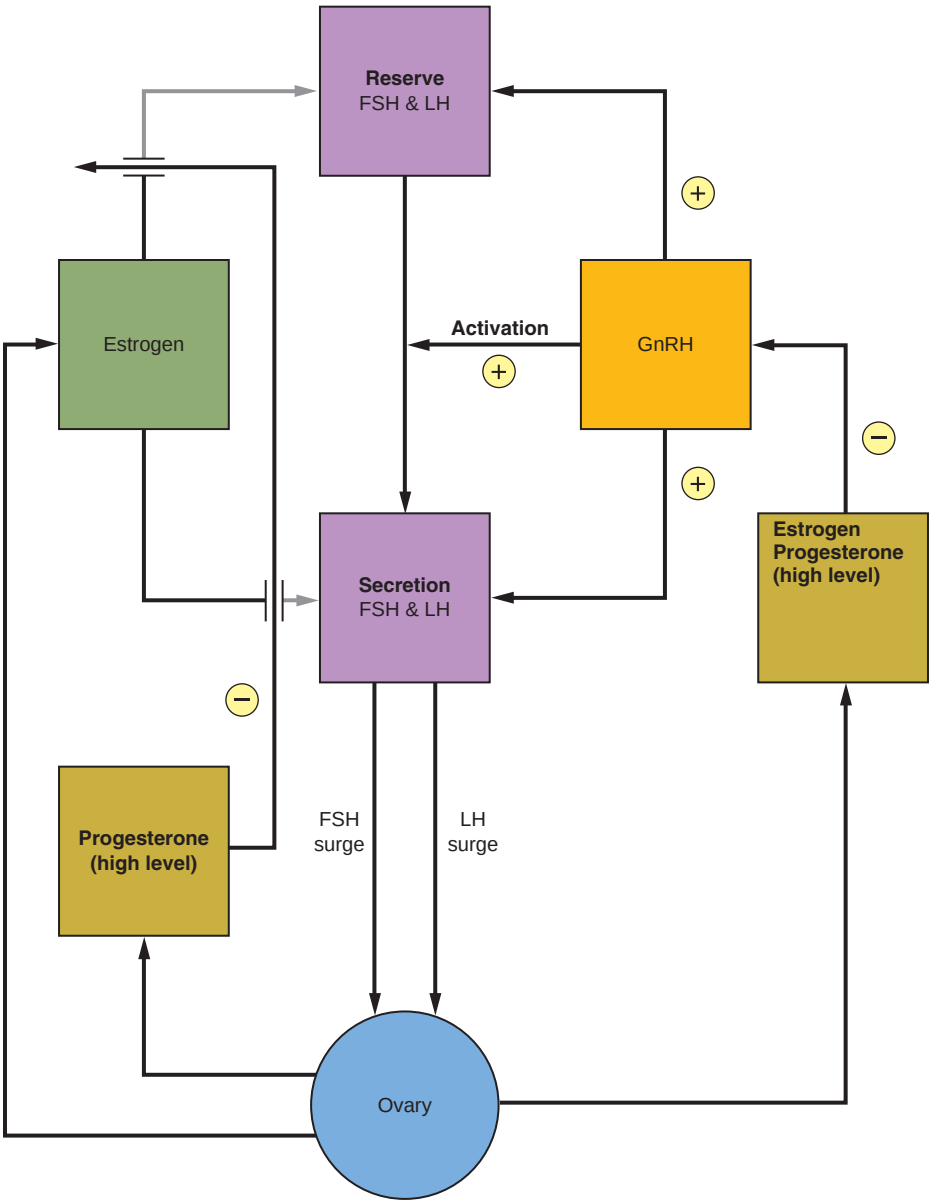
Another aspect of gonadotropin secretion is clinically important. A disparity exists between the quantity of gonadotropins measured during the midcycle surge as determined by immunoassay and bioassay. More FSH and LH are secreted at midcycle in a molecular form with greater biologic activity.[176, 177] There is a well-established relationship between the activity and half-life of glycoprotein hormones and molecular composition (see Chapter 2, under "Heterogeneity" of tropic hormones). The estrogen influence on gonadotropin synthesis is an additional method for maximizing the biologic effects of the midcycle surge. The bioactivity is also very dependent on pulsatile stimulation by GnRH. In addition to the change at midcycle that favors gonadotropin activity at the ovarian follicle, FSH isoforms with greater biologic activity also increase during the late luteal phase, a change that is appropriately geared toward propelling new ovarian follicle growth for the next cycle.[178]

The midcycle surge of FSH has an important clinical purpose. A normal corpus luteum requires the induction of an adequate number of LH receptors on granulosa cells, a specific action of FSH. In addition, FSH accomplishes important intrafollicular changes necessary for the physical expulsion of the ovum. The midcycle surge of FSH, therefore, plays a critical role in ensuring ovulation and a normal corpus luteum. Emerging progesterone secretion, just prior to ovulation, is the key.

Progesterone, at low levels and in the presence of estrogen, augments the pituitary secretion of LH and is responsible for the FSH surge in response to GnRH.[179–182] As the rising levels of LH produce the morphologic change of luteinization in the ovulating follicle, the granulosa layer begins to secrete progesterone directly into the bloodstream. The process of luteinization is inhibited by the presence of the oocyte; therefore, progesterone secretion is relatively suppressed, ensuring that only low levels of progesterone reach the brain.

After ovulation, rapid and full luteinization is accompanied by a marked increase in progesterone levels, which, in the presence of estrogen, exercise a profound negative feedback action to suppress gonadotropin secretion. This action of progesterone takes place in two locations, the hypothalamus and the pituitary.[183–185] There definitely is a central action to decrease GnRH.[186] An important role for progesterone is to mediate the slowing of GnRH pulses in the late luteal phase, favoring the rise in FSH necessary to initiate the next cycle.[51] Progesterone fails to block estradiol-induced gonadotropin discharges in monkeys with hypothalamic lesions if pulsatile GnRH replacement is provided. Therefore, high levels of progesterone inhibit ovulation at the hypothalamic level. In contrast, the facilitatory action of low levels of progesterone is exerted only at the pituitary on the response to GnRH.


Reserve
FSH & LH
Estrogen
Activation
GnRH
Secretion
FSH & LH
Progesterone
(low level)
FSH
surge
LH
surge
Ovary



SUMMARY: Key Points

1. Pulsatile GnRH secretion must be within a critical range for frequency and concentration (amplitude). This is absolutely necessary for normal reproductive function.

2. GnRH has only positive actions on the anterior pituitary: synthesis and storage, activation, and secretion of gonadotropins. The gonadotropins are secreted in a pulsatile fashion in response to the similar pulsatile release of GnRH.

3. Lower GnRH pulse frequencies favor FSH secretion, and higher GnRH pulse frequencies favor LH secretion.

4. Low levels of estrogen enhance FSH and LH synthesis and storage, have little effect on LH secretion, and inhibit FSH secretion.

5. High levels of estrogen induce the LH surge at midcycle, and high steady levels of estrogen lead to sustained elevated LH secretion.

6. Low levels of progesterone acting at the level of the pituitary gland enhance the LH response to GnRH and are responsible for the FSH surge at midcycle.

7. High levels of progesterone inhibit pituitary secretion of gonadotropins by inhibiting GnRH pulses at the level of the hypothalamus. In addition, high levels of progesterone can antagonize pituitary response to GnRH by interfering with estrogen action.

The Pineal Gland

Although no physiologic role has been firmly established in the human, the reproductive functions of the hypothalamus may also be under inhibitory control of the brain via the pineal gland. The pineal arises as an outgrowth of the roof of the third ventricle, but soon after birth it loses all afferent and efferent neural connections with the brain. Instead, the parenchymal cells receive a new and unusual sympathetic innervation that allows the pineal gland to be an active neuroendocrine organ that responds to photic and hormonal stimuli and exhibits circadian rhythms.[187–189]

The neural pathway begins in the retina and passes through the suprachiasmatic and paraventricular nuclei in the hypothalamus to the inferior accessory optic tracts and the medial forebrain bundle to the upper spinal cord. Preganglionic fibers terminate at the superior cervical ganglion, and postganglionic sympathetic nerves terminate directly on pineal cells. Interruption of this pathway gives the same effect as darkness, which is an increase in pineal biosynthetic activity.

Hydroxyindole-*o*-methyltransferase (HIOMT), an enzyme essential for melatonin synthesis, is found mainly in pineal parenchymal cells, and its products are essentially unique to the pineal. Norepinephrine stimulates tryptophan entry into the pineal cell and also adenylate cyclase activity in the membrane. The resulting increase in cyclic AMP leads to *N*-acetyltransferase activity, the rate-limiting step in melatonin synthesis. Tryptophan is converted by the combined action of *N*-acetyltransferase and HIOMT into melatonin (*N*-acetyl-5-methoxytryptamine). Thus, melatonin synthesis is controlled by norepinephrine stimulation of adenylate cyclase, and the norepinephrine is liberated by sympathetic stimulation due to the absence of light. HIOMT is also found in the retina where melatonin may serve to regulate the pigment in retinal cells and in the intestine. However, pinealectomy completely eliminates detectable levels of melatonin in the circulation. Calcification of the pineal gland is common. It is frequently present in young children, and almost all elderly people have pineal calcification.

The association of hyperplastic pineal tumors with decreased gonadal function and destructive tumors with precocious puberty suggested that the pineal is the source of gonadal inhibiting substances. However, pineal mechanisms cannot be absolutely essential for gonadal function. Normal reproductive function returns to the pinealectomized rat several weeks after pinealectomy; blind women have normal fertility, and pinealectomy in a primate did not affect pubertal development.[190]

DARKNESS → INCREASED MELATONIN → DECREASED GnRH

A rat in constant light develops a small pineal with decreased HIOMT and melatonin, while the ovarian weight increases. A rat in constant dark has the opposite result, increased pineal size, HIOMT, and melatonin, with decreased ovarian weight and pituitary function. A rhythm is established in pineal HIOMT activity by the presence or absence of light. Short days and long nights with increased melatonin secretion result in gonadal atrophy, and this is the major mechanism governing seasonal breeding.[191, 192] In humans, melatonin secretion increases after darkness peaks in the middle of the night, and then decreases. This rhythm is endogenous, originating in the suprachiasmatic nucleus. Light does not cause the rhythm, but influences its timing.

Possible roles in humans may be to give circadian rhythmicity to other functions such as temperature and sleep. In all vertebrates tested so far, there is a daily and seasonal rhythm in melatonin secretion: high values during the dark and low during light, greater secretion in the winter compared with the summer. Desynchronization with travel across time zones may contribute to the symptom complex known as jet lag. Melatonin ingestion improves both the duration and quality of sleep, but the optimal timing of administration is unknown.[189, 193]

The pineal, therefore, serves as an interface between the environment and hypothalamic-pituitary function. In order to correctly interpret day length, animals require a daily rhythm in melatonin secretion. This coordination of temporal, environmental information is especially important in seasonal breeders. This pineal rhythm appears to require the suprachiasmatic nucleus, perhaps the site at which pineal function and light changes are coordinated.

Melatonin is synthesized and secreted by the pineal gland and circulates in the blood like a classical hormone. It affects distant target organs, especially the neuroendocrine centers of the central nervous system. Whether melatonin is secreted primarily into the CSF or blood is still debated, but most evidence favors blood. Melatonin may reach the hypothalamus from the CSF by way of tanycyte transport.

The gonadal changes associated with melatonin are mediated via the hypothalamus and suggest a general suppressive effect on GnRH pulsatile secretion and reproductive function.[194] In humans, melatonin blood levels are highest in the first year of life (with highest levels at night), and then decrease with age, eventually releasing, some claim, the suppression of GnRH prior to puberty.[191] This hypothesis is challenged by the association of blindness in human females with an age of menarche that is earlier than normal.[195] Furthermore, pinealectomy in monkeys does not affect puberty.[190] Others have failed to find a decrease in melatonin levels with aging.[196]

Pineal activity can be viewed as the net balance between hormone- and neuron-mediated influences. The pineal contains receptors for the active sex hormones, estradiol, testosterone, dihydrotestosterone, progesterone, and prolactin. Furthermore, the pineal converts testosterone and progesterone to the active 5α-reduced metabolites, and androgens are aromatized to estrogens. The pineal also appears to be unique in that a catecholamine neurotransmitter (norepinephrine), interacting with cell membrane receptors, stimulates cellular synthesis of estrogen and androgen receptors. In general, however, the sympathetic activity producing the circadian rhythm takes precedence over hormonal effects.

Despite a variety of suggestive leads, there is no definitive evidence for a role of the pineal in humans. Nevertheless the important relationship between light exposure and circadian rhythms continues to focus attention on the pineal gland as a coordinator.[197] There is a seasonal distribution in human conception in northern countries with a decrease in ovarian

activity and conception rates during the dark winter months.[198, 199] In addition, the pineal can disrupt normal gonadal function. A male with delayed puberty due to hypogonadotropism has been described, who had an enlarged, hyperfunctional pineal gland.[200] Over time, his melatonin levels spontaneously decreased, and normal pituitary-gonadal function developed. Elevated nocturnal levels of melatonin have been reported in patients with hypothalamic amenorrhea and in women with anorexia nervosa, but this increase is probably a consequence of low estrogen levels, and not etiologic.[194, 201]

A possible influence of the pineal gland may be the synchronization of menstrual cycles noted among women who spend time together. A significant increase in synchronization of cycles was first reported in 1971 among roommates and among closest friends in the first 4 months of residency in a dormitory of a women's college.[32] A similar increase in synchrony has been observed in women coworkers in occupations characterized by levels of interdependency that were equal to or greater than the levels of encountered job stress and in Bedouin families in which women live together for many years.[33, 35] However, efforts to replicate these results have not always been successful.[34, 202] There is some evidence that the timing of ovulation can be affected by axillary human pheromones.[31, 203]

Melatonin is available in 1- and 5-mg doses that produce blood levels that are 10 to 100 times higher than normal nighttime peaks.[189, 193] The effects include increased sleepiness and decreased alertness. No data are available regarding long-term consequences on reproductive function.

A number of other indoles (also derivatives of tryptophan) have been identified in the pineal gland. Biologic roles for these indoles remain elusive, but one in particular has been extensively investigated. Arginine vasotocin differs from oxytocin by a single amino acid in position 8 and from vasopressin by a single amino acid in position 3. In general, arginine vasotocin has an inhibitory action on the gonads and pituitary secretion of prolactin and LH. Nevertheless a precise role continues to be evasive.

Gonadotropin Secretion Through Fetal Life, Childhood, and Puberty

We have often considered the endocrine events during puberty as an awakening, a beginning. However, endocrinologically, puberty is not a beginning but just another stage in a development that began at conception. The development of the anterior pituitary in the human starts between the fourth and fifth weeks of fetal life, and by the 12th week of gestation the vascular connection between the hypothalamus and the pituitary is functional. Gonadotropin production has been documented throughout fetal life, during childhood, and into adult life.[204] Remarkable levels of FSH and LH, similar to postmenopausal levels, can be measured in the fetus. GnRH is detectable in the hypothalamus by 10 weeks of gestation, and by 10–13 weeks when the vascular connection is complete, FSH and LH are produced in the pituitary. The peak pituitary concentrations of FSH and LH occur at about 20–23 weeks of intrauterine life, and peak circulating levels occur at 28 weeks.

The increasing production rate of gonadotropins until midgestation reflects the growing ability of the hypothalamic-pituitary axis to perform at full capacity. Beginning at midgestation, there is an increasing sensitivity to inhibition by steroids and a resultant decrease in gonadotropin secretion. Full sensitivity to steroids is not reached until late in infancy. The rise in gonadotropins after birth reflects loss of the high levels of placental steroids. Thus, in the first year of life there is considerable follicular activity in the ovaries in contrast to

later in childhood when gonadotropin secretion is suppressed. Furthermore, the postnatal rise in gonadotropins is even greater in infants born prematurely.

Testicular function in the fetus can be correlated with the fetal hormone patterns. Initial testosterone production and sexual differentiation are in response to the fetal levels of hCG, whereas further testosterone production and masculine differentiation are maintained by the fetal pituitary gonadotropins. Decreased testosterone levels in late gestation reflect the decrease in gonadotropin levels. The fetal generation of Leydig cells somehow avoids down-regulation and responds to high levels of hCG and LH by increased steroidogenesis and cell multiplication. This generation of cells is replaced by the adult generation that becomes functional at puberty and responds to high levels of hCG and LH with down-regulation and decreased steroidogenesis.

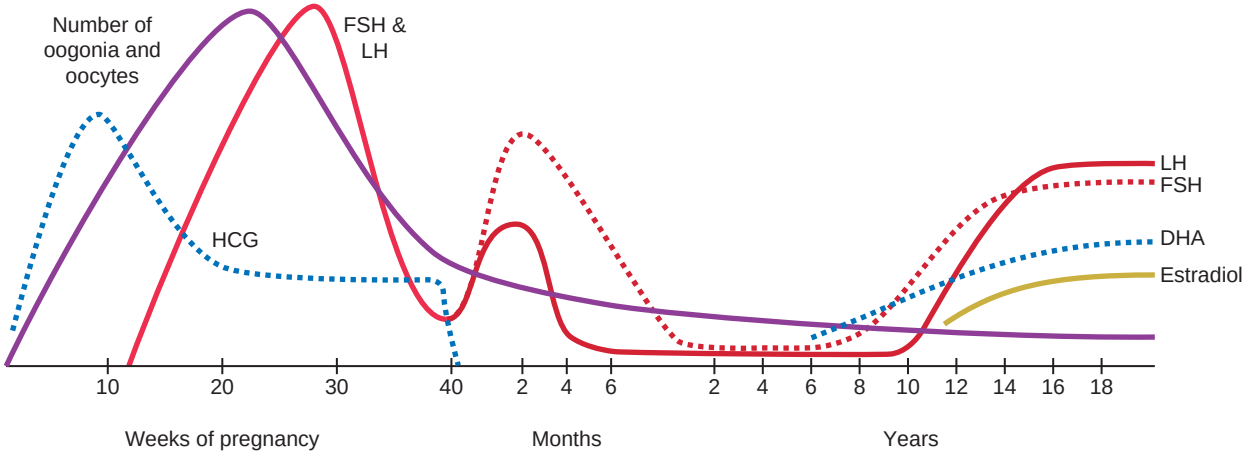
There is a sex difference in fetal gonadotropin levels. There are higher pituitary and circulating FSH and LH levels in female fetuses. The lower male levels are due to testicular testosterone and inhibin production. In infancy, the postnatal FSH rise is more marked and more sustained in females, whereas LH values are not as high. This early activity is accompanied by inhibin levels comparable to the low range observed during the follicular phase of the menstrual cycle.[205] After the postnatal rise, gonadotropin levels reach a nadir during early childhood (by about 6 months of age in males and 1–2 years in females) and then rise slightly between 4 and 10 years. The difference between males and females is notable; in the female the suppression of gonadotropins begins later in childhood, is less intense than in the male, and the duration of suppression is shorter in females. The difference is believed to reflect the presence of testosterone in males.

This childhood period is characterized by low levels of gonadotropins in the pituitary and in the blood, little response of the pituitary to GnRH, and maximal hypothalamic suppression. This low level of activity is not maintained by the ovaries or testes because gonadal removal results in little change. Children without gonads experience this same low level of activity.[206] A central inhibitory force must be operating within the brain, awaiting a signal to initiate puberty.

The central inhibitory force restrains GnRH pulsatile secretion. The mechanism involves several neurotransmitters: GABA (γ-amino butyric acid), neuropeptide Y, and kisspeptin. GABA release in the hypothalamus declines as GnRH secretion increases at the onset of puberty, and blockade of GABA synthesis or activity can initiate puberty in monkeys.[207, 208] Neuropeptide Y is recognized as an inhibitory neurotransmitter on GnRH pulsatility.[209] Kisspeptins is a newly appreciated neuroendocrine regulator of reproduction.[210] Mutations of the G-protein receptor for kisspeptins results in hypogonadotropic hypogonadism and a failure to enter puberty; these patients respond to either exogenous gonadotropins or GnRH. Kisspeptin expression is found in the arcuate nucleus of the hypothalamus and increases just prior to the resurgence of GnRH pulsatile secretion at puberty.[211] These are some of the signals involved, but the precise trigger that senses maturation and initiates the events of puberty remains elusive.

In girls, the first steroids to rise in the blood are dehydroepiandrosterone (DHEA) and its sulfate (DHEAS) beginning at 6–8 years of age, shortly before FSH begins to increase. Estrogen levels, as well as LH, do not begin to rise until 9–12 years of age. If the onset of puberty is triggered by the first hormone to increase in the circulation, then a role for adrenal steroids must be considered. However, there is no evidence to suggest that the adrenal steroids are necessary for the proper timing of puberty, and adrenarche is independent; it is not controlled by the same mechanism that turns on the gonads.[212] Indeed, a longitudinal study concluded that increasing adrenal steroid activity reflects gradually increasing maturation with aging, without a sudden change associated with puberty.[213] Neither is there a definite relationship demonstrated between melatonin secretion and puberty. Because the studies have focused on the amount of melatonin secreted rather than the rhythm of

secretion, this question remains open. Nutritional status influences reproductive function, and the leptin system of communication probably contributes to the onset of puberty, but it is unlikely that it is the primary signal (discussed in Chapter 19). Leptin circulating levels can indicate readiness for puberty, in this case, an indication of an adequate amount of fat tissue in the body to support the metabolic requirements of reproduction.[214] Indeed, a more logical concept is to view puberty as the coming together of multiple systems and influences, including genetic, metabolic, and hormonal factors.



Prior to puberty, gonadotropin levels are low but still associated with pulses (although quite irregular).[215] This prepubertal restraint involves brain peptides, especially γ-aminobutyric acid (GABA) and neuropeptide Y.[209, 216] The clinical onset of puberty is preceded by an increase in pulse frequency, amplitude, and regularity, especially during the night.[215, 217–220] At the time of appearance of secondary sex characteristics, the mean LH levels are 2 to 4 times higher during sleep than during wakefulness. This pattern is not present before or after puberty and is an early sign of changes taking place in the hypothalamus, where there is increasing coordination of GnRH neurons with increasing GnRH pulsatile secretion. This pattern can be detected in individuals who develop increasing and decreasing degrees of hypothalamic suppression (e.g., in individuals with worsening and improving anorexia nervosa). FSH levels plateau by midpuberty, whereas LH and estradiol levels continue to rise until late puberty. Biologically active LH has been found to rise proportionately more than immunoreactive LH with the onset of puberty.

The rise of gonadotropins at puberty must be independent of the gonads because the same response can be observed in patients with gonadal dysgenesis (who lack functional steroid-producing gonadal tissue). Adolescent girls with Turner syndrome (45,X) also demonstrate augmented gonadotropin secretion during sleep.[221] Thus, maturation at puberty must involve changes in the hypothalamus that are independent of ovarian steroids.

The maturational change in the hypothalamus is followed by an orderly and predictable sequence of events. Increased GnRH secretion leads to increased pituitary responsiveness to GnRH (a combination of steroid influence on the pituitary and a frequency effect of GnRH pulses on GnRH receptor numbers), leading to increasing production and secretion of gonadotropins. Increased gonadotropins are responsible for follicular growth and development in the ovary and increased sex steroid levels. The rising estrogen contributes to achieving an adult pattern of pulsatile GnRH secretion, finally leading to cyclic menstrual patterns.

The trend toward lowering of the menarcheal age and the period of acceleration of growth has slowed, but continued. In a 10-year prospective study of middle class

American girls in the 1970s, the mean age of menarche was 12.83.[222] Studies analyzing data from the National Health and Nutrition Examination Survey (NHANES) have observed a 2.3 month decrease in the average age of menarche between surveys for the years 1988–1994 (12.53 years) and 1999–2002 (12.34 years), and an overall 4.9 month decrease since 1960.[223] The decrease in age of menarche has been observed in all ethnic groups, declining from 12.57 to 12.52 years in non-Hispanic white girls, from 12.09 to 12.06 years for non-Hispanic black girls, and from 12.24 to 12.09 for Hispanic American girls.0

Ethnic Group	Mean Age of Menarche
Black girls	12.06 years
Mexican-American girls	12.09 years
White girls	12.52 years

The age of onset of puberty is variable and influenced by genetic factors, socioeconomic conditions, and general health. The earlier menarche today compared with the past is undoubtedly due to improved nutrition and better health. It has been suggested that initiation of growth and menarche occur at a particular body weight (48 kg) and percent of body fat (17%).[224] It is thought that this relationship reflects a required stage of metabolism. Although this hypothesis of a critical weight is a helpful concept, the extreme variability in onset of menarche indicates that there is no particular age or size at which an individual girl should be expected to experience menarche.

In the female, the typical sequence of events is growth initiation, thelarche, pubarche, and finally menarche. This generally begins sometime between 8 and 14 years of age. The length of time involved in this evolution is usually 2–4 years. During this time span, puberty is said to occur. Individual variation in the order of appearance of this sequence is great. For example, growth of pubic hair and breast development are not always correlated.

Puberty is due to the reactivation of the hypothalamic-pituitary axis, once fully active during fetal life but suppressed during childhood. If the systems are potentially responsive, what holds function in check until puberty? The hypothalamic-pituitary-gonadal system is operative prior to puberty but is extremely sensitive to sex steroids and, therefore, suppressed. However, the typical "diphasic" pattern of gonadotropin secretion from infancy to puberty results primarily from changing levels of central inhibition of pulsatile GnRH secretion, and to a lesser extent, from a high sensitivity to low levels of gonadal steroid feedback. Negative feedback of steroids cannot be the sole explanation for the low gonadotropin levels in children because agonadal children show the same decline in gonadotropins from age 2 to 6 as do normal children.[225] This indicates an intrinsic CNS inhibitory mechanism independent of gonadal steroids. Therefore, the restraint of puberty can be viewed as the result of two forces:

1. A CNS inhibitory force, a mechanism suppressing GnRH pulsatile secretion.

2. A very sensitive negative feedback of gonadal steroids (6–15 times more sensitive before puberty).

Because agonadal children show a rise in gonadotropins at pubertal age following suppression to a nadir during childhood, the dominant mechanism must be a CNS inhibitory force. The initial maturational change in the hypothalamus would then be a decrease in this inhibitory influence. A search for this mechanism continues. Some have argued that, rather than a chronic state of inhibition prior to puberty, the GnRH neurons exist in an unrestrained but uncoordinated pattern of activity that prevents adequate secretion.

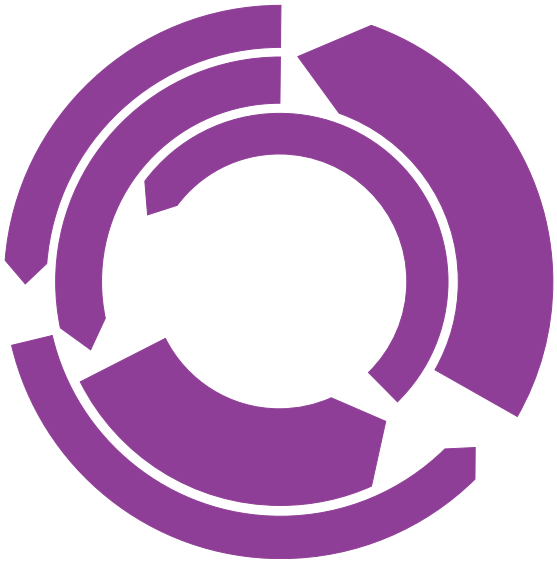
The pubertal changes result in increasing GnRH pulsatile secretion, leading to increasing gonadotropin production and ovarian stimulation, and finally to increasing estrogen levels. The reason that FSH is the first gonadotropin to rise at puberty is that arcuate activity begins with a low frequency of GnRH pulses. This is associated with a rise in FSH and little change in LH. With acceleration of frequency, FSH and LH reach adult levels. In addition, there is a qualitative change as a greater increase occurs in the bioactive forms of the gonadotropins.

The development of the positive feedback response to estrogen occurs later. This explains the well-known finding of anovulation in the first months (as long as 18 months) of menstruation. There are frequent exceptions, however, and ovulation can occur even at the time of menarche.

Don't think of puberty as being turned on by a controlling center in the brain but rather as a functional confluence of all factors. This is more a concept than an actual locus of action. The overall result of this change in the hypothalamus is the development of secondary sex characteristics, attainment of adult set-point levels, and the ability to reproduce. Neoplastic and vascular disorders that alter hypothalamic sensitivity can reverse the prepubertal threshold restraint and lead to precocious puberty.

All references are available online at: http://www.clinicalgynendoandinfertility.com

6 Regulation of the Menstrual Cycle



Many superstitious beliefs have surrounded menstruation throughout recorded history. Indeed, attitudes and ideas about this aspect of female physiology have changed slowly. Hopefully, the scientific progress of the last few decades, which has revealed the dynamic relationships between the pituitary and gonadal hormones and the cyclic nature of the normal reproductive process, will yield a better understanding. The hormone changes, correlated with the morphologic and autocrine-paracrine events in the ovary, make the coordination of this system one of the most remarkable events in biology.

The diagnosis and management of abnormal menstrual function must be based on an understanding of the physiologic mechanisms involved in the regulation of the normal cycle. To understand the normal menstrual cycle, it is helpful to divide the cycle into three phases: the follicular phase, ovulation, and the luteal phase. We will examine each of these phases, concentrating on the changes in ovarian and pituitary hormones, what governs the pattern of hormone changes, and the effects of these hormones on the ovary, pituitary, and hypothalamus in regulating the menstrual cycle.

The Follicular Phase

During the follicular phase an orderly sequence of events takes place that ensures the proper number of follicles is ready for ovulation. In the human ovary the end result of

this follicular development is (usually) one surviving mature follicle. This process, which occurs over the space of 10–14 days, features a series of sequential actions of hormones and autocrine-paracrine peptides on the follicle, leading the follicle destined to ovulate through a period of initial growth from a primordial follicle through the stages of the preantral, antral, and preovulatory follicle.

The Primordial Follicle

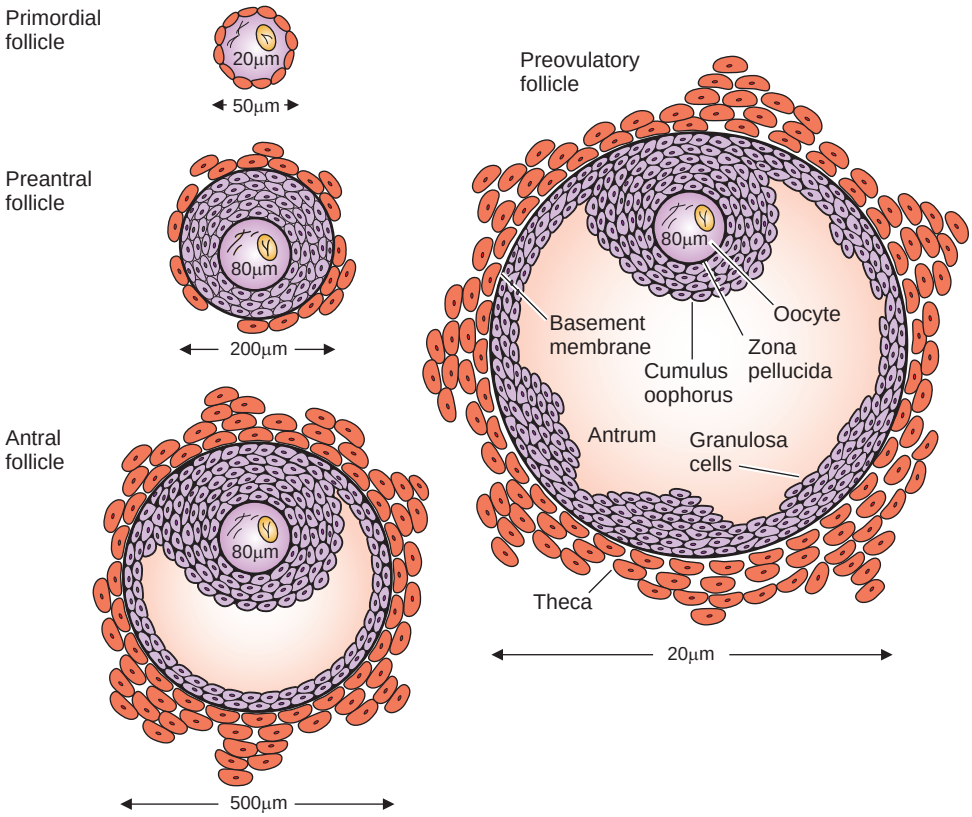
The primordial germ cells originate in the endoderm of the yolk sac, allantois, and hindgut of the embryo, and by 5–6 weeks of gestation, they have migrated to the genital ridge. A rapid mitotic multiplication of germ cells begins at 6–8 weeks of pregnancy, and by 16–20 weeks, the maximum number of oocytes is reached: a total of 6–7 million in both ovaries.[1] The primordial follicle is non-growing and consists of an oocyte, arrested in the diplotene stage of meiotic prophase, surrounded by a single layer of spindle-shaped granulosa cells.

Until their numbers are exhausted, follicles begin to grow and undergo atresia under all physiologic circumstances. Growth and atresia are not interrupted by pregnancy, ovulation, or periods of anovulation. This dynamic process continues at all ages, including infancy and around the menopause. From the maximum number at 16–20 weeks of pregnancy, the number of oocytes will irretrievably decrease. The rate of decrease is proportional to the total number present; thus, the most rapid decrease occurs before birth, resulting in a decline from 6–7 million to 2 million at birth and to 300,000 at puberty. From this large reservoir, about 400 follicles will ovulate during a woman's reproductive years.

The mechanism for determining which follicles and how many will start growing during any one cycle is unknown. The number of follicles that starts growing each cycle appears to be dependent upon the size of the residual pool of inactive primordial follicles.[2, 3] Reducing the size of the pool (e.g., unilateral oophorectomy) causes the remaining follicles to redistribute their availability over time. It is possible that the follicle which is singled out to play the leading role in a particular cycle is the beneficiary of a timely match of follicle "readiness" (perhaps prepared by autocrine-paracrine actions in its microenvironment) and appropriate tropic hormone stimulation. The first follicle able to respond to stimulation may achieve an early lead that it never relinquishes. Nevertheless, each cohort of follicles that begins growth is engaged in a serious competition that ends with only one follicle succeeding.

Rescue From Atresia (Apoptosis)

The follicle destined to ovulate is recruited in the first few days of the cycle.[4] The early growth of follicles occurs over the time span of several menstrual cycles, but the ovulatory follicle is one of a cohort recruited at the time of the luteal-follicular transition.[5, 6] The total duration of time to achieve preovulatory status is approximately 85 days. The majority of this time (until a late stage) involves responses that are independent of hormonal regulation.[7] Eventually, this cohort of follicles reaches a stage where, unless recruited (rescued) by follicle-stimulating hormone (FSH), the next step is atresia. Thus, follicles are continuously available (2–5 mm in size) for a response to FSH. An increase in FSH is the critical feature in rescuing a cohort of follicles from atresia, the usual fate of most follicles, eventually allowing a dominant follicle to emerge and pursue a path to ovulation. In addition, maintenance of this increase in FSH for a critical duration of time is essential.[8] Without the appearance and persistence of an increase in the circulating FSH level, the cohort is doomed to the process of apoptosis, programmed physiologic cell death to eliminate superfluous cells.[9] "Apoptosis" is derived from Greek and means falling off, like leaves from a tree.



"Recruitment" has been traditionally used to describe the continuing growth of antral follicles in response to FSH. A more useful concept is that the cohort of follicles responding to FSH at the beginning of a cycle is *rescued* from apoptosis. Remember that the very early development of follicles begins continuously and independently from gonadotropin influence. The fate of almost all of these follicles is apoptosis; only those exposed to an increase in FSH stimulation because of the juxtaposition of their readiness to respond and the increase in FSH during the luteal-follicular transition have the good fortune to compete for selection as a dominant follicle.

The first visible signs of follicular development are an increase in the size of the oocyte and the granulosa cells becoming cuboidal rather than squamous in shape. These changes are better viewed as a process of maturation rather than growth. At this same time, small gap junctions develop between the granulosa cells and the oocyte. Gap junctions are channels that when open permit the exchange of nutrients, ions, and regulatory molecules. Thus, the gap junctions serve as the pathway for nutritional, metabolite, and signal interchange between the granulosa cells and the oocyte. In one direction, inhibition of the final maturation of the oocyte (until the LH surge) is maintained by factors derived from the granulosa cells. In the other direction, the process of follicular growth is influenced by regulatory factors that originate in the oocyte.

The molecular events that regulate primordial follicle formation involve a variety of factors, all locally produced and regulated, including members of the transforming growth factor β (TGF-β) superfamily of proteins and another family of trophic factors called neurotrophins. Activins, inhibins, antimüllerian hormone (AMH) and bone morphogenetic proteins (BMPs) are members of the TGF-β family of proteins. Activins promote and inhibins retard primordial follicle development, and their relative local concentrations in the fetal ovary during the time of follicle assembly may determine the size of the ovarian follicular pool.[10] AMH is an important inhibitor of primordial follicle growth, and

BMPs exert the opposite effect.[10] Neurotrophins and their receptors are essential for the differentiation and survival of various neuronal populations in the central and peripheral nervous systems, but their presence in the developing ovary suggests they also play a role in ovarian development. Four mammalian neurotrophins have been identified, including nerve growth factor (NGF), brain-derived neurotropic factor (BDNF), neurotrophin-3 (NT-3), and neurotrophin 4/5 (NT-4/5), all of which exert their actions via binding to high-affinity trans-membrane tyrosine kinase receptors encoded by members of the *trk* proto-oncogene family (NGF to TrkA, BDNF and NT-4/5 to TrkB, and NT-3 to TrkC).[11] Observations in NGF- and TrkA-null mice indicate that NGF stimulates the proliferation of ovarian mesenchymal cells during the early stages of follicular assembly and promotes differentiation and synthesis of FSH receptors in granulosa cells. Similar experiments with TrkB-null mice suggest that TrkB signaling is required for oocyte survival after follicular assembly and for preantral follicular development.[11] The specific signaling mechanisms that mediate the effects of activins, inhibins, BMPs and neurotrophins remain to be established.

Other paracrine factors mediate a bi-directional communication between oocytes and their surrounding granulosa cells. Oocytes are linked to their investment of granulosa cells via gap junctions which allow passage of small molecules such as ions (e.g., calcium), metabolites (e.g., pyruvate, nucleic acids, inositol), amino acids (e.g., L-alanine), cholesterol, and intracellular signaling molecules (e.g., cyclic adenosine monophosphate, cAMP) between granulosa cells and oocytes. In mice, targeted deletions of gap junction proteins (known as connexins), disrupt follicular and oocyte development.[12] Oocytes are unable to use glucose as an energy source to support meiotic maturation, cannot transport certain amino acids, and lack both the enzymes necessary for cholesterol synthesis and the receptors for its uptake from carrier-borne sources. Consequently, oocytes are dependent on adjacent granulosa cells to metabolize glucose into a usable energy substrate, such as pyruvate, for transport of essential amino acids, such as L-alanine, and for synthesis and transfer of cholesterol.[13] ***To meet their needs, oocytes stimulate glycolysis, amino acid transport, and cholesterol synthesis in granulosa cells via paracrine and juxtacrine signals that promote expression of transcripts involved in these metabolic processes, at least in some species.***[13] Candidate signaling molecules include closely related members of the TGF-β family, growth differentiation factor 9 (GDF9) and BMP15; both are expressed robustly in oocytes and appear crucial for normal ovarian follicle development in mammalian species.[14]

Mice that are genetically deficient in growth differentiation factor-9 (GDF-9), a peptide synthesized only in the oocyte after the primordial follicle becomes a preantral follicle, are infertile because follicular development cannot proceed beyond the primordial follicle stage.[15, 16] Mutations in GDF-9 are rare causes of ovarian failure, but confirmed BMP-15 mutations have not been reported.[17, 18]

Mutations in *FOXL2*, a gene encoding a transcription factor, cause blepharophimosis/ptosis/epicanthus inversus syndrome that is a disorder affecting the eyelid and producing premature ovarian failure.[19, 20] This transcription factor has been demonstrated to be essential for granulosa cell differentiation; indeed, mutations are associated with an absence of the very first sign of follicular development, the change to a cuboidal shape by the granulosa cells.[21]

The gap junction is composed of channels formed by an arrangement of proteins known as connexins, and more recently as GJAs. The connexin gap junctions are essential for growth and multiplication of the granulosa cells, and for the nutrition and regulation of oocyte development.[22] Connexin expression in ovarian follicles is up-regulated by FSH and down-regulated by LH.[23] In addition, FSH maintains an open channel in the gap junctions, a pathway that is closed by LH.[24] After ovulation, the gap junctions are important again in the corpus luteum, when their function is regulated by locally produced oxytocin.[25]

With multiplication of the cuboidal granulosa cells (to approximately 15 cells), the primordial follicle becomes a primary follicle. The granulosa layer is separated from the stromal

cells by a basement membrane called the basal lamina. The surrounding stromal cells differentiate into concentric layers designated the theca interna (closest to the basal lamina) and the theca externa (the outer portion). The theca layers appear when granulosa proliferation produces 3–6 layers of granulosa cells.[6]

The belief that the initiation of follicular growth is independent of gonadotropin stimulation is supported by the persistence of this initial growth in gonadotropin-deficient mutant mice and in anencephalic fetuses.[26, 27] In the vast majority of instances this growth is limited and rapidly followed by atresia. In studies of human ovarian follicles, expression of the gene for the FSH receptor could not be detected until after primordial follicles began to grow.[28] Furthermore, in a woman with an inactivating mutation in the beta subunit FSH gene, antral follicular activity was present, although successful growth and ovulation were impossible.[29] Treatment of FSH-deficient women with exogenous FSH results in follicular growth, ovulation, and pregnancy, demonstrating that oocytes and growth of follicules until the appearance of FSH are normal.[29, 30]

The general pattern of limited growth and quick atresia is interrupted at the beginning of the menstrual cycle when a group of follicles (after approximately 70 days of development) responds to a hormonal change and is propelled to grow. In young women, this cohort numbers 3–11 per ovary.[31] The decline in luteal phase steroidogenesis and inhibin-A secretion allows a rise in FSH, beginning a few days before menses.[32, 33] The timing of this important event was based on data derived from the immunoassay of FSH. Using a sensitive measurement of FSH bioactivity, it has been suggested that increasing bioactivity of FSH begins in the mid- to late luteal phase.[34]

The Preantral Follicle

Once growth is accelerated, the follicle progresses to the preantral stage as the oocyte enlarges and is surrounded by a membrane, the zona pellucida. The granulosa cells undergo a multilayer proliferation as the theca layer continues to organize from the surrounding stroma. This growth is dependent upon gonadotropins and is correlated with increasing production of estrogen. Molecular studies indicate that all of the granulosa cells in mature follicles are derived from as few as three precursor cells.[35]

The granulosa cells of the preantral follicle have the ability to synthesize all three classes of steroids; however, significantly more estrogens than either androgens or progestins are produced. An aromatase enzyme system acts to convert androgens to estrogens and is a factor limiting ovarian estrogen production. Aromatization is induced or activated through the action of FSH. The binding of FSH to its receptor and activation of the adenylate cyclase-mediated signal is followed by expression of multiple mRNAs, which encode proteins responsible for cell proliferation, differentiation, and function. Thus, FSH both initiates steroidogenesis (estrogen production) in granulosa cells and stimulates granulosa cell growth.[36]

Specific receptors for FSH are not detected on granulosa cells until the preantral stage,[28] and the preantral follicle requires the presence of FSH in order to aromatize androgens and generate its own estrogenic microenvironment.[37] Estrogen production is, therefore, limited by FSH receptor content. The administration of FSH will raise and lower the concentration of its own receptor on granulosa cells (up- and down-regulation) both in vivo and in vitro.[38] This action of FSH is modulated by growth factors.[39] FSH receptors quickly reach a concentration of approximately 1,500 receptors per granulosa cell.[40]

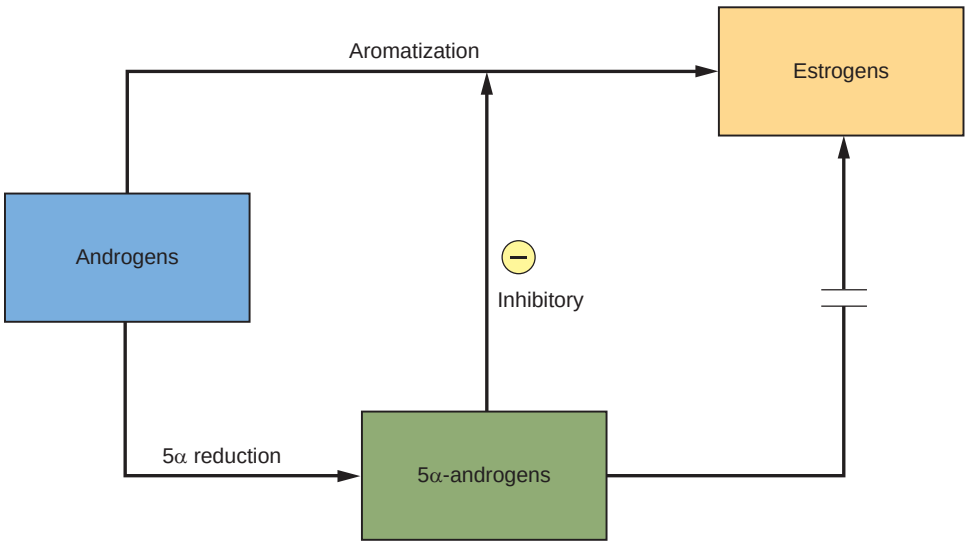
FSH operates through the G protein, adenylate cyclase system (described in Chapter 2), which is subject to down-regulation and modulation by many factors, including a calcium-calmodulin

intermediary. Although steroidogenesis in the ovarian follicle is mainly regulated by the gonadotropins, multiple signaling pathways are involved that respond to many factors besides the gonadotropins. Besides the adenylate cyclase enzyme system, these pathways include ion gate channels, tyrosine kinase receptors, and the phospholipase system of second messengers. These pathways are regulated by a multitude of factors, including growth factors, nitric oxide, prostaglandins, and peptides such as gonadotropin-releasing hormone (GnRH), angiotensin II, tissue necrosis factor-α, and vasoactive intestinal peptide. The binding of luteinizing hormone (LH) to its receptor in the ovary is also followed by activation of the adenylate cyclase-cyclic AMP pathway via the G protein mechanism.

FSH combines synergistically with estrogen to exert (at least in the nonprimate) a mitogenic action on granulosa cells to stimulate their proliferation. Together, FSH and estrogen promote a rapid accumulation of FSH receptors, reflecting in part the increase in the number of granulosa cells. The early appearance of estrogen within the follicle allows the follicle to respond to relatively low concentrations of FSH, an autocrine function for estrogen within the follicle. As growth proceeds, the granulosa cells differentiate into several subgroups of different cell populations. This appears to be determined by the position of the cells relative to the oocyte.

There is a system of communication that exists within follicles. Not every cell has to contain receptors for the gonadotropins. Cells with receptors can transfer a signal (by gap junctions), which causes protein kinase activation in cells that lack receptors.[41] Thus, hormone-initiated action can be transmitted throughout the follicle despite the fact that only a subpopulation of cells binds the hormone. This system of communication promotes a coordinated and synchronous performance throughout the follicle, a system that continues to operate in the corpus luteum.

The role of androgens in early follicular development is complex. Specific androgen receptors are present in the granulosa cells.[42] The androgens serve not only as substrate for FSH-induced aromatization but, in low concentrations, can further enhance aromatase activity. When exposed to an androgen-rich environment, preantral granulosa cells favor the conversion of androgens to more potent 5α-reduced androgens rather than to estrogens.[43] These androgens cannot be converted to estrogen and, in fact, inhibit aromatase activity.[44] They also inhibit FSH induction of LH receptor formation, another essential step in follicular development.[45]



The fate of the preantral follicle is in delicate balance. At low concentrations, androgens enhance their own aromatization and contribute to estrogen production. At higher levels,

the limited capacity of aromatization is overwhelmed, and the follicle becomes androgenic and ultimately atretic.[46] Follicles will progress in development only if emerging when FSH is elevated and LH is low. Those follicles arising at the end of the luteal phase or early in the subsequent cycle would be favored by an environment in which aromatization in the granulosa cell can prevail. ***The success of a follicle depends upon its ability to convert an androgen-dominated microenvironment to an estrogen-dominated microenvironment.***[47, 48]

Summary of Events in the Preantral Follicle

1. Initial follicular development occurs independently of hormone influence.
2. FSH stimulation propels follicles to the preantral stage.
3. FSH-induced aromatization of androgen in the granulosa results in the production of estrogen.
4. Together, FSH and estrogen increase the FSH receptor content of the follicle.

The Antral Follicle

Under the synergistic influence of estrogen and FSH there is an increase in the production of follicular fluid that accumulates in the intercellular spaces of the granulosa, eventually coalescing to form a cavity, as the follicle makes its gradual transition to the antral stage. The accumulation of follicular fluid provides a means whereby the oocyte and surrounding granulosa cells can be nurtured in a specific endocrine environment. The granulosa cells surrounding the oocyte are now designated the **cumulus oophorus**. The differentiation of the cumulus cells is believed to be a response to signals originating in the oocyte.[49] The follicular fluid, rich in hormones, growth factors, and cytokines, provides the milieu that is required for the orderly maturation and development of the oocyte and its surrounding cells.

In the presence of FSH, estrogen becomes the dominant substance in the follicular fluid. Conversely, in the absence of FSH, androgens predominate.[50, 51] LH is not normally present in follicular fluid until the midcycle. If LH is prematurely elevated in plasma and antral fluid, mitotic activity in the granulosa decreases, degenerative changes ensue, and intrafollicular androgen levels rise. Therefore, the dominance of estrogen and FSH is essential for sustained accumulation of granulosa cells and continued follicular growth. Antral follicles with the greatest rates of granulosa proliferation contain the highest estrogen concentrations and the lowest androgen/estrogen ratios, and are the most likely to house a healthy oocyte.[52] An androgenic milieu antagonizes estrogen-induced granulosa proliferation and, if sustained, promotes degenerative changes in the oocyte.

The steroids present in follicular fluid can be found in concentrations several orders of magnitude higher than those in plasma and reflect the functional capacity of the surrounding granulosa and theca cells. The synthesis of steroid hormones is functionally compartmentalized within the follicle: the two-cell system.[40, 46, 51, 53, 54]

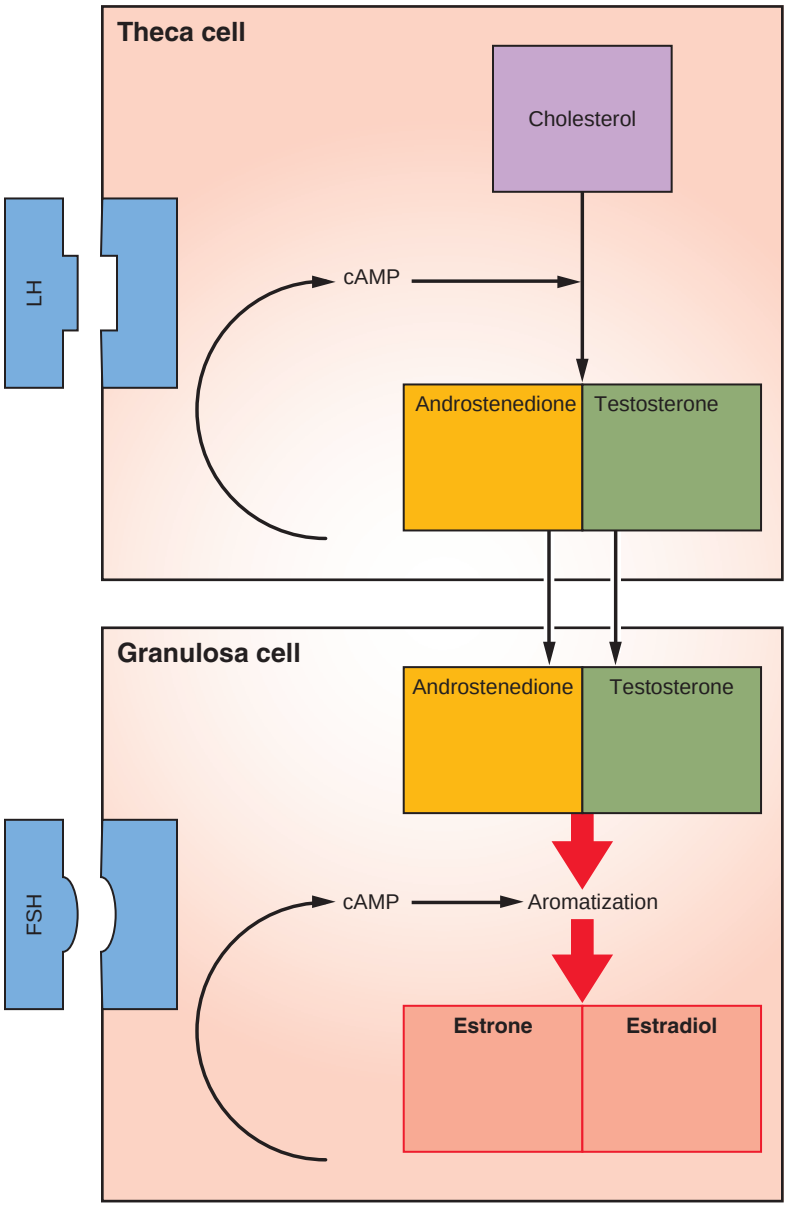
The Two-Cell, Two-Gonadotropin System

The aromatase activity of the granulosa far exceeds that observed in the theca. In human preantral and antral follicles, LH receptors are present only on the theca cells and FSH

receptors only on the granulosa cells.[55, 56] Theca interstitial cells, located in the theca interna, have approximately 20,000 LH receptors in their cell membranes. In response to LH, theca tissue is stimulated to produce androgens that can then be converted, through FSH-induced aromatization, to estrogens in the granulosa cells.

The interaction between the granulosa and theca compartments, with resulting accelerated estrogen production, is not fully functional until later in antral development. Like preantral granulosa cells, the granulosa of small antral follicles exhibits an in vitro tendency to convert significant amounts of androgen to the more potent 5α-reduced form. In contrast, granulosa cells isolated from large antral follicles readily and preferentially metabolize androgens to estrogens. The conversion from an androgen microenvironment to an estrogen microenvironment (a conversion essential for further growth and development) is dependent upon a growing sensitivity to FSH brought about by the action of FSH and the enhancing influence of estrogen.

As the follicle develops, theca cells begin to express the genes for LH receptors, P450scc, and 3β-hydroxysteroid dehydrogenase.[57] The separately regulated (by LH) entry of



cholesterol into mitochondria, utilizing internalization of LDL-cholesterol, is essential for steroidogenesis. ***Therefore, ovarian steroidogenesis is LH-dependent to a significant degree***. Human ovarian granulosa cells, after luteinization and vascularization that occur following ovulation, can use HDL-cholesterol in a system that differs from the LDL-cholesterol pathway. The HDL lipoproteins are not internalized, but rather, the cholesteryl esters are extracted from the lipoproteins at the cell surface and then transferred into the cell.[58]

As the follicle emerges, the theca cells are characterized by their expression of P450c17, the enzyme step which is rate-limiting for the conversion of 21-carbon substrate to androgens.[59] Granulosa cells do not express this enzyme and thus are dependent upon androgens from the theca in order to make estrogen. Increasing expression of the aromatization system (P450arom) is a marker of increasing maturity of granulosa cells. The presence of P450c17 only in theca cells and P450arom only in granulosa cells is impressive evidence confirming the two-cell, two-gonadotropin explanation for estrogen production.[60]

The importance of the two-cell, two-gonadotropin system in the primate is supported by the response of women with a deficiency in gonadotropins to treatment with recombinant (pure) FSH.[61–63] Follicles developed in these women (confirming the essential role of FSH, and the lesser role for LH, in recruitment and initial growth), but estradiol production was limited. Some aromatization occurred, perhaps using androgens originating in the adrenal glands, producing early follicular phase estradiol levels, but the usual robust steroidogenesis was impossible without the presence of LH to provide theca production of androgen substrate. This same response has been observed in experiments that use a GnRH antagonist to produce LH-deficient monkeys, followed by the administration of recombinant, pure human FSH.[64, 65] ***These results indicate that only FSH is required for folliculogenesis, and that in the primate, autocrine-paracrine peptides have assumed the important intraovarian role of modulating gonadotropin response. However, the final stages of maturation are optimized by LH, increasing the amount of androgen substrate for estrogen production and promoting the growth of the dominant follicle while simultaneously hastening the regression of smaller follicles.***[66]

Selection of the Dominant Follicle

The successful conversion to an estrogen dominant follicle marks the "selection" of a follicle destined to ovulate, the process whereby, with rare exception, only a single follicle succeeds.[67] This selection process is to a significant degree the result of two estrogen actions: (1) a local interaction between estrogen and FSH within the follicle, and (2) the effect of estrogen on pituitary secretion of FSH. While estrogen exerts a positive influence on FSH action within the maturing follicle, its negative feedback relationship with FSH at the hypothalamic-pituitary level serves to withdraw gonadotropin support from the other less developed follicles. The fall in FSH leads to a decline in FSH-dependent aromatase activity, limiting estrogen production in the less mature follicles. Even if a lesser follicle succeeds in achieving an estrogen microenvironment, decreasing FSH support would interrupt granulosa proliferation and function, promote a conversion to an androgenic microenvironment, and thereby induce irreversible atretic change. Indeed, the first event in the process of atresia is a reduction in FSH receptors in the granulosa layer.

The loss of oocytes (and follicles) through atresia is a response to changes in many factors. Certainly gonadotropin stimulation and withdrawal are important, but ovarian steroids and autocrine-paracrine factors are also involved. The consequence of these unfavorable changes, atresia, in the process called ***apoptosis***, programmed cell death, is heralded by alterations in mRNAs required for cell proteins that maintain follicle integrity.[68] This type of "natural death" is a physiologic process, in contrast to the pathologic cell death of necrosis.

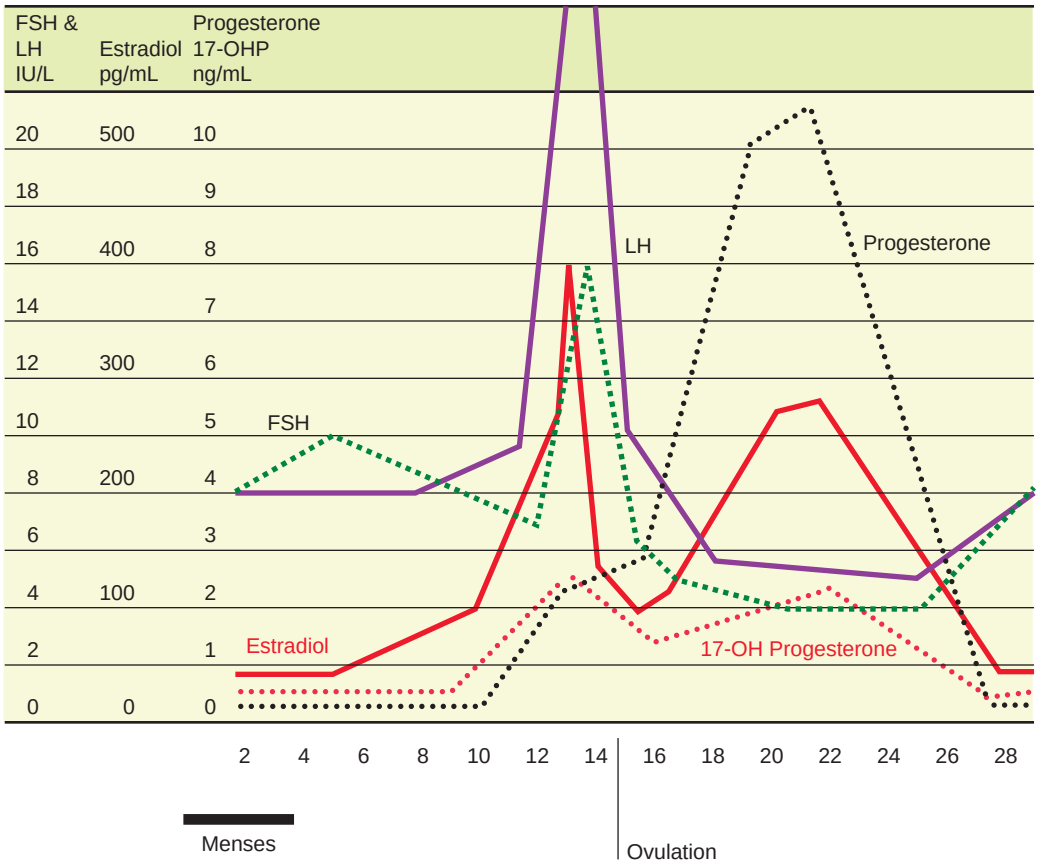
Once cells have entered the process of apoptosis, their response to FSH is modulated by local growth factors. Tumor necrosis factor (TNF), produced in the granulosa cells, inhibits FSH stimulation of estradiol secretion, except in the dominant follicle.[69] An inverse relationship exists between TNF expression and gonadotropin stimulation of granulosa cells. Thus, as the successful follicle increases its response to gonadotropins, its TNF production decreases. Those follicles with a failing response to gonadotropins increase their TNF production, hastening their demise.

Although the principal function of anti-Müllerian hormone (AMH) is to cause Müllerian duct regression during male sexual differentiation, it is detected in the granulosa cells of early primordial follicles and reaches peak concentrations in small antral follicles.[70] Its secretion appears to be regulated by the mature oocyte, and AMH decreases when FSH-stimulated follicular growth and estrogen production occur.[71, 72] Studies with knock-out model mice have indicated that AMH inhibits the recruitment of primordial follicles.[73] The paracrine activity of AMH inhibits FSH-stimulated follicle growth, thus suppressing the growth of lesser follicles and allowing the dominant follicle to emerge. Because of these activities, the circulating level of AMH reflects the number of growing follicles, and the blood concentration of AMH can be a measure of ovarian aging and prognosis for fertility.[74] Because AMH levels are not affected by gonadotropins or the sex steroids, measurement of AMH is reliable on any day in an individual's menstrual cycle, even in women on steroid contraception.[75]

An asymmetry in ovarian estrogen production, an expression of the emerging dominant follicle, can be detected in ovarian venous effluent on day 5 of the cycle, corresponding with the gradual fall of FSH levels observed at the midfollicular phase and preceding the increase in diameter that marks the physical emergence of the dominant follicle.[76] This is a crucial time in the cycle. Exogenous estrogen, administered even after selection of the dominant follicle, disrupts preovulatory development and induces atresia by reducing FSH levels below the sustaining level. Because the lesser follicles have entered the process of atresia, loss of the dominant follicle during this period of time requires beginning over, with recruitment of another set of preantral follicles.[77]

The negative feedback of estrogen on FSH serves to inhibit the development of all but the dominant follicle. The selected follicle remains dependent upon FSH and must complete its preovulatory development in the face of declining plasma levels of FSH. The dominant follicle, therefore, must escape the consequences of FSH suppression induced by its own accelerating estrogen production. ***The dominant follicle has two significant advantages, a greater content of FSH receptors acquired because of a rate of granulosa proliferation that surpasses that of its cohorts and enhancement of FSH action because of its high intrafollicular estrogen concentration, a consequence of local autocrine-paracrine molecules.*** Thus, the dominant follicle is more sensitive to FSH, and as long as a critical duration of FSH exposure was initially present, the dominant follicle continues to develop.[8] As a result, the stimulus for aromatization, FSH, can be maintained, while at the same time it is being withdrawn from among the less developed follicles. A wave of atresia among the lesser follicles, therefore, is seen to parallel the rise in estrogen.

The accumulation of a greater mass of granulosa cells is accompanied by advanced development of the theca vasculature. By day 9, theca vascularity in the dominant follicle is twice that of other antral follicles.[78] This allows a preferential delivery of gonadotropins to the follicle, permitting the dominant follicle to retain FSH responsiveness and sustain continued development and function despite waning gonadotropin levels. The monkey ovary expresses a potent growth factor (vascular endothelial growth factor) that induces angiogenesis, and this expression is observed at the two development points when proliferation of capillaries is important: the emerging dominant follicle and the early corpus luteum.[79, 80]



In order to respond to the ovulatory surge and to become a successful corpus luteum, the granulosa cells must acquire LH receptors. FSH induces LH receptor development on the granulosa cells of the large antral follicles. Here again estrogen and local autocrine-paracrine peptides (primate), serve as the chief coordinators. With increasing concentrations of estrogen within the follicle, FSH changes its focus of action, from up-regulating its own receptor to generation of the LH receptors.[81] The combination of a capacity for continued response despite declining levels of FSH and a high local estrogen environment in the dominant follicle provides optimal conditions for LH receptor development. LH can induce the formation of its own receptor in FSH-primed granulosa cells, but the primary mechanism utilizes FSH stimulation and estrogen enhancement.[82, 83] The role for estrogen goes beyond synergism and enhancement; it is obligatory.

Evidence from ovarian stimulation for in vitro fertilization indicates that LH plays a critical role in the late stages of follicle development, providing support for the final maturation and function of the dominant follicle.[66, 84] At least one LH contribution in the late follicular phase is the LH stimulation of androgen production in the theca to provide for the large amounts of estrogen required at this point in the cycle. In addition, the theca androgens may have a direct beneficial effect on essential growth factors within the follicle. LH presence in the follicle prior to ovulation, therefore, is an important contributor to optimal follicular development that ultimately provides a healthy oocyte.[85, 86]

The local action of estrogen within the ovarian follicle was questioned when initial studies failed to detect estrogen receptors in any of the significant ovarian compartments.[87] Subsequently, it was discovered that human granulosa cells and primate oocytes contain only mRNA for estrogen receptor-β.[88–91] The dynamic expression of estrogen receptor-β is consistent with an important local role for estrogen in ovarian follicle and corpus luteum growth and function.

Although prolactin is always present in follicular fluid, there is no evidence to suggest that prolactin is important during normal ovulatory cycles in the primate.

The Feedback System

Through its own estrogen and peptide production, the dominant follicle assumes control of its own destiny. By altering gonadotropin secretion through feedback mechanisms it optimizes its own environment to the detriment of the lesser follicles.

As reviewed in Chapter 5, gonadotropin-releasing hormone (GnRH) plays an obligatory role in the control of gonadotropin secretion, but the pattern of gonadotropin secretion observed in the menstrual cycle is the result of feedback modulation of steroids and peptides originating in the dominant follicle, acting directly on the hypothalamus and anterior pituitary.[4] An increase in GnRH accompanying the LH surge, indicating that estrogen positive feedback operates at both pituitary and hypothalamic sites, has been reported in monkeys, but not in women.[92, 93] Positron emission tomography studies in women indicated that the positive feedback effects of estrogen on LH occur at the pituitary.[94]

Estrogen exerts its inhibitory effects in both the hypothalamus and the anterior pituitary, decreasing both GnRH pulsatile secretion and GnRH pituitary response.[95] Progesterone also operates in two sites. Its inhibitory action is at the hypothalamic level, and its positive action is directly on the pituitary.[96] As determined by positron emission tomography, the primary site of estrogen negative feedback on LH is the hypothalamus.[94]

The secretion of FSH is very sensitive to the negative inhibitory effects of estrogen even at low levels. At higher levels, estrogen combines with inhibin for a suppression of FSH that is profound and sustained. In contrast, the influence of estrogen on LH release varies with concentration and duration of exposure. At low levels, estrogen imposes a negative feedback relationship with LH. At higher levels, however, estrogen is capable of exerting a positive stimulatory feedback effect on LH release.

The transition from suppression to stimulation of LH release occurs as estradiol rises during the midfollicular phase. There are two critical features in this mechanism: (1) the concentration of estradiol and (2) the length of time during which the estradiol elevation is sustained. In women, the estradiol concentration necessary to achieve a positive feedback is more than 200 pg/mL, and this concentration must be sustained for approximately 50 hours.[97] This level of estrogen essentially never occurs until the dominant follicle has reached a diameter of 15 mm.[98] The estrogen stimulus must be sustained beyond the initiation of the LH surge until after the surge actually begins. Otherwise, the LH surge is abbreviated or fails to occur at all.

Within the well-established monthly pattern, the gonadotropins are secreted in a pulsatile fashion with a frequency and magnitude that vary with the phase of the cycle. The pulsatile pattern is directly due to a similar pulsatile secretion of GnRH, but amplitude and frequency modulation are the consequence of steroid feedback on both hypothalamus and anterior pituitary.[99–101] ***Pulsatile secretion is more frequent but smaller in amplitude during the follicular phase compared to the luteal phase, with a slight increase in frequency observed as the follicular phase progresses to ovulation***.

The pulsatile pattern of FSH is not easily discerned because of its relatively longer half-life compared to LH, but the experimental data indicate that FSH and LH are secreted simultaneously and that GnRH stimulates the secretion of both gonadotropins. Even as late as only 36–48 hours before menses, gonadotropin secretion is still characterized by infrequent LH pulses and low FSH levels typical of the late luteal phase.[100] During the transition from the

previous luteal phase to the next follicular phase, GnRH and the gonadotropins are released from the inhibitory effects of estradiol, progesterone, and inhibin. A progressive and fairly rapid increase in GnRH pulse secretion is associated with a preferential secretion of FSH compared to LH. The frequency of GnRH and LH pulses increases 4.5-fold during this period of time, accompanied by a 3.5-fold increase in the circulating levels of FSH, and a lesser 2-fold increase in LH levels.[102]

The GnRH pulse frequency changes in the luteal phase correlate with duration of exposure to progesterone, while pulse amplitude changes appear to be influenced by changes in progesterone levels.[99] Both estradiol and progesterone are required to achieve the low, suppressed secretory pattern of GnRH during the luteal phase.[103] The evidence suggests that steroids influence the hypothalamic release of GnRH for frequency changes and the pituitary for action on amplitude of the gonadotropin pulses. The inhibitory action of luteal phase steroids appears to be mediated by an increase in hypothalamic endogenous opioid peptides. Both estrogen and progesterone can increase endogenous opiates, and administration of clomiphene (an estrogen antagonist) during the luteal phase increases the LH pulse frequency with no effect on amplitude.[104] Thus, estrogen appears to enhance the stimulatory action of progesterone in the luteal phase on endogenous opioid peptides, creating relatively high levels of endogenous opiates during the luteal phase.

Plasma endorphin begins to rise in the 2 days before the LH peak, coinciding with the midcycle gonadotropin surge.[105] The maximal level is reached just after the LH peak, coinciding with ovulation. Levels then gradually decline until the nadir is reached during menses and the early follicular phase. Monkeys have their highest beta-endorphin levels in the hypophyseal portal blood at midcycle.[106] ***Normal cyclicity requires sequential periods of high (midcycle and luteal phase) and low (during menses) hypothalamic opioid activity.***

There is another important action of estrogen. A disparity exists between the patterns of FSH and LH secretion as determined by immunoassay and bioassay, indicating that more biologically active gonadotropins are secreted at midcycle than at other times in the cycle.[107, 108] This quality, bioactivity vs immunoreactivity, is determined by the molecular structure of the gonadotropin molecule, a concept referred to in Chapter 2 as heterogeneity of the tropic hormones. There is a well-established relationship between the activity and half-life of glycoprotein hormones and their sialic acid content. The feedback effects of estrogen include modulation of sialylation and the size and activity of the gonadotropins subsequently released, as well as an augmentation of GnRH-stimulated secretory release of biologically active gonadotropin. It certainly makes sense to intensify the gonadotropin effect at midcycle. The positive feedback action of estrogen, therefore, increase both the quantity and the quality (the bioactivity) of FSH and LH. In addition to the change at midcycle that favors gonadotropin activity at the ovarian follicle, FSH isoforms with greater biologic activity also increase during the late luteal phase, a change that is appropriately geared toward propelling new ovarian follicle growth for the next cycle.[109]

There is a diurnal rhythm in FSH and LH secretion.[110] In contrast to the nocturnal rise seen with ACTH, thyroid-stimulating hormone (TSH), growth hormone, and prolactin, FSH and LH exhibit nocturnal decline, probably mediated by endogenous opiates. This diurnal rhythm for LH is present only in the early follicular phase, while FSH maintains a circadian rhythm throughout the menstrual cycle (and thus it is not influenced by steroid hormone feedback) and even in the postmenopausal period of life.

Inhibin, Activin, Follistatin

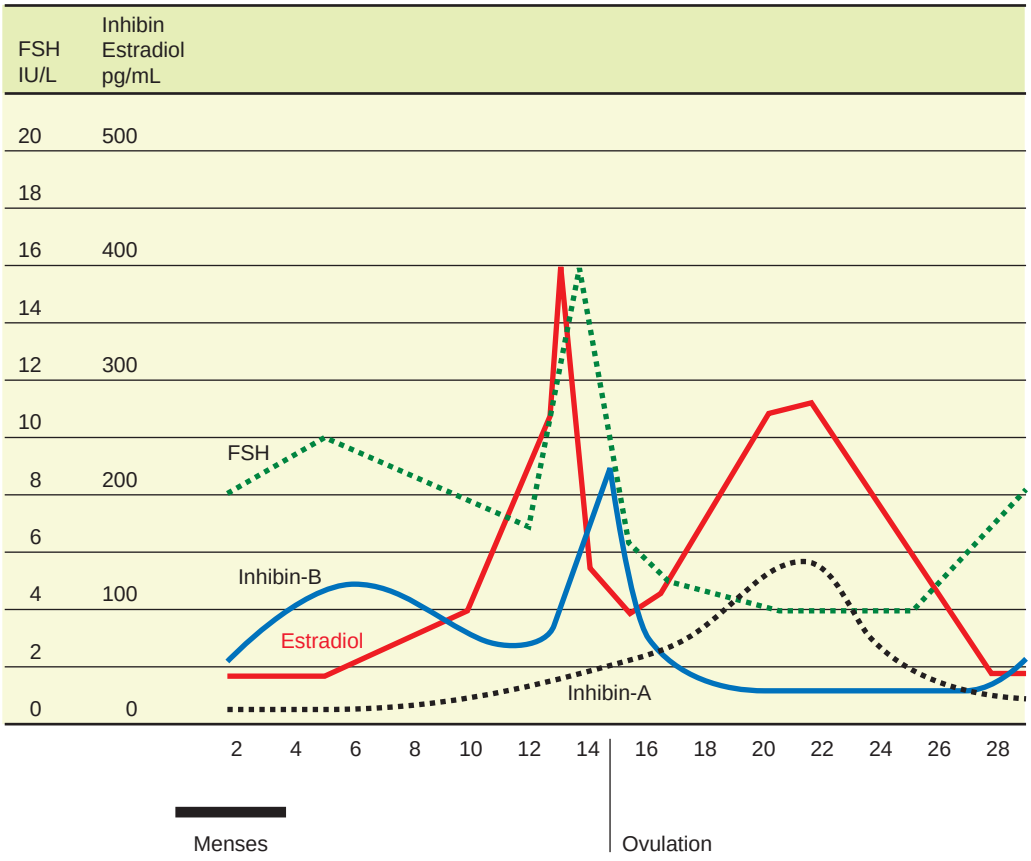
This family of peptides is synthesized by granulosa cells in response to FSH and secreted into the follicular fluid and ovarian venous effluent.[111–114] The expression of these peptides

is not limited to the ovary; they are present in many tissues throughout the body serving as autocrine-paracrine regulators. Inhibin is an important inhibitor of FSH secretion. Activin stimulates FSH release in the pituitary and augments FSH action in the ovary. Follistatin suppresses FSH activity, by binding activin.

Inhibin consists of two dissimilar peptides (known as alpha- and beta-subunits) linked by disulfide bonds. Two forms of inhibin (inhibin-A and inhibin-B) have been purified, each containing an identical alpha-subunit and distinct but related beta-subunits. Thus, there are three subunits for inhibins: alpha, beta-A, and beta-B. Each subunit is a product of different messenger RNA, each derived from its own precursor molecule.

The 2 Forms of Inhibin:
Inhibin-A: Alpha-BetaA
Inhibin-B: Alpha-BetaB

FSH stimulates the secretion of inhibin from granulosa cells and, in turn, is suppressed by inhibin—a reciprocal relationship.[115, 116] Refinements in assay techniques have revealed that inhibin-B is the form of inhibin predominantly secreted by granulosa cells in the follicular phase of the cycle.[117, 118] The secretion of inhibin is further regulated by local autocrine-paracrine control. GnRH and epidermal growth factor diminish FSH stimulation of inhibin secretion, whereas insulin-like growth factor-I enhances inhibin production. The inhibitory effects of GnRH and epidermal growth factor are consistent with their known ability to decrease FSH-stimulated estrogen production and LH receptor formation. The two forms of GnRH (GnRH-I and GnRH-II) along with their receptor are expressed in granulosa cells.[119, 120]



The secretion of inhibin-B into the circulation further amplifies the withdrawal of FSH from other follicles, another mechanism by which an emerging follicle secures dominance. Inhibin-B rises slowly but steadily, in a pulsatile fashion (60–70 minute periodicity) reaching peak levels in the early and midfollicular phases, and then decreasing in the late follicular phase before ovulation to reach a nadir in the midluteal phase.[33, 117, 121, 122] An inhibin-B peak the day after ovulation is probably the result of release from the ruptured follicle. This relationship of inhibin-B and FSH is supported by the demonstration that inhibin-B levels are lower and FSH levels are higher in the follicular phase in women 45–49 years old compared to younger women.[121, 123] An ovarian fibrothecoma secreting inhibin-B predictably was associated with secondary amenorrhea and infertility due to suppression of FSH secretion.[124]

With the appearance of LH receptors on the granulosa cells of the dominant follicle and the subsequent development of the follicle into a corpus luteum, inhibin expression comes under the control of LH, and expression changes from inhibin-B to inhibin-A.[125] The circulating levels of inhibin-A rise in the late follicular phase to reach a peak level at the midluteal phase.[33, 126] Inhibin-A, therefore, contributes to the suppression of FSH to nadir levels during the luteal phase, and to the changes at the luteal-follicular transition.

Inhibin has multiple, diverse inhibitory effects on gonadotropin secretion. Inhibin can block the synthesis and secretion of FSH, prevent the up-regulation of GnRH receptors by GnRH, reduce the number of GnRH receptors present, and, at high concentrations, promote the intracellular degradation of gonadotropins.

Activin, derived from granulosa cells, but present as well in the pituitary gonadotropes, contains two subunits that are identical to the beta subunits of inhibins A and B. In addition, activins have been identified with variants of the beta subunits, designated as beta-C, beta-D, and beta-E.[127] The activin beta-C and beta-E genes have been demonstrated to be nonessential in mouse knockout models.[128] Activin augments the secretion of FSH and inhibits prolactin, ACTH, and growth hormone responses.[129–132] Activin increases pituitary response to GnRH by enhancing GnRH receptor formation.[133, 134] The effects of activin are blocked by inhibin and follistatin.[135] The structure of the activin genes is homologous to that of transforming growth factor-β, indicating that these products all come from the same gene family.[136] Another important member of this family is the antimüllerian hormone, as well as a protein active during insect embryogenesis, and a protein active in frog embryos.

The 3 Forms of Activin:

Activin A:	**BetaA-BetaA**
Activin AB:	**BetaA-BetaB**
Activin B:	**BetaB-BetaB**

Activin is present in many cell types, regulating growth and differentiation. In the ovarian follicle, activin increases FSH binding in granulosa cells (by regulating receptor numbers) and augments FSH stimulation of aromatization and inhibin production.[113] Considerable evidence derived from human cells exists to indicate that inhibin and activin act directly on theca cells to regulate androgen synthesis.[137–139] Inhibin enhances the stimulatory action of LH and/or IGF-I, while activin suppresses this action. Inhibin in increasing doses can overcome the inhibitory action of activin. Prior to ovulation, activin suppresses granulosa progesterone production, perhaps preventing premature luteinization. There is a repertoire of cell transmembrane kinase receptors for activin, with differing binding affinities and domain structures.[140] This receptor heterogeneity allows the many different responses elicited by a single peptide. Both activin A and inhibin-A have been demonstrated to be very potent in stimulating in vitro maturation of oocytes that subsequently yield a high rate of fertilization.[141]

In the male, activin inhibits and inhibin facilitates LH stimulation of androgen biosynthesis in Leydig cells. In addition, activin stimulates and inhibin decreases spermatogonial

proliferation; inhibin is produced in the Sertoli cell, the locus that has the principal role in modulating spermatogenesis. Thus, activin and inhibin play similar autocrine-paracrine roles in both the male and female gonads.

The anterior pituitary expresses the inhibin-activin subunits, and locally produced activin B augments FSH secretion. Activin A has been demonstrated to directly stimulate the synthesis of GnRH receptors in pituitary cells.[142] Follistatin is a glycopeptide secreted by a variety of pituitary cells, including the gonadotropes.[143] This peptide has also been called FSH-suppressing protein because of its main action: inhibition of FSH synthesis and secretion and the FSH response to GnRH by binding to activin and in that fashion decreasing the activity of activin.[144, 145] Activin stimulates follistatin production, and inhibin prevents this response. Follistatin is also expressed by granulosa cells in response to FSH, and, therefore, follistatin, like inhibin and activin, functions locally in the follicle and in the pituitary.[146] Circulating levels of activin increase in the late luteal phase to peak at menses; however, activin A is highly bound in the circulation, and it is not certain it has an endocrine role.[147] Nevertheless, the timing is right for activin to contribute to the rise in FSH during the luteal-follicular transition.

In summary, the pituitary secretion of FSH can be significantly regulated by the balance of activin and inhibin, with follistatin playing a role by inhibiting activin and enhancing inhibin activity. Within the ovarian follicle, activin and inhibin influence growth and development by modulating theca and granulosa responses to the gonadotropins.

The inhibin-activin family of peptides (also including antimüllerian hormone and transforming growth factor-β) inhibits cell growth and can be considered as a class of tumor-suppressor proteins. Mice have been generated that are deficient in the inhibin-alpha-subunit gene.[114] The mice that are homozygous and lack inhibin are susceptible to the development of gonadal stromal tumors that appear after normal sexual differentiation and development. Thus, the alpha-inhibin gene is a specific tumor-suppressor gene for the gonads. A contributing factor to this tumor development could be the high FSH levels associated with the deficiency in inhibin.

Growth Factors

Growth factors are polypeptides that modulate cell proliferation and differentiation, operating through binding to specific cell membrane receptors. They are not classic endocrine substances; they act locally and function in paracrine and autocrine modes. There are multiple growth factors, and most cells contain multiple receptors for the various growth factors.

Insulin-Like Growth Factors

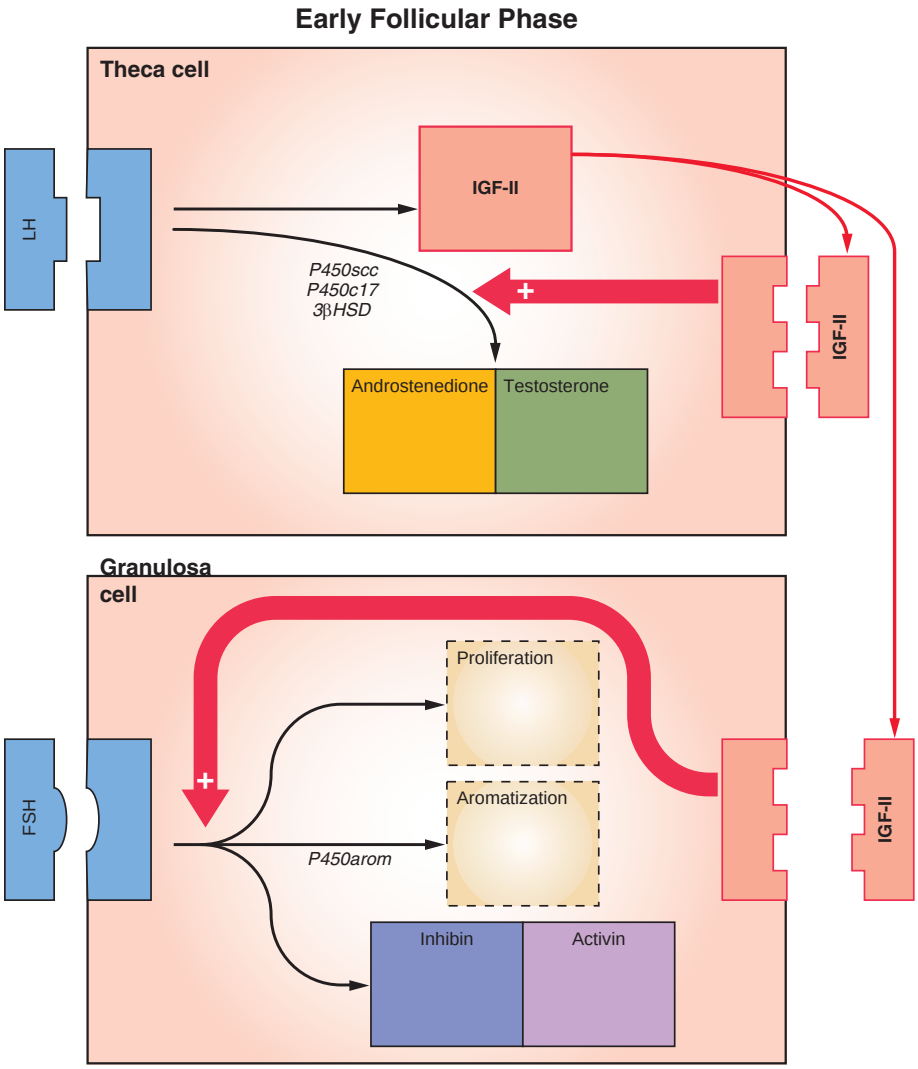
The insulin-like growth factors (also called somatomedins) are peptides that have structural and functional similarity to insulin and mediate growth hormone action.[148] Insulin-like growth factor-I (IGF-I) and insulin-like growth factor-II (IGF-II) are single chain polypeptides containing three disulfide bonds. IGF-I is encoded on the long arm of chromosome 12 and IGF-II on the short arm of chromosome 11 (which also contains the insulin gene). The genes are subject to a variety of promoters, and thus differential regulation can govern ultimate actions.

IGF-I mediates the growth-promoting actions of growth hormone. The majority of circulating IGF-I is derived from growth hormone-dependent synthesis in the liver. However, IGF-I is synthesized in many tissues where production can be regulated in conjunction with growth hormone or ***independently*** by other factors.

IGF-II has little growth hormone dependence. It is believed to be important in fetal growth and development. Both IGFs induce the expression of cellular genes responsible for cellular proliferation and differentiation.

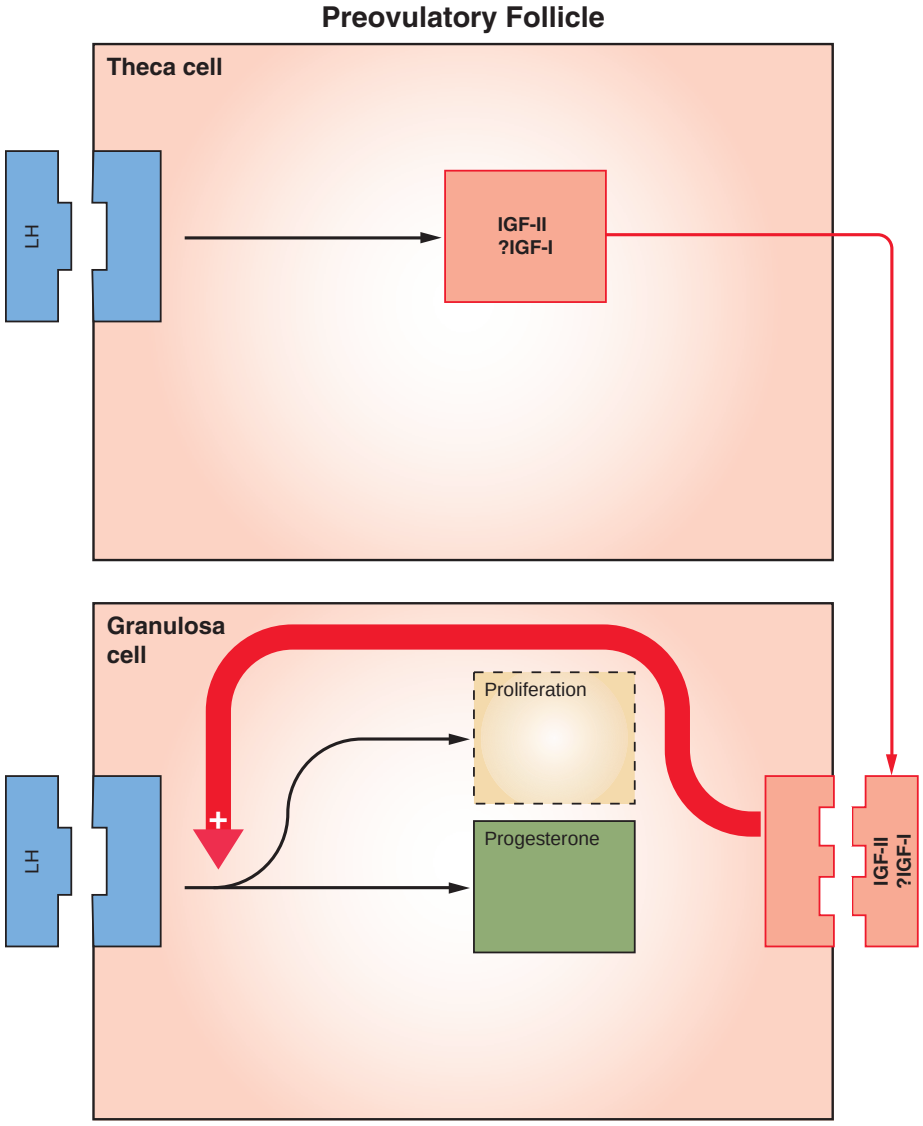
Insulin-like Growth Factor Binding Proteins. There are six known nonglycosylated peptides that function as IGF binding proteins, IGFBP-1 to IGFBP-6.[149] These binding proteins serve to carry the IGFs in serum, prolong half-lives, and regulate tissue effects of the IGFs. The regulating action appears to be due to binding and sequestering of the IGFs, preventing their access to the cell membrane surface receptors, and, thus, not permitting the synergistic actions that result when gonadotropins and growth factors are combined. The IGFBPs may also exert direct actions on cellular functions, independently of growth factor functions. IGFBP-1 is the principal BP in amniotic fluid; IGFBP-3 is the main BP in serum and its synthesis, primarily in the liver, is dependent on growth hormone. Circulating levels of IGFBP-3 reflect the total IGF concentration (IGF-I plus IGF-II) and carry at least 90% of the circulating IGFs. These BPs do not bind insulin. The BPs change with age (decreasing levels of IGFBP-3) and during pregnancy (decreasing IGFBP-3 due to a circulating protease unique to pregnancy).

The IGF Receptors. The Type I receptor preferentially binds IGF-1 and can be called the IGF-I receptor. The Type II receptor in a similar fashion can be called the IGF-II receptor. IGF-I also binds to the insulin receptor but with low affinity. Insulin binds to the IGF-I receptor with moderate affinity. The IGF-I receptor and the insulin receptor are similar in



structure: tetramers composed of two α-subunits and two β-subunits linked by disulfide bonds. The intracellular component of the β-subunit is a tyrosine kinase that is activated by autophosphorylation. The IGF-II receptor does not bind insulin. It is a single chain glycoprotein, with 90% of its structure extending extracellularly. This receptor functions as a receptor coupled to a G protein. The physiologic effects of IGF-I are mediated by its own receptor, but IGF-II can exert its actions via both receptors. Indeed, the IGF-I receptor binds IGF-I and IGF-II with equal affinity. In human cells, the IGF-I receptor and IGF-II receptor are present in theca and granulosa cells and in luteinized granulosa cells. Ovarian stromal tissue contains IGF-I receptors.

The Ovarian Actions of IGFs. IGF-I has been demonstrated to stimulate the following events in ovarian theca and granulosa cells: DNA synthesis, steroidogenesis, aromatase activity, LH receptor synthesis, and inhibin secretion. IGF-II stimulates granulosa mitosis. In human ovarian cells, IGF-I, in synergy with FSH, stimulates protein synthesis and steroidogenesis. After LH receptors appear, IGF-I enhances LH-induced progesterone synthesis and stimulates proliferation of granulosa-luteal cells. IGF-I, in synergy with FSH, is very active in stimulating aromatase activity in preovulatory follicles. Thus, IGF-I can be involved in both estradiol and progesterone synthesis.



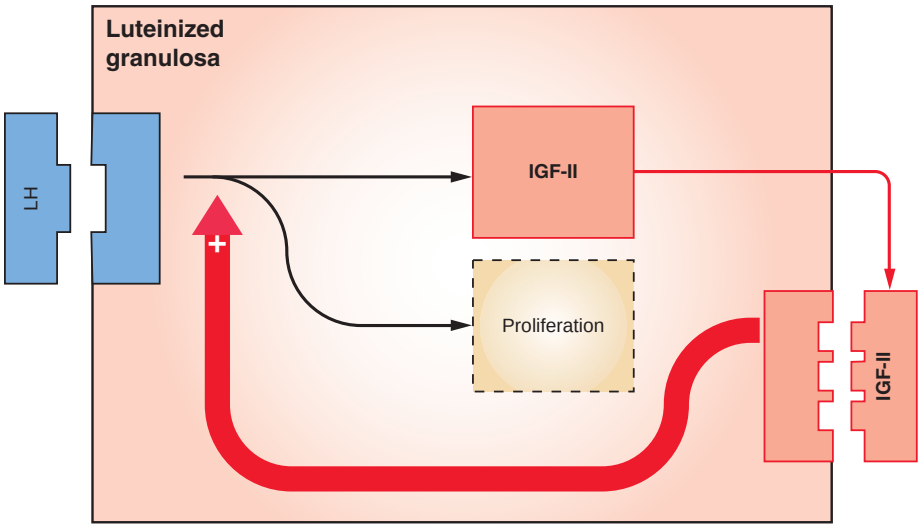
In animal experiments, the synthesis of IGF-I by granulosa cells is dependent upon FSH but enhanced by estradiol. Growth hormone also acts synergistically with FSH and estradiol to increase IGF synthesis. The story becomes confused when various growth factors and regulators are studied, because of their various stimulating and inhibiting effects. In the rat, the granulosa cell is the major site for IGF-I gene expression, which is active only prior to ovulation. It is not detected in atretic follicles or in corpora lutea. Again in the rat, IGF-II gene expression appears to be limited to the theca and interstitial cells. However, the site of IGF expression is different in primates.

In studies with human ovarian tissue, IGF-II is highly expressed in both theca cells and granulosa cells; however, the level is highest in the granulosa and increases with growth of the follicle.[150, 151] IGF-II is also synthesized by luteinized granulosa and appears to function locally in an autocrine fashion.[152] These findings indicate that IGF-II is the primary IGF in the human ovary. Nevertheless, IGF-I is still a significant product of human theca cells.[153]

Human theca cells express mRNA transcripts that encode receptors for both IGF-I and insulin.[154] Because insulin and IGF-II can both activate the receptor for IGF-I, this pathway provides a method for the exertion of a paracrine influence on granulosa cells and autocrine activity in the theca (augmenting LH stimulation of androgen production). In vitro studies confirm that IGF-II is capable of stimulating steroidogenesis and proliferation in human theca and granulosa cells.[155–157] These actions are augmented by growth hormone, which increases IGF production and, thus, indirectly enhances gonadotropin stimulation of ovarian follicles.[158]

This primate scenario is supported by finding higher levels of IGF-II, but not IGF-I, in the follicular fluid of developing follicles, with the highest levels present in dominant follicles.[159] The IGF levels in follicular fluid correlate with estradiol levels and undergo a further short increase after the LH surge. There are no menstrual cycle changes in the circulating levels of IGF-I, IGF-II, IGFBP-1, or IGFBP-3; high levels in the dominant follicle are not associated with an increase in circulating levels.[160]

In studies with human tissue, IGFBP-1 inhibits IGF-1 mediated steroidogenesis and proliferation of luteinized granulosa cells. The synthesis of IGFBPs by human granulosa is inhibited by FSH, IGF-I, and IGF-II.[161, 162] These findings fit with the overall idea that the BPs counteract the synergistic results of gonadotropins and growth factors. In general, IGFBP-1 expression is found in granulosa cells of growing follicles; IGFBP-3 in theca cells and the granulosa of the dominant follicle; IGFBP-2, -4, and -5 in theca and granulosa of antral and atretic follicles; and IGFBP-6 has not been found in the ovary.[150] The predominant binding protein in preovulatory follicles is IGFBP-2 in the granulosa and IGFBP-3



in the theca, that increase progressively in the follicle that emerges as the dominant follicle, and then decrease in the late follicular phase.[151, 163, 164] This suggests that –1, –2, and –3 play a role in growing follicles, –2, –4, and –5 in atretic and failing follicles. IGFBP expression in polycystic ovaries is similar to that seen in atretic follicles. The decrease in IGFBP-3 that occurs in dominant follicles should allow an increase in IGF levels and activity. The increase in IGFBP-2 in failing follicles probably correlates with sequestering of IGF, depriving the follicle of an important force in gonadotropin augmentation.

Circulating levels of IGFBP-1 decrease in response to insulin, and thus circulating levels are decreased in women with anovulation and polycystic ovaries who have elevated levels of insulin.[165] These patients also have increased circulating levels of IGF-I, probably a consequence of LH-stimulated synthesis and secretion in theca cells. The level of IGFBP-1 in follicular fluid from polycystic ovaries is decreased; thus this BP is not playing a role inhibiting the action of IGF-I in polycystic ovaries. The levels of IGFBPs -2 and -4 in the follicular fluid from follicles in anovulatory patients are increased (as in atretic follicles).[150, 166] Even though these changes may play a role in anovulatory pathophysiology, they are consistent with failure in development and thus may not be etiologic factors.

IGF activity may also be modulated by the proteases that regulate the activity of the IGF binding proteins.[167] Estrogen-dominant follicular fluid contains very low levels of IGFBP-4, in contrast to the high levels present in androgen-dominant follicular fluid. The low level of IGFBP-4 in estrogen-dominant follicular fluid is associated with the presence of an IGFBP-4 specific protease. This protease would decrease IGFBP activity and enhance IGF activity, another mechanism for ensuring the success of the dominant follicle.

The insulin-like growth factor story is at once complex, fascinating, and compelling. However, its contribution may be facilitatory and not essential. Laron-type dwarfism is characterized by a deficiency in IGF-I due to an abnormality in the growth hormone receptor. Despite low levels of IGF-I and high levels of IGFBP, a woman with Laron-type dwarfism responded to exogenous gonadotropin stimulation with the production of multiple, mature follicles with good estrogen production and fertilizable oocytes.[168] Another explanation for this observation is that IGF-II, rather than IGF-I, is the important factor in the human dominant follicle. This possibility is supported by evidence indicating that IGF-II is the most abundant IGF in human ovarian follicles.[150, 151] Another possibility is that the Laron-type dwarf is deficient only in growth hormone-dependent IGF-I, and ovarian IGFs are not totally dependent on growth hormone.

Summary of Insulin-Like Growth Factor Action in the Ovary

1. IGF-II stimulates granulosa cell proliferation, aromatase activity, and progesterone synthesis.
2. IGF-II is produced in theca cells, granulosa cells, and luteinized granulosa cells In the pig and rat, the primary IGF is IGF-I.
3. Gonadotropins stimulate IGF production, and in animal experiments, this stimulation is enhanced by estradiol and growth hormone.
4. IGF-I receptors are present in theca and granulosa cells, and only IGF-II receptors are present in luteinized granulosa. IGF-II activates both IGF-I and IGF-II receptors.
5. The most abundant IGF in human follicles is IGF-II.
6. FSH inhibits binding protein synthesis, and thus maximizes growth factor availability.

Epidermal Growth Factor

Epidermal growth factor is a mitogen for a variety of cells, and its action is potentiated by other growth factors. Granulosa cells, in particular, respond to this growth factor in a variety of ways related to gonadotropin stimulation, including proliferation. Epidermal growth factor suppresses the up-regulation of FSH on its own receptor.[39] Amphiregulin and Epiregulin, ligands that are similar to epidermal growth factor, are produced in luteinized granulosa cells in response to LH and induce progesterone synthesis in the corpus luteum.[169]

Transforming Growth Factor

TGF-α is a structural analog of epidermal growth factor and can bind to the epidermal growth factor receptor. TGF-β utilizes a receptor distinct from the epidermal growth factor receptor. These factors are thought to be autocrine growth regulators. Inhibin and activin are derived from the same gene family. TGF-β, secreted by theca cells, enhances FSH induction of LH receptors on granulosa cells, an action which is opposite that of epidermal growth factor.[170] While this action can be viewed as a positive impact on granulosa cells, in the theca, TGF-β has a negative action, inhibiting androgen production.[171] Growth differentiation factor 9 (GDF-9) is a member of the TGF-β family that originates in the oocyte and is essential for normal growth and development of the ovarian follicle.[16]

Fibroblast Growth Factor

This factor is a mitogen for a variety of cells and is present in all steroid-producing tissues. Important roles in the ovarian follicle include stimulation of mitosis in granulosa cells, stimulation of angiogenesis, stimulation of plasminogen activator, inhibition of FSH up-regulation of its own receptor, and inhibition of FSH-induced LH receptor expression and estrogen production.[39, 172] These actions are opposite of those of transforming growth factor-β.

Platelet-Derived Growth Factor

This growth factor modifies cyclic AMP pathways responding to FSH, especially those involved in granulosa cell differentiation. Both platelet-derived growth factor and epidermal growth factor may also modify prostaglandin production within the follicle.

Angiogenic Growth Factors

Vascularization of the follicle is influenced by peptides in the follicular fluid, especially vascular endothelial growth factor (VEGF), a cytokine produced in granulosa cells in response to LH.[173, 174] Luteal cells respond to human chorionic gonadotropin (hCG) with greater VEGF output, a probable mechanism contributing to the increased vascular permeability associated with ovarian hyperstimulation that can occur with exogenous gonadotropin administration (Chapter 31).[175] Angiopoietins bind to an endothelial receptor (Tie-2) and provide an inhibitory influence on angiogenesis. Angiopoietin-1 is the active agent, opposed by angiopoietin-2, which competes for the Tie-2 receptor on endothelial cells. Differential expression of these angiogenic factors is involved in the coordinated growth and regression of follicles and the corpus luteum.[176–178] The injection of VEGF and angiopoietin antagonists directly into dominant follicles in the monkey interferes with both the physical process of ovulation and the subsequent function of the corpus luteum.[179]

The Interleukin-1 System

Leukocytes are a prominent component of the ovarian follicle and a major source of interleukins. Interleukin-1 is a member of the cytokine family of immunomediators. The human ovary contains the complete interleukin-1 system (ligand and receptor). In the rat, interleukin-1 stimulates ovarian prostaglandin synthesis and perhaps plays a role in ovulation.[180]

Tumor Necrosis Factor-a (TNF-a)

TNF-α is also a product of leukocytes (macrophages). It very likely is a key player in the process of apoptosis, a feature of follicular atresia as well as luteolysis of the corpus luteum.

Other Peptides

The follicular fluid is a veritable protein soup! It is composed of exudates from plasma and secretions from follicular cells. A variety of hormones can be found in the follicular fluid, as well as enzymes and peptides, which play important roles in follicular growth and development, ovulation, and modulation of hormonal responses.

Follicular fluid contains ***prorenin,*** the inactive precursor of renin, in a concentration that is about 12 times higher than plasma levels.[181] LH stimulates its synthesis in the follicle, and there is a midcycle peak in prorenin plasma levels. The circulating levels of prorenin also increase (10-fold) during the early stages of pregnancy, the result of ovarian stimulation by the rise in human chorionic gonadotropin (hCG). These increases in prorenin from the ovary are not responsible for any significant changes in the plasma levels of the active form, renin. Possible roles for this ovarian prorenin-renin-angiotensin system include stimulation of steroidogenesis to provide androgen substrate for estrogen production, regulation of calcium and prostaglandin metabolism, and stimulation of angiogenesis. This system may affect vascular and tissue functions both within and outside the ovary.

Members of the proopiomelanocortin family are found in human follicular fluid.[182] Follicular levels of ***ACTH*** and ***β-lipotropin*** remain constant throughout the cycle, but ***β-endorphin*** levels peak just before ovulation. In addition, enkephalin is present in relatively unchanging concentrations. The ***corticotropin-releasing hormone (CRH)*** system is present in theca cells, but not in granulosa cells, complete with CRH, the CRH receptor, and CRH-binding protein.[183] CRH inhibits LH-stimulated androgen production in theca cells, apparently by suppressing P450c17 gene expression.[184]

Antimüllerian hormone (AMH), a member of the transforming growth factor-β family like inhibin and activin, is produced by granulosa cells and may play a role in oocyte maturation (it inhibits oocyte meiosis) and follicular development.[185, 186] Antimüllerian hormone directly inhibits proliferation of granulosa and luteal cells, as well as epidermal growth factor-stimulated proliferation. Its paracrine function may be to suppress growth of all but the dominant follicle in each cycle.[70] The circulating level of AMH is highest in the late follicular phase, peaking simultaneously with inhibin-A just before ovulation.[187] Experimental evidence suggests that the source of the AMH is the entire cohort of follicles except for the dominant follicle, and thus, the circulating level correlates with follicle number and potential for fertility.[188, 189] With aging and a decrease in the number of follicles, AMH levels decline. AMH cam be measured on any day in an individual's menstrual cycle, even in women on steroid contraception, because AMH secretion is affected by gonadotropins or sex steroids insufficiently to produce clinically meaningful changes.[75]

Follicular fluid prevents resumption of meiosis until the preovulatory LH surge either overcomes or removes this inhibition. This action is attributed to ***oocyte maturation inhibitor (OMI). Pregnancy-associated plasma protein A,*** found in the placenta, is also present in follicular fluid. It may inhibit proteolytic activity within the follicle before ovulation. ***Endothelin-1*** is a peptide, produced in vascular endothelial cells, which may be the substance previously known as luteinization inhibitor; endothelin gene expression is induced by the hypoxia associated with the avascular granulosa, and it inhibits LH-induced progesterone production.[190] It is uncertain whether ***GnRH-like peptides*** have a follicular role or represent sequestered GnRH. ***Oxytocin*** is found in preovulatory follicles and the corpus luteum. Growth hormone-binding protein is present in follicular fluid and similar in characteristics to the same binding protein in serum.

Summary of Events in the Antral Follicle

1. Follicular phase estrogen production is explained by the two-cell, two-gonadotropin mechanism.
2. Selection of the dominant follicle is established during days 5–7, and consequently, peripheral levels of estradiol begin to rise significantly by cycle day 7.
3. Estradiol levels, derived from the dominant follicle, increase steadily and, through negative feedback effects, exert a progressively greater suppressive influence on FSH release.
4. While directing a decline in FSH levels, the midfollicular rise in estradiol exerts a positive feedback influence on LH secretion.
5. The positive action of estrogen also includes modification of the gonadotropin molecule, increasing the quality (the bioactivity) as well as the quantity of FSH and LH at midcycle.
6. LH levels rise steadily during the late follicular phase, stimulating androgen production in the theca.
7. A unique responsiveness to FSH allows the dominant follicle to utilize the androgen as substrate and further accelerate estrogen production.
8. FSH induces the appearance of LH receptors on granulosa cells.
9. Follicular response to the gonadotropins is modulated by a variety of growth factors and autocrine-paracrine peptides.
10. Inhibin-B, secreted by the granulosa cells in response to FSH, directly suppresses pituitary FSH secretion.
11. Activin, originating in both pituitary and granulosa, augments FSH secretion and action.

Follicular Growth and Development in the Primate Ovary

Evidence strongly indicates that autocrine-paracrine peptides, not estrogen, play the major role in regulating ovarian follicle growth and development in the primate. In monkey experiments, no reduction in total number or size of follicles resulted when estradiol production was effectively suppressed by treatment with an inhibitor of the aromatase enzyme system or with an inhibitor of 3β-hydroxysteroid dehydrogenase.[191–193] *Oocyte development was not altered, although the subsequent fertilization rate was reduced by this treatment. Another argument against a major role for estrogen in follicular growth and development is the successful stimulation with gonadotropins of normal follicular growth and development in*

women with 17α-hydroxylase deficiency (an inherited disorder that prevents the production of androgens and estrogens).[194, 195]

A reduced role for estrogen is further supported by the response of women with a deficiency in gonadotropins to treatment with recombinant (pure) FSH.[61–63] *Some aromatization occurred, perhaps using androgens originating in the adrenal glands, producing early follicular phase estradiol levels, but the usual robust steroidogenesis was impossible without the presence of LH to provide theca production of androgen substrate. Nevertheless, oocytes were retrieved, and with in vitro fertilization, pregnancy was achieved. This same response was observed in experiments that used a GnRH antagonist to produce LH-deficient monkeys and then the administration of recombinant, pure human FSH.*[64, 65]

These results indicate that only FSH is required for early folliculogenesis, and that in the primate, autocrine-paracrine peptides have replaced estrogen in the important role of modulating gonadotropin response. Consider the following actions that have been documented in primate ovaries:

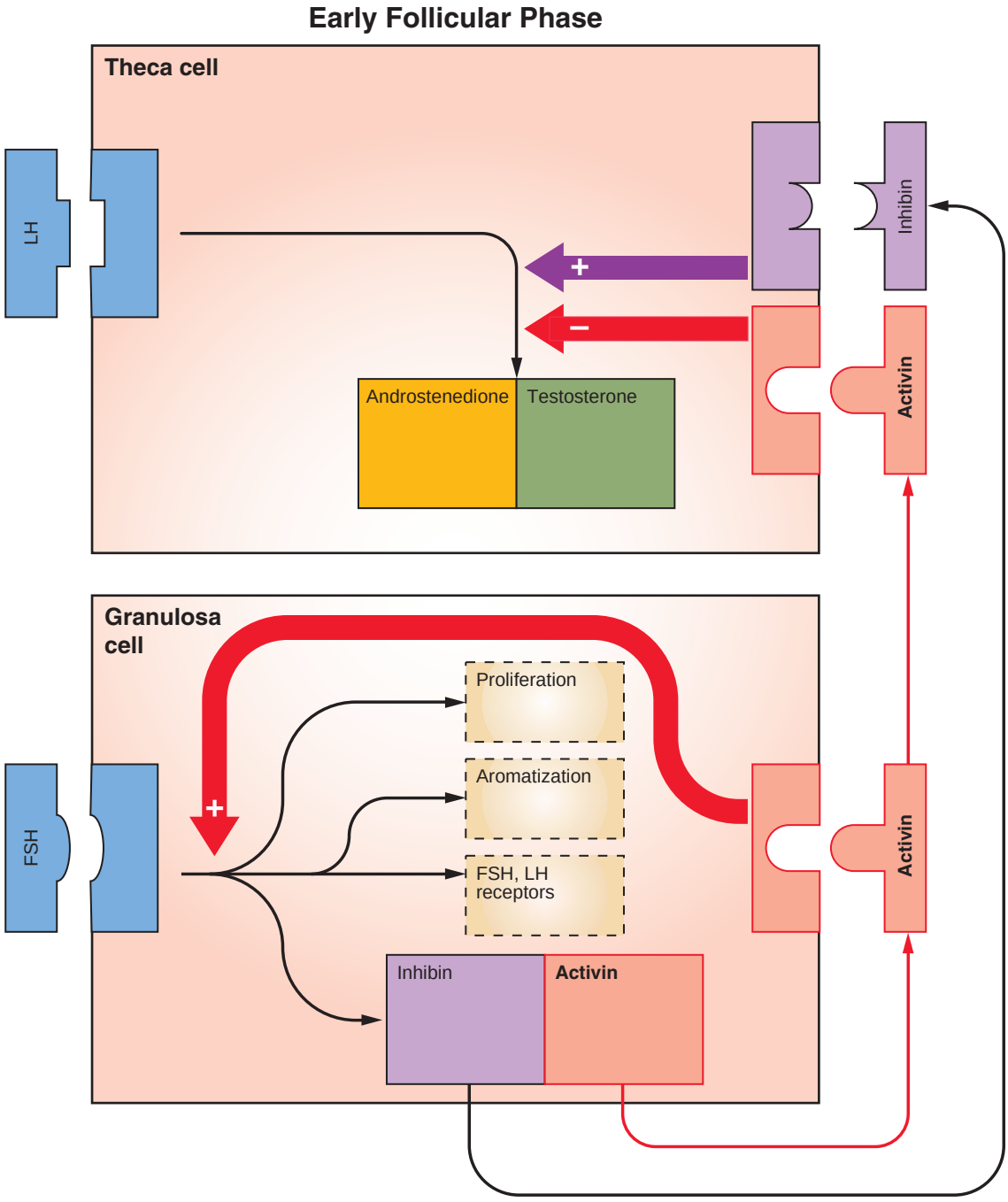
1. *Inhibin and activin regulate androgen synthesis in human theca cells. Inhibin enhances and activin suppresses the stimulatory action of LH and/or IGF-I, and inhibin can overcome the inhibitory action of activin on theca cells.*[137–139]

2. *In immature granulosa cells, activin augments all FSH activities, especially aromatase activity (estrogen production).*[113, 196]

3. *In luteinizing granulosa cells, activin has direct mitogenic activity and suppresses steroidogenesis in response to LH, while inhibin has no effect on LH-dependent aromatase in mature granulosa cells.*[196, 197]

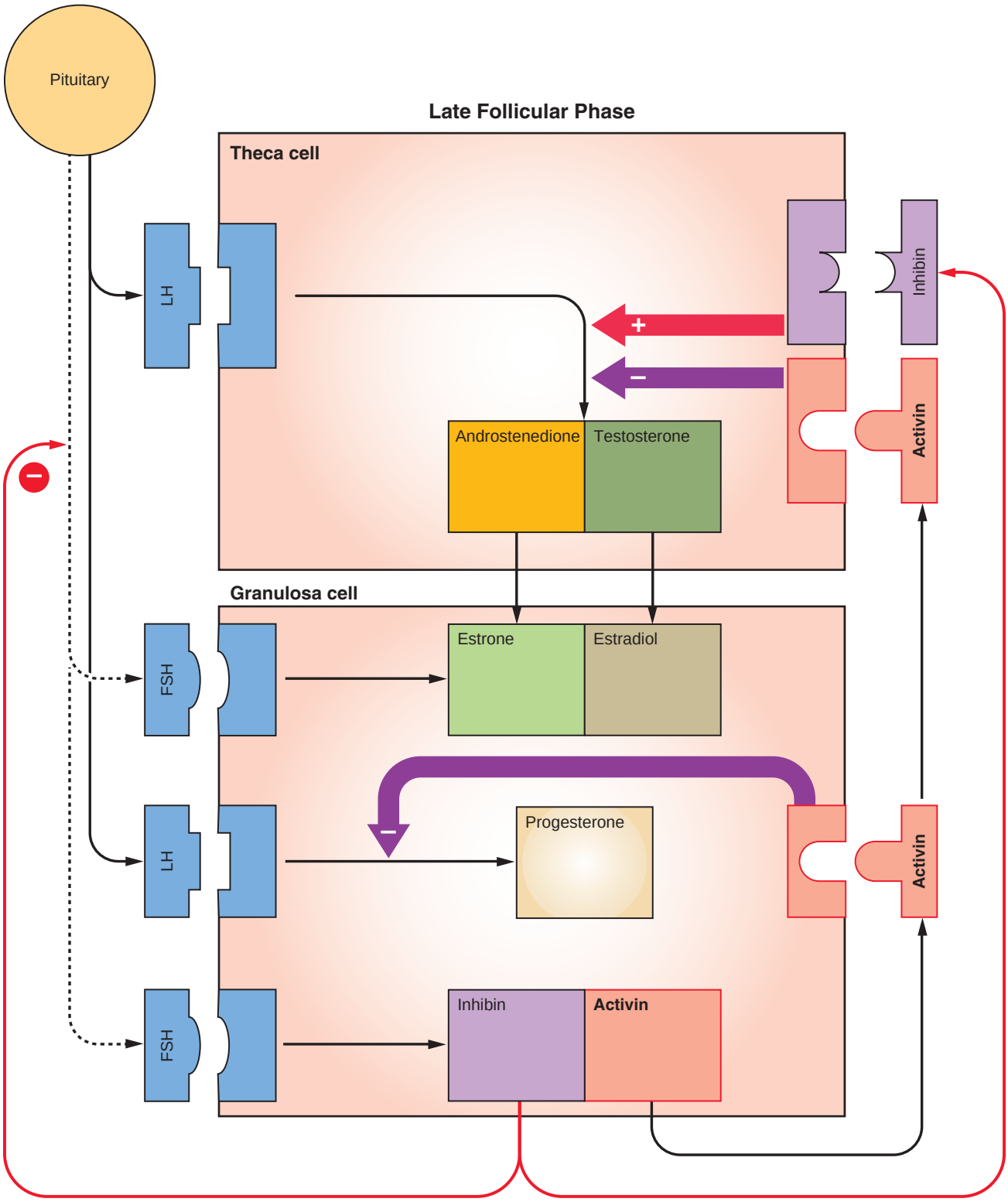
4. *In the follicular phase, granulosa production of inhibin is under the control of FSH, but during the late follicular phase a change occurs, culminating in LH control of luteal synthesis of inhibin.*[198, 199]

5. *As the follicle grows, activin production decreases and inhibin production increases.*[200, 201] *In addition, follistatin levels increase in follicular fluid with increasing growth of the follicle, a mechanism for decreasing activin activity.*[202] *In the early follicular phase, FSH and estradiol enhance inhibin-B secretion, probably indirectly by increasing granulosa cell numbers, whereas late in the follicular phase, when LH levels increase, inhibin-A secretion is favored.*[203]

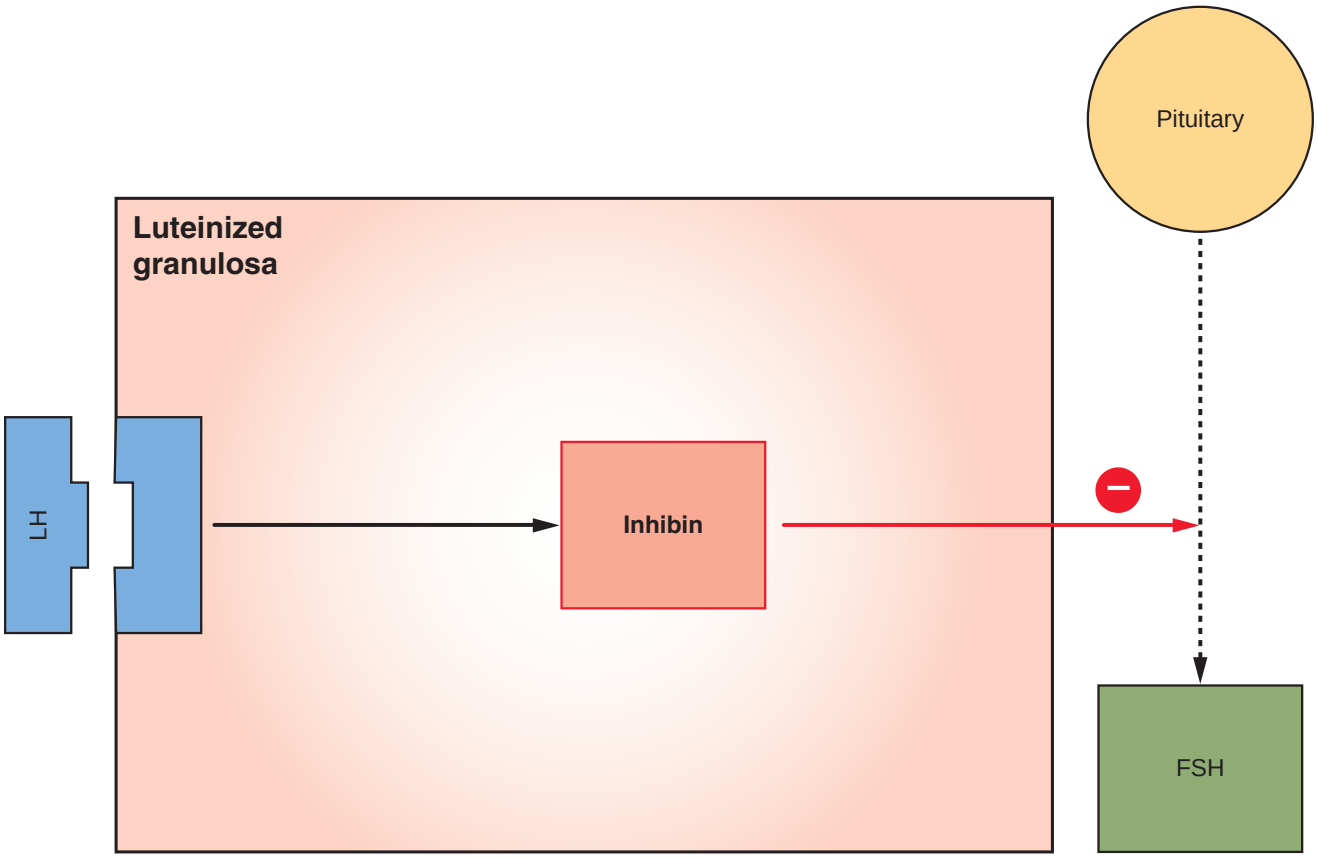


These actions come together as follows. In the early follicular phase, activin produced by granulosa in immature follicles enhances the action of FSH on aromatase activity and FSH and LH receptor formation, while simultaneously suppressing theca androgen synthesis. In the late follicular phase, increased production of inhibin (specifically inhibin-B) by the granulosa (and decreased activin) promotes androgen synthesis in the theca in response to LH and IGF-II to provide substrate for even greater estrogen production in the granulosa. In the mature granulosa of the dominant preovulatory follicle, activin serves to prevent premature luteinization and progesterone production.

The successful follicle is the one that acquires the highest level of aromatase activity and LH receptors in response to FSH. The successful follicle is characterized by the



highest estrogen (for central feedback action) and the greatest inhibin production (for both local and central actions). This accomplishment occurs in synchrony with the appropriate activin expression. The highest level of gene activity encoding activin is found in immature antral follicles and the lowest level in preovulatory follicles. Thus



the activin proteins (which enhance FSH activity) are produced in greatest amounts early in follicular development to enhance follicle receptivity to FSH. As with circulating levels of inhibin, inhibin-B is the predominant inhibin in the follicular fluid of preantral follicles and inhibin-A increases when follicles become large and mature.[204–206] *Inhibin synthesis and secretion during the follicular phase are regulated by FSH and growth factors.*[207]

The right concentration of androgens in granulosa cells promotes aromatase activity and inhibin production and, in turn, inhibin promotes LH stimulation of theca androgen synthesis. With development of the follicle, inhibin expression (specifically inhibin-A) comes under control of LH. A key to successful ovulation and luteal function is conversion of the inhibin production to LH responsiveness to maintain FSH suppression centrally and enhancement of LH action locally.

Responses of ovarian follicles to exogenous FSH and LH stimulation for in vitro fertilization indicate that the final maturation and function of the dominant follicle prior to ovulation are significantly influenced by LH.[84] *Final maturation of the dominant follicle and the health of the oocyte are optimized by the required presence of a threshold level of LH.*[66, 85, 86, 208]

A lesser role is assigned to the insulin-like growth factors in view of the successful production of multiple, estrogen-producing follicles which yielded fertilizable oocytes in a woman with IGF-I deficiency treated with gonadotropins.[168] *The growth factors assume an important, but perhaps not essential, role as facilitating agents. However, the successful pregnancy in a woman with IGF-I deficiency may indicate the greater importance of IGF-II.*

Summary of Events in the Primate Ovarian Follicle

1. FSH has multiple activities in the granulosa cell: stimulating aromatization of androgens to estrogens, increasing granulosa cell content of FSH and LH receptors, stimulating proliferation of granulosa cells, and producing autocrine-paracrine factors, especially activin and inhibin.

2. In the granulosa of the early follicular phase, activin augments FSH activities: FSH receptor expression, aromatization, inhibin/activin production, and LH receptor expression. In the theca, activin suppresses androgen production, allowing the emergence of an estrogen microenvironment.

3. Later in the follicular phase, inhibin enhances LH stimulation of androgen synthesis in the theca to provide substrate for aromatization to estrogen in the granulosa, making available the large amount of estrogen necessary for local follicular actions and to trigger the LH surge.

4. Inhibin-B is secreted by the granulosa cells into the circulation, where it acts in a classic endocrine fashion to suppress FSH secretion by the pituitary gland, an important method to ensure the dominance of a single follicle.

5. With the appearance of LH receptors, inhibin production is maintained as it comes under control of LH.

6. Late in the follicular phase, final follicular maturation to yield the most favorable level of steroidogenesis and an oocyte with the best viability requires the presence of a threshold level of LH.

7. All functions are modulated by an army of growth factors, and IGF-II may be especially important.

The Preovulatory Follicle

Granulosa cells in the preovulatory follicle enlarge and acquire lipid inclusions while the theca becomes vacuolated and richly vascular, giving the preovulatory follicle a hyperemic appearance. The oocyte proceeds in meiosis, approaching completion of its reduction division.

Approaching maturity, the preovulatory follicle produces increasing amounts of estrogen. During the late follicular phase, estrogens rise slowly at first, then rapidly, reaching a peak approximately 24–36 hours prior to ovulation.[209] The onset of the LH surge occurs when the peak levels of estradiol are achieved.[210] In providing the ovulatory stimulus to the selected follicle, the LH surge seals the fate of the remaining follicles, with their lower estrogen and FSH content, by further increasing androgen superiority.

Acting through its own receptors, LH promotes luteinization of the granulosa in the dominant follicle, resulting in the production of progesterone. The LH receptor, once expressed, inhibits further cell growth and focuses the cell's energy on steroidogenesis (actions enhanced by IGF).[211] An increase in progesterone can be detected in the venous effluent of the ovary bearing the preovulatory follicle as early as day 10 of the cycle.[76] This small but significant increase in the production of progesterone in the preovulatory period has immense physiologic importance. Prior to the emergence of this follicular progesterone, the circulating level of progesterone was derived from the adrenal gland.[212]

Progesterone receptors begin to appear in the granulosa cells of the dominant follicle in the periovulatory period.[87] The traditional view has been that progesterone receptors are expressed in response to estrogen through an estrogen-receptor-mediated mechanism; however, this is not the case. Experimental data in the monkey provide excellent evidence that LH stimulates progesterone receptor expression in granulosa cells.[213] In vitro data with human cells suggest that the preovulatory progesterone and progesterone receptor expression directly inhibit granulosa cell mitosis, probably explaining the limitation of granulosa cell proliferation as these cells gain LH receptors.[214]

Progesterone affects the positive feedback response to estrogen in both a time and dose dependent manner. When introduced after adequate estrogen priming, progesterone facilitates the positive feedback response, in a direct action on the pituitary, and in the presence of subthreshold levels of estradiol can induce a characteristic LH surge.[215, 216] Hence, the surprising onset of ovulation occasionally observed in an anovulatory, amenorrheic woman administered a progestin challenge. When administered before the estrogen stimulus, or in high doses (achieving a blood level greater than 2 ng/mL), progesterone blocks the midcycle LH surge.

Appropriately low levels of progesterone derived from the maturing follicle contribute to the precise synchronization of the midcycle surge. In addition to its facilitator action on LH, progesterone at midcycle is significantly responsible for the FSH surge.[216] This action of progesterone can be viewed as a further step in ensuring completion of FSH action on the follicle, especially making sure that a full complement of LH receptors is in place in the granulosa layer. In certain experimental situations, incremental estradiol alone can elicit simultaneous surges of LH and FSH, suggesting that progesterone certainly enhances the effect of estradiol but may not be obligatory.[217] Nevertheless, blockade of midcycle progesterone synthesis or activity in the monkey impaired the ovulatory process and luteinization.[218] These actions of estrogen and progesterone require the presence and continuous action of GnRH.

The preovulatory period is associated with a rise in plasma levels of 17α-hydroxyprogesterone. This steroid does not appear to have a role in cycle regulation, and its appearance in the blood simply represents the secretion of an intermediate product. As such, however, it signals the LH stimulation of P450scc and P450c17, important enzyme activity for the production of theca androgens, the substrate for granulosa estrogen. After ovulation, some theca cells become luteinized as part of the corpus luteum and lose the ability to express P450c17. Other luteinized theca cells retain P450c17 activity and are believed to continue to produce androgens for aromatization to estrogens.

When the lesser follicles fail to achieve full maturity and undergo atresia, the theca cells return to their origin as a component of stromal tissue, retaining, however, an ability to respond to LH with P450 activity and steroid production. Because the products of theca tissue are androgens, the increase in stromal tissue in the late follicular phase is associated with a rise in androgen levels in the peripheral plasma at midcycle, a 15% increase in androstenedione and a 20% increase in testosterone.[219] This response is enhanced by the rise in inhibin, known to augment LH stimulation of androgen production in theca cells.

Androgen production at this stage in the cycle may serve two purposes: (1) a local role within the ovary to enhance the process of atresia, and (2) a systemic effect to stimulate libido.

Intraovarian androgens accelerate granulosa cell death and follicular atresia. The specific mechanism for this action is unclear, although it is attractive to suspect an interference with estrogen and the autocrine-paracrine factors in enhancing FSH activity. Therefore, androgens may play a regulatory role in ensuring that only a dominant follicle reaches the point of ovulation.

It is well known that libido can be stimulated by androgens. If the midcycle rise in androgens affects libido, then an increase in sexual activity should coincide with this rise. Early studies failed to demonstrate a consistent pattern in coital frequency in women because of the effect of male partner initiation. If only sexual behavior initiated by women is studied, a peak in female-initiated sexual activity is seen during the ovulatory phase of the cycle.[220] The coital frequency of married couples has also been noted to increase at the time of ovulation.[221] Therefore, the midcycle rise in androgens may serve to increase sexual activity at the time most likely to achieve pregnancy.

Summary of Events in the Preovulatory Follicle

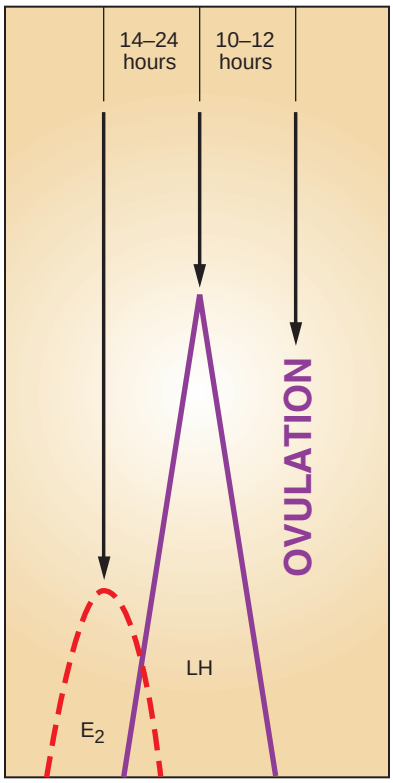
1. Estrogen production becomes sufficient to achieve and maintain peripheral threshold concentrations of estradiol that are required in order to induce the LH surge.
2. Acting through its receptors, LH initiates luteinization and progesterone production in the granulosa layer.
3. The preovulatory rise in progesterone facilitates the positive feedback action of estrogen and may be required to induce the midcycle FSH peak.
4. A midcycle increase in local and peripheral androgens occurs, derived from the theca tissue of lesser, unsuccessful follicles.

Ovulation

The preovulatory follicle, through the elaboration of estradiol, provides its own ovulatory stimulus. Considerable variation in timing exists from cycle to cycle, even in the same woman. A reasonable and accurate estimate places ovulation approximately 10–12 hours after the LH peak and 24–36 hours after peak estradiol levels are attained.[209, 222] The onset of the LH surge appears to be the most reliable indicator of impending ovulation, occurring 34–36 hours prior to follicle rupture.[223] A threshold of LH concentration must be maintained for 14–27 hours in order for full maturation of the oocyte to occur.[224] Usually the LH surge lasts 48–50 hours.[223]

Because of the careful timing involved in in vitro fertilization programs, we have available some interesting data.[225] The LH surge tends to occur at approximately 3 A.M., beginning between midnight and 8:00 A.M. in over two-thirds of women.[98] Ovulation occurs primarily in the morning during Spring, and primarily in the evening during Autumn and Winter. From July to February in the Northern Hemisphere, about 90% of women ovulate between 4 and 7 P.M.; during Spring, 50% of women ovulate between midnight and 11 A.M.

Most of the studies have concluded that ovulation occurs more frequently (about 55% of the time) in the right ovary compared with the left, and oocytes from the right ovary have a higher potential for pregnancy.[226] The side of ovulation does not affect cycle characteristics, but cycles with short follicular phases tend to be followed by contralateral ovulation, and ovulation occurs randomly following cycles with a long follicular phase.[227, 228] Ovulation alternating between the two ovaries predominates in younger women, but after age 30 ovulations occur more frequently from the same ovary; however, throughout the reproductive years more ovulations occur from the right ovary.[229] Contralateral ovulation favors pregnancy more than ipsilateral ovulation, and ipsilateral ovulation increases with increasing age and decreasing fertility.[229]



The gonadotropin surge stimulates a large collection of events that ultimately leads to ovulation, the physical release of the oocyte and its cumulus mass of granulosa cells.[230] This is not an explosive event; therefore, a complex series of changes must occur which cause the final maturation of the oocyte and the decomposition of the collagenous layer of the follicular wall.[231]

The LH surge initiates the resumption of meiosis in the oocyte (meiosis is not completed until after the sperm has entered and the second polar body is released), luteinization of granulosa cells and progesterone production, expansion of the cumulus, and the synthesis of prostaglandins and other eicosanoids essential for follicle rupture. Premature oocyte maturation and luteinization are prevented by local factors.

An LH-induced increase in cyclic AMP occurs within the follicle just prior to ovulation. Cyclic AMP is transferred from the granulosa cells to the oocyte via the gap junction network and thus a reduction in cyclic AMP occurs when LH causes a breakdown of the gap junctions. This results in a decrease in the local inhibitory action of oocyte maturation inhibitor (OMI) and luteinization inhibitor (LI). LI may be endothelin-1, a product of vascular endothelial cells.[190] OMI originates from the granulosa cells, and its activity depends upon an intact cumulus oophorous. Locally produced activin suppresses progesterone production by luteal cells, providing yet another means of preventing premature luteinization.[232, 233] The propagation of the LH-induced changes throughout the follicle depends on growth factors and their receptors, especially members of the epidermal growth factor-like family, specifically LH-induced factors named amphiregulin, epiregulin, and betacellulin.[234] Disruption of this pathway interferes with oocyte resumption of meiosis and ovulation.

There is abundant evidence that the oocyte exerts control over granulosa functions, affecting both metabolism and proliferation through the secretion of proteins in the transforming growth factor-β family.[49, 235–238] These proteins include inhibin, activin, AMH, bone morphogenetic proteins (BMPs), and GDF9, which must be secreted in their active forms

after processing of precursor proteins by proteases. The production of the active proteins is regulated by an interaction of the signaling proteins from the oocyte and the granulosa cells, determined by changing responsiveness to FSH as the components of the ovarian follicle develop and differentiate.[239] The differentiation and maintenance of the cumulus cells from the preantral granulosa cells are under the direction of the oocyte.[240, 241]

The cumulus oophorus differs from other granulosa cells, lacking in LH receptors and progesterone production; FSH-induced LH receptor expression is suppressed in the contiguous granulosa cells by the oocyte. The oocyte enables cumulus cells to respond to the gonadotropin-induced physical and biochemical changes just before ovulation. The local factors that prevent premature oocyte maturation and luteinization are probably under control of the oocyte. One mediator of this control system is nitric oxide, which maintains the gap junction system of communication.[242] Nitric oxide resists LH-induced resumption of oocyte meiosis and breakdown of the gap junction network until the massive LH surge overcomes this resistance and communication between the oocyte and the follicular cells is interrupted.

With the LH surge, levels of progesterone in the follicle continue to rise up to the time of ovulation. The progressive rise in progesterone may act to terminate the LH surge as a negative feedback effect is exerted at higher concentrations. In addition to its central effects, progesterone increases the distensibility of the follicle wall. A change in the elastic properties of the follicular wall is necessary to explain the rapid increase in follicular fluid volume, which occurs just prior to ovulation, unaccompanied by any significant change in intrafollicular pressure. The escape of the ovum is associated with degenerative changes of the collagen in the follicular wall so that just prior to ovulation the follicular wall becomes thin and stretched. FSH, LH, and progesterone stimulate the activity of proteolytic enzymes, resulting in digestion of collagen in the follicular wall and increasing its distensibility. The gonadotropin surge also releases histamine, and histamine alone can induce ovulation in some experimental models.

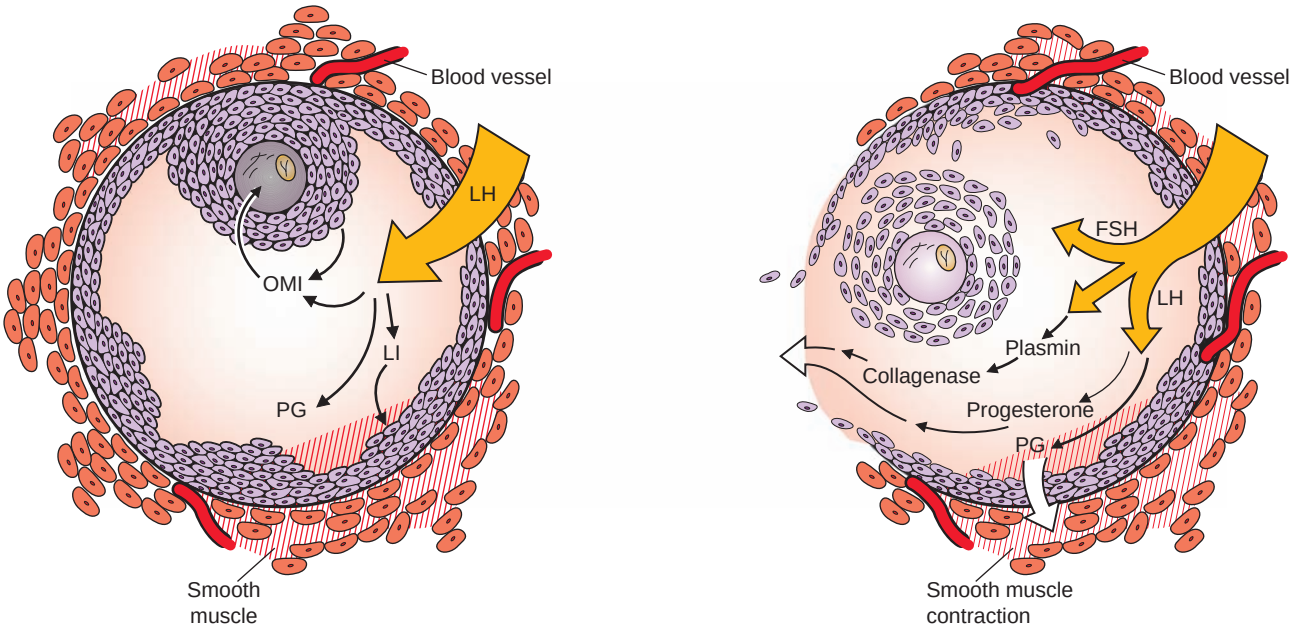
The proteolytic enzymes are activated in an orderly sequence.[243] The granulosa and theca cells produce plasminogen activator in response to the gonadotropin surge. Plasminogen is activated by either of two plasminogen activators: tissue-type plasminogen activator and urokinase-type plasminogen activator. These activators are encoded by separate genes and are also regulated by inhibitors.

Plasminogen activators produced by granulosa cells activate plasminogen in the follicular fluid to produce plasmin. Plasmin, in turn, generates active collagenase to disrupt the follicular wall. In rat models, plasminogen activator synthesis is triggered by LH stimulation (as well as growth factors and FSH), while plasminogen inhibitor synthesis is decreased.[244] Thus, before and after ovulation, the inhibitor activity is high, while just at ovulation, activator activity is high and the inhibitors are at a nadir. The molecular regulation of these factors is necessary for the coordination that leads to ovulation. Plasminogen activator synthesis in granulosa cells is expressed only at the right preovulatory stage in response to LH. The inhibitor system, which is very active in the theca and interstitial cells, prevents inappropriate activation of plasminogen and disruption of growing follicles. The inhibitor system has been demonstrated to be present in human granulosa cells and preovulatory follicular fluid and to be responsive to paracrine substances, epidermal growth factor and interleukin-1β.[245–247] Movement of the follicle destined to ovulate to the surface of the ovary is important in that the exposed surface of the follicle is now prone to rupture because it is separated from cells rich in the plasminogen inhibitor system. Ovulation is the result of proteolytic digestion of the follicular apex, a site called the stigma. The matrix metalloproteinase enzymes and their endogenous inhibitors, TIMPs, increased in response to LH and progesterone, are also involved in this event.[248]

In the rat, the gene that encodes for plasminogen activator contains a promoter region which has several sequences for known transcription factors, such as the cyclic AMP-responsive element (CRE). The activation of this CRE (which involves the CRE binding protein) requires FSH stimulation. Thus, both gonadotropins appear to be involved in this process. Studies in the monkey indicate that the activation of plasminogen activator is mediated by prostaglandin E_2.[249]

Prostaglandins E_2 and $F_{2\alpha}$, but mainly prostaglandin E_2, and other eicosanoids (especially HETEs, hydroxyeicosatetraenoic acids) increase markedly in the preovulatory follicular fluid in response to the LH surge, reaching a peak concentration at ovulation.[250–252] Prostaglandin synthesis is stimulated by interleukin-1β, implicating this cytokine in ovulation.[253] Inhibition of cyclooxygenase-2 (COX-2) synthesis of these products from arachidonic acid blocks follicle rupture without affecting the other LH-induced processes of luteinization and oocyte maturation.[254–256]

Prostaglandins act to free proteolytic enzymes within the follicular wall, and the HETEs may promote angiogenesis and hyperemia (an inflammatory-like response).[249, 251, 257] LH and PGE_2 both activate the epidermal growth factor-like signaling pathway that leads to cumulus expansion and resumption of oocyte meiosis.[258] Prostaglandins may also contract smooth muscle cells that have been identified in the ovary, thereby aiding the extrusion of the oocyte-cumulus cell mass. ***This ovulatory role of prostaglandins is so well demonstrated that infertility patients should be advised to avoid the use of drugs that inhibit prostaglandin synthesis.***[256, 259, 260]



A large number of leukocytes enter the follicle prior to ovulation. Neutrophils are a prominent feature in the theca compartment of both healthy and atretic antral follicles.[261] The accumulation of leukocytes is mediated by chemotactic mechanisms of the interleukin system.[262] However, ovulation does not depend on these invading immune cells for the expression of the inflammatory-like response associated with ovulation. Ovarian follicular cells themselves in response to LH express the genes involved with immune responses, resulting in the release of the host of products that affect the cellular reactions associated with ovulation and the remodeling process that leads to the corpus lutem.[263]

Estradiol levels plunge as LH reaches its peak. This may be a consequence of LH down-regulation of its own receptors on the follicle. Theca tissue derived from healthy antral follicles exhibits marked suppression of steroidogenesis when exposed to high levels of LH whereas exposure over a low range stimulates steroid production. The low midcycle levels of progesterone exert an inhibitory action on further granulosa cell multiplication, and the drop in estrogen may also reflect this local follicular role for progesterone. Finally, estrogen can exert an inhibitory effect on P450c17, a direct action on the gene that is not receptor-mediated.

The granulosa cells that are attached to the basement membrane and enclose the follicle become luteal cells. The cumulus granulosa cells attach to the oocyte. In the mouse, the cumulus cells are metabolically linked to the oocyte and respond to the FSH surge by secreting hyaluronic acid that disperses the cumulus cells prior to ovulation. This hyaluronic acid response depends upon maintenance of the link with the oocyte, indicating the secretion of a supporting factor. The oocyte further secretes factors that promote granulosa cell proliferation and maintain the structural organization of the follicle.[264] Proliferation of the cumulus cells is suppressed by FSH, while FSH stimulates mural granulosa cell proliferation, supported by the oocyte factor or factors.

The FSH peak, partially and perhaps totally dependent on the preovulatory rise of progesterone, has several functions. Plasminogen activator production is sensitive to FSH as well as LH. Expansion and dispersion of the cumulus cells allows the oocyte-cumulus cell mass to become free-floating in the antral fluid just before follicle rupture. The process involves the deposition of a hyaluronic acid matrix, the synthesis of which is stimulated by FSH. Finally, an adequate FSH peak ensures an adequate complement of LH receptors on the granulosa layer. It should be noted that a shortened or inadequate luteal phase is observed in cycles when FSH levels are low or selectively suppressed at any point during the follicular phase.

The mechanism that shuts off the LH surge is unknown. Within hours after the rise in LH, there is a precipitous drop in the plasma estrogens. The decrease in LH may be due to a loss of the positive stimulating action of estradiol or to an increasing negative feedback of progesterone. The abrupt fall in LH levels may also reflect a depletion in pituitary LH content due to down-regulation of GnRH receptors, either by alterations in GnRH pulse frequency or by changes in steroid levels.[265, 266] LH may further be controlled by "short" negative feedback of LH upon the hypothalamus. Direct LH suppression of hypothalamic-releasing hormone production has been demonstrated. However, in the sheep, the LH surge ends before the GnRH signal begins to decline.[267] Another possibility has been suggested, a so-called gonadotropin surge-inhibiting factor (GnSIF) originating in the ovary.[268, 269] GnSIF is produced in granulosa cells under the control of FSH and reaches a peak level in the circulation in the midfollicular phase. Its major role is believed to be prevention of premature luteinization. It is likely that a combination of all of these influences cause the rapid decline in gonadotropin secretion.

The many contributions of progesterone to ovulation are highlighted by the results of experiments in the monkey. Suppression of steroidogenesis at midcycle prevented ovulation, but not resumption of oocyte meiosis.[218] The administration of a progestin agonist to this experimental model restored ovulation. Mice bearing the knock-out for the progesterone receptor gene fail to ovulate, although maturation of the oocyte and luteinization are not impeded.[270, 271] These experiments indicate that progesterone receptor-A is the critical isoform necessary for normal ovulation.

An adequate gonadotropin surge does not ensure ovulation. The follicle must be at the appropriate stage of maturity in order for it to respond to the ovulating stimulus. In the normal cycle, gonadotropin release and final maturation of the follicle coincide because the timing of the gonadotropin surge is controlled by the level of estradiol, which in turn is

a function of follicular growth and maturation. Therefore, gonadotropin release and morphological maturity are usually coordinated and coupled in time. In the majority of human cycles, the requisite feedback relationships in this system allow only one follicle to reach the point of ovulation. Nonidentical multiple births may, in part, reflect the random statistical chance of more than one follicle fulfilling all the requirements for ovulation.

Summary of the Key Ovulatory Events

1. The LH surge initiates the continuation of meiosis in the oocyte, luteinization of the granulosa, and synthesis of progesterone and prostaglandins within the follicle.

2. Progesterone enhances the activity of proteolytic enzymes responsible, together with prostaglandins, for digestion and rupture of the follicular wall.

3. The progesterone-influenced midcycle rise in FSH serves to free the oocyte from follicular attachments, to convert plasminogen to the proteolytic enzyme, plasmin, and to ensure that sufficient LH receptors are present to allow an adequate normal luteal phase.

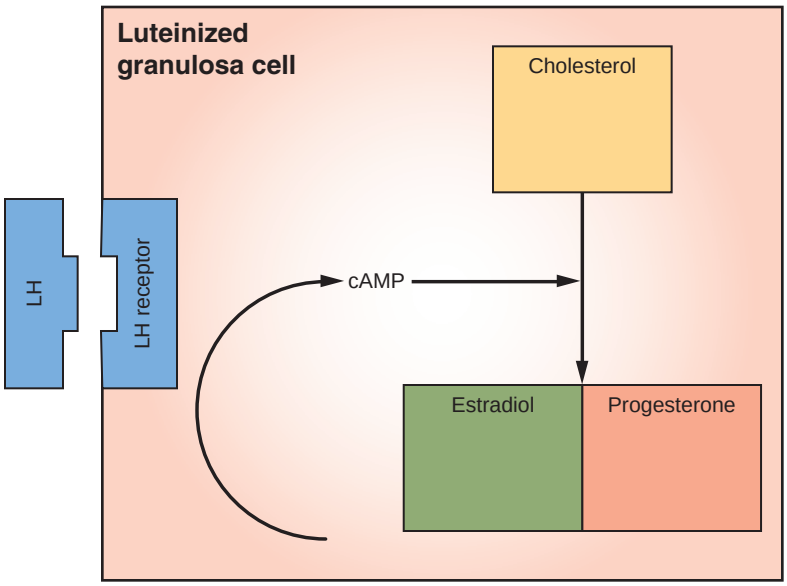
The Luteal Phase

Before rupture of the follicle and release of the ovum, the granulosa cells begin to increase in size and assume a characteristic vacuolated appearance associated with the accumulation of a yellow pigment, lutein, which lends its name to the process of luteinization and the anatomical subunit, the corpus luteum. During the first 3 days after ovulation, the granulosa cells continue to enlarge. In addition, theca lutein cells may differentiate from the surrounding theca and stroma to become part of the corpus luteum. Dissolution of the basal lamina and rapid vascularization and luteinization make it difficult to distinguish the origin of specific cells.

Capillaries begin to penetrate into the granulosa layer after the cessation of the LH surge, reach the central cavity, and often fill it with blood.[272] Angiogenesis is an important feature of the luteinization process, a response to LH that is mediated by factors such as vascular endothelial growth factor (VEGF) and angiopoietins produced in luteinized granulosa cells.[173, 174, 273] In the early luteal phase, angiogenesis accompanies an increased expression of VEGF, with stabilization of vessel growth maintained by angiopoietin-1 binding to the endothelial Tie-2 receptor.[177, 274] With regression of the corpus luteum, VEGF and angiopoietin-1 expression decrease allowing greater occupancy of the Tie-2 receptor by angiopoietin-2, leading to the vascular breakdown that accompanies luetolysis.

By day 8 or 9 after ovulation, a peak of vascularization is reached, associated with peak levels of progesterone and estradiol in the blood. The corpus luteum has one of the highest blood flows per unit mass in the body. On occasion, this ingrowth of vessels and bleeding will result in unchecked hemorrhage and an acute surgical emergency that can present at any time during the luteal phase. Indeed, this is a significant clinical risk in women who are anticoagulated; such women should receive medication to prevent ovulation.

Normal luteal function requires optimal preovulatory follicular development. Suppression of FSH during the follicular phase is associated with lower preovulatory estradiol levels, depressed midluteal progesterone production, and a decrease in luteal cell mass.[275]



Experimental evidence supports the contention that the accumulation of LH receptors during the follicular phase predetermines the extent of luteinization and the subsequent functional capacity of the corpus luteum. The successful conversion of the avascular granulosa of the follicular phase to the vascularized luteal tissue is also of importance. Because steroid production is dependent upon low-density lipoprotein (LDL) transport of cholesterol, the vascularization of the granulosa layer is essential to allow LDL-cholesterol to reach the luteal cells to provide sufficient substrate for progesterone production. One of the important jobs for LH is to regulate LDL receptor binding, internalization, and postreceptor processing; the induction of LDL receptor expression occurs in granulosa cells during the early stages of luteinization in response to the midcycle LH surge.[276, 277] This mechanism supplies cholesterol to the mitochondria for utilization as the basic building block in steroidogenesis.

The lifespan and steroidogenic capacity of the corpus luteum are dependent on continued tonic LH secretion. Studies in hypophysectomized women have demonstrated that normal corpus luteum function requires the continuous presence of small amounts of LH.[278] The dependence of the corpus luteum on LH is further supported by the prompt luteolysis that follows the administration of GnRH agonists or antagonists or withdrawal of GnRH when ovulation has been induced by the administration of pulsatile GnRH.[279, 280] There is no evidence that other luteotropic hormones, such as prolactin, play a role in primates during the menstrual cycle.[281]

The corpus luteum is not homogeneous. Besides the luteal cells, also present are endothelial cells, leukocytes, and fibroblasts. The nonsteroidogenic cells form the bulk, about 70%, of the total cell population. The leukocyte immune cells produce several cytokines, including interleukin-1β and tumor necrosis factor-α.[282] The many different leukocytes in the corpus luteum are also a rich resource for cytolytic enzymes, prostaglandins, and growth factors involved in angiogenesis, steroidogenesis, and luteolysis.

The corpus luteum is one of the best examples of communication and cross talk in biology. For example, endothelial cells contribute vasoactive compounds, and, in turn, steroidogenic cells contribute factors that influence angiogenesis. The harmonious function of this system is in inverse proportion to its complexity.

Endothelial cells constitute about 35% of the cells in a mature corpus luteum.[283] As elsewhere in the body, endothelial cells participate in immune reactions and endocrine functions. The endothelial cells are a source of endothelin-1, expressed in response to changes in blood flow, blood pressure, and oxygen tension. Studies have indicated that endothelin-1

may be a mediator of luteolysis.[284, 285] Inhibition of vascular endothelial growth factor (VEGF) prevents luteal angiogenesis.[286]

Even the luteal cell population is not homogeneous, being composed of at least two distinct cell types, large and small cells.[287] Some believe that the large cells are derived from granulosa cells and the small cells from theca cells. The small cells are the most abundant. Despite the fact that greater steroidogenesis takes place in the large cells, it is the small cells that contain LH and hCG receptors.[288, 289] The absence of LH/hCG receptors on the large cells, presumably derived from granulosa cells that acquire LH receptors in the late follicular phase, requires explanation. Perhaps large cells are functioning at a maximal level with receptors totally occupied and functional, or because of intercellular communication through gap junctions, the large cells do not require direct gonadotropin support. Thus, the large cells can be functioning at a high level, under the control of regulating factors that originate in the small cells in response to gonadotropins. In addition, the overall function is influenced by autocrine-paracrine signals from endothelial and immune cells.

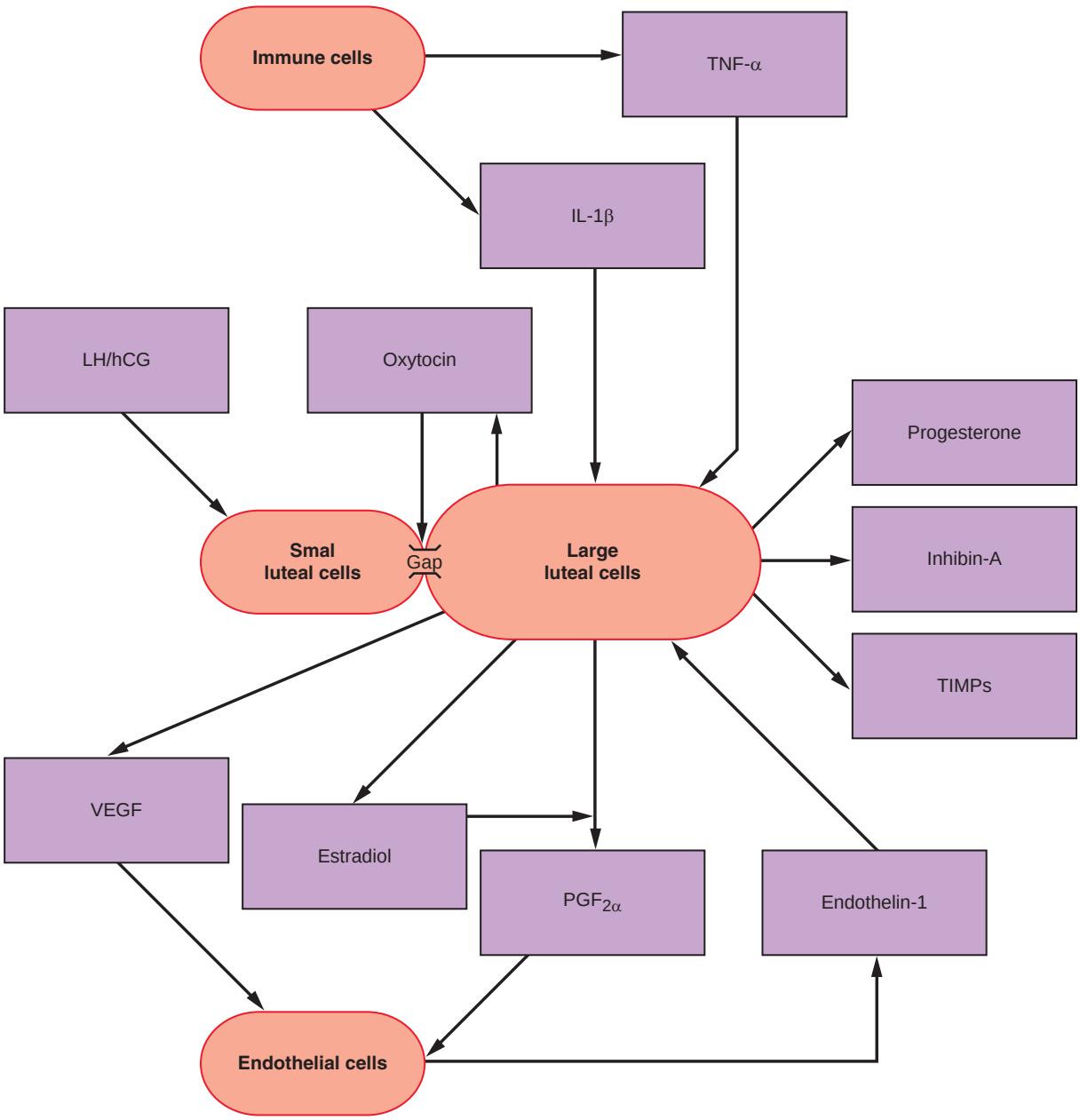
Large luteal cells produce peptides (oxytocin, relaxin, inhibin, GnRH, growth factors, and prostaglandins) and are more active in steroidogenesis, with greater aromatase activity and more progesterone synthesis than small cells.[290, 291] Human granulosa cells (already luteinizing when recovered from in vitro fertilization patients) contain minimal amounts of P450c17 mRNA. This is consistent with the two-cell explanation, which assigns androgen production (and P450c17) to the cells derived from theca cells. With luteinization, expression of StAR, P450scc and 3β-hydroxysteroid dehydrogenase markedly increases as expected, to account for the increasing production of progesterone, and the continued expression of these essential factors requires LH.[292–294] The aromatase system (P450arom), of course, continues to be active in luteinized granulosa cells.

Progesterone levels normally rise sharply after ovulation, reaching a peak approximately 8 days after the LH surge. Initiation of new follicular growth during the luteal phase is further inhibited by the low levels of gonadotropins due to the negative feedback actions of estrogen, progesterone, and inhibin-A. With the appearance of LH receptors on the granulosa cells of the dominant follicle and the subsequent development of the follicle into a corpus luteum, inhibin expression comes under the control of LH, and expression changes from inhibin-B to inhibin-A.[125, 288, 295] The circulating levels of inhibin-A rise in the late follicular phase to reach a peak level at the midluteal phase.[33, 126, 187] Inhibin-A, therefore, contributes to the suppression of FSH to nadir levels during the luteal phase, and to the changes at the luteal-follicular transition. There is a wave of small follicule growth during the luteal phase, probably in response to the FSH surge at midcycle; however, luteal phase FSH suppression ensures that a mature, large follicle will not emerge.[296, 297]

The secretion of progesterone and estradiol during the luteal phase is episodic, and the changes correlate closely with LH pulses.[99, 298] Because of this episodic secretion, relatively low midluteal progesterone levels, which some inappropriately believe are indicative of an inadequate luteal phase, can be found in the course of totally normal luteal phases. The corpus luteum of the primate is unique in its production of estrogen; however, unlike the follicular phase, luteal estrogen synthesis is dependent on LH. Within the corpus luteum, progesterone acts locally to enhance the LH-induced luteinization of granulosa cells, to support its own LH-stimulated synthesis, and to inhibit apoptosis.[299–301]

In the normal cycle the time period from the LH midcycle surge to menses is consistently close to 14 days. For practical purposes, luteal phases lasting between 11 and 17 days can be considered normal.[302] The incidence of short luteal phases is about 5–6%. It is well known that significant variability in cycle length among women is due to the varying number of days required for follicular growth and maturation in the follicular phase. The luteal phase cannot be extended indefinitely even with progressively increasing LH exposure, indicating that the demise of the corpus luteum is due to an active luteolytic mechanism.

The corpus luteum rapidly declines 9–11 days after ovulation, and the mechanism of the degeneration remains unknown. In certain nonprimate mammalian species, a luteolytic factor originating in the uterus and stimulated by estrogen (prostaglandin $F_{2\alpha}$) regulates the lifespan of the corpus luteum. No definite luteolytic factor has been identified in the primate menstrual cycle, and removal of the uterus in the primate does not affect the ovarian cycle. The morphological regression of luteal cells may be induced by the estradiol produced by the corpus luteum.[303] The premature elevation of circulating estradiol levels in the early luteal phase results in a prompt fall in progesterone concentrations. Direct injections of estradiol into the ovary bearing the corpus luteum induce luteolysis while similar treatment of the contralateral ovary produces no effect.[304] This action of estrogen may be mediated by nitric oxide. Nitric oxide stimulates luteal prostaglandin synthesis and decreases progesterone production.[305] Nitric oxide and hCG have opposing actions in the human corpus luteum; nitric oxide is associated with apoptosis of luteal cells.[306] The final signal for luteolysis, however, is prostaglandin $F_{2\alpha}$, produced within the ovary in response to



the locally synthesized luteal estrogen.[304, 307] These relationships are supported by genome studies delineating prostaglandin $F_{2\alpha}$ and hCG effects on gene expression.[308] The early primate luteal phase is dominated by the intraluteal synthesis of the luteotropic prostaglandin, PGE_2; late in the luteal phase, intraluteal prostaglandin synthesis shifts to $PGF_{2\alpha}$.[291]

There is another possible role for the estrogen produced by the corpus luteum. In view of the known estrogen requirement for the synthesis of progesterone receptors in endometrium, luteal phase estrogen may be necessary to allow the progesterone-induced changes in the endometrium after ovulation. Inadequate progesterone receptor content due to inadequate estrogen priming of the endometrium is an additional possible mechanism for infertility or early miscarriage, another form of luteal phase deficiency.

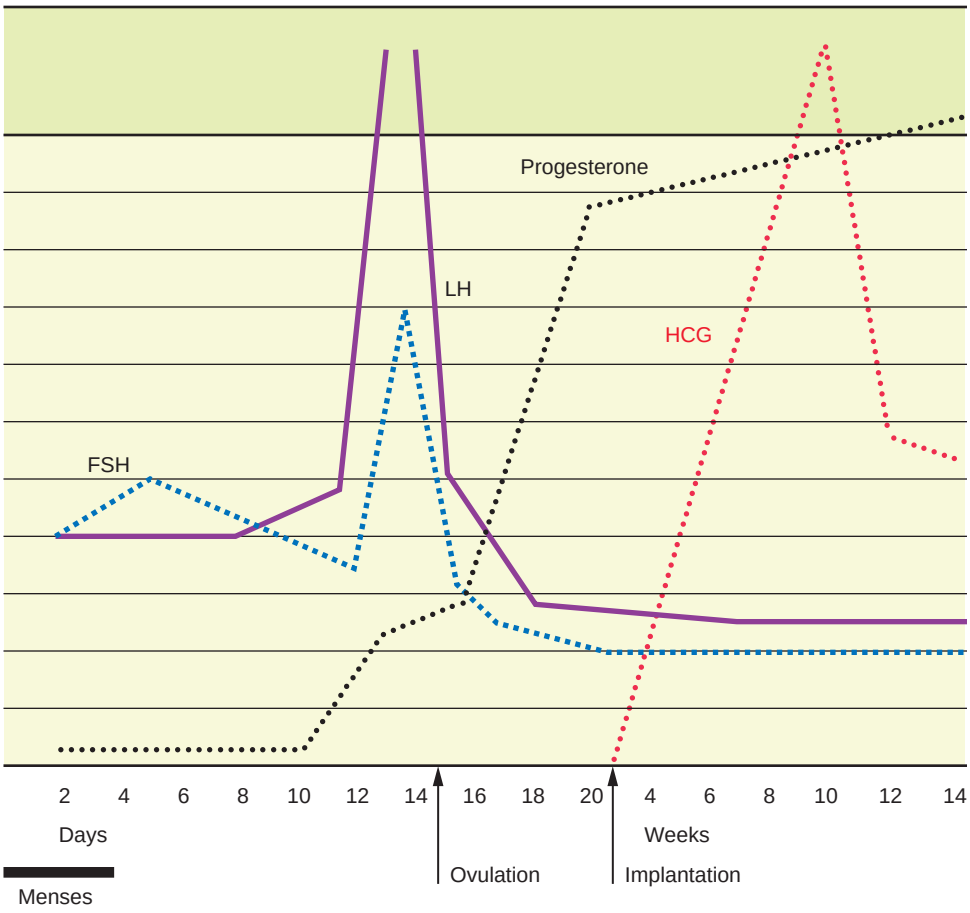
Experimental evidence indicates that the luteolytic effect of prostaglandin $F_{2\alpha}$ is partially mediated by endothelin-1.[284, 285] Prostaglandin $F_{2\alpha}$ stimulates the synthesis of endothelin-1; endothelin-1 inhibits luteal steroidogenesis, and in turn, endothelin-1 stimulates prostaglandin production in luteal cells.[309] In addition, endothelin-1 stimulates the release of tumor necrosis factor-α, a growth factor known to induce apoptosis, and members of the tumor necrosis factor family, including their receptors, are expressed in the corpus lueum with a peak at the time of luteolysis.[310, 311]

The corpus luteum involves cellular interactions that require cell-to-cell contact. Gap junctions are a prominent feature of luteal cells, just as they are in the follicle before ovulation. When the various cell types of the corpus luteum are studied together, the performance is different compared with studies of single cell types, greater steroidogenesis more closely approximating the total function of the corpus luteum.[312] It is believed that communication and exchange of signals takes place through the gap junction structures, explaining how the small cells respond to LH and hCG, but the large cells are the main site of steroidogenesis. Regulation of the gap junction system is influenced by oxytocin, a paracrine role for oxytocin in the corpus luteum.[25]

When ovulation is induced by the administration of GnRH, normal luteal phase demise occurs despite no change in treatment, arguing against a change in LH as the luteolytic mechanism. In addition, LH receptor binding affinity does not change throughout the luteal phase; thus the decline in steroidogenesis must reflect deactivation of the system (producing a refractoriness of the corpus luteum to LH), perhaps through the uncoupling of the G protein adenylate cyclase system. This is supported by studies in the monkey in which alteration in LH pulse frequency or amplitude did not provoke luteolysis.[313]

The process of luteolysis involves proteolytic enzymes, especially the matrix metalloproteinases (MMPs). These enzymes are held under inhibitory control by tissue inhibitors of metalloproteinases (TIMPs) secreted by the steroidogenic luteal cells, and because TIMP levels do not change in luteal tissue, luteolysis is believed to involve a direct increase in MMP expression. An important part of the rescue mission for human chorionic gonadotropin (hCG) is to prevent this increase in MMP expression.[314] hCG can increase TIMP production, and this, too, would inhibit MMP activity and luteolysis.[315] The source of the metalloproteinases is the fibroblast cell, and because luteal fibroblasts do not contain LH/hCG receptors, the release of metalloproteinases depends on another signal. One such signal can be locally produced activin-A that acts upon the fibroblasts to synthesize and release metalloproteinases.[316] Emerging hCG from a pregnancy can inhibit this activin-A system by increasing follistatin, the glycopeptide that binds activin. In addition, the human ovary contains the complete interleukin-1 system, providing another resource for cytolytic enzymes.

The survival of the corpus luteum is prolonged by the emergence of a new stimulus of rapidly increasing intensity, hCG. Blastocysts grown in culture produce and secrete human chorionic gonadotropin (hCG), beginning days 7–8 after fertilization.[317] Messenger RNA



for hCG can be found in 6- to 8-cell human embryos.[318] Because the 8- to 12-cell stage is achieved about 3 days after fertilization, it is believed that the human embryo begins to produce hCG before implantation when it can be detected in the mother (about 6–7 days after ovulation). The embryo is capable, therefore, of preimplantation signaling, and higher levels of estradiol and progesterone can be measured in the maternal circulation even before maternal hCG is detectable, presumably because of stimulation of the corpus luteum by hCG delivered directly from the uterine cavity to the ovary.[319] Function of the corpus luteum is crucial during the first 7–9 weeks of pregnancy, and luteectomy early in pregnancy can precipitate abortion.[320] Similarly, early pregnancy loss in primates can be induced by injections of anti-hCG serum.[321] The rescue of the corpus luteum by an early pregnancy with hCG is associated with maintenance of the vascular system (not new vessel growth), a process dependent on the angiogenic factors VEGF and angiopoietin-2.[176, 177, 274, 322]

Unlike the biphasic pattern demonstrated by the circulating level of progesterone (a decrease after ovulation and then a new higher peak at the midluteal phase), the mRNA levels for the two major enzymes involved in progesterone synthesis (cholesterol side-chain cleavage and 3β-hydroxysteroid dehydrogenase) are maximal at ovulation and decline throughout the luteal phase.[323] This suggests that the lifespan of the corpus luteum is established at the time of ovulation, and luteal regression is inevitable unless the corpus luteum is rescued by the hCG of pregnancy. ***Therefore, primates have developed a system that requires rescue of the corpus luteum in contrast to lower animals that use a mechanism that actively causes the demise of the corpus luteum (luteolysis).***

Summary of Events in the Luteal Phase

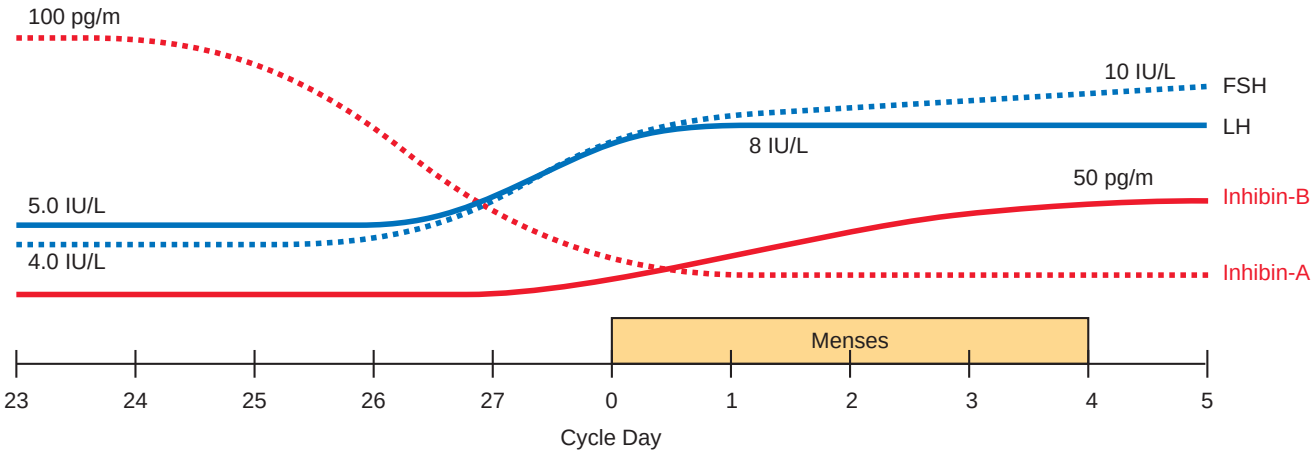
1. Normal luteal function requires optimal preovulatory follicular development (especially adequate FSH stimulation) and continued tonic LH support.
2. The early luteal phase is marked by active angiogenesis mediated by VEGF. New vessel growth is held in check by angiopoietin-1 working through its receptor Tie-2 on endothelial cells. Regression of the corpus luteum is associated with a decrease in VEGF and angiopoietin-1 expression and an increase in angiopoietin-2 activity.
3. Progesterone, estradiol, and inhibin-A act centrally to suppress gonadotropins and new follicular growth.
4. Regression of the corpus luteum may involve the luteolytic action of its own estrogen production, mediated by an alteration in local prostaglandin and involving nitric oxide, endothelin, and other factors.
5. In early pregnancy, hCG rescues the corpus luteum, maintaining luteal function until placental steroidogenesis is well established.

The Luteal-Follicular Transition

The interval extending from the late luteal decline of estradiol and progesterone production to the selection of the dominant follicle is a critical and decisive time, marked by the appearance of menses, but less apparent and very important are the hormone changes that initiate the next cycle. The critical factors include GnRH, FSH, LH, estradiol, progesterone, and inhibin.

Given the important role for FSH-mediated actions on the granulosa cells, it is appropriate that the recruitment of a new ovulating follicle is directed by a selective increase in FSH that begins approximately 2 days before the onset of menses.[324–326] Using a sensitive FSH bioassay, an increase in FSH bioactivity can be measured beginning as early as the mid-luteal phase.[34] There are at least two influential changes that result in this important increase in FSH: a decrease in luteal steroids and inhibin and a change in GnRH pulsatile secretion.

The Luteal-Follicular Transition



Inhibin-B, originating in the granulosa cells of the corpus luteum and now under the regulation of LH, reaches a nadir in the circulation at the midluteal period.[187] Inhibin-A reaches a peak in the luteal phase, and, thus, may help to suppress FSH secretion by the pituitary to the lowest levels reached during a menstrual cycle.[33, 187] The process of luteolysis, whatever the mechanism, with the resulting demise of the corpus luteum, affects inhibin-A secretion as well as steroidogenesis. The administration of inhibin-A to monkeys effectively suppresses circulating FSH.[327] Thus, an important suppressing influence on FSH secretion is removed from the anterior pituitary during the last days of the luteal phase. The selective action of inhibin on FSH (and not LH) is partly responsible for the greater rise in FSH seen during the luteal-follicular transition, compared to the change in LH. The administration of recombinant (pure) FSH to gonadotropin-deficient women has demonstrated that the early growth of follicles requires FSH, and that LH is not essential during this period of the cycle.[61, 62]

Inhibin-B levels begin to rise shortly after the increase in FSH (a consequence of FSH stimulation of granulosa cell secretion of inhibin) and reach peak levels about 4 days after the maximal increase in FSH.[33, 117, 187] Thus, suppression of FSH secretion during the follicular phase is an action exerted by Inhibin-B, whereas escape of FSH inhibition during the luteal-follicular transition is partly a response to decreasing inhibin-A secretion by the corpus luteum.

Circulating levels of activin increase before ovulation to a peak in the luteal phase; however, activin A is highly bound in the circulation, and it is not certain it has an endocrine role.[147, 187] Nevertheless, the timing is right for activin to contribute to the rise in FSH during the luteal-follicular transition. Activin enhances and follistatin suppresses GnRH activity. Evidence in vivo and in vitro indicates that gonadotropin response to GnRH requires activin activity.[328] Activin specifically acts synergistically with GnRH to stimulate gene expression in the pituitary for the FSH beta-subunit.[329]

The selective rise in FSH is also significantly influenced by a change in GnRH pulsatile secretion, previously strongly suppressed by the high estradiol and progesterone levels of the luteal phase exerting a negative feedback effect at the hypothalamus.[103, 330] A progressive and rapid increase in GnRH pulses (as assessed by the measurement of LH pulses) occurs during the luteal-follicular transition.[102] From the midluteal peak to menses, there is a 4.5-fold increase in LH pulse frequency (and presumably GnRH) from approximately 3 pulses/24 hours to 14 pulses/24 hours.[102] During this time period, the mean level of LH increases approximately 2-fold, from approximately a mean of 4.8 IU/L to 8 IU/L. The increase in FSH is, as noted, greater than that of LH. FSH pulse frequency increases 3.5-fold from the midluteal period to the time of menses, and FSH levels increase from a mean of approximately 4 IU/L to 15 IU/L.

An increase in GnRH pulse frequency from a low level of secretion has been associated with an initial selective increase in FSH in several experimental models, including the ovariectomized monkey with destruction of the hypothalamus. Treatment of hypogonadal women with pulsatile GnRH results first in predominance of FSH secretion (over LH). This experimental response and the changes during the luteal-follicular transition are similar to that observed during puberty, a predominance of FSH secretion as GnRH pulsatile secretion begins to increase.

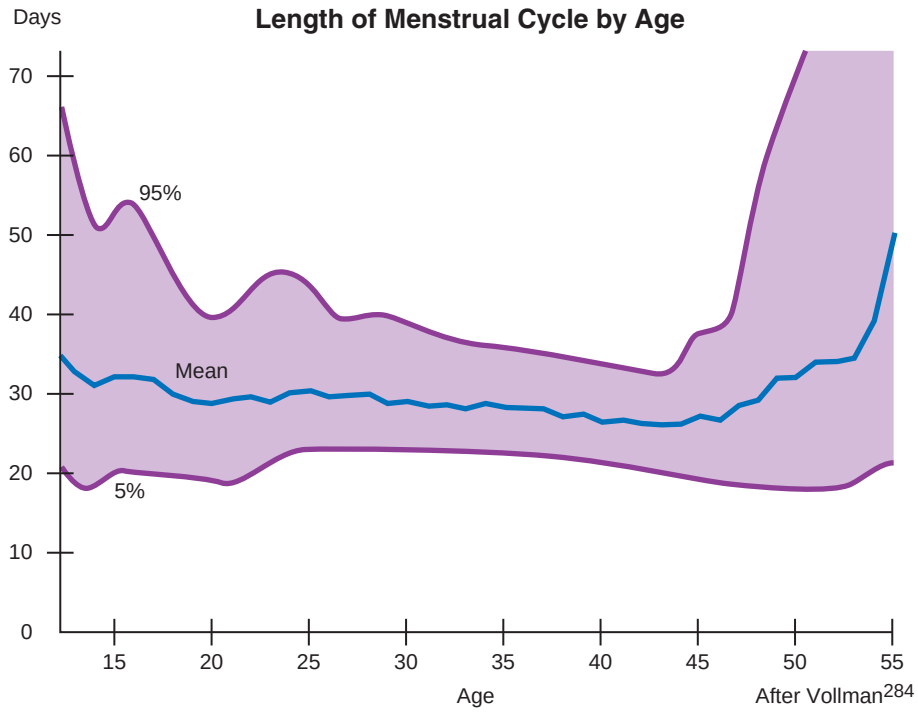
The pituitary response to GnRH is also a factor. Estradiol suppresses FSH secretion by virtue of its classic negative feedback relationship at the pituitary level. The decrease in estradiol in the late luteal phase restores the capability of the pituitary to respond with an increase in FSH secretion.[331]

Summary of Events in the Luteal-Follicular Transition

1. The demise of the corpus luteum results in a nadir in the circulating levels of estradiol, progesterone, and inhibin.
2. The decrease in inhibin-A removes a suppressing influence on FSH secretion in the pituitary.
3. The decrease in estradiol and progesterone allows a progressive and rapid increase in the frequency of GnRH pulsatile secretion and a removal of the pituitary from negative feedback suppression.
4. The removal of inhibin-A and estradiol and increasing GnRH pulses combine to allow greater secretion of FSH compared with LH, with an increase in the frequency of the episodic secretion.
5. The increase in FSH is instrumental in rescuing approximately a 70-day-old group of ready follicles from atresia, allowing a dominant follicle to begin its emergence.

The Normal Menstrual Cycle

Menstrual cycle length is determined by the rate and quality of follicular growth and development, and it is normal for the cycle to vary in individual women.[332, 333] Cycle lengths are the shortest (with the least variability) in the late 30s, a time when subtle but real increases in FSH and decreases in inhibin are occurring.[123, 302, 334–337] This can be pictured as accelerated follicular growth (because of the changes in FSH and inhibin-B). At the same time,



fewer follicles grow per cycle as a woman ages.[338] Approximately 2–4 years prior to menopause, the cycles lengthen again. In the last 10–15 years before menopause, there is an acceleration of follicular loss.[3] This accelerated loss begins when the total number of follicles reaches approximately 25,000, a number reached in normal women at age 37–38.290 Eventually menopause occurs because the supply of follicles is depleted.[339]

The changes in the later reproductive years reflect either lesser follicular competence as the better primordial follicles respond early in life, leaving the lesser follicles for later, or the fact that the total follicular pool is reduced in number (or both factors).[340] Arguing in favor of a role for a reduced follicular pool is the observation that follicular fluid obtained from preovulatory follicles of older women contains amounts of inhibin-A and -B that are similar to that measured in follicular fluid from young women.[341]

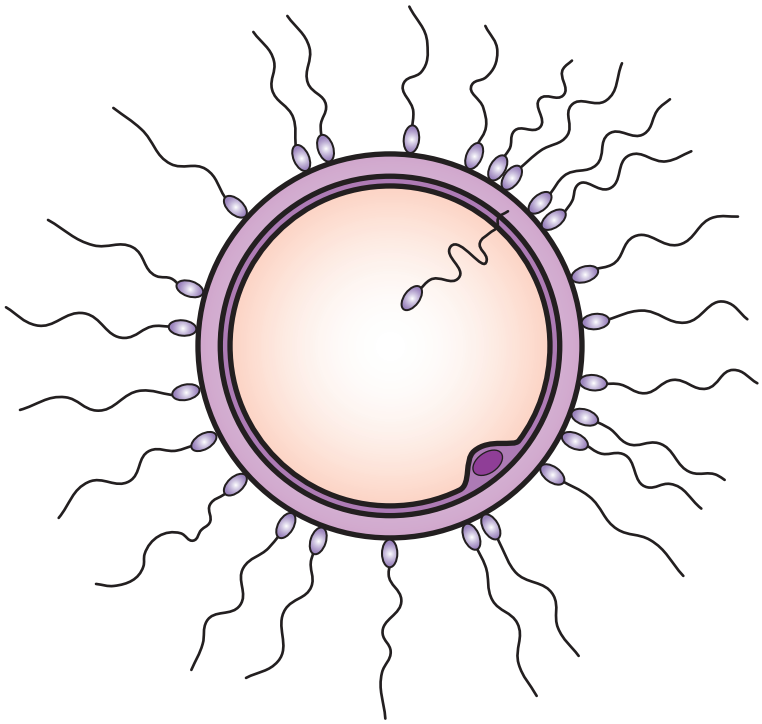
Variations in menstrual flow and cycle length are common at the extremes of reproductive age, during the early teenage years and the years preceding the menopause. The prevalence of anovulatory cycles is highest in women under age 20 and over age 40.[342, 343] Menarche is typically followed by approximately 5–7 years of relatively long cycles that gradually decrease in length and become more regular. Although menstrual cycle characteristics generally do not change appreciably during the reproductive years,[344] overall cycle length and variability slowly decrease. On average, mean cycle length and variability reach their lows at about age 40–42.[333, 344] Over the subsequent 8–10 years before the menopause, the trend is reversed; both average cycle length and variability steadily increase as ovulations become less regular and frequent.[332, 333, 345, 346] Mean cycle length is greater in women at the extremes of body mass and composition; both high and low body mass index are associated with an increased mean cycle length.[347, 348]

In general, variations in cycle length reflect differences in the length of the follicular phase of the ovarian cycle. Women who have a 25-day cycle ovulate on or about cycle day 10–12, and those with a 35-day cycle ovulate approximately 10 days later. Within a few years after menarche, the luteal phase becomes extremely consistent (13–15 days) and remains so until the perimenopause.[332, 333] At age 25, over 40% of cycles are between 25 and 28 days in length; from age 25 to 35, over 60% are. Although it is the most often reported intermenstrual interval, only approximately 15% of cycles in reproductive aged women are actually 28 days in length. Less than 1% of women have a regular cycle lasting less than 21 days or more than 35 days.[349] Most women have cycles that last from 24 to 35 days, but at least 20% of women experience irregular cycles.[344]

All references are available online at: http://www.clinicalgynendoandinfertility.com

7

Sperm and Egg Transport, Fertilization, and Implantation

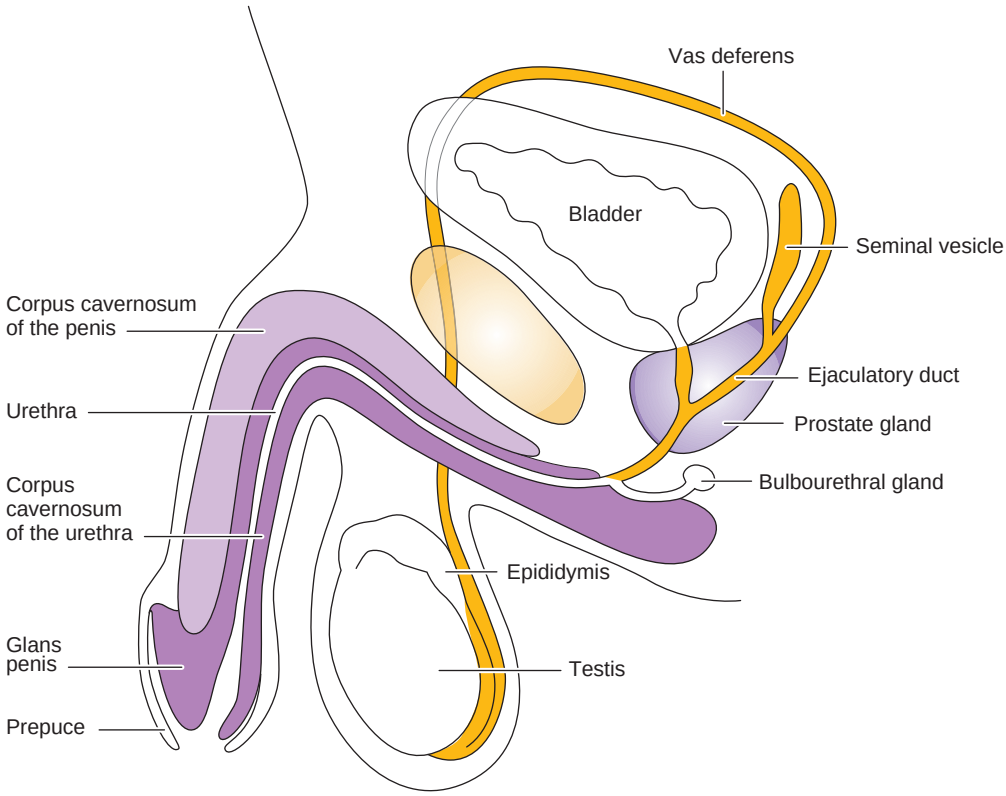


Among his many accomplishments, Galileo Galilei gave to science, in 1609, two important instruments, the telescope and the microscope.[1] Anton van Leeuwenhoek of Delft, Holland, was fascinated by Galileo's microscope. Leeuwenhoek was a draper and had no medical or scientific training, yet he became a fellow of the Royal Society of London to which he submitted 375 scientific papers. In 1677, Leeuwenhoek described (fairly accurately) the "little animals of the sperm." It was another 198 years before Wilhelm August Oscar Hertwig, in Germany, demonstrated the union of sperm and egg, fertilization, in the sea urchin.

The coming together of sperm and egg is one of the essentials of reproduction; however, the remote site of this event and the enclosed origins of the participants made fertilization a difficult subject for study. This changed with the advent of in vitro fertilization. Greater understanding of sperm and egg development and union is one of the major benefits of the clinical application of the assisted reproductive technologies. This chapter examines the mechanisms involved in sperm and egg transport, fertilization, and implantation.

Sperm Transport

The evolution of scrotal mammals and the adoption of internal fertilization are associated with sperm maturation that occurs outside of the testes. This includes epididymal maturation in the male and capacitation in the female before fertilization. The need for capacitation (the final step required to acquire the ability to fertilize) may be an evolutionary consequence of the development of a storage system for inactive sperm in the caudal epididymis.[2]



The epididymis is divided into four segments, the initial segment, the caput where the sperm begin their process of maturation, the corpus where maturation continues, and the cauda, the site of final maturation and storage.[3] The sperm reach the caudal epididymis approximately 72 days after the initiation of spermatogenesis. At this time, the head of the sperm contains a membrane-bound nucleus capped by the acrosome, a large vesicle of proteolytic enzymes. The inner acrosomal membrane is closely apposed to the nuclear membrane, and the outer acrosomal membrane is next to the surface plasma membrane. The flagellum is a complex structure of microtubules and fibers, surrounded at the proximal end by mitochondria. Motility and the ability to fertilize are acquired gradually as the sperm pass into the epididymis.

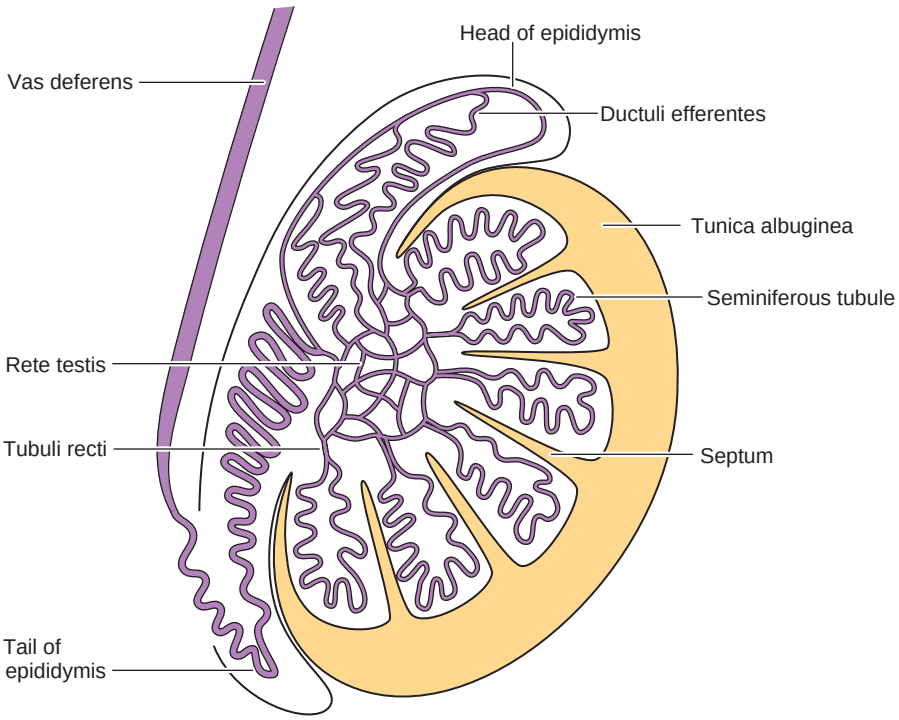
The caudal epididymis stores sperm available for ejaculation. The ability to store functional sperm provides a capacity for repetitive fertile ejaculations. Preservation of optimal sperm function during this period of storage requires adequate testosterone levels in the circulation and maintenance of the normal scrotal temperature.[4] The importance of temperature is emphasized by the correlation of reduced numbers of sperm associated with episodes of body fever. The evolution of the scrotum served the purpose of achieving the cooler temperatures required for effective sperm storage.

Surface proteins are acquired by the sperm in the epididymis, proteins that must be removed in the process of capacitation, discussed later. It can be argued that the epididymis is limited to a storage role because sperm that have never passed through the epididymis and that have been obtained from the vasa efferentia in men with a congenital absence of the vas deferens can fertilize the human oocyte in vitro and result in pregnancy with live birth.[5] Indeed, the injection of sperm obtained by testicular biopsy directly into an oocyte (intracytoplasmic sperm injection) is very successful in achieving fertilization and pregnancy.[6] Epididymal functions should not be dismissed, however, because with direction injection into the oocyte, surface protein action at the oocyte membrane is bypassed.

The use of sperm from men with sperm abnormalities should be pursued with some caution. Sperm obtained from men with Y chromosome microdeletions involving the AZFc region of Yq11 can transmit the deletion to male children, who then will also likely be infertile. In addition, men with certain Y microdeletions, including a portion of AZFc, may have an increased susceptibility for developing testicular germ cell tumors. The outcome in subsequent generations must be assessed and appropriate genetic screening must be developed to avoid the transmission of subtle but important genetic alterations. Meanwhile, men with severe oligospermia or azoospermia should receive appropriate genetic counseling and should be offered testing for Y microdeletions before their sperm are used for intracytoplasmic sperm injection (ICSI).

Semen forms a gel almost immediately following ejaculation but then is liquefied in 20–30 minutes by enzymes derived from the prostate gland. The alkaline pH of semen provides protection for the sperm from the acid environment of the vagina. This protection is transient, and most sperm left in the vagina are immobilized within 2 hours. The more fortunate sperm, by their own motility, gain entrance into the tongues of cervical mucus that layer over the ectocervix. These are the sperm that enter the uterus; the seminal plasma is left behind in the vagina. This entry is rapid, and sperm have been found in mucus within 90 seconds of ejaculation.[7] The destruction of all sperm in the vagina 5 minutes after ejaculation does not interfere with fertilization in the rabbit, further attesting to the rapidity of transport.[8]

Contractions of the female reproductive tract occur during coitus, and these contractions may be important for entry of sperm into the cervical mucus and further transport.



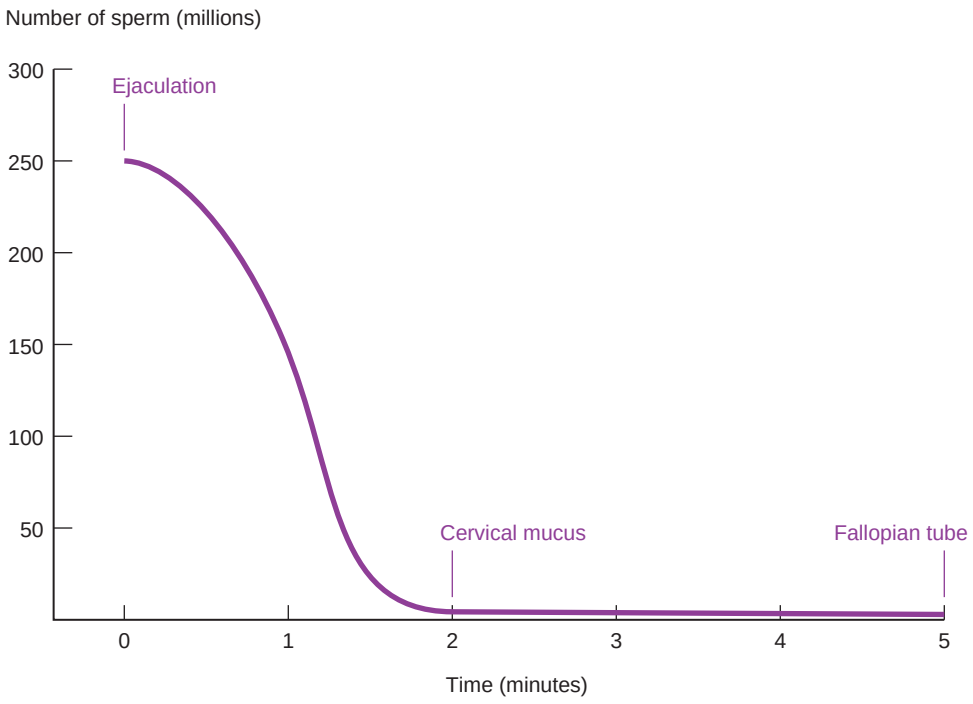
Presumably successful entry is the result of combined female and male forces (the flagellar activity of the sperm). The success of therapeutic insemination, however, indicates that coitus and female orgasm are not essential for sperm transport.

The sperm swim and migrate through pores in the cervical mucus that are smaller than the sperm head; therefore, the sperm must actively push their way through the mucus.[9] One cause of infertility, presumably, is impaired sperm movement that prevents this transport through the mucus. This movement is probably also influenced by the interaction between the mucus and the surface properties of the sperm head; for example, sperm antibodies on the sperm head inhibit sperm movement in the mucus.[10] Abnormal morphology of the sperm head is often associated with impaired flagellar function; however, abnormal head morphology alone can be a cause of poor mucus penetration.[11, 12] A sperm coating protein (beta-defensin 126), acquired in the caudal epididymis, carries a high negative charge and is critical for movement through the cervical mucus.[13] It is generally believed that the cervical mucus has a filtering action; abnormal and less "capable" sperm have difficulty getting through.[14]

Uterine contractions and sperm motility propel the sperm upward, and in the human, sperm can be found in the tube 5 minutes after insemination.[15] Labeled albumin is present in the tubes within 30 seconds after intrauterine instillation.[16] It is possible that the first sperm to enter the tube are at a disadvantage. In the rabbit these early sperm have only poor motility, and there is frequent disruption of the head membranes.[17] The sperm in this vanguard are unlikely to achieve fertilization. Other sperm that have colonized the cervical mucus and the cervical crypts then make their way more slowly to the ampulla of the tube in order to meet the egg. The number of sperm in the cervical mucus is relatively constant for 24 hours after coitus, and after 48 hours there are relatively few remaining in the mucus.[18] Although the isthmic region of the tube functions as a sperm reservoir in many species, this does not appear to be the case in human fallopian tubes.[19]

Human sperm have been found in the fallopian tube as long as 80 hours after intercourse, and these sperm can still perform normally with zona-free hamster oocytes.[20] In animals, the fertilizable lifespan is usually one-half the motile lifespan.

The ***attrition in sperm numbers*** from vagina to tube is substantial.[21] Of an average of 200 to 300 million sperm deposited in the vagina, at most only a few hundred (rarely reaching



1,000), and often less, achieve proximity to the egg.[19] Greater numbers are observed in the tubal ampulla at the time of ovulation. The major loss occurs in the vagina, with expulsion of semen from the introitus playing an important role. Other causes for loss are digestion of sperm by vaginal enzymes and phagocytosis of sperm along the reproductive tract. There are also reports of sperm burrowing into or being engulfed by endometrial cells. Many sperm continue past the oocyte to be lost into the peritoneal cavity. The cervix serves as a reservoir, providing a supply of sperm for up to 72 hours.

Within the fallopian tube, sperm that are not yet capacitated are bound to the epithelial cells. When these sperm release and undergo capacitation, they display a new pattern of movement that has been called ***hyperactivated motility***.[22, 23] This motility may be influenced by an interaction with the tubal epithelium that results in greater speed and better direction as well as prevention of attachment and entrapment.

Structure of the Cervical Mucus

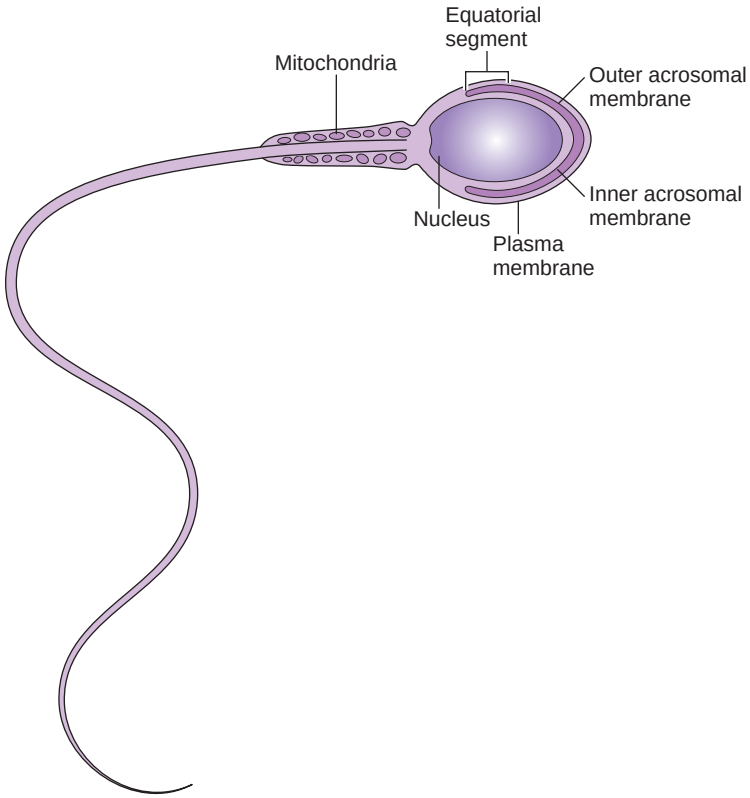
The cervical mucus is a complex structure that is not homogeneous.[24] The mucus is secreted in granular form, and a networked structure of the mucus is formed in the cervical canal. Thus, not all areas of the cervical mucus are equally penetrable by the sperm. It is proposed, based on animal studies, that the outward flow of the cervical mucus establishes a linear alignment of parallel strands that direct the sperm upward. Pressurization of the mucus by contractions of the uterus further aid this alignment and may contribute to the speed of sperm transport. Responding to the midcycle estrogen peak, cervical mucus production, water content, and space between its large glycoproteins reach a maximum in the immediate preovulatory period. The process of capacitation is initiated, but not completed during the sperm's passage through the cervix.

Capacitation

The discovery, in 1951, that rabbit and rat spermatozoa must spend some hours in the female tract before acquiring the capacity to penetrate ova stimulated intensive research efforts to delineate the environmental conditions required for this change in the sperm to occur.[25, 26] The process by which the sperm were transformed was called ***capacitation, the cellular changes that ejaculated spermatozoa must undergo in order to fertilize***.[27] Attention was focused on the hormonal and time requirements and the potential for in vitro capacitation. Capacitation occurs while the sperm are in the fallopian tube and is characterized by three accomplishments:

1. The ability to undergo the acrosome reaction.
2. The ability to bind to the zona pellucida.
3. The acquisition of hypermotility.

Capacitation changes the surface characteristics of sperm, as exemplified by removal of seminal plasma factors that coat the surface of the sperm, modification of the surface charge, and restriction of receptor mobility. The protein beta-defensin 126 (DEFB126), derived from the epididymis, is the major coating protein of sperm that facilitates movement through the cervical mucus; its release from the sperm head is essential in order for the sperm to bind to the zona pellucida of the ovum.[13, 28] Sperm proteomics have demonstrated a huge collection of receptors coated on the surface of sperm.[29, 30] The purpose of the coating



proteins may be to produce a reservoir of sperm in the fallopian tubes by promoting the binding of sperm to tubal epithelial cells.[31]

The surface changes are associated with modifications of sperm cell membrane sterols, lipids, and glycoproteins that cause decreased stability of the plasma membrane and the membrane lying immediately under it, the outer acrosomal membrane. The membranes undergo further, more striking modifications when capacitated sperm reach the vicinity of an oocyte or when they are incubated in follicular fluid. There is a breakdown and merging of the plasma membrane and the outer acrosomal membrane, the ***acrosome reaction.***[32] This allows egress of the enzyme contents of the acrosome, the caplike structure that covers the sperm nucleus. These enzymes, which include hyaluronidase, a neuraminidase-like factor, cumulus-dispersing enzyme, and a protease called acrosin, are all thought to play roles in sperm penetration of the egg investments. The changes in the sperm head membranes also prepare the sperm for fusion with the egg membrane. It is the inner acrosomal membrane that fuses with the oocyte plasma membrane. The acrosome reaction can be induced by zona pellucida proteins of the oocyte and by human follicular fluid in vitro.[33, 34] In addition, capacitation endows the sperm with hypermotility, and the increased velocity of the sperm is a very critical factor in achieving zona penetration.[22]

The events that constitute the process of capacitation are regulated by the redox status of the sperm cell.[35, 36] Redox reactions induce tyrosine phosphorylation, an absolute requirement for capacitation. These reactions are dependent on a critical increase in intracellular calcium concentrations due to an influx of extracellular calcium, believed to be induced by progesterone. Sperm are stimulated to undergo capacitation when they encounter the alkaline change in pH at the time of ovulation, a response of the fallopian tubes to the midcycle hormonal changes.

Although capacitation classically has been defined as a change sperm undergo in the female reproductive tract, specifically in the fallopian tubes, it is apparent that sperm of

some species, including the human, can acquire the ability to fertilize after a short incubation in defined media and without residence in the female reproductive tract. Therefore, success with assisted reproductive technologies is possible. In vitro capacitation requires a culture medium that is a balanced salt solution containing energy substrates such as lactate, pyruvate, and glucose and a protein such as albumin, or a biologic fluid such as serum or follicular fluid. Sperm-washing procedures probably remove factors that coat the surface of the sperm, one of the initial steps in capacitation. The removal of cholesterol from the sperm membrane is believed to prepare the sperm membrane for the acrosome reaction.[37] The loss of cholesterol regulates the expression of sperm cell membrane surface lectins that are involved in sperm surface receptors for the zona pellucida.[38] The time required for in vitro capacitation is approximately 2 hours.[39]

The final dash to the oocyte is aided by the increased motility due to the state of hyperactivity. This change in motility can be measured by an increase in velocity and flagellar beat amplitude. Perhaps the increase in thrust gained by this hyperactivity is necessary for avoiding attachment to tubal epithelium and achieving penetration of the cumulus and zona pellucida.

Key Steps in Sperm Transport

1. Approximately 72 days are required to produce spermatozoa, a time period followed by storage in the epididymis prior to ejaculation.
2. Sperm enter the cervical mucus and then the fallopian tubes within minutes, but only a few hundred sperm or less reach the oocyte. The cervix serves as a reservoir of sperm for up to 72 hours.
3. Capacitation, a process initiated during the sperm's passage through the cervix and completed in the fallopian tube or during in vitro incubation in an appropriate medium, is characterized by the acquired ability of sperm to undergo the acrosome reaction, to bind to the zona pellucida, and to acquire hyperactivated motility.
4. The acrosome reaction is due to the modification and breakdown, followed by a merger, of the sperm cell membrane and the outer acrosomal membrane, allowing the release of enzymes and changes in the inner acrosomal membrane, necessary for fusion with the oocyte cell membrane.

Egg Transport

The oocyte, at the time of ovulation, is surrounded by granulosa cells (the ***cumulus oophorus***) that attach the oocyte to the wall of the follicle. The ***zona pellucida***, a noncellular porous layer of glycoproteins secreted by the oocyte, separates the oocyte from the granulosa cells. The granulosa cells communicate metabolically with the oocyte by means of ***gap junctions*** between the oocyte plasma membrane and the cumulus cells. In response to the midcycle surge in luteinizing hormone (LH), maturation of the oocyte proceeds with the resumption of meiosis as the oocyte completes the first meiotic division, enters into the second meiotic division, and arrests in the second metaphase. Just before ovulation, the cumulus cells retract their cellular contacts from the oocyte. The disruption of the gap

junctions induces maturation and migration of the cortical granules to the outer cortex of the oocyte.[40] Prior to ovulation, the oocyte and its cumulus mass of cells prepare to leave their long residence in the ovary by becoming detached from the follicular wall.

Egg transport encompasses the period of time from ovulation to the entry of the morula into the uterus. The egg can be fertilized only during the early stages of its sojourn in the fallopian tube. Within 2–3 minutes of ovulation in some animals, the cumulus and oocyte are in the ampulla of the fallopian tube. It takes longer in humans.

In rats and mice the ovary and distal portion of the tube are covered by a common fluid-filled sac. Ovulated eggs are carried by fluid currents to the fimbriated end of the tube. In contrast, in primates, including humans, the ovulated eggs adhere with their cumulus mass of follicular cells to the surface of the ovary. The fimbriated end of the tube sweeps over the ovary in order to pick up the egg. Entry into the tube is facilitated by muscular movements that bring the fimbriae into contact with the surface of the ovary. Variations in this pattern surely exist, as evidenced by women who achieve pregnancy despite having only one ovary and a single tube located on the contralateral side. Furthermore, eggs deposited in the cul-de-sac by transvaginal injection are picked up by the tubes.[41]

Although there can be a small negative pressure in the tube in association with muscle contractions, oocyte pickup is not dependent on a suction effect secondary to this negative pressure. Ligation of the tube just proximal to the fimbriae does not interfere with pickup.[42]

The fallopian tubes are lined by an epithelium that undergoes cyclic changes comparable to the endometrium, in response to the hormonal changes of the menstrual cycle.[43] The epithelium is composed of nonciliated cells and ciliated cells. The nonciliated cells are characterized by major secretory activity during the follicular phase of the cycle, culminating in the release of cytoplasmic components during the passage of the egg, perhaps providing important metabolic factors for transport and implantation. The cilia on the surface of the fimbriae (where they are present in greater concentrations) display adhesive sites, and these seem to have prime responsibility for the initial movement of the egg into the tube. This movement is dependent on the presence of follicular cumulus cells surrounding the egg, because removal of these cells prior to egg pickup prevents effective egg transport.

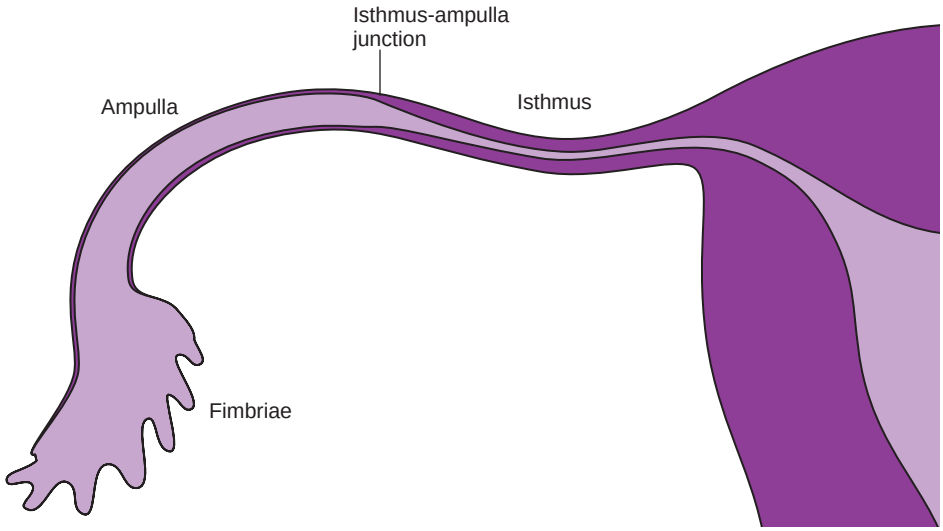
In the ampulla of the tube the many cilia beat synchronously in the direction of the uterus, and in the fimbria the ciliary beat is faster in the secretory phase of the menstrual cycle.[44] In women and monkeys, this unidirectional beat is also found in the isthmus of the tube. The specific contribution of the cilia to egg transport in the ampulla and isthmus is an unresolved question. Most investigators have credited muscular contractions of the tubes as the primary force for moving the egg.[45] However, interference with muscle contractility in the rabbit did not block egg transport.[46] Reversing a segment of the ampulla of the tube so that the cilia in this segment beat toward the ovary interferes with pregnancy in the rabbit without blocking fertilization. The fertilized ova are arrested when they come in contact with the transposed area.[47] This suggests that ciliary beating is crucial for egg transport. Spontaneous pregnancies have been reported in women who suffer from Kartagener's syndrome in which there is a congenital absence of dynein arms (a protein structure associated with motility) in all bodily cilia, and thus the cilia do not beat.[48] However, motility of cilia in the tube may be disordered and not totally absent. Nevertheless, the pregnancies in women with Kartagener's syndrome indicate the importance of uterine and fallopian tube muscular peristalsis.[49] Human fallopian tube muscular contractions are stimulated by prostaglandins E_2 and $F_{2\alpha}$, and decreased by progestins, hCG, and oxytocin.[50]

Transvaginal endoscopic observation of actual ovum and cumulus oophorous pickup in women revealed that the process is relatively slow (more than 15 minutes), the fimbriae on the ovulating side are distinguished by being erect (probably due to engorged blood

vessels and suggesting a local ovarian influence), and the only observable active mechanism involved ciliary movement.[51, 52] It is probable, therefore, that in normal circumstances, smooth muscle contractions and the flow of secretory fluid in response to ciliary activity work together to accomplish egg transport.

In most species, transport of the ovum (the fertilized oocyte) through the tube requires approximately 3 days.[53] The time spent within the various parts of the tube varies from one species to another. Transport through the ampulla is rapid in the rabbit, whereas in women the egg spends about 80 hours in the tube, 90% of which is in the ampulla at the junction of the ampulla with the isthmus. It is in this location that fertilization and dispersion of the cumulus cells are completed.

Attempts to modify tubal function as a method for understanding its physiology have involved three major pharmacologic approaches: (1) altering levels of steroid hormones, (2) interference with or supplementation of adrenergic stimuli, and (3) treatment with prostaglandins. Although there is abundant literature on the effects of estrogen and progesterone on tubal function, it is clouded by the use of different hormones, different doses, and different timing of injections. Because of these variations, it is difficult to obtain a coherent picture and to relate the experimental results to the in vivo situation. In general, pharmacologic doses of estrogen favor retention of eggs in the tube. This "tube-locking" effect of estrogen can be partially reversed by treatment with progesterone.



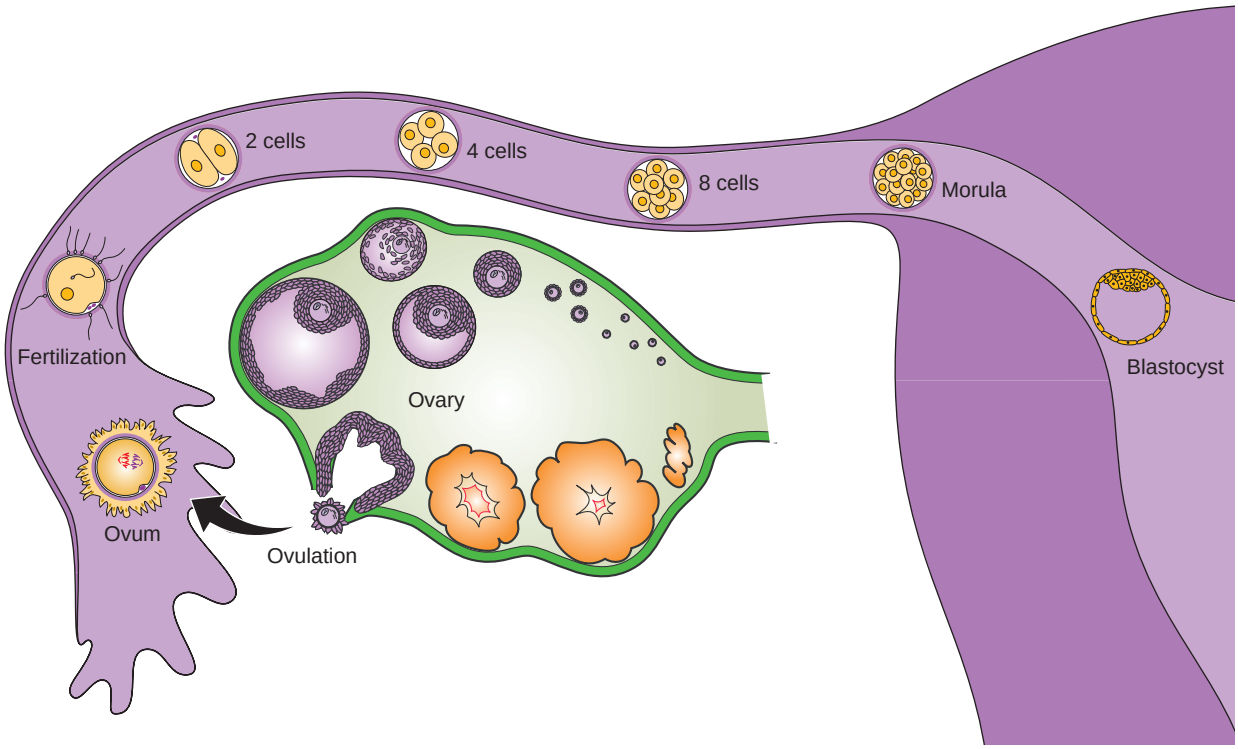
The isthmus of the tube has an extensive adrenergic innervation. Surgical denervation of the tube, however, does not disrupt ovum transport. Prostaglandins (PG) of the F series stimulate muscle activity of the tube. Although $PGF_{2\alpha}$ stimulates human oviductal motility in vivo, it does not cause acceleration of ovum transport.

Is there an essential anatomic segment of the tube? Excision of the ampullary-isthmus junction in rabbits does not prevent fertility.[54] This is equally true if small segments of the ampulla are removed, and pregnancy can occur even if the entire isthmus and uterotubal junction are excised. Although the fimbriae are thought to play a crucial role in fertility, spontaneous pregnancies have been reported following sterilization by fimbriectomy or following surgical repair of tubes whose fimbriated ends had been excised.[55, 56] The fallopian tube appears to readily adapt to anatomic changes and restrictions.

In most species, a period of residence in the tube appears to be a prerequisite for full development. Rabbit eggs can be fertilized in the uterus, but they do not develop unless transferred to the tubes within 3 hours of fertilization.[57] This implies that there may be a component in uterine fluid during the first 48 hours following ovulation that is toxic to the egg.[57] Indirect evidence of an inhospitable environment is also provided by studies indicating that there must be synchrony between development of the endometrium and the egg for successful pregnancy to occur.[58, 59] If the endometrium is in a reduced or advanced stage of development compared with the egg, fertility is compromised. In addition, the blastocyst must undergo cleavage and development in order to gain the capability to implant in the uterus. ***Thus, it is conceptually useful to view the fallopian tube not as an active transport mechanism, but as a structure that provides an important holding action. This functional behavior is coordinated by the increasing estrogen and progesterone levels after ovulation, although local embryonic signals may also be involved.***

Successful pregnancies have occurred in the human following the Estes procedure, in which the ovary is transposed to the uterine cornua.[60] Eggs are ovulated directly into the uterus, completely bypassing the tube. Moreover, when fertilized donor eggs are transferred to women who are receiving hormone supplementation, there are several days during the treatment cycle when the blastocysts will implant. This crucial difference between animal and human physiology is of more than academic importance. There has been speculation about the use of drugs that could accelerate tubal transport as a means of providing contraception by ensuring that the egg would reach the uterus when it was in an unreceptive state. Although this may work in animals, it is of doubtful value in the human because perfect synchrony is not required.

Animal and human reproduction also differ in the occurrence of ectopic pregnancy. Ectopic pregnancies are rare in animals, and in rodents they are not induced even if the uterotubal junction is occluded immediately following fertilization. The embryos reach the blastocyst stage and then degenerate.



Key Steps in Egg Transport

1. After ovulation, the oocyte and its surrounding cumulus are in the ampulla of the fallopian tube within 15–20 minutes.
2. Tubal transport depends on smooth muscle contractions and ciliary-induced flow of secretory fluid.
3. The fallopian tube provides an important holding action to allow time for the endometrium to become receptive and the blastocyst to become capable of implantation, a time period of approximately 80 hours, 90% of which is in the ampulla.

Oocyte Maturation

Oocyte maturation, as reviewed in Chapter 6, is regulated by the sex hormones and the complex interaction among an array of growth factors and cytokines in the follicular fluid. In nonmammalian species, a nongenomic action of progesterone causes an increase in intracellular calcium concentrations. In human oocytes, an influx of extracellular calcium occurs in response to estradiol, followed by secondary rises in calcium ions from intracellular stores, characterized by wavelike oscillations.[61] This is a nongenomic response to estradiol at the cell surface, and the transient increases in intracellular calcium improve the quality of the oocyte and contribute to the capability for fertilization.

Calcium oscillations are a property common to mammalian oocytes and are also an early reaction to the fertilizing spermatozoan.[62] Neither the presence of estradiol nor the calcium oscillations are required for oocytes to resume meiosis. However, improved fertilization following estradiol-induced calcium increases indicates an important role for intrafollicular estradiol in overall oocyte maturation.

Fertilization

The fertilizable life of the human oocyte is unknown, but estimates range between 12 and 24 hours. However, immature human eggs recovered for in vitro fertilization can be fertilized even after 36 hours of incubation. Equally uncertain is knowledge of the fertilizable lifespan of human sperm. The most common estimate is 48–72 hours, although motility can be maintained after the sperm have lost the ability to fertilize. The extreme intervals that have achieved pregnancy documented after a single act of coitus are 6 days prior to and 3 days after ovulation.[63] The great majority of pregnancies occur when coitus takes place within the 3-day interval just prior to ovulation.[64]

Contact of sperm with the egg, which occurs in the ampulla of the tube, may not be random; there is some evidence for sperm-egg communication that attracts sperm to the oocyte.[65–67] This chemotactic responsiveness of sperm requires the changes that take place in the capacitation process.[68] Thus, this may be a system to select a sperm that is fully capable of fertilization.

The cumulus oophorus undergoes a preovulatory expansion that has at least two important roles. The ampullary space of the human fallopian tube is relatively large (compared with the oocyte), and the expanded cumulus may serve to increase the chances of an encounter with one of the few spermatozoa that have reached the far section of the tube. In addition, this change may facilitate sperm passage through the cumulus. Sperm pass through the cumulus without the release of acrosomal enzymes.[69] It has been suggested, based on in vitro experiments, that the cumulus is essential for full development of the fertilizing ability of sperm; however, removal of the cumulus does not prevent sperm penetration and fertilization.

Despite the evolution from external to internal fertilization over a period of about 100 million years, many of the mechanisms have remained the same.[70–72] The acellular zona pellucida that surrounds the egg at ovulation and remains in place until implantation has two major functions in the fertilization process:

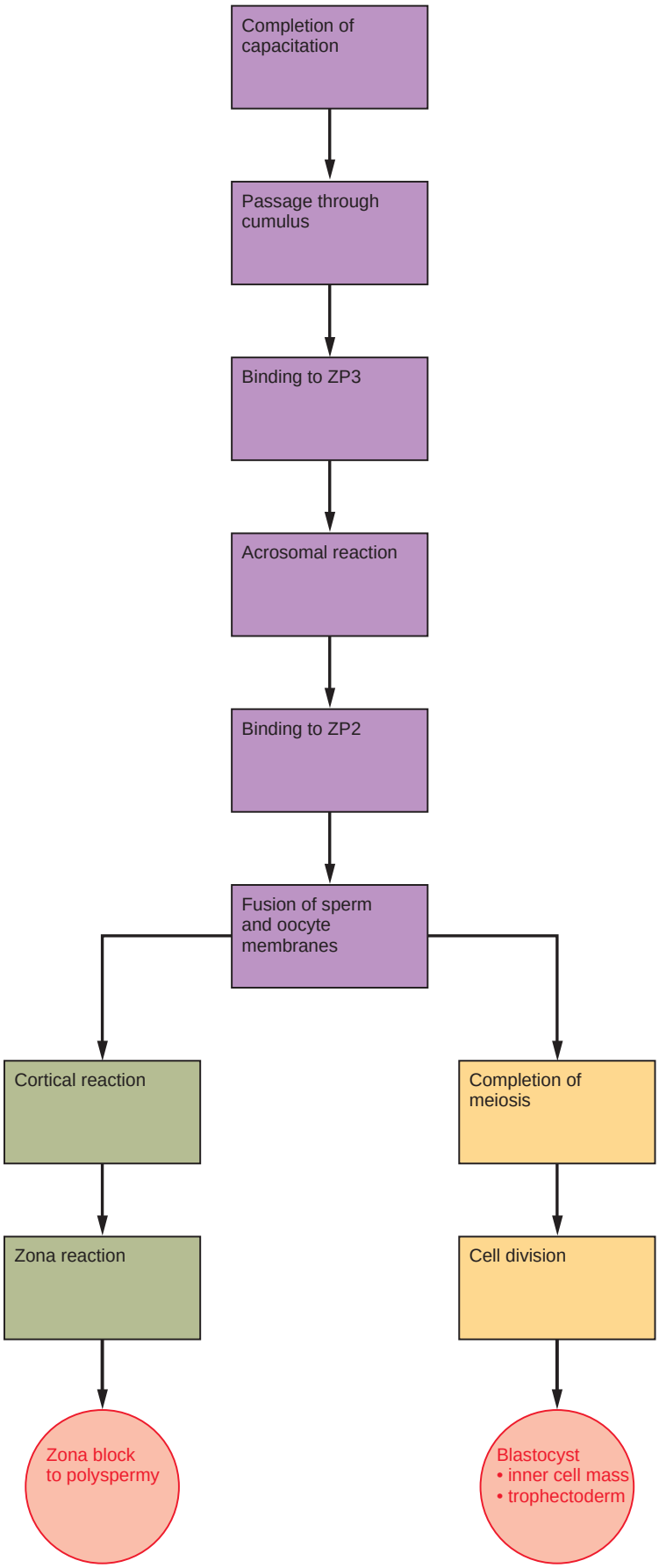
1. The zona pellucida contains ligands for sperm, which are, with some exceptions, relatively species-specific.

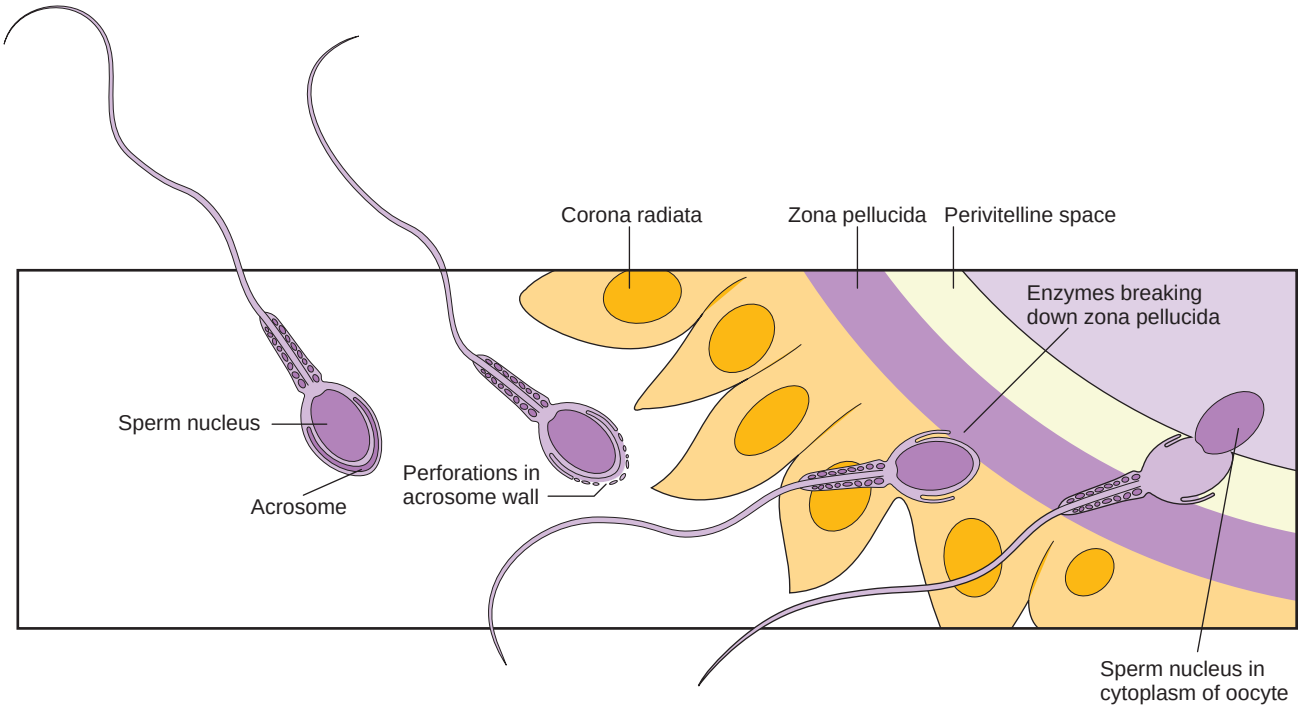
2. The zona pellucida undergoes the ***zona reaction*** in which the zona becomes impervious to other sperm once the fertilizing sperm penetrates, and thus it provides a bar to polyploidy.[73]

Sperm bind to the zona pellucida for about a minute and then penetrate it rapidly, mediated by acrosin, a trypsin-like proteinase that is bound to the inner acrosomal membrane of the sperm.[72, 74, 75] The pivotal role assigned to acrosin has been disputed. For example, manipulations that increase the resistance of the zona to acrosin do not interfere with sperm penetration, and thus sperm motility may be the critical factor. The zona pellucida is a porous structure due to the many glycoproteins assembled into long, interconnecting filaments. Nevertheless, a preponderance of evidence favors tenacious binding of capacitated spermatozoa to the zona pellucida as a requirement for penetration, although it is clear that penetration requires physical thrust with active motility not only of the tail but of the head as well. Indeed, sperm undergo rapid lateral oscillations of the head about a fulcrum at the head-tail junction, suggesting a scythe-like action on the zona.[2, 72]

The acrosome is a lysosome-like organelle in the anterior region of the sperm head, lying just beneath the plasma membrane like a cap over the nucleus. The lower part of the two arms is called the equatorial segment. The acrosome contains many enzymes that are exposed by the ***acrosome reaction, the loss of the acrosome immediately before fertilization. This reaction is one of exocytosis, the fusion of an intracellular storage vesicle with the inner surface of the cell membrane, followed by release of the vesicle contents.*** The acrosome reaction requires an influx of calcium ions, the efflux of hydrogen ions, an increase in pH, and fusion of the plasma membrane with the outer acrosomal membrane, leading to the exposure and escape of the enzymes contained on the inner acrosomal membrane. Binding to the zona pellucida is required to permit a component of the zona to induce the acrosomal reaction. This component is believed to be a glycoprotein sperm receptor, which thus serves dual functions: binding of sperm and induction of the acrosomal reaction.

The initial contact between the sperm and the oocyte is a receptor-mediated process. The zona pellucida is composed of glycoproteins *secreted by the oocyte*, known as ZP1, ZP2, ZP3, and ZP4, with ZP3 being the most abundant.[76–79] In humans, ZP3 and ZP4 are the primary ligands for sperm and ZP2 binding occurs after the acrosome reaction, participating in the zona reaction to prevent polyspermy.[80] Structural alteration of these glycoproteins leads to a loss of activity; inactivation of these ligands after fertilization is probably accomplished by one or more cortical granule enzymes. The *ZP* gene is expressed only in growing oocytes. DNA sequence similarities of the *ZP3* gene in various mammals indicate that this gene has been evolutionarily conserved and that the sperm-ligand interaction is a


Completion of capacitation
Passage through cumulus
Binding to ZP3
Acrosomal reaction
Binding to ZP2
Fusion of sperm and oocyte membranes
Cortical reaction
Zona reaction
Zona block to polyspermy
Completion of meiosis
Cell division
Blastocyst
• inner cell mass
• trophectoderm



common mechanism among mammals.[81] Mice with a disrupted *ZP3* gene produce oocytes lacking a zona pellucida and are unable to become pregnant.[82, 83] A vaccine against zona pellucida glycoproteins from pigs is used to control reproduction in a variety of female animals, including female elephants and deer.[84] Human use of such a vaccine has been hampered by the difficulty in preparing pure glycoproteins, but this is now possible using recombinant techniques.[85]

The initial binding of the sperm to the zona requires recognition on the part of the sperm of the carbohydrate component of the species-specific glycoprotein ligand molecule.[79, 86] Once binding is accomplished, the acrosome reaction is triggered by the peptide chain component of the receptor glycoprotein. At least one receptor on the sperm head is a tyrosine kinase that is activated by binding to the ZP3 glycoprotein and is an initiator of the acrosome reaction.[87, 88] This interaction is analogous to the general principle of behavior for hormone-receptor binding and activity. In the case of sperm and oocyte, recognition of the oocyte zona ligand involves an enzyme on the surface of the sperm that becomes exposed during capacitation. Formation of the ZP3-enzyme complex, therefore, not only produces binding but also induces the acrosome reaction. The G protein signaling system is present on sperm heads, and activation at this point in time by progesterone, in an extragenomic mechanism, opens calcium channels to increase intracellular levels of calcium ions, a requirement for the acrosome reaction.[89–91] Thus, the initial sperm-zona interaction depends on binding of acrosome-intact spermatozoa, followed by a process mediated by the enzymes released by the zona-induced acrosome reaction. Protein kinase C activation is an important step in the acrosomal reaction, leading to phosphorylation of sperm proteins involved in the process.[92, 93]

Glycodelin is a glycoprotein with many isoforms, found in endometrium, the fallopian tubes, follicular fluid, and in seminal fluid. The various forms of glycodelin modulate sperm function and fertilization by maintaining sperm in the uncapacitated state and inhibiting binding of sperm to the zona pellucida by competing for zona receptors. A specific receptor for glycodelin is present on sperm, and thus it makes sense that down-regulation of glycodelin expression would be associated with the hormonal changes at ovulation coinciding with the opening of the fertilization window.[94]

Spermatozoa enter the perivitelline space at an angle. The oocyte is a spherical cell covered with microvilli. The sperm head is like a flat dish, and the thickness of the head is a little less than the distance between the oocyte microvilli.[95] The region of the equatorial segment of the sperm head, the distal portion of the acrosome, makes initial contact with the vitelline membrane (the egg plasma membrane or oolemma). At first, the egg membrane engulfs the sperm head, and, subsequently, there is fusion of egg and sperm membranes. Bedford has opined that the tangential trajectory of this process allows closure with expansion of the blastocyst, preventing the persistence of a hole that would allow herniation or interfere with the normal hatching that occurs later within the uterus.[96]

Sperm-egg fusion is mediated by specific proteins. Two membrane proteins from the sperm head have been sequenced; one (PH-20, also called SPAM1) is involved in binding to the zona pellucida, and the other (PH-30, also called fertilin) is involved in fusion with the oocyte.[97, 98] PH-20, with hyaluronidase activity, is also active in dispelling the cumulus.[99] The cell membrane of the unfertilized oocyte contains integrin adhesion/fusion molecules that recognize peptides such as fibronectin, laminin, and collagen.[100] Fibronectin appears on the spermatozoa, but it is disputed whether it appears with caudal maturation or after capacitation. Vitronectin is a sperm protein that is activated after capacitation and the acrosome reaction and may be the key peptide interacting with oocyte cell membrane integrins.[101] These steps in the fusion process will occur only with sperm that have undergone the acrosome reaction. Multiple sperm-associated surface proteins are involved in binding to the oocyte membrane, but no single peptide has been identified as absolutely essential for fertilization, implying redundancy.[102, 103]

Fusion of the sperm and oocyte membrane to form a zygote is followed by the cortical reaction and metabolic activation of the oocyte. An increase in intracellular free calcium in a periodic, oscillatory pattern always precedes the cortical reaction and oocyte activation at fertilization, and this is believed to be the mechanism by which the spermatozoon triggers these developmental events.[62, 104, 105] It is believed that calcium signaling in fertilization is initiated by the introduction of a sperm factor into the egg, a phospholipase protein that activates inositol 1,4,5-trisphosphate leading to calcium release.[105] An analysis of failed fertilizations in one couple after intracytoplasmic sperm injections (ICSI) indicated a high prevalence of failed oocyte activation; repeated ICSI after initiating oocyte activation with a calcium ionophore resulted in a successful pregnancy.[106]

The initiation of the block to penetration of the zona by other sperm is mediated by the ***cortical reaction,*** another example of exocytosis with the release of materials from the ***cortical granules***, lysosome-like organelles that are found just below the egg surface.[107] As with other lysosome-like organelles, these materials include various hydrolytic enzymes. Changes brought about by these enzymes lead to the ***zona reaction, the hardening of the extracellular layer by cross-linking of structural proteins, and inactivation of ligands for sperm receptors***.[108] Thus, the zona block to polyspermy is accomplished. The initial change in this zona block is a rapid depolarization of the oocyte membrane associated with a release of calcium ions from calmodulin.[109, 110] The increase in intracellular calcium acts as a signal or trigger to activate protein synthesis in the oocyte. The depolarization of the membrane initiates only a transient block to sperm entry. The permanent block is a consequence of the cortical reaction and the release of enzymes, also apparently triggered by the increase in calcium.

Approximately 3 hours after insemination, meiosis is completed.[111] The second polar body is released, leaving the egg with a haploid complement of chromosomes. The addition of chromosomes from the sperm restores the diploid number to the now fertilized egg. The chromatin material of the sperm head decondenses, and the male pronucleus is formed. The male and the female pronuclei migrate toward each other, and as they move into close proximity the limiting membranes break down, and a spindle is formed on which the chromosomes become arranged. Thus, the stage is set for the first cell division.

Embryonic genome activity in the human begins early; DNA synthesis activity can be detected 9–10 hours after insemination.[112] Human gene expression (transcription) begins between the 4- and 8-cell stages of preimplantation cleavage, 2–3 days after fertilization.[113] Earlier embryonic signals may be derived from a store of maternal messenger RNAs, termed the "maternal legacy."[114, 115] In addition, proteomics have identified RNAs and transcription factors within sperm that suggest a mechanism for a paternal contribution to the early development of the embryo.[30, 116]

Clinicians are interested not only in how normal fertilization takes place but also in the occurrence of abnormal events that can interfere with pregnancy. It is worthwhile, therefore, to consider the failures that occur in association with in vivo fertilization. Studies in the nonhuman primate have involved monkeys and baboons. A surgical method was used to flush the uterus of regularly cycling rhesus monkeys, and 9 preimplantation embryos and 2 unfertilized eggs were recovered from 22 flushes. Two of the nine embryos were morphologically abnormal and probably would not have implanted.[117] Hendrickx and Kraemer used a similar technique in the baboon and recovered 23 embryos, of which 10 were morphologically abnormal.[118] This suggests that, in nonhuman primates, some ovulated eggs are not fertilized and that many early embryos are abnormal and, in all likelihood, will be aborted. Similar findings have been reported in the human in the classic study of Hertig et al.[119] They examined 34 early embryos recovered by flushing and examination of reproductive organs removed at surgery. Ten of these embryos were morphologically abnormal, including 4 of the 8 preimplantation embryos. Because the 4 preimplantation losses would not have been recognized clinically, there would have been 6 losses recorded in the remaining 30 pregnancies.

By using sensitive pregnancy tests, it has been suggested that the total rate of pregnancy loss after implantation is approximately 30%.[120] When the loss of fertilized oocytes before implantation is included, approximately 46% of all pregnancies end before the pregnancy is clinically perceived.[121]

In the postimplantation period, if only clinically diagnosed pregnancies are considered, the generally accepted figure for spontaneous miscarriage in the first trimester in young women is 15%. Approximately 50–60% of these abortuses have chromosome abnormalities.[122] This suggests that a minimum of 7.5% of all human conceptions are chromosomally abnormal. The fact that only 1 in 200 newborns has a chromosome abnormality attests to the powerful selection mechanisms operating in early human gestation. In each ovulatory cycle, only 20–30% of normally fertile couples can achieve a pregnancy.[123] Once conception is achieved, only 30% survive to birth.[64]

Key Steps in Fertilization

1. Sperm penetration of the zona pellucida depends on a combination of sperm motility, an acrosomal proteinase, and binding of sperm head receptors to zona ligands.

2. Binding of sperm head receptors and zona ligands produces an enzyme complex that induces the acrosome reaction, releasing enzymes essential for the fusion of the sperm and oocyte membranes.

3. Fusion of the sperm and oocyte membranes triggers the cortical reaction, the release of substances from the cortical granules, organelles just below the egg cell membrane.

4. The cortical reaction leads to the enzyme-induced zona reaction, the hardening of the zona and the inactivation of ligands for sperm receptors, producing an obstacle to polyspermy.

5. Cell division begins promptly after fertilization; human gene expression begins between the 4- and 8-cell stages.

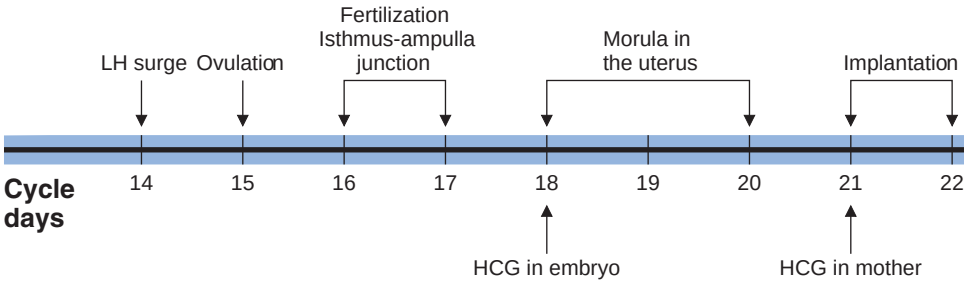
Implantation and Placentation

A normal pregnancy is, of course, impossible without successful implantation and placentation. Because there are differences among the various species, we will focus on the physical and biochemical events that are relevant in human reproduction.[124, 125] Shortly after the 8-cell morula enters the uterine cavity about 4 days after the gonadotropin surge and 3 days after ovulation, a blastocyst (a preimplantation embryo of varying cell number, from 32 to 256) is formed. Implantation (the embedding of the blastocyst in the endometrial stroma) begins with the loss of the zona pellucida (hatching) about 1–3 days after the morula enters the uterine cavity.

Preparation for Implantation

The change from proliferative to secretory endometrium, described in detail in Chapter 4, is an essential part of achieving the receptive conditions required for implantation. The primary endocrine requirement is the presence of progesterone; in the monkey, implantation and pregnancy can be achieved in the absence of luteal phase estrogen.[126] This change is the histologic expression of many biochemical and molecular events. The endometrium is 10–14 mm thick at the time of implantation in the midluteal phase. By this time, secretory activity has reached a peak, and the endometrial cells are rich in glycogen and lipids. Indeed, nourishment of the human fetus is dependent on contributions from the endometrial glands until the end of the first trimester of pregnancy, when a high level of maternal blood flow is reached within the placenta.[127]

Understanding the dynamic endocrine behavior of the endometrium (Chapter 4) increases the appreciation for its active participation in the implantation process. The window of endometrial receptivity is restricted to days 16–22 of a 28-day normal cycle (5 to 10 days after the LH surge), and days 16–19 of cycles stimulated by exogenous gonadotropins.[59, 128–130] The harmonious synchronization of a large cast of biochemical and molecular players is a complex achievement required for normal implantation. It is not surprising that gene expression studies have begun to identify the presence of dysregulated endometrial genes in women with repeated implantation failures.[131] The maximal chance of a normal implantation is only about 40% per cycle under optimal conditions.[132]



Endometrial receptivity is heralded by the progesterone-induced formation of ***pinopodes (also called uterodomes)***, surface epithelial cells that lose their microvilli and develop smooth protrusions, appearing and regressing during the window of receptivity.[133] The pinopodes may serve to absorb fluid from the uterine cavity forcing the blastocyst to be in contact with the endometrial epithelium. Blastocysts adhere at sites with pinopodes where the cell surface loses its nonadhesive character.[134] The most critical feature of the pinopode is the removal of adhesion-inhibiting mucin during the window of implantation.[135] Pinopodes appear around day 21 and are present only for a few days during implantation, following the peak in progesterone levels and marked by a decrease in progesterone receptor B in the endometrium.[136] However, this limited appearance of pinopodes is controversial as others have described their appearance throughout the luteal phase and into pregnancy.[137]

Even before the blastocyst adheres to the surface epithelium, but after hatching from the zona pellucida, a dialogue between the mother and the early embryo has begun. Early pregnancy factor (EPF) can be detected in the maternal circulation within 1–2 days after fertilization.[138] EPF prior to implantation is apparently produced by the ovary in response to a signal from the embryo. After implantation, EPF is no longer secreted by the ovary but is derived from the embryo. EPF has immunosuppressive properties and is associated with cell proliferation and growth. Indeed, there is reason to believe that endometrial receptivity for implantation requires appropriate signals from the embryo. One such signal is human chorionic gonadotropin.

Blastocysts grown in culture produce and secrete human chorionic gonadotropin (hCG), beginning days 7–8 after fertilization.[139] Messenger RNA for hCG can be found in 6- to 8-cell human embryos.[140] Because the 8- to 16-cell stage is achieved about 3 days after fertilization, it is believed that the human embryo begins to produce hCG before implantation when it can be detected in the mother (about 6–7 days after ovulation). The embryo is capable, therefore, of preimplantation signaling, and higher levels of estradiol and progesterone can be measured in the maternal circulation even before maternal hCG is detectable, presumably because of stimulation of the corpus luteum by hCG delivered directly from the uterine cavity to the ovary.[141] Function of the corpus luteum is crucial during the first 7–9 weeks of pregnancy, and luteectomy early in pregnancy can precipitate abortion.[142] Similarly, early pregnancy loss in primates can be induced by injections of anti-hCG serum.[143] Another substance secreted very early by the preimplantation embryo is platelet-activating factor, perhaps part of the immunosuppressive activity required to induce maternal tolerance of the embryo. In the rabbit, platelet-activating factor also induces the production of early pregnancy factor.[144] Indeed, many growth factors are produced by the early embryo.[145]

In rodents and rabbits, implantation can be interrupted by injection of prostaglandin inhibitors.[146, 147] Indomethacin prevents the increase in endometrial vascular permeability normally seen just prior to implantation. Additional evidence for a role by prostaglandins in the earliest stages of implantation is the finding of increased concentrations of prostaglandins at implantation sites, similar to any inflammatory response.[148] The blastocysts of mice, rabbits, sheep, and cows produce prostaglandins, and prostaglandin E_2 release from human blastocysts and embryos has been demonstrated.[149]

The secretory endometrial epithelial cells are also a source of prostaglandin E_2 (but not prostaglandin $F_{2\alpha}$), and its synthesis may be stimulated by the tissue response that accompanies implantation. However, decidual synthesis of prostaglandins is significantly reduced compared with proliferative and secretory endometrium, apparently a direct effect of progesterone activity and perhaps a requirement in order to maintain the pregnancy.[148] Nevertheless, prostaglandin E_2 synthesis is increased at the implantation site, perhaps in response to blastocyst factors, e.g., platelet-activating factor, and correlates with an increase in vascular

permeability.[148, 150] It is now well-accepted that decidua-derived prostaglandin E_2 is one of the major regulators of trophoblastic invasion, activating other signaling proteins.[151]

As discussed in Chapter 4, the many cytokines, peptides, and lipids secreted by the endometrium are interrelated through the stimulating and inhibiting actions of estrogen and progesterone, as well as the autocrine/paracrine activities of these substances on each other. The response to implantation certainly involves the many members of the growth factor and cytokine families.

Angiogenesis, the growth of blood vessels from pre-existing vessels, is a key feature of the endometrial cycle and implantation. This process is regulated indirectly by the sex steroids and directly by growth factors, especially members of the fibroblast growth factor family, the angiopoietins and the vascular endothelial cell growth factor (VEFG) family. There are at least five VEGF isoforms and four receptors. Two angiopoietins, Ang-1 and Ang-2, share a common tyrosine kinase receptor, Tie2. VEGF-A seems critical for vascular growth and is up-regulated in the presence of reduced oxygen. The angiopoietins also promote the growth of blood vessels, and act synergistically with VEGF. Controlled growth as well as appropriate regression reflect the balance between an ever-increasing number of stimulatory and inhibitory factors discovered by scientists in this field.[152, 153]

Lipids	*Cytokines*	*Growth Factors*
Prostaglandins	Interleukin-1α	Epidermal growth factor family
Thromboxanes	Interleukin-1β	EGF
	Interleukin-6	Heparin-binding EGF
	Interleukin-11	TGF-α
	Colony-stimulating factor-1	Insulin-like growth factor family
	Tumor necrosis factor-α	IGF-I
	Leukemia-inhibiting factor	IGF-II
	Interferon-γ	IGFBPs 1–6
		Platelet-derived growth factor
		Transforming growth factor-β
		Fibroblast growth factors
		Vascular endothelial growth factors
		Angiopoietins

Implantation

Implantation is defined as the process by which an embryo attaches to the uterine wall and penetrates first the epithelium and then the circulatory system of the mother to form the placenta. The embryo completely invades the endometrium only in great apes and humans. Implantation is a process that is limited in both time and space, begining 2–3 days after the fertilized egg enters the uterus usually on day 18 or 19 of the cycle (3 or 4 days after ovulation).[130] Thus, implantation occurs 5–7 days after fertilization. A careful study of women attempting to conceive documented that the first hormonal evidence of implantation (the appearance of hCG) occurred on 8, 9, or 10 days after ovulation; the earliest was 6 days and the latest 12 days.[154] The risk of spontaneous early miscarriage markedly increases with late implantations (later than 9 days after ovulation). Implantation consists of three stages: apposition, adhesion, and invasion (also called migration to denote its benign nature).

Apposition and Adhesion

The human blastocyst remains in the uterine secretions for approximately 1 to 3 days and then hatches from its zona pellucida in preparation for attachment. The implantation site in the human uterus is usually in the upper, posterior wall in the midsagittal plane. Implantation is marked initially by apposition of the blastocyst to the uterine epithelium, usually about 2–4 days after the morula enters the uterine cavity. A prerequisite for this contact is a loss of the zona pellucida, which, in vitro, can be ruptured by contractions and expansions of the blastocyst. In vivo, this activity is less critical, because the zona can be lysed by components of the uterine fluid. Nevertheless, blastocyst movement and escape from the zona pellucida appear to involve cytoplasmic projections (this leads to penetrations of the zona by the trophectoderm prior to zona hatching).[155] By this time, the blastocyst has differentiated into an inner cell mass (embryo) and trophectoderm (placenta), both essential for implantation.

The endometrium produces at least three cytokines involved in implantation.[156] These are colony-stimulating factor-1 (CSF-1), leukemia-inhibitory factor (LIF), and interleukin-1 (IL-1). CSF-1 expression and receptors for CSF-1 are found in both the human endometrium (peaking in decidua) and the preimplantation embryo. Mice with an inactivating mutation in the CSF-1 gene are infertile because of low rates of implantation and fetal viability.[157] LIF displays the same pattern of expression as CSF-1, and mice with an LIF gene mutation have a failure of blastocyst implantation.[158, 159] Blocking the interleukin-1 receptor in mice also prevents implantation.[156] Interleukin-1 stimulates hCG release from human trophoblast cells, and in the endometrium increases VEGF expression and regulates the tissue inhibitor of metalloproteinases; GnRH is produced in the human blastocyst and stimulates endometrial expression of interleukin-1.[160] Perhaps the first maternal change in the implantation process, increased permeability of the capillaries near the adherent blastocyst, is due to a blastocyst-directed change in heparin-binding epidermal growth factor (HB-EGF) expression in the surface epithelium.[161] In addition, the blastocyst contains receptors for epidermal growth factor that respond to HB-EGF and promote growth and zona hatching.

The adhesion process further involves a collection of adhesion molecules, including integrins, selectins, and trophinin.[162] The decidualized endometrium and the early embryo express extracellular matrix components, especially laminin and fibronectin, which mediate cell adhesion by binding to the adhesion molecules.[163] Cells are fixed and supported by the extracellular matrix utilizing components such as laminin and fibronectin with attachments to these components via cell surface receptors, especially the integrins. An increase in specific isoforms of laminin in decidua at the time of implantation suggests an important interaction with the invading trophoblast.[164] Thus, implantation starts with adhesion due to binding with endometrial proteins, followed by invasion (migration) of the trophoblast by proteinase degradation of the extracellular matrix.

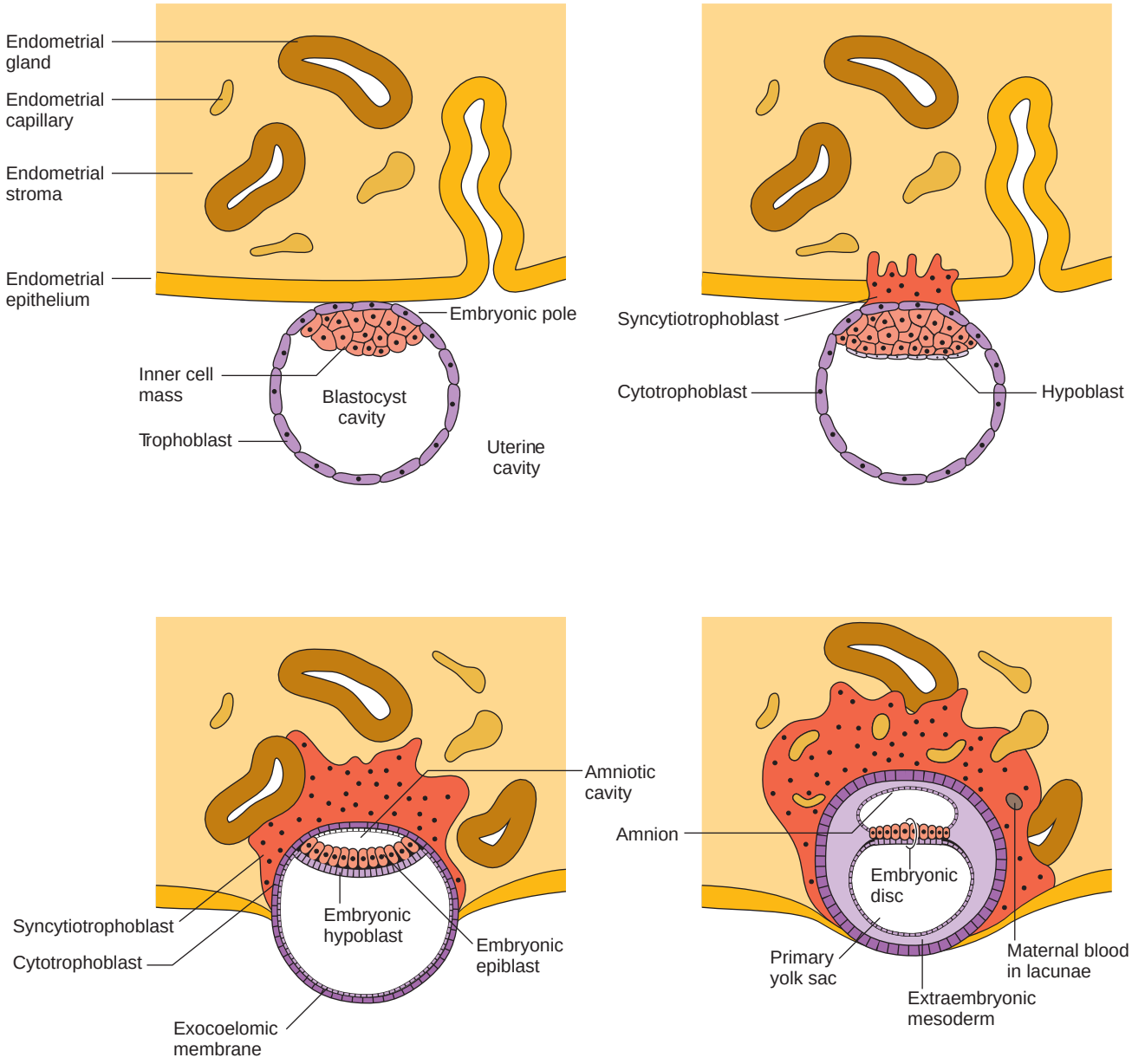
Integrins are members of a family of transmembrane cell surface receptors for collagen, fibronectin, and laminin. Integrins are utilized in cell-cell and cell-matrix interactions, contributing to cell migration, cell differentiation, and tissue structure. A cyclic change in integrin expression in the endometrial epithelial cells indicates peak expression at the time of implantation.[165] It has been suggested that a lack of integrin expression during the implantation window can be a cause of infertility.[166] The blastocyst also expresses integrins in a time sequence and at a site (outgrowing trophoblast cells) that are appropriate for key activity during implantation.[167] The integrins are a diverse collection of receptors, reflecting various combinations of the α and β subunits in the structure of the receptor, as well as variations in the cytoplasmic domain in the subunits. Stimulation and inhibition of cellular proliferation in the endometrium and decidua are influenced by specific expression of the appropriate subunits.[168] Mechanisms that control variation in the structure through splicing would account for the expression of an integrin variant appropriate for proliferation

early in the endometrial cycle and prevention of proliferation in the decidua, and perhaps prevention of trophoblastic invasion.

Ephrins are peptides that bind to tyrosine kinase cell membrane receptors. Ephrin expression can be detected in both endometrial epithelial cells and in blastocysts.[169] This is another system for the cell-to-cell communication involved in trophoblastic migration.

The process of tissue disruption is accompanied by an increase in lymphocytes, another source for cytokines and growth factors in addition to trophoblast and endometrial cells. The distinction between cytokines and growth factors is not always clear, but T lymphocytes and macrophages are significant secretors of cytokines.

In general, cytokines, growth factors, and their receptors have been identified in virtually all tissues associated with implantation. The cataloging is lengthy and often confusing.[156, 170] It is helpful to simply view these various substances as the biochemical tools by which the physical process of trophoblast adhesion and invasion is accomplished.



Even if the hormonal milieu and protein composition of the uterine fluid are hospitable to the implantation, it may not occur if the embryo is not at the proper stage of development. It has been inferred from this information that there must be developmental maturation of the surface of the embryo before it is able to achieve attachment and implantation.

Reports of changes in the surface charge of preimplantation embryos differ in their findings, and it is unlikely that changes in surface charge are solely responsible for adherence of the blastocyst to the surface of epithelial cells. Binding of the lectin concanavalin A to the blastocyst changes during the preimplantation period, an indication that the surface glycoproteins of the blastocyst are in transition.[171] It is reasonable to assume that these changes in configuration on the surface occur in order to enhance the ability of the early embryo to adhere to the maternal surface.

As the blastocyst comes into close contact with the endometrium, the microvilli on the surface flatten and interdigitate with those on the luminal surface of the epithelial cells. A stage is reached where the cell membranes are in very close contact and junctional complexes are formed. The early embryo can no longer be dislodged from the surface of the epithelial cells by flushing the uterus with physiologic solutions.

Eventually, the characterization of an endometrium that is normally receptive to implantation will allow medical interventions. It is not far-fetched to consider therapeutic manipulations that will either improve implantation rates or provide contraception.

Invasion and Placentation

In the second week after ovulation, the placenta is formed.[163] By this time, the trophoblasts at the implantation site have formed masses of cytotrophoblasts and syncytiotrophoblasts, and invasion of maternal blood vessels has begun. The walls of the spiral arteries are destroyed, as sinusoidal sacs are formed lined with endovascular trophoblast. The fundamental change is a replacement of maternal vascular cells with cytotrophoblast cells. The purpose of placental invasion is to remodel the uterine vasculature, establishing a structure that will allow and maintain a high blood-flow interchange between mother and fetus, replacing small, high-resistance vessels with large, low-resistance vessels. ***The invading placental cells are special trophoblast cells, known as the extravillous trophoblast, that arise by proliferation and differentiation of cytotrophoblast stem cells within the chorionic villi.***

Three types of interactions between the implanting trophoblast and the uterine epithelium have been described.[172] First, trophoblast cells intrude between uterine epithelial cells on their path to the basement membrane. In the second type of interaction, the epithelial cells lift off the basement membrane, an action that allows the trophoblast to insinuate itself underneath the epithelium. Last, fusion of the trophoblast with individual uterine epithelial cells has been identified by electron microscopy in the rabbit.[173] This latter method of gaining entry into the epithelial layer raises interesting questions about the immunologic consequences of mixing embryonic and maternal cytoplasm.

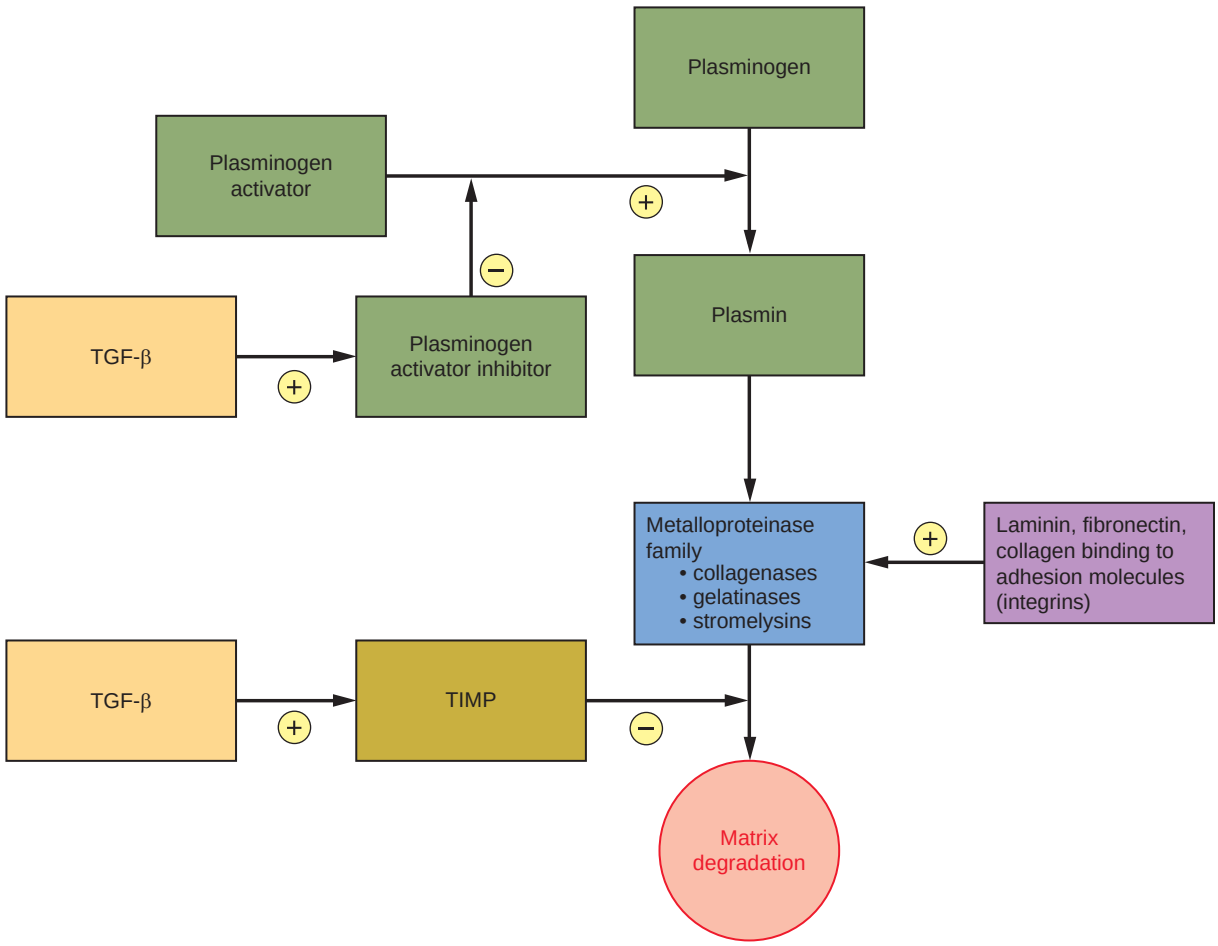
Trophoblast has the ability to phagocytose a variety of cells, but, in vivo, this activity seems largely confined to removal of dead endometrial cells, or cells that have been sloughed from the uterine wall. Similarly, despite the invasive nature of the trophoblast, destruction of maternal cells by enzymes secreted by the embryo does not play a major role in implantation; there is virtually no necrosis. The early embryo does secrete a variety of enzymes (e.g., collagenase and plasminogen activators), and these are important for digesting the intercellular matrix that holds the epithelial cells together. Studies in vitro have demonstrated the presence of plasminogen activator in mouse embryos and in human trophoblast, and its activity is important in the attachment and early outgrowth stages of implantation.[174, 175]

Urokinase and proteases, trophoblastic enzymes that convert plasminogen to plasmin, are inhibited by hCG, indicating regulation of this process by the embryo.[176]

The trophoblast at a somewhat later stage of implantation can digest, in vitro, a complex matrix composed of glycoproteins, elastin, and collagen, all of which are components of the normal intercellular matrix.[177, 178] Studies in vitro indicate that cells move away from trophoblast in a process called "contact inhibition."[179] Trophoblast then spreads to fill the spaces vacated by the cocultured cells. Once the extracellular matrix has been lysed, this movement of epithelial cells away from trophoblast would allow space for the implanting embryo to move through the epithelial layer. Trophoblast movement is aided by the fact that only parts of its surface are adhesive, and the major portion of the surface is nonadhesive to other cells.

The highly proliferative phase and migration of trophoblastic tissue during early embryogenesis are regulated by the many growth factors and cytokines produced in both fetal and maternal tissues.[180] VEGF is important for the growth of new blood vessels, and the angiopoietins recruit perivascular cells to provide vascular stability. The interaction of VEGF and angiopoietins is important in the remodeling of maternal vessels necessary to develop the uteroplacental circulation. Another signal from the fetus to induce maternal blood vessels to grow is hCG, which of course is available even before implantation to bind to its receptor in the endometrium and stimulate vessels directly as well as the expression of angiogenic factors, such as VEGF.[181, 182]

Invasion of the early trophoblast requires the expression of integrins, stimulated by trophoblast-derived insulin-like growth factor-II and decidua-derived IGF-binding



protein-1, and inhibited by decidua-derived transforming growth factor-β.[183, 184] Actively migrating trophoblast cells have a different integrin profile than nonmigrating cells, specifically cell surface receptors that preferentially bind laminin.[163, 185] The controlling mechanism (not yet known) for this change in integrin expression must be a key regulator of trophoblast invasion. The specific nature of integrin expression can determine binding to matrix components, a requirement for migration.

Integrin cell surface binding for the matrix components can be also regulated by activating and inactivating the integrins. This would allow trophoblast cells to alternate between adhesive and nonadhesive states, thus achieving directional cell migration.[186] The role of integrin cell surface receptors is not simply to bind to a structural component. Binding activates cellular signaling pathways (similar to the classic endocrine tropic hormone-cell membrane receptor pathway) that activate enzymes that ultimately produce adhesion as well as cellular gene transcription.[163] IGF-binding protein-1 can stimulate trophoblastic cell migration independently of the IGF system by binding to an integrin receptor and activating kinase pathways.[187]

The uterine spiral arterioles are invaded by cytotrophoblasts, and the maternal endothelium is replaced by cytotrophoblast tissue as far as the first third of the myometrium. The maternal vascular invasion by trophoblast cells and replacement of vascular endothelium with endovascular trophoblast may utilize a different class of surface molecules, the selectin family.[163] The selectins have been demonstrated to be present in decidual vascular endothelial cells, but only at the site of implantation. The selectins are responsive to inflammatory mediators, including cytokines. As the trophoblast cells replace maternal endothelium, the receptor profile for adhesion peptides of the trophoblast changes to resemble endothelial cells.[188] ***It has been long recognized that this invasion process is limited in pregnancies with preeclampsia, and this is the fundamental cause of the poor placental perfusion associated with preeclampsia and intrauterine growth retardation***. The relative failure in this process in preeclampsia is characterized by insufficient conversion to endothelial adhesion receptors as well as low levels of IGF-binding protein-1 and matrix metalloproteinases.[189–191]

The matrix metalloproteinases, significantly involved in the process of menstruation (Chapter 4), are also key players in matrix degradation during trophoblast invasion. The metalloproteinases include collagenases, gelatinases, and the stromelysins. Integrin-mediated adhesion can activate this family of proteolytic enzymes, which then accomplish the degradation of matrix proteins that is necessary in order for trophoblast migration to take place. Production of the metalloproteinases is regulated by the combined actions of plasminogen activators, cytokines, and tissue inhibitors (TIMPs). Early trophoblastic invasion is enhanced by trophoblast-derived GnRH that suppresses the expression of TIMPs, the inhibitors of matrix metalloproteinases.[192]

Further penetration and survival depend on factors that are capable of suppressing the maternal immune response to fetal antigens. The endometrial tissue makes a significant contribution to growth factor activity and immune suppression by synthesizing proteins in response to the blastocyst even before implantation.[193, 194] One of the great mysteries associated with implantation is the mechanism by which the mother rejects a genetically abnormal embryo or fetus. It is possible that the abnormal embryo cannot produce a signal in early pregnancy that can be recognized by the mother. This may be a failure of the trophoblast to produce the proteins required to alter the immune environment of the decidua in order to tolerate the process of implantation.

The embryonic signals will be effective only in a proper hormone milieu. Much of the knowledge concerning the hormone requirements for implantation in animals has been gained from studies of animals with delayed implantation. In a number of species, preimplantation embryos normally lie dormant in the uterus for periods of time, which may

extend for as long as 15 months before implantation is initiated. In other species, delayed implantation can be imposed by postpartum suckling or by performing ovariectomy on day 3 of pregnancy. This produces a marked decrease in synthesis of DNA and protein by the blastocyst. The embryo can be maintained at the blastocyst stage by injecting the mother with progesterone. Using this model, hormonal requirements for implantation have been determined. In mice, there is a requirement for estrogen and progesterone. In other species, including the primate, the nidatory stimulus of estrogen is not required, and progesterone alone is sufficient.[126] However, genomic profiling indicates that the presence of some estrogen facilitates the progesterone-induced gene expression associated with implantation.[195]

Although it is known that the hormone milieu of delayed implantation renders the embryo quiescent, it is not known whether this represents a direct effect on the embryo or whether there is a metabolic inhibitor present in uterine secretions that acts on the embryo. Removal of the embryo from the uterus to culture dishes allows rapid resumption of normal metabolism, suggesting that in fact, there has been a release from the inhibitory effects of a uterine product.

Limitation of Invasion

Unlike the tissue invasion associated with cancer, trophoblast invasion must be limited, confining the placenta to its intrauterine location and within the time constraint of a pregnancy.

Invasion of the endometrial stromal compartment, breaching of the basement membrane, and penetration of maternal blood vessels are mediated by serine proteases and metalloproteinases. The serine proteases are plasminogen activators that provide plasmin for proteolytic degradation of the extracellular matrix and activation of the metalloproteinase family. Trophoblast cells contain plasminogen activator receptors. Binding of plasminogen activator to this receptor is believed to be a method by which plasmin proteolysis is exerted in a controlled and limited site.[196]

Many components of the inflammatory response play roles in the process of implantation. Cytokine secretion from the lymphocyte infiltrate in the endometrium activates cellular lysis of trophoblast, perhaps an important process in limiting invasion.[197] The decidua at the time of implantation contains a large number of natural killer cells (large granular lymphocytes). It has been proposed that an interaction between these cells and a human leukocyte antigen uniquely present in invading trophoblast limits invasion by producing appropriate cytokines.[198]

Invasion by the trophoblast is limited by the formation of the decidual cell layer in the uterus. Fibroblast-like cells in the stroma are transformed into glycogen and lipid-rich cells. In the human, decidual cells surround blood vessels late in the nonpregnant cycle, but extensive decidualization does not occur until pregnancy is established. Ovarian steroids govern decidualization, and in the human a combination of estrogen and progesterone is critical. Tissue factor (TF), expressed in decidualized endometrium, is a receptor for factor VII and its active form, VIIa. Concentrated in perivascular sites, TF forms what Lockwood calls a “hemostatic envelope,” to promote hemostasis by providing fibrin.[199, 200] TF is appropriately positioned to counteract the threat of hemorrhage associated with trophoblastic invasion.

Limitation of trophoblastic invasion is attributed to the balance of promoting and restraining growth factors, cytokines, and enzymes. Plasminogen activator inhibitor-1 (PAI-1) is a major product of decidual cells, inhibiting excessive bleeding during menses and restraining

trophoblast invasion in early pregnancy.[199, 201] PAI-1 binds plasminogen activator with a high affinity and is regulated by cytokines and growth factors. The metalloproteinases that degrade the extracellular matrix components, such as collagens, gelatins, fibronectin, and laminin, are restrained by tissue inhibitors of metalloproteinases (TIMPs). In addition, metalloproteinase degradation can be suppressed by inhibiting trophoblast production of these enzymes and by preventing conversion from an inactive to an active form.[202] Decidual TGF-β is a key growth factor involved in limitation of trophoblast invasion by inducing the expression of both TIMP and PAI-1. In addition, TGF-β can inhibit integrin expression and influence cytotrophoblasts to differentiate into noninvasive syncytiotrophoblasts.[183, 203] Decorin is a decidua-derived binding proteoglycan that can bind TGF-β, storing it for when it is needed to limit invasion when it is freed and activated by plasmin. In addition, decorin exerts antiproliferative, antimigratory, and antiinvasive effects on trophopblast independently of TGF-β.[204] Even human chorionic gonadotropin (hCG) may exert a governing force by inhibiting protease activity.[176, 205]

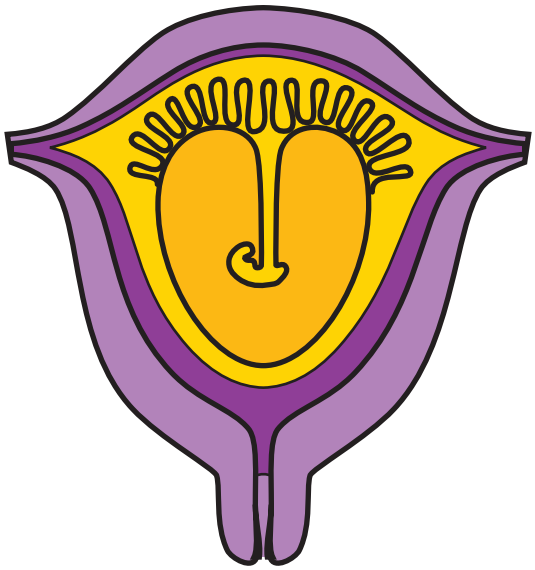
In the baboon, the lower estrogen levels in early pregnancy permit trophoblast invasion, but the increasing estrogen levels later in pregnancy suppress maternal spiral artery invasion, an effect mediated by estrogen-induced decreases in cytotrophoblast VEGF expression.[206, 207]

Key Steps in Implantation

1. The early embryo enters the uterine cavity as an 8-cell morula and becomes a 32 to 256-cell blastocyst before implantation.
2. Implantation begins with hatching from the zona pellucida about 1–3 days after the morula entered the uterine cavity.
3. The endometrium is prepared for implantation by the complex activity of cytokines, growth factors, and lipids modulated by the sex hormones, especially progesterone. The endometrium is receptive for implantation for only a few days.
4. The process of implantation begins with apposition and adhesion of the blastocyst to the uterine epithelium, about 2–4 days after the morula enters the uterine cavity. This process is mediated by cytokines and involves adhesion molecules (integrins) that interact with extracellular components, especially laminin and fibronectin.
5. Trophoblastic invasion rapidly follows adhesion of the blastocyst, mediated by proteinase degradation of the extracellular matrix. The placenta is formed in the second week after ovulation. Limitation of trophoblastic invasion is due to a restraint imposed by proteinase inhibitors, especially plasminogen activator inhibitor and tissue inhibitors of metalloproteinases.

All references are available online at: http://www.clinicalgynendoandinfertility.com

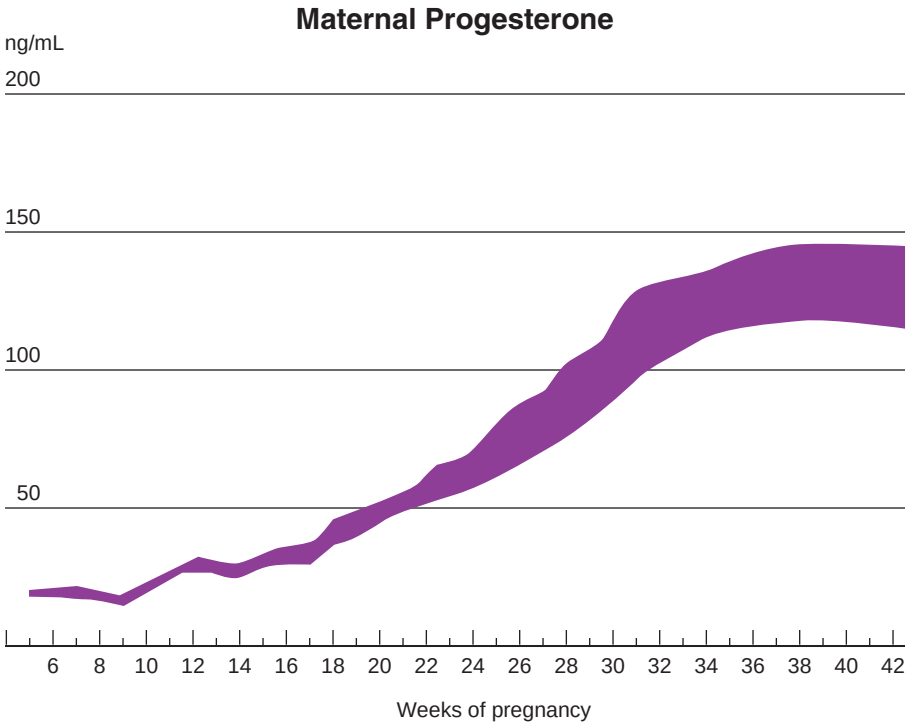
8 The Endocrinology of Pregnancy



Who is in charge of pregnancy, the mother or her fetus? From the vantage point of an outsider looking in, it seems as if the mother is in charge. But from the fetal point of view, it is overwhelmingly logical that the maternal adaptations of pregnancy are controlled by the fetus. For the fetus, one of the crucial aspects of intrauterine life is its dependency on the effective exchange of nutritive and metabolic products with the mother. It is logical that mechanisms exist by which a growing fetus can influence or control the exchange process and, hence, its environment. The methods by which a fetus can influence its own growth and development involve a variety of messages transmitted, in many cases, by hormones. Hormonal messengers from the conceptus can affect metabolic processes, uteroplacental blood flow, and cellular differentiation. Furthermore, a fetus may signal its desire and readiness to leave the uterus by hormonal initiation of parturition. This chapter reviews the mechanisms by which the fetus establishes influence over important events during pregnancy. The important process of lactation is discussed in Chapter 16.

Steroid Hormones in Pregnancy

Steroidogenesis in the fetoplacental unit does not follow the conventional mechanisms of hormone production within a single organ. Instead, the final products result from critical interactions and interdependence of separate organ systems that individually do not possess the necessary enzymatic capabilities. It is helpful to view the process as consisting of a fetal compartment, a placental compartment (specifically the syncytiotrophoblast),



and a maternal compartment. Separately, the fetal and placental compartments lack certain steroidogenic activities. Together, however, they are complementary and form a complete unit that utilizes the maternal compartment as a source of basic building materials and as a resource for clearance of steroids.

Progesterone

In its key location as a way station between mother and fetus, the placenta can use precursors from either mother or fetus to circumvent its own deficiencies in enzyme activity. The placenta converts little, if any, acetate to cholesterol or its precursors. Cholesterol and pregnenolone are obtained from the maternal bloodstream for progesterone synthesis. The fetal contribution is negligible because progesterone levels remain high after fetal demise. Thus, the massive amount of progesterone produced in pregnancy depends on placental-maternal cooperation, although some have argued that the fetal liver is an important source of cholesterol (discussed later).

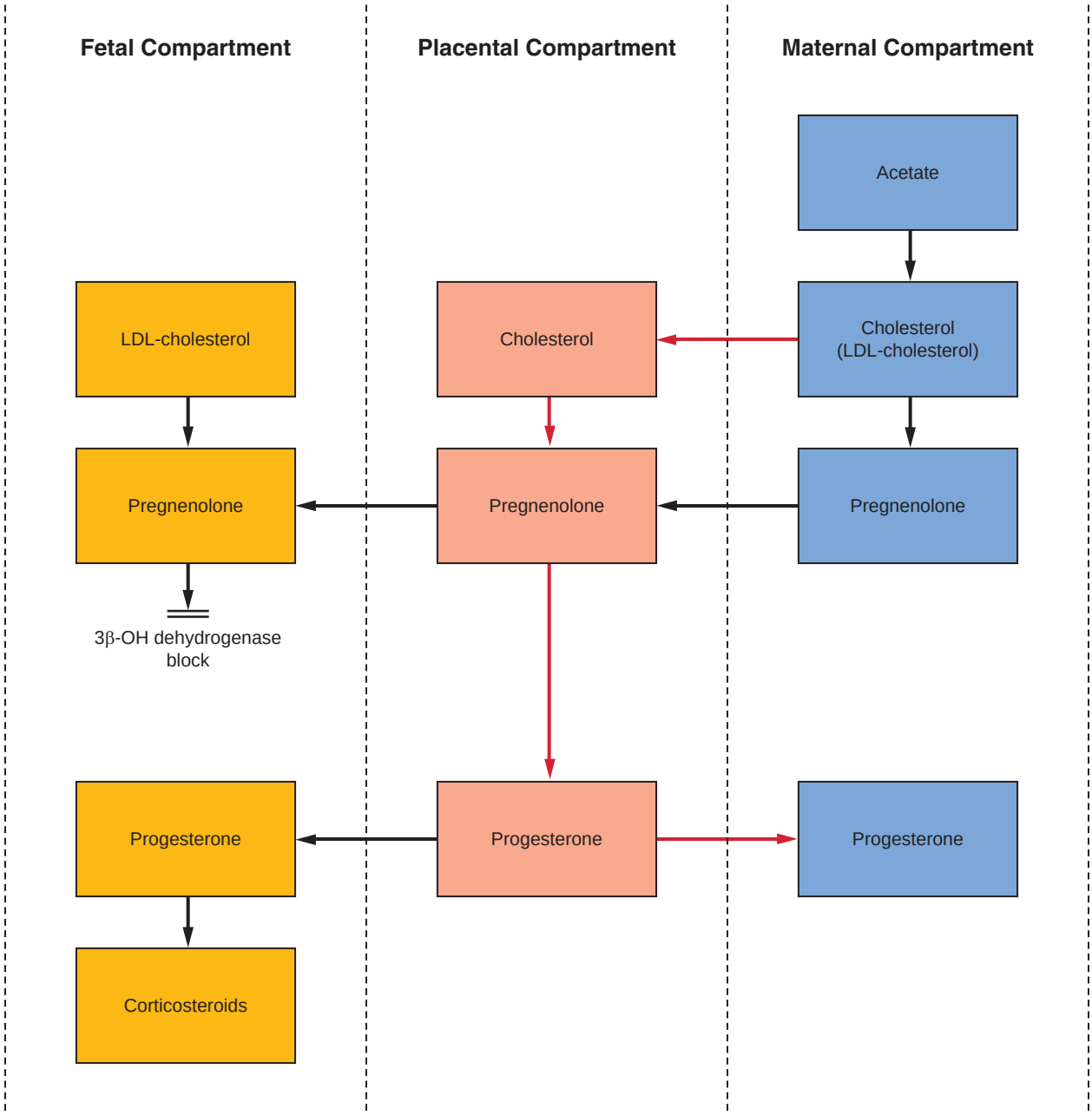
Progesterone is largely produced by the corpus luteum until about 10 weeks of gestation. Indeed, until approximately the seventh week, the pregnancy is dependent on the presence of the corpus luteum.[1] Exogenous support for an early pregnancy (until 10 weeks) requires 100 mg of progesterone daily, associated with a maternal circulating level of approximately 10 ng/mL.[2] Despite this requirement, patients pregnant after ovarian stimulation with one of the techniques of assisted reproductive technology have concluded a successful pregnancy after experiencing extremely low progesterone levels.[3,4] Thus, individual variation is great, and very low circulating levels of progesterone can be encountered occasionally in women who experience normal pregnancies. The predictive value, therefore, of progesterone measurements is limited.

After a transition period of shared function between the seventh week and tenth week, during which there is a slight decline in circulating maternal progesterone levels, the placenta

emerges as the major source of progesterone synthesis, and maternal circulating levels progressively increase.[2, 5, 6] At term, progesterone levels range from 100 to 200 ng/mL, and the placenta produces about 250 mg/day. Most of the progesterone produced in the placenta enters the maternal circulation.

In contrast to estrogen, progesterone production by the placenta is largely independent of the quantity of precursor available, the uteroplacental perfusion, fetal well being, or even the presence of a live fetus. This is because the fetus contributes essentially no precursor. The majority of placental progesterone is derived from maternal cholesterol that is readily available. At term a small portion (3%) is derived from maternal pregnenolone.

The cholesterol utilized for progesterone synthesis enters the trophoblast from the maternal bloodstream as low-density lipoprotein (LDL)-cholesterol, by means of the process of endocytosis (internalization, as described in Chapter 2) i nvolving the LDL cell membrane



receptors, a process enhanced in pregnancy by estrogen.[7, 8] Hydrolysis of the protein component of LDL may yield amino acids for the fetus, and essential fatty acids may be derived from hydrolysis of the cholesteryl esters. Unlike steroidogenesis elsewhere, it is not clear whether placental progesterone production requires the control of tropic hormones. Although some evidence suggests tropic hormone support is not necessary, other evidence indicates that a small amount of human chorionic gonadotropin (hCG) must be present.[9, 10]

There is evidence in the baboon that estrogen (estradiol) regulates progesterone production in the placenta.[11] The fetoplacental units in human and baboon pregnancies are virtually identical. Estradiol increases LDL-cholesterol uptake in baboon trophoblastic tissue by increasing LDL receptor gene transcription, and in human syncytiotrophoblast, estradiol increases progesterone production by means of an increase in LDL uptake.[11, 12] Estrogen also stimulates cholesterol production in the human fetal liver to provide circulating LDL-cholesterol substrate for steroidogenesis.[13] In addition, estrogen increases placental P450scc enzyme activity that converts cholesterol to pregnenolone, the immediate precursor for progesterone. Because estrogen production ultimately depends on the fetal adrenal gland for precursors, the influence of estrogen on progesterone production would be another example of fetal direction and control in the endocrinology of pregnancy. The proponents of this interaction and dependence of progesterone production on fetal precursors argue that the lack of impact by conditions of estrogen deficiency (e.g., anencephaly, fetal demise) on progesterone production is due to the fact that active, unbound estrogen remains within a critical, effective range, and what is lost reflects the degree of excess production in pregnancy.[11]

The human decidua and fetal membranes also synthesize and metabolize progesterone.[14] In this case, neither cholesterol nor LDL-cholesterol are significant substrates; pregnenolone sulfate may be the most important precursor. This local steroidogenesis may play a role in regulating parturition.

Amniotic fluid progesterone concentration is maximal between 10 and 20 weeks and then decreases gradually. Myometrial levels are about 3 times higher than maternal plasma levels in early pregnancy, remain high, and are about equal to the maternal plasma concentration at term.

In early pregnancy, the maternal levels of 17α-hydroxyprogesterone rise, marking the activity of the corpus luteum. By the tenth week of gestation, this compound has returned to baseline levels, indicating that the placenta has little 17α-hydroxylase activity. However, beginning about the 32nd week there is a second, more gradual rise in 17α-hydroxyprogesterone due to placental utilization of fetal precursors.

There are two active metabolites of progesterone that increase significantly during pregnancy. There is about a 10-fold increase of the 5α-reduced metabolite, 5α-pregnane-3-20-dione.[15] This compound contributes to the resistance in pregnancy against the vasopressor action of angiotensin II. The circulating level, however, is the same in normal and hypertensive pregnancies. The maternal blood concentration of deoxycorticosterone (DOC) at term is 1,200 times the nonpregnant levels. Some of this is due to the 3–4-fold increase in cortisol-binding globulin during pregnancy, but a significant amount is due to 21-hydroxylation of circulating progesterone in the kidney.[16] This activity is significant during pregnancy because the rate is proportional to the circulating concentration of progesterone. The fetal kidney is also active in 21-hydroxylation of the progesterone secreted by the placenta into the fetal circulation. Currently, there is no known physiologic role for DOC during pregnancy.

Progesterone has a role in parturition as discussed later in this chapter. It has been suggested that progesterone is also important in suppressing the maternal immunologic response to fetal antigens, thereby preventing maternal rejection of the trophoblast. And, of course, progesterone prepares and maintains the endometrium to allow implantation. The human corpus luteum makes significant amounts of estradiol, but it is progesterone and

not estrogen that is required for successful implantation.[17] Because implantation normally occurs about 5–6 days after ovulation, and human chorionic gonadotropin (hCG) must appear by the 10th day after ovulation to rescue the corpus luteum, the blastocyst must successfully implant and secrete hCG within a narrow window of time. In the first 5–6 weeks of pregnancy, hCG stimulation of the corpus luteum results in the daily secretion of about 25 mg of progesterone and 0.5 mg of estradiol. Although estrogen levels begin to increase at 4–5 weeks due to placental secretion, progesterone production by the placenta does not significantly increase until about 10–11 weeks after ovulation.

Progesterone serves as the substrate for fetal adrenal gland production of glucocorticoids and mineralocorticoids; however, cortisol synthesis is also derived from low-density lipoprotein cholesterol (LDL-cholesterol) synthesized in the fetal liver and obtained from the fetal circulation.[13, 18] The fetal zone in the adrenal gland is extremely active but produces steroids with a 3β-hydroxy-Δ^5 configuration like pregnenolone and dehydroepiandrosterone (DHEA), rather than 3-keto-Δ^4 products such as progesterone. The fetus, therefore, lacks significant activity of the 3β-hydroxysteroid dehydrogenase, Δ^{4-5} isomerase system. Thus, the fetus must borrow progesterone from the placenta to circumvent this lack in order to synthesize the biologically important corticosteroids. In return, the fetus supplies what the placenta lacks: 19-carbon compounds to serve as precursors for estrogens.

Steroid levels have been compared in maternal blood, fetal blood, and amniotic fluid obtained at fetoscopy in women undergoing termination of pregnancy at 16–20 weeks gestation.[19] Cortisol, corticosterone, and aldosterone are definitely secreted by the fetal adrenal gland independently of the mother. The fetal arterial-venous differences confirm that placental progesterone is a source for fetal adrenal cortisol and aldosterone.

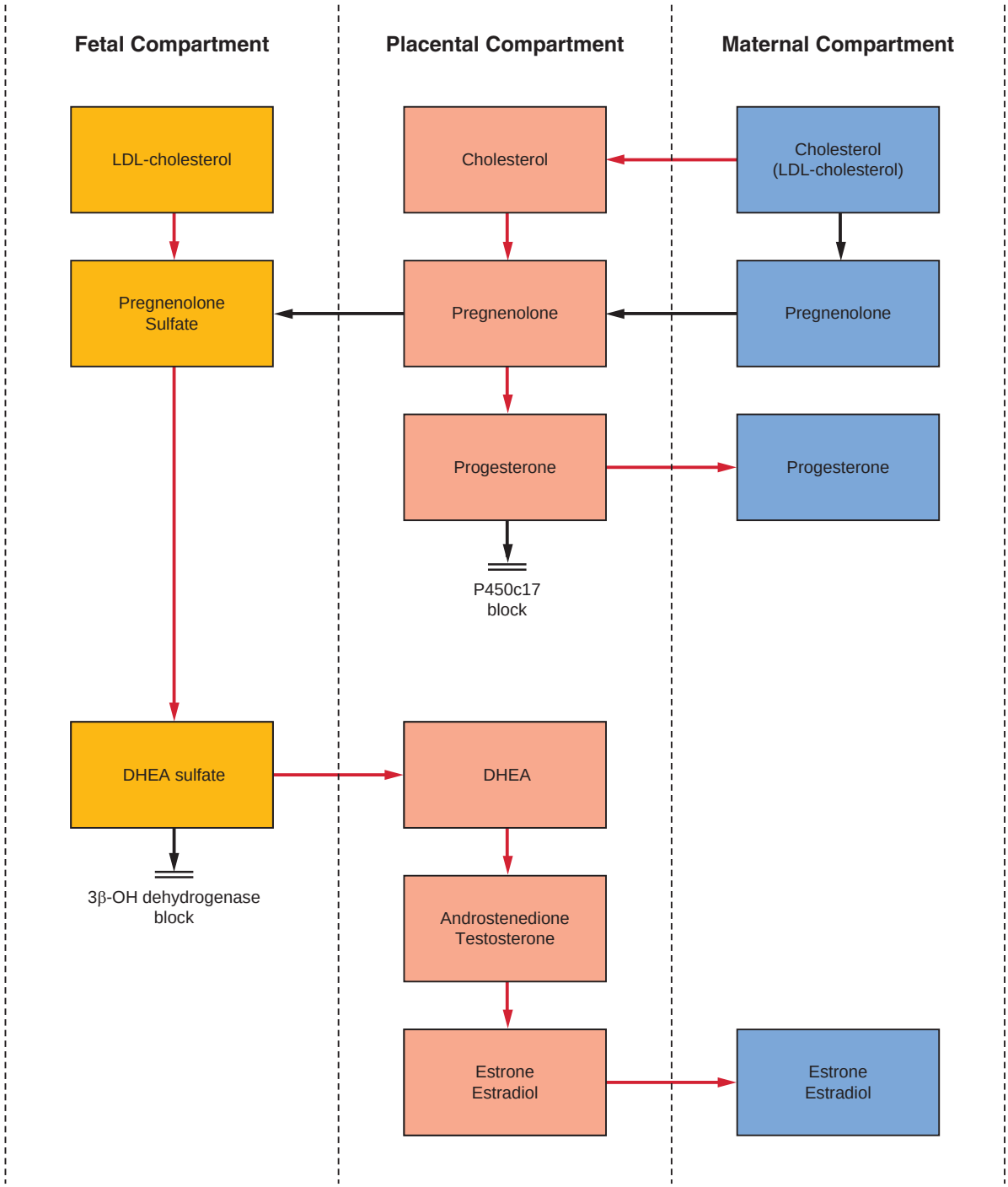
Estrogens

Estrogen production in pregnancy is under the control of the fetus and is a fundamental signaling method by which the fetus directs important physiologic processes that affect fetal well-being. Estrogen influences progesterone production, uteroplacental blood flow, mammary gland development, and fetal adrenal gland function.[11]

The basic precursors of estrogens are 19-carbon androgens. However, there is a virtual absence of 17α-hydroxylation and 17–20 desmolase (lyase) activity (P450c17) in the human placenta. As a result, 21-carbon products (progesterone and pregnenolone) cannot be converted to 19-carbon steroids (androstenedione and dehydroepiandrosterone). Like progesterone, estrogen produced by the placental aromatase (P450arom) enzyme system must derive precursors from outside the placenta.[20]

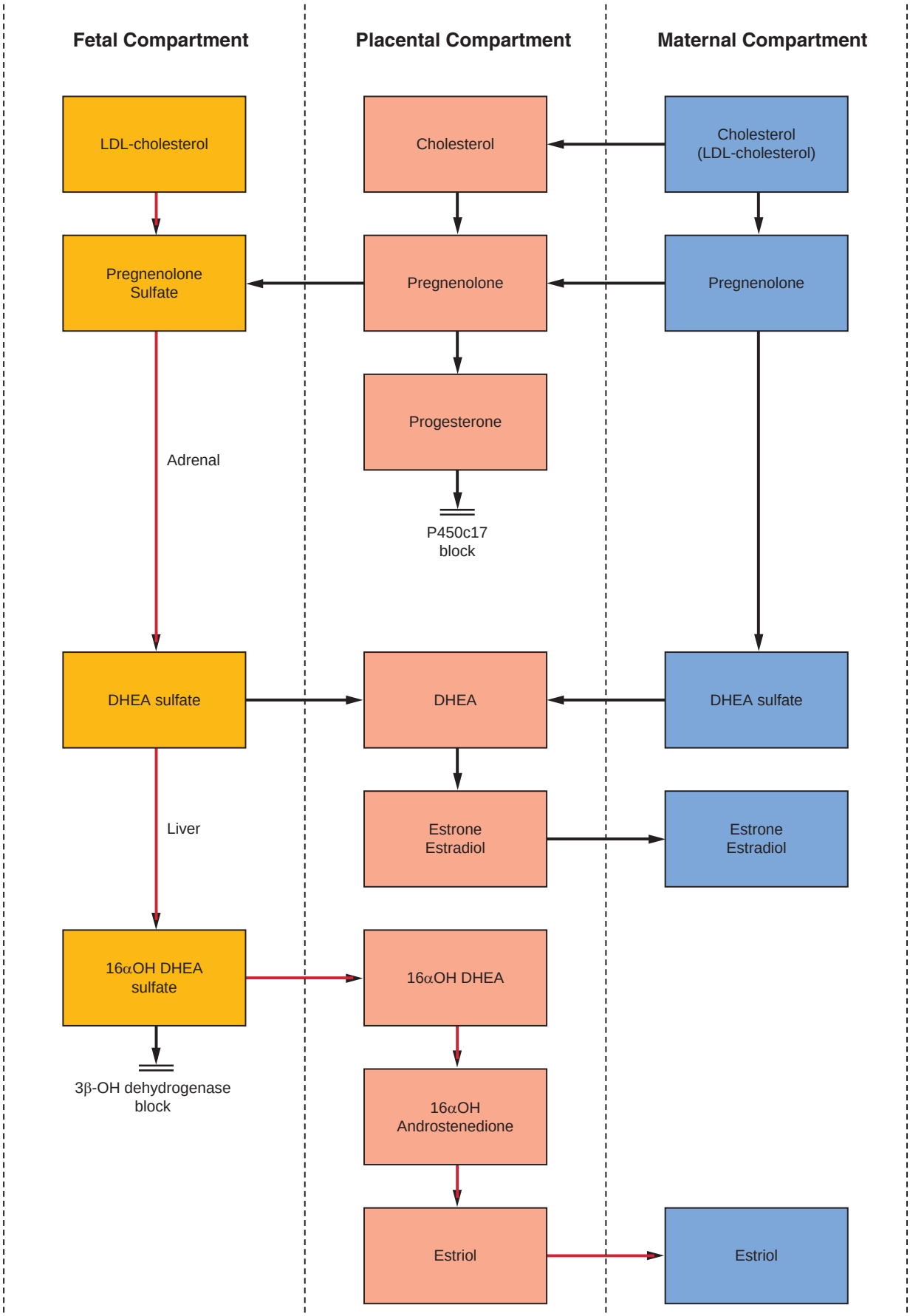
The androgen compounds utilized for estrogen synthesis in human pregnancy are, in the early months of gestation, derived from the maternal bloodstream. By the 20th week of pregnancy, the vast majority of estrogen excreted in the maternal urine is derived from fetal androgens. In particular, approximately 90% of estriol excretion can be accounted for by dehydroepiandrosterone sulfate (DHEAS) production by the fetal adrenal gland.[20, 21] The high output of DHEAS by the fetal zone is due to low 3β-hydroxysteroid dehydrogenase gene expression.[22] Removed into cell culture conditions, this gene becomes active in response to adrenocorticotropic hormone (ACTH).

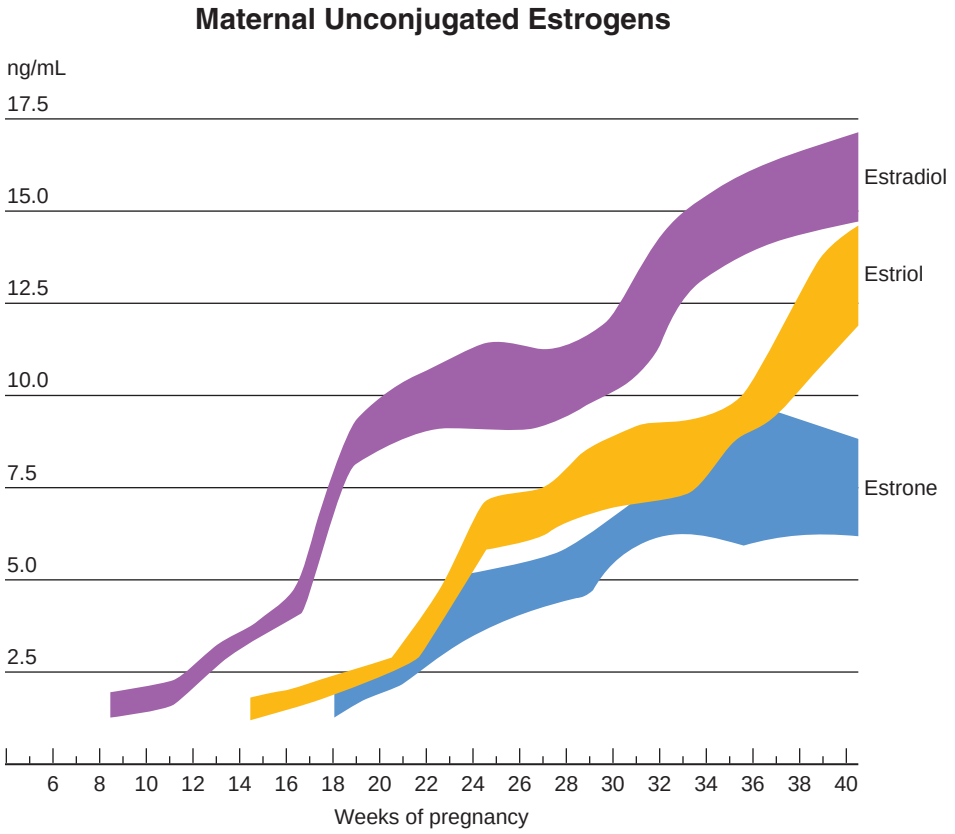
The fetal endocrine compartment is characterized by rapid and extensive conjugation of steroids with sulfate. This is a protective mechanism, blocking the biologic effects of potent steroids present in such great quantities. In order to utilize fetal precursors, the placenta must be extremely efficient in cleaving the sulfate conjugates brought to it via the



fetal bloodstream. Indeed, the sulfatase activity in the placenta is rapid and quantitatively very significant. It is recognized that a deficiency in placental sulfatase is associated with low estrogen excretion, giving clinical importance to this metabolic step. This syndrome is discussed in greater detail later in this chapter.

The fetal adrenal provides DHEAS as precursor for placental production of estrone and estradiol. However, the placenta lacks 16α-hydroxylation ability, and estriol with its 16α-hydroxyl group must be derived from an immediate fetal precursor. The fetal adrenal, with the aid of 16α-hydroxylation in the fetal liver, provides the 16α-hydroxydehydroepiandrosterone


Fetal Compartment
Placental Compartment
Maternal Compartment
LDL-cholesterol
Cholesterol
Cholesterol (LDL-cholesterol)
Pregnenolone Sulfate
Pregnenolone
Pregnenolone
Progesterone
Adrenal
P450c17 block
DHEA sulfate
DHEA
DHEA sulfate
Liver
Estrone Estradiol
Estrone Estradiol
16αOH DHEA sulfate
16αOH DHEA
3β-OH dehydrogenase block
16αOH Androstenedione
Estriol
Estriol



sulfate for placental estriol formation. After birth, neonatal hepatic 16α-hydroxylation activity rapidly disappears. The maternal contribution of DHEAS to total estrogen synthesis must be negligible because, in the absence of normal fetal adrenal glands (as in an anencephalic infant), maternal estrogen levels and excretion are extremely low. The fetal adrenals secrete more than 200 mg of DHEAS daily, about 10 times more than the mother.[23] Estriol is the estrogen produced in greatest quantity during pregnancy; estrone and estradiol are derived equally from fetal and maternal precursors.[21] The profiles of the unconjugated compounds in the maternal compartment for the three major estrogens in pregnancy are as follows:

1. A rise in estrone begins at 6–10 weeks, and individual values range from 2 to 30 ng/mL at term.[24] This wide range in normal values precludes the use of estrone measurements in clinical applications.

2. A rise in estradiol begins in weeks 6–8 when placental function becomes apparent.[2] Individual estradiol values vary between 6 and 40 ng/mL at 36 weeks of gestation and then undergo an accelerated rate of increase.[24] At term, an equal amount of estradiol arises from maternal DHEAS and fetal DHEAS, and its importance in fetal monitoring is negligible.

3. Estriol is first detectable at 9 weeks when the fetal adrenal gland secretion of precursor begins. Estriol concentrations plateau at 31–35 weeks and then increase again at 35–36 weeks.[25]

During pregnancy, estrone and estradiol production is increased about 100 times over nonpregnant levels. However, the increase in maternal estriol excretion is about a thousand-fold. The traditional view that estriol in pregnancy is a weak estrogen metabolite is not accurate. A weak estrogen provided in high concentrations can produce a biologic response

equivalent to that of estradiol.[26] Because of its high production rate and concentration, estriol is an important hormone in pregnancy. The maternal level of estradiol is higher than in the fetus; in contrast, the estriol level in the fetus is greater than in the mother.

The maternal cardiovascular adaptations to pregnancy that are so necessary to serve the fetus are appropriately under the influence of the fetus and significantly regulated by estrogen.[27] Blood volume is increased by estrogen stimulation of the maternal and trophoblastic renin-angiotensin systems, and uteroplacental blood flow, which is so critical for the fetus, is influenced by the vasodilatory effects of estrogen.

The enzyme responsible for estrogen synthesis is the cytochrome P450 aromatase enzyme (P450arom), the product of the *CYP19* gene.[28] The *CYP19* gene is regulated in various tissues by tissue-specific promoters. The placenta, with its huge capacity for estrogen synthesis, uses a powerful, unique promoter that allows specific regulation. An autosomal-recessive disorder due to mutations in the P450arom gene is associated with a failure to convert androgen precursors to estrogen by placental aromatase.[29] Consequently, a female fetus and the mother can undergo virilization. Nevertheless, growth and development of the fetus are not impaired, and this disorder raises the question: How much, if any, estrogen is essential in human pregnancy? Is this another example of backup mechanisms operating to achieve the goal?

Normally, placental aromatization is so efficient that little androgen presented to the placenta escapes.[30] For this reason, fetuses are well protected against masculinization, and even in the presence of an androgen-secreting tumor, extremely large amounts of aromatizable androgens or the secretion of nonaromatizable androgens are required to produce unwanted virilization.

The estrogens presented to the maternal bloodstream are rapidly metabolized by the maternal liver prior to excretion into the maternal urine as a variety of more than 20 products. The bulk of these maternal urinary estrogens is composed of glucosiduronates conjugated at the 16-position. Significant amounts of the 3-glucosiduronate and the 3-sulfate-16-glucosiduronate are also excreted. Only approximately 8–10% of the maternal blood estriol is unconjugated.

The Fetal Adrenal Cortex

The fetal adrenal cortex is unique, differentiating by 8–9 weeks gestational age into a thick inner fetal zone and a thin outer definitive zone, which is the source of cortisol and the forerunner of the adult cortex.[31] Early in pregnancy, adrenal growth and development are remarkable, and the gland achieves a size equal to or larger than that of the kidney by the end of the first trimester. After 20–24 weeks, the adrenal glands slowly decrease in size until a second spurt in growth begins at about 34–35 weeks. The gland remains proportionately larger than the adult adrenal glands. After delivery, the fetal zone (about 85% of the bulk of the gland) rapidly involutes to be replaced by simultaneous expansion of the outer definitive zone to form the zona glomerulosa, and the transitional zone to form the zona fasciculata and the zona reticularis (which expands again during adrenarche at puberty). By age 1, the fetal zone is gone, replaced by the adult adrenal cortex. Thus, the specific steroidogenic characteristics of the fetus are associated with a specific adrenal morphology that is dependent on specific factors present during intrauterine life.

Fetal dehydroepiandrosterone (DHEA) and DHEAS production rises steadily concomitant with the increase in the size of the fetal zone and adrenal weight.[32] DHEA and DHEAS are the major secretory products of the fetal zone because 3β-hydroxysteroid dehydrogenase-isomerase activity and the expression of this enzyme's gene are suppressed.[22, 33] The well-known increase in maternal estrogen levels is significantly influenced by the increased availability of fetal DHEAS as a precursor. Indeed, the accelerated rise in maternal estrogen levels near term can be explained, in part, by an

increase in fetal DHEAS. The stimulus for the substantial adrenal growth and steroid production has been a puzzle.

Early in pregnancy, the adrenal gland can grow and function without ACTH, perhaps in response to hCG.[31] After 15–20 weeks, fetal ACTH is required. However, during the last 12–14 weeks of pregnancy when fetal ACTH levels are declining, the adrenal quadruples in size.[34] Because pituitary prolactin is the only fetal pituitary hormone to increase throughout pregnancy, paralleling fetal adrenal gland size changes, it was proposed that fetal prolactin is the critical tropic substance. In experimental preparations, however, only ACTH exerts a steroidogenic effect. There is no fetal adrenal response to prolactin, hCG, growth hormone, melanocyte-stimulating hormone (MSH), or thyrotropin-releasing hormone (TRH).[35, 36] Furthermore, in patients treated with bromocriptine, fetal blood prolactin levels are suppressed, but DHEAS levels are unchanged.[37] Nevertheless, interest in prolactin persists because both ACTH and prolactin can stimulate steroidogenesis in vivo in the fetal baboon.[38]

There is no question that, in the second half of pregnancy, ACTH is essential for the morphologic development and the steroidogenic mechanism of the fetal adrenal gland.[39, 40] ACTH activates adenylate cyclase, leading to steroidogenesis. Soon the supply of cholesterol becomes rate limiting. Further ACTH action results in an increase in LDL receptors, leading to an increased uptake of circulating LDL-cholesterol, largely derived from the fetal liver.[18] With internalization of LDL-cholesterol, hydrolysis by lysosomal enzymes of the cholesteryl ester makes cholesterol available for steroidogenesis. For this reason, fetal plasma levels of LDL are low, and after birth newborn levels of LDL rise as the fetal adrenal involutes. In the presence of low levels of LDL-cholesterol, the fetal adrenal is capable of synthesizing cholesterol de novo.[41] Thus, near term, both de novo synthesis and utilization of circulating LDL-cholesterol are necessary to sustain the high rates of DHEAS and estrogen formation. In addition, ACTH increases adrenal response by increasing the expression of its own receptor.[42]

The tropic support of the fetal adrenal gland by ACTH from the fetal pituitary is protected by placental estrogen. The placenta prevents cortisol that is present in higher levels in the mother from reaching the fetus by converting cortisol to cortisone. The converting enzyme, 11β-hydroxysteroid dehydrogenase, is abundantly expressed in syncytiotrophoblast at the interface between fetal tissue and maternal blood, and is stimulated by placental estrogen.[43, 44] Regulation of this enzyme by estrogen thus influences fetal ACTH secretion. With increasing estrogen levels in late gestation, even greater 11β-hydroxysteroid dehydrogenase activity would result in even less maternal cortisol reaching the fetal circulation. Thus, it is proposed that near labor, fetal ACTH secretion increases, the fetal adrenal gland undergoes greater maturation, and fetal cortisol synthesis from endogenous cholesterol increases.[45] A relative deficiency in 11β-hydroxysteroid dehydrogenase type 2 (the high affinity isoform) would expose the fetus to excessively high cortisol levels and is correlated with low birth weight, which in turn is correlated with insulin resistance, abnormal lipids, and hypertension in adult life.[46–48] A reduction of type 2 11β-hydroxysteroid dehydrogenase activity has been reported in pregnancies complicated by smoking and preeclampsia, conditions known to be associated with intrauterine growth retardation.[49, 50] A similar reduction in activity has been documented with idiopathic intrauterine growth retardation, accompanied by a decreased ratio of cortisone to cortisol in umbilical artery blood.[48]

An interaction has been demonstrated in vitro between progesterone and the lipoxygenase pathways that lead to the products of arachidonic acid other than prostaglandins in regards to the regulation of 11β-hydroxysteroid dehydrogenase activity.[51] Progesterone down-regulates 11β-hydroxysteroid dehydrogenase expression, as do the products of lipoxygenase activity. Because the lipoxygenase products increase progesterone output by trophoblast cells, an increase in lipoxygenase activity because of infection could increase progesterone levels, which in turn would decrease 11β-hydroxysteroid dehydrogenase activity, resulting in higher cortisol levels in the fetus with the consequences of stress and growth retardation.

It has been suggested that the increase in fetal cortisol secretion during normal pregnancy competes with progesterone for the glucocorticoid receptor in the placenta, thus blocking the inhibitory action of progesterone on corticotropin-releasing hormone (CRH) synthesis, leading to an increase in CRH.[52] Placental production of CRH and the size of the fetal adrenal gland are closely correlated in several primates, both reaching a peak in humans at the time of parturition. The increase in CRH would augment fetal ACTH secretion, producing adrenal growth and even more fetal cortisol in a positive feedback relationship, as well as more DHEAS to serve as precursor for the increase in estrogen that occurs prior to parturition. However, fetal ACTH levels in the last half of pregnancy are not increasing, but slightly decreasing. It is significant that CRH, as demonstrated in vitro, also directly stimulates DHEAS synthesis by the fetal adrenal gland.[53, 54]

This is an important fetal placental-adrenal cycle. Cortisol from the adrenal gland increases placental CRH production; CRH induces ACTH receptor expression in the definitive zone of the fetal adrenal gland, leading to even greater adrenal cortisol secretion, and that in turn increases placental CRH biosynthesis as gestation advances.[55] Direct stimulation of the fetal zone by CRH, supported by the presence of ACTH, augments the increasing production of DHEA and DHEAS required for estrogen synthesis in late gestation.[54]

Adrenal gland steroidogenesis involves autocrine and paracrine regulation by various growth factors.[31] Fetal adrenal cells produce inhibin, and the alpha subunit (present only in inhibin) is preferentially increased by ACTH.[56, 57] In the fetal adrenal, the beta subunit is not expressed; thus, inhibin-A and activin-A are the principal forms.

Activin enhances ACTH-stimulated steroidogenesis while inhibiting mitogenesis in human fetal zone adrenal cells.[57] This effect on steroidogenic activity is not present in adult adrenal cells. In vitro, activin enhances a shift in fetal adrenal cells from ACTH stimulation of DHEAS production to cortisol production. This shift is analogous to the shift that occurs after birth. Perhaps activin plays this role in the remodeling of the fetal zone in the newborn. A specific action for inhibin in fetal adrenal cells has not been described.

We should not expect the fetal adrenal gland to be an exception to the ubiquitous presence and actions of all growth factors.[31] Basic fibroblast growth factor has potent mitogenic activity mediating the growth response of the fetal adrenal cortex to ACTH. Evidence indicates that the epidermal growth factor receptor is activated in the fetal adrenal, but the ligand using this receptor is probably transforming growth factor-α. Like activin, transforming growth factor-β inhibits fetal zone cellular proliferation and, in addition, suppresses steroidogenesis.

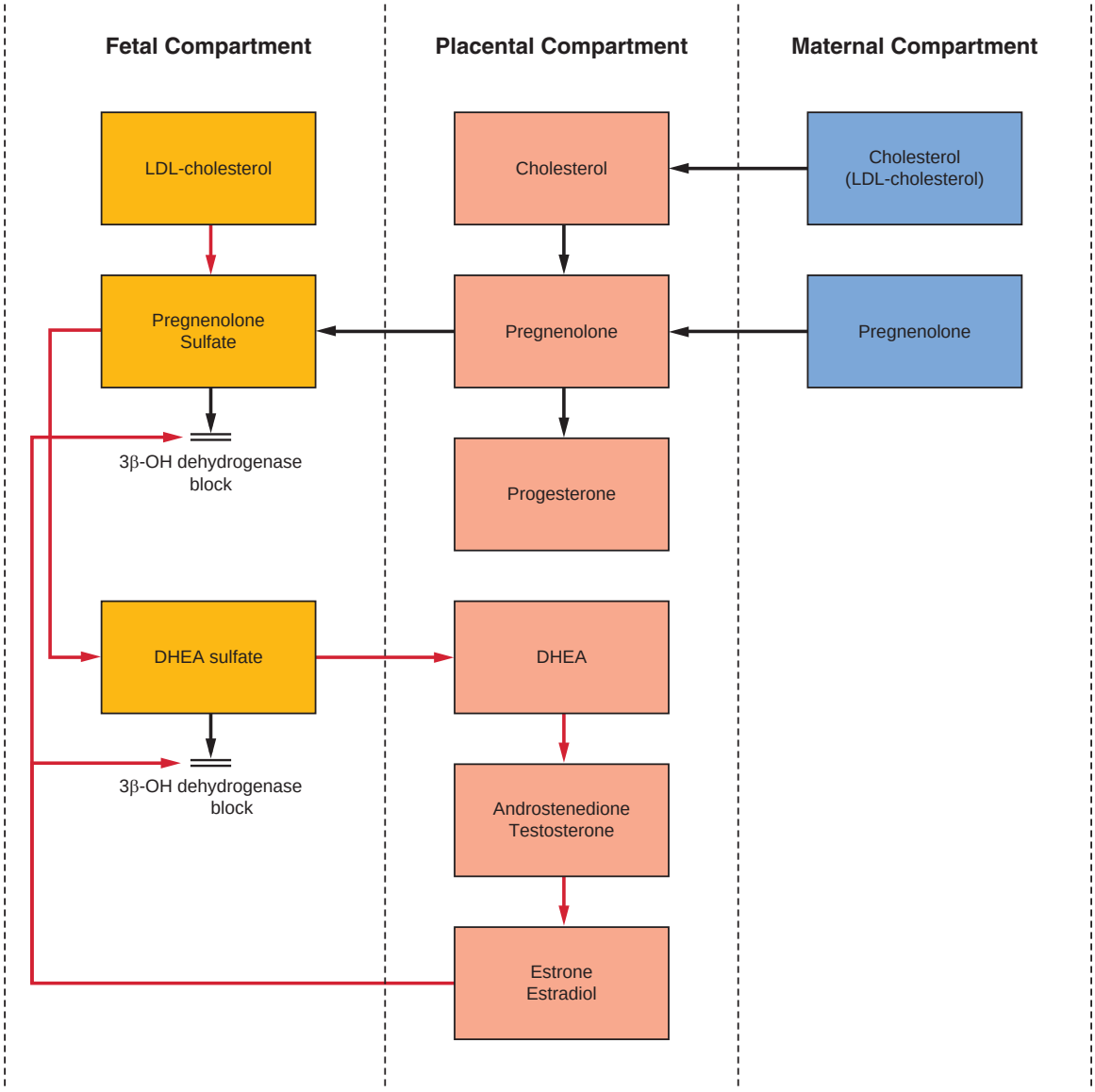
The insulin-like growth factors (IGF-I and IGF-II) are important in mediating the tropic effects of ACTH, particularly increasing adrenal responsiveness to ACTH in the second half of pregnancy.[58] IGF-II production in the fetal adrenal is very significant and is stimulated by ACTH. IGF-II is believed to be important in prenatal growth.[59] The abundance of IGF-II in the fetal adrenal gland implicates this growth factor as a mediator of ACTH-induced growth.[60] Both IGF-I and IGF-II are equally mitogenic in a cell culture system of fetal adrenal cells and enhance the proliferation stimulated by basic fibroblast growth factor and epidermal growth factor.[60] However, only transcription of IGF-II is stimulated by ACTH. IGF-II augments ACTH-stimulated steroidogenesis in the fetal adrenal, specifically by increasing the expression of P450c17.[58] Thus, the growth-promoting and steroidogenic effects of ACTH are mediated by various growth factors, with a principal role played by IGF-II. In this regard, the fetal adrenal differs from the adult adrenal where IGF-I is predominant; however, IGF-II is able to modulate responsiveness to ACTH in the fetal adrenal by activating the IGF-I receptor.

Steroidogenic factor-1 (SF-1) and DAX-1 (named for the location of its gene on the X chromosome) are nuclear receptors for which specific ligands have not been identified ("orphan receptors"). SF-1 influences the expression of genes that encode steroidogenic enzymes, and when genetic expression of SF-1 is disrupted in mice, gonads and adrenal glands fail to develop.[61, 62]

Mutations in the DAX-1 gene result in adrenal hypoplasia, and DAX-1 is believed to work with SF-1 in regulating development and function of steroid-producing tissues.[63]

The production of DHEA is dependent on the *CYP17* gene that is responsible for both 17α-hydroxylase and 17,20-lyase enzyme activity. Differential regulation of these two activities with an increase in 17,20-lyase could account for the increase in DHEA in the fetal zone of the adrenal gland. The *SULT2A1* gene is responsible for the sulfation and production of DHEAS. Modulation of this gene would also contribute to the steroidogenic activity of the fetal zone.

The unique features of the fetal adrenal gland can be ascribed to its high-estrogen environment. Tissue culture studies have demonstrated that hormonal peptides of pituitary or placental origin are not the factors that are responsible for the behavior of the fetal adrenal gland.[64–66] Estrogens at high concentration inhibit 3β-hydroxysteroid dehydrogenase-isomerase activity in the fetal adrenal gland and, in the presence of ACTH in conjunction with IGF-II, enhance the secretion of DHEA.[67] Estradiol concentrations of 10–100 ng/mL are required to inhibit cortisol secretion.[68] The total estrogen concentrations in the fetus are easily in this range. A study of the kinetics of 3β-hydroxysteroid dehydrogenase activity in human adrenal microsomes reveals that all steroids are inhibitory, and most notably, estrone and



estradiol at levels found in fetal life cause almost total inhibition.[69] In a study utilizing a human adrenocortical cell line, estradiol in high concentrations inhibited 3β-hydroxysteroid dehydrogenase and the mechanism appeared to be independent of the estrogen receptor.[70] The increase in DHEAS secretion by the fetal zone is a consequence of suppression of the gene (*HSD3B2*) that controls 3β-hydroxysteroid dehydrogenase expression; transcriptional factors necessary for the activity of this gene are absent in the fetal zone.[71]

The development of the adrenal gland during human fetal life and during the neonatal period is paralleled in the baboon.[72] The adrenal cortex of the fetal baboon is characterized by the same deficiency in 3β-hydroxysteroid dehydrogenase as that seen in the human, with the same diversion of steroidogenesis into production of DHEA and DHEAS. Experimental studies in the baboon suggest that placental estrogen maintains the production of DHEA and DHEAS by the fetal adrenal cortex, but excessive adrenal growth and steroidogenesis is at the same time suppressed by the increasing estrogen levels in late pregnancy.[73, 74]

Tissue growth in mammals is a consequence of cellular proliferation promoted by the cell regulators, cyclin D1 and cyclin E, which dimerize with kinases to form enzymes that carry out key phosphorylations during cell cycles. These key regulators are expressed in increasingly high concentrations in the baboon fetal adrenal cortex, beginning in early to mid gestation, especially in the outer definitive zone (destined to be the adult adrenal cortex and the source of cortisol).[75] This early increase in the definitive zone is followed by a progressive decrease in these factors required for cellular proliferation in the definitive zone in late gestation. The cells in the fetal zone are derived from the definitive zone, and these changes indicate that the continued growth of the fetal adrenal cortex during gestation predominantly reflects cellular hypertrophy. Furthermore, because there is a progressive increase in the expression of 3β-hydroxysteroid dehydrogenase within the fetal definitive zone, the decline in proliferation is associated with functional differentiation as the definitive zone acquires the ability to produce mineralocorticoids and glucocorticoids.[75] Here again, the key modulator of this change may be estrogen, specifically an estrogen-induced decrease in cyclin expression with advancing gestation.

The explanation that estrogen regulates 3β-hydroxysteroid dehydrogenase is challenged by in vitro studies of human fetal zone cells indicating that estradiol and IGF-II combine to direct steroidogenesis to DHEAS in a mechanism not due to inhibition of 3β-hydroxysteroid dehydrogenase.[67] ***Nevertheless, it is an attractive and useful hypothesis to view the principal mission of the fetal adrenal as providing DHEAS as the basic precursor for placental estrogen production. Estrogen, in turn, feeds back to the adrenal to direct steroidogenesis along the Δ^5 pathway to provide even more of its precursor, DHEAS. Thus far, this is the only known function for DHEAS. With birth and loss of exposure to estrogen, the fetal adrenal gland quickly changes to the adult type of gland. It seems reasonable to conclude that this complex change in the fetal adrenal cortex is orchestrated by the interplay among fetal pituitary ACTH, placental estrogen, and placental growth factors.***

Measurement of Estrogen in Pregnancy

Because pregnancy is characterized by a great increase in maternal estrogen levels and estrogen production is dependent on fetal and placental steroidogenic cooperation, the amount of estrogen present in the maternal blood or urine reflects both fetal and placental enzymatic capability and, hence, well-being. Attention focused on estriol because 90% of maternal estriol is derived from fetal precursors. The end product to be assayed in the maternal blood or urine is influenced by a multitude of factors. Availability of precursor from the fetal adrenal gland is a prime requisite, as well as the ability of the placenta to perform its conversion steps. Maternal metabolism of the product and the efficiency of maternal renal excretion of the product can modify the daily amount of estrogen in the urine. Blood

flow to any of the key organs in the fetus, placenta, and mother becomes important.[76, 77] Fetal hypoxemia due to acute reductions in uteroplacental blood flow is associated with a marked increase in adrenal androgen production in response to an increase in fetal ACTH and, in response to the availability of androgen precursors, an increase in maternal estrogen levels.[78] The response to acute stress is in contrast to the effect of chronic uteroplacental insufficiency, which is associated with a reduction in fetal androgens and maternal estrogens. In addition, drugs or diseases can affect any level in the cascade of events leading up to the assay of estrogen.

For years, measurement of estrogen in a 24-hour urine collection was the standard hormonal method of assessing fetal well-being. This was replaced by immunoassay of unconjugated estriol in the plasma.[79] Because of its short half-life (5–10 minutes) in the maternal circulation, unconjugated estriol has less variation than urinary or total blood estriol. However, assessment of maternal estriol levels has been superseded by various biophysical fetal monitoring techniques such as nonstress testing, stress testing, and measurement of fetal breathing and activity. Modern screening for fetal aneuploidy (discussed later in the chapter) utilizes three markers in the maternal circulation: alpha fetoprotein, human chorionic gonadotropin, and unconjugated estriol.

Amniotic Fluid Estrogen Measurements

Amniotic fluid estriol is correlated with the fetal estrogen pattern rather than the maternal. Most of the estriol in the amniotic fluid is present as 16-glucosiduronate or as 3-sulfate-16-glucosiduronate. A small amount exists as 3-sulfate. Very little unconjugated estriol is present in the amniotic fluid because free estriol is rapidly transferred across the placenta and membranes. Estriol sulfate is low in concentration because the placenta and fetal membranes hydrolyze the sulfated conjugates, and the free estriol is then passed out of the fluid. Because the membranes and the placenta have no glucuronidase activity, the glucosiduronate conjugates are removed slowly from the fetus. The glucosiduronates, therefore, predominate in the fetal urine and the amniotic fluid. Because of the slow changes in glucosiduronates, measurements of amniotic fluid estriol have wide variations in both normal and abnormal pregnancies. An important clinical use for amniotic fluid estrogen measurements has not emerged.

Estetrol

Estetrol (15α-hydroxyestriol) is formed from a fetal precursor and is very dependent on 15α-hydroxylation activity in the fetal liver. The capacity for 15α-hydroxylation of estrogens increases during fetal life, reaching a maximum at term. This activity then declines during infancy and is low, absent, or undetectable in adults. Estetrol may contribute to the estrogen effects taking place during pregnancy as maternal estetrol levels steadily increase with advancing gestation, and fetal levels are higher than maternal levels.[80] Because of wide variations within and between individuals, there is no clinical use for maternal blood or urine estetrol measurements during pregnancy. However, estetrol, given in sufficient doses, is a potent, orally active estrogen that has potential for pharmacologic therapy.[81]

Placental Sulfatase Deficiency

There is an X-linked metabolic disease characterized by a ***placental sulfatase deficiency*** in the syncytiotrophoblast and postnatal ichthyosis, occurring in about 1 in 2,000–3,000

newborn males.[82] Patients with the placental sulfatase disorder are unable to hydrolyze DHEAS or 16α-hydroxy-DHEAS; therefore, the placenta cannot form normal amounts of estrogen. A deficiency in placental sulfatase is usually discovered when patients go beyond term and are found to have extremely low estriol levels and no evidence of fetal distress. The patients usually fail to go into labor and require delivery by cesarean section. Most striking is the failure of cervical softening and dilation; thus, cervical dystocia occurs that is resistant to oxytocin stimulation. There are many case reports of this deficiency, almost all detected by finding low estriol levels. It was suggested that mothers who are carriers of this disorder are at increased risk for intrauterine growth retardation and perinatal complications even if the fetus is not affected.[83] However, a careful analysis of unexplained low estriol levels concluded that this is a rare occurrence (about 3 per 10,000 pregnancies) and that perinatal complications in pregnancies at risk for placental sulfatase deficiency are not increased (other than a higher cesarean section rate).[84] All newborn children, with a few exceptions, have been male. The steroid sulfatase X-linked recessive ichthyosis locus (the steroid sulfatase gene) has been mapped to the distal short arm portion of the X chromosome, Xp22.32. 90% of cases of ichthyosis have a complete deletion of this gene plus flanking genes. There are no known geographic or racial factors that affect the gene frequency.

The characteristic steroid findings are as follows: extremely low estriol and estetrol levels in the mother with extremely high amniotic fluid DHEAS and normal amniotic fluid DHEA and androstenedione. The normal DHEA and androstenedione with a high DHEAS rule out congenital adrenal hyperplasia. The small amount of estriol that is present in these patients probably arises from 16α-hydroxylation of DHEAS in the maternal liver, thus providing 16α-hydroxylated DHEA to the placenta for aromatization to estriol. Maternal estrone and estradiol are also low but not as markedly reduced because of their utilization of maternal precursors. Measurement in maternal urine of steroids derived from fetal sulfated compounds is a simple and reliable means of prenatal diagnosis. Demonstration of a high level of DHEAS in the amniotic fluid is confirming. To establish the diagnosis with certainty, a decrease in sulfatase activity should be demonstrated in an in vitro incubation of placental tissue. The clinician should keep in mind that fresh tissue is needed for this procedure because freezing lowers enzyme activity. Alternatively, steroid sulfatase activity can be assayed in leukocytes.

It is now recognized that steroid sulfatase deficiency is present in other tissues and can persist after birth. These children develop ichthyosis beginning between birth and 6 months of age, characterized by hyperkeratosis (producing scales on the neck, trunk, and palms) and associated with mild corneal opacities, pyloric stenosis, and cryptorchidism. The skin fibroblasts have a low activity of steroid sulfatase, and scale formation that occurs early in the first year of life is thought to be due to an alteration in the cholesterol:cholesteryl ester ratio (due to the accumulation of cholesterol sulfate). This inherited disorder, thus, represents a single entity: placental sulfatase deficiency and X-linked ichthyosis, both reflecting a deficiency of microsomal sulfatase. More extensive deletions include contiguous genes and result in attention deficit hyperactivity disorder, autism, and mental retardation.[85]

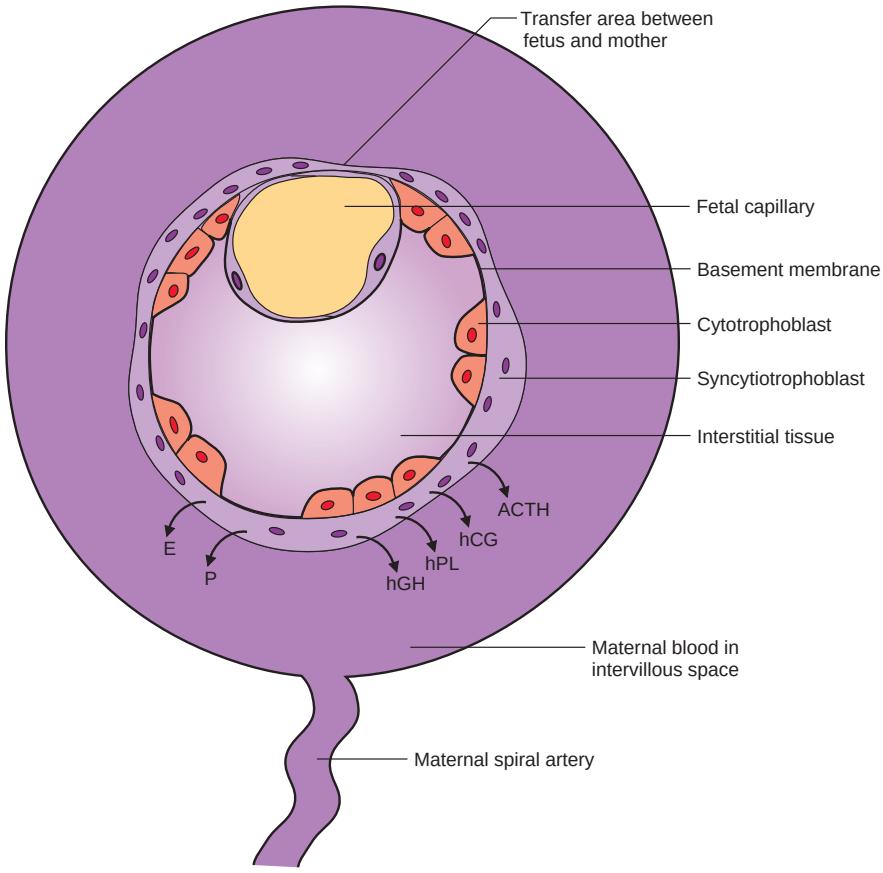
A family history of scaling in males (as well as repeated postdate pregnancies and cesarean sections) should prompt a consideration for prenatal diagnosis. Because the clinical use of estriol measurements has declined, there is no effective method to identify the presence of this problem in women with normal obstetrical histories. However, a low maternal level of unconjugated estriol can be encountered with multiple marker screening (discussed later in this chapter). Furthermore, consideration should be given to antenatal screening by estriol measurement in pregnancies in which a male fetus is present and there is a previous history of a growth-retarded or stillborn male. However, perinatal outcome is good even when placental sulfatase deficiency is not known to be present and only a very small number of affected boys have serious manifestations of the disorder; therefore, it is difficult to justify the need for antenatal diagnosis.[84]

The Differential Diagnosis of an Extremely Low Estriol

1. **Impending or present fetal demise.**
2. **Adrenal hypofunction.**
3. **Placental sulfatase deficiency.**
4. **Placental aromatase deficiency.**
5. **Drug-related effects.**

Protein Hormones of Pregnancy

The placental villus is composed of trophoblast, mesenchymal cells, and fetal blood vessels. The two main trophoblastic layers consist of the cytotrophoblast, separate mononuclear cells prominent early in pregnancy and sparse late in pregnancy, and the syncytiotrophoblast, a continuous multinuclear layer on the surface of the villi. The cytotrophoblast is the basic placental stem cell from which the syncytiotrophoblasts arise by differentiation. The syncytiotrophoblast is, therefore, the functional cell of the placenta, the major site of hormone and protein production. Control of this important cellular differentiation is still not understood; however, the process is influenced by hCG and, undoubtedly, a variety of growth factors.[86] The protein hormone system is complicated because individual peptides



can have multiple functions.[87] The surface of the syncytiotrophoblast is in direct contact with the maternal blood in the intervillous space. This may be a reason why placental proteins are secreted preferentially into the mother.

Proteins Associated with Pregnancy

Fetal Compartment	*Placental Compartment*	*Maternal Compartment*
Alpha-fetoprotein	Hypothalamic-like hormones	Decidual proteins
	GnRH	Prolactin
	CRH	Relaxin
	TRH	Progesterone-associated endometrial protein
	Somatostatin	
	GHRH	IGFBP-1
	Neuropeptide Y	Interleukin-1
		Colony-stimulating factor-1
	Pituitary-like hormones	Corpus luteum proteins
	hCG	Relaxin
	hPL	Prorenin
	hGH	
	hCT	
	ACTH	
	Oxytocin	
	Prolactin	
	Growth factors	
	IGF-I	
	IGF-II	
	Epidermal growth factor	
	Platelet-derived growth factor	
	Fibroblast growth factor	
	Transforming growth factor-a	
	Transforming growth factor-b	
	Inhibin	
	Activin	
	Follistatin	
	Cytokines	
	Interleukins	
	Interferons	
	Tissue necrosis factor-a	
	Colony-stimulating factor-1	
	Other proteins	
	Opiates	
	Prorenin	
	Pregnancy-specific β1-glycoprotein	
	Pregnancy-associated plasma protein A	

Hypothalamic-like Releasing Hormones

The human placenta contains many releasing and inhibiting hormones, including gonadotropin-releasing hormone (GnRH), corticotropin-releasing hormone (CRH), thyrotropin-releasing hormone (TRH) and somatostatin.[88] Because of the presence of hypothalamic-like releasing hormones in an organ that produces tropic hormones, we are motivated to construct a system of regulation analogous to the hypothalamic-pituitary axis. However, as we shall see, this proves to be very difficult.

Immunoreactive GnRH can be localized in the cytotrophoblast and syncytiotrophoblast. Evidence indicates that placental GnRH regulates placental steroidogenesis and release of prostaglandins, as well as hCG.[88–92] In some studies, the highest amount of GnRH was present early in pregnancy when the number of cytotrophoblasts is greatest and hCG secretion reaches its peak; however, others report relatively constant levels throughout pregnancy.[93, 94] All isoforms of GnRH are expressed in the human placenta, but GnRH-I is the predominant form.[95, 96]

The placental receptors for GnRH have a lower affinity than that of GnRH receptors in the pituitary, ovary, and testis.[97, 98] This reflects the situation in which the binding site is in close proximity to the site of secretion for the regulatory hormone. A higher affinity is not necessary because of the large amount of GnRH available in the placenta, and the low-affinity receptors avoid response to the low levels of circulating GnRH. GnRH receptors, present in both cytotrophoblasts and syncytiotrophoblasts, are produced in a pattern that parallels the curve of hCG secretion, further evidence that placental GnRH and its receptor regulate hCG secretion.[99] GnRH release is increased by estrogen, activin-A, insulin, and prostaglandins, and inhibited by progesterone, endogenous opiates, inhibin, and follistatin.[87] The GnRH receptor is highly expressed in the fetal zone of the adrenal gland, raising the possibility of another pathway by which the placenta can influence fetal adrenal function.[96]

CRH, identical in structure to hypothalamic CRH, is produced in the trophoblast, the fetal membranes, and the decidua.[87] Its production is regulated by steroids, decreased by progesterone, and, in contrast to the usual negative feedback action in the hypothalamus, increased by glucocorticoids.[100] These interactions are consistent with the increase in fetal cortisol associated with the last weeks of pregnancy and the increase in ACTH with labor. Placental CRH is further regulated (as in the hypothalamus) by an array of substances such as vasopressin, norepinephrine, angiotensin II, prostaglandins, neuropeptide Y, and oxytocin. CRH release is stimulated by activin and interleukin and inhibited by inhibin and nitric oxide. The progressive increase in maternal CRH levels during pregnancy is due to the secretion of intrauterine CRH into the maternal circulation. The highest levels are found at labor and delivery. A binding protein for CRH exists in the human circulation, and it is produced in placenta, membranes, and decidua.[101] Maternal levels of this binding protein are not different in pregnancy until a slight increase at 35 weeks, followed by a major decrease until term, increasing the bioavailability of CRH in late gestation. Maternal CRH levels are elevated in women with pregnancies under stress, e.g., with preeclampsia and preterm labor.[87] An increase in placental CRH may be a response to the activation of fetal pituitary ACTH and adrenal cortisol secretion in the presence of hypoxemia. Placental CRH has multiple roles, including stimulation of the fetal adrenal plus parturition and regulation of blood flow. It is not certain to what extent CRH contributes to the increase in maternal adrenal secretion during pregnancy.

Trophoblast, amnion, chorion, and decidua also produce a peptide similar to CRH, named urocortin, that binds to CRH receptors and CRH-binding protein.[102] Urocortin displays activities similar to CRH, such as inducing the secretion of prostaglandins and matrix metalloproteinases in placental cells and fetal membrane cells, and directly stimulating steroidogenesis in cells derived from the fetal zone of the adrenal gland.[54, 103, 104]

Human Chorionic Gonadotropin (hCG)

Human chorionic gonadotropin is a glycoprotein, a peptide framework to which carbohydrate side chains are attached.[105] Alterations in the carbohydrate component (about one-third of the molecular weight) change the biologic properties. For example, the long half-life of hCG is approximately 24 hours as compared with 2 hours for luteinizing hormone (LH), a 12-fold difference, which is due mainly to the greater sialic acid content of hCG. As with the other glycoproteins, follicle-stimulating hormone (FSH), LH, and thyroid-stimulating hormone (TSH), hCG consists of two subunits, noncovalently linked by disulfide bonds, called alpha (α) and beta (β).[106] The α subunit in these glycoprotein hormones is identical, consisting of 92 amino acids. Unique biologic activity as well as specificity in immunoassays is attributed to the molecular and carbohydrate differences in the β subunits (see "Heterogeneity" in Chapter 2).

β-hCG is the largest β subunit, containing a larger carbohydrate moiety and 145 amino acid residues, including a unique carboxyl terminal tailpiece of 24 amino acid groups. It is this unique part of the hCG structure that allows the production of highly specific antibodies and the utilization of highly specific immunologic assays. The extended sequence in the carboxyl-terminal region of β-hCG contains four sites for glycosylation, the reason why hCG is glycosylated to a greater extent than LH, a difference that is responsible for the longer circulating half-life for hCG.

All human tissues appear to make hCG, but the placenta is different in having the ability to glycosylate the protein, thus reducing its rate of metabolism and giving it biologic activity through a long half-life. The carbohydrate components of the glycoproteins are composed of fructose, galactose, mannose, galactosamine, glucosamine, and sialic acid. Although the other sugars are necessary for hormonal function, sialic acid is the critical determinant of biologic half-life. Removal of sialic acid residues in hCG, FSH, and LH leads to very rapid elimination from the circulation.

Genes for tropic hormones contain promoter and enhancer or inhibitor regions located in the 5′ flanking regions upstream from the transcription site. These sites respond to second messengers (cyclic AMP) as well as steroids and other yet unknown regulators. Differences in hCG structure are associated with a different promoter and transcriptional site that is located upstream in the hCG β subunit gene compared with the transcriptional site in the LH β subunit gene. The hCG β subunit promoter does not contain steroid hormone response elements, allowing hCG secretion to escape feedback regulation by the sex steroids, in contrast to FSH and LH.

The protein cores of the two glycoprotein subunits are the products of distinct genes.[107] Using recombinant DNA technology, it has been demonstrated that there is a single human gene for the expression of the α subunit. The gene for the α subunit shared by FSH, LH, hCG, and TSH is located on chromosome 6p21.1–23. A single promoter site subject to multiple signals and hormones regulates transcription of the α-gene in both placenta and pituitary. The α subunit gene is expressed in several different cell types, but the β subunit genes are restricted in cell type. The TSH β-gene is expressed only in thyrotrophs regulated by thyroid hormone; the FSH β-gene is expressed in gonadotrophs regulated by GnRH, activin, inhibin, and gonadal steroids; the LH β-gene, also expressed in gonadotrophs, is regulated by GnRH and is unaffected by activin and inhibin.[108]

The α subunit gene requires the activation of distinct regulatory elements in thyrotroph and gonadotroph cells, as well as in the placenta. It is the activation of these cell-specific elements that produces tissue specificity for α-gene expression. In gonadotrophs, the GnRH-signaling pathway for α-gene transcription utilizes phosphorylase stimulation of diacylglycerol (DAG) and inositol triphosphate (IP_3) that lead to a release of intracellular calcium

stores. GnRH also stimulates the influx of calcium at the cell membrane. DAG, IP_3, and calcium work together to stimulate protein kinase C activity. Protein kinase regulation of the α-promoter is a principal part of the overall mechanism. This pituitary process is influenced by multiple factors including growth factors and gonadal steroids. In the placenta, the mechanism also utilizes specific regulatory elements, but the primary signal is mediated by the cyclic AMP-protein kinase A pathway.

The genes that encode for the β subunits of LH, hCG, and TSH are located in a cluster on chromosome 19q13.3. There are six genes for the β subunit of hCG, and only one for β-LH.[109] Transcription for the six hCG genes, each with different promoter activity, varies, and it is not certain why hCG requires multigenic expression (perhaps this is necessary to reach the extremely high level of production in early pregnancy). It is thought that β-hCG evolved relatively recently from β-LH, and the unique amino acid terminal extension of β-hCG arose by a read-through mutation of the translation stop codon in the β-LH gene; the DNA sequences of the β-hCG genes and the β-LH gene are 96% identical.[109] Gene studies have indicated that the β-hCG gene originated in the common ancestor of monkeys, apes, and humans after the anthropoids diverged from tarsiers, about 35 to 55 million years ago.[110, 111]

Only primates and equine species have been demonstrated to have genes for the β subunit of chorionic gonadotropin. In contrast to human chorionic gonadotropin, equine chorionic gonadotropin exerts both LH and FSH activities in many mammalian species because it contains peptide sequences in its β subunit that are homologous to those in the pituitary gonadotropins of other species. The equine β-chorionic gonadotropin gene is identical to the equine β-LH gene, and although the primate β-hCG gene evolved from the same ancestral β-LH gene, the equine chorionic gonadotropin gene evolved in a different way. The β-LH gene is not expressed in the placenta.

The genetic complexity for the transcription of β-hCG raises the possibility for mutations of these genes as causes of reproductive problems. A search for β-hCG gene deletions in women with recurrent miscarriage or unexplained infertility and for duplications in women with gestational trophoblastic neoplasia found only normal gene structures.[112]

hCG production and secretion are the result of complex interactions among the sex steroids, cytokines, GnRH, and growth factors. GnRH is synthesized by placental cells; GnRH receptors are present on placental cells; and GnRH stimulates the secretion of hCG and the steroid hormones in in vitro studies of placental cells.[113–115] Similar responses can be demonstrated with other peptides, such as interleukin-1β.[116] Similar to opiate action in the hypothalamus, the endorphins are a major inhibiting influence on hCG secretion.[117] Also similar to the pituitary secretion of gonadotropins, inhibin restrains and activin enhances the GnRH-hCG system, with a positive influence of estrogen and a negative impact by progesterone.[118, 119] Follistatin, by binding activin, prevents the stimulatory activity of activin. Other growth factors, specifically IGF-I, IGF-II, TGF-α, and EGF, also influence hCG secretion.

Although a relatively clear story can be constructed into a working concept regarding the autocrine/paracrine interactions in the regulation of the menstrual cycle (Chapter 6), placental function is more complex, and a simple presentation of the many interactions cannot be produced. For example, epidermal growth factor stimulates hCG secretion, but also stimulates inhibin secretion in placental cells, and inhibin suppresses GnRH stimulation of hCG.[120] Inhibin secretion in the placenta is further stimulated by prostaglandins.[121]

Can the cytotrophoblast-syncytiotrophoblast relationship be compared with the hypothalamic-pituitary axis? It does appear that hypothalamic-like peptides (CRH, GnRH) originate in the cytotrophoblast and influence the syncytiotrophoblast to secrete pituitary-like hormones (hCG, hPL, ACTH). Unraveling the interaction is made more difficult by the incredible complexity of the syncytiotrophoblast, a tissue that produces and responds to steroid and peptide hormones, growth factors, and neuropeptides. The best we can say is

that locally produced hormones, growth factors, and peptides work together to regulate placental function.

To this day, the only definitely known function for hCG is support of the corpus luteum, taking over for LH on about the eighth day after ovulation, 1 day after implantation, when β-hCG first can be detected in maternal blood. hCG has been detected at the 8-cell stage in the embryo using molecular biology techniques.[122] Continued survival of the corpus luteum is totally dependent on hCG, and, in turn, survival of the pregnancy is dependent on steroids from the corpus luteum until the seventh week of pregnancy.[1] From the seventh week to the tenth week, the corpus luteum is gradually replaced by the placenta, and by the tenth week, removal of the corpus luteum will not be followed by steroid withdrawal abortion.

It is very probable, but not conclusively proven, that hCG stimulates steroidogenesis in the early fetal testes, so that androgen production will ensue, and masculine differentiation can be accomplished.[123] However, normal masculine differentiation occurs in mouse models lacking LH receptors, and molecular evidence indicates that fetal Leydig cells (but not adult cells) respond to ACTH as well as hCG.[124] It is also possible that the function of the inner fetal zone of the adrenal cortex depends on hCG for steroidogenesis early in pregnancy. The β-hCG gene is expressed in fetal kidney and fetal adrenal, suggesting that hCG may affect the development and function of these organs.[125] In addition, hCG may regulate placental development by influencing the differentiation of cytotrophoblasts.[126]

hCG gene expression is present in both cytotrophoblast and syncytiotrophoblast, but it is synthesized mainly in the syncytiotrophoblast.[127] The maternal circulating hCG concentration is approximately 100 IU/L at the time of the expected but missed menses. A maximal level of about 100,000 IU/L in the maternal circulation is reached at 8–10 weeks of gestation. Why does the corpus luteum involute at the time that hCG is reaching its highest levels? One possibility is that a specific inhibitory agent becomes active at this time. Another is down-regulation of receptors by the high levels of hCG. In early pregnancy, down-regulation may be avoided because hCG is secreted in an episodic fashion.[128] For unknown reasons, the fetal testes escape desensitization; no receptor down-regulation takes place.[123]

hCG levels decrease to about 10,000–20,000 IU/L by 18–20 weeks and remain at that level to term. It is not certain why hCG levels are decreased in the second half of pregnancy. Advancing gestation is associated with increasing amounts of "nicked" hCG molecules in the maternal circulation.[129] These molecules are missing a peptide linkage on the β-subunit, and, therefore, they dissociate into free α and β subunits. At any one point in time, the maternal circulation contains hCG, nicked hCG, free subunits, and fragments of hCG. In addition, the carbohydrate content of hCG varies throughout pregnancy, with more glycosylation present in early pregnancy (hyperglycosylated hCG). Overall, there are about 20–30 isoforms in the maternal blood, and the production of normal molecules is maximal in early gestation when the biologic actions of hCG are so important.[130] A major route of clearance for hCG is renal metabolism in which a final reduced fragment of the β subunit is produced, known as the β-core fragment.

In the complex process of hCG regulation, several inhibiting factors have been identified, including inhibin and progesterone. The decline in hCG occurs at the time of increasing placental progesterone production, and a direct inhibition by this steroid could explain the lower levels of hCG after the tenth week of gestation.[131]

hCG levels close to term are higher in women bearing female fetuses. This is true of serum levels, placental content, urinary levels, and amniotic fluid concentrations. The mechanism and purpose of this difference are not known. Women who have markedly elevated levels of hCG in the second trimester, with no apparent explanation, have increased risks of spontaneous miscarriage, small-for-gestational-age infants, preeclampsia, and preterm delivery.[132]

There are two clinical conditions in which blood hCG titers are especially helpful: trophoblastic disease and ectopic pregnancies. Early pregnancy is characterized by the sequential appearance of hCG, followed by β-hCG and then α-hCG. The ratio of β-hCG to whole hCG remains constant after early pregnancy. Trophoblastic disease is distinguished by very high β-hCG levels (3–100 times higher than normal pregnancy). Ectopic production of α- and β-hCG by nontrophoblastic tumors is rare, but does occur.

In the United States, hydatidiform moles occur in approximately 1 in 600 induced abortions and 1 in 1,000–2,000 pregnancies. About 20% of patients with hydatidiform moles will develop malignant complications. Following molar pregnancies, the hCG titer should fall to a nondetectable level by 16 weeks in patients without persistent disease. Patients with trophoblastic disease show an abnormal curve (a titer greater than 500 IU/L) frequently by 3 weeks and usually by 6 weeks.[133, 134] A diagnosis of gestational trophoblastic disease is made when the β-HCG plateaus or rises over a 2-week period, or a continued elevation is present 16 weeks after evacuation. In the United States, the rare occurrence of this disease mandates consultation with a certified subspecialist in gynecologic oncology. Following treatment, hCG should be measured monthly for at least a year, then twice yearly for 5 years. To avoid missing the diagnosis of nonmolar trophoblastic disease, abnormal bleeding after any pregnancy should be evaluated with an hCG measurement, and all patients with elevated hCG levels and early pregnancy losses should be followed with serial hCG testing.

Choriocarcinoma is associated with the increased secretion of β-hCG that is glycosylated to a greater degree, so-called hyperglycosylated hCG, sometimes called invasive trophoblast antigen.[135, 136] Hyperglycosylated hCG detected in mothers in the first weeks of normal pregnancies is the major circulating form of hCG, but the levels decrease rapidly to be replaced by the usual hCG isoform by the second trimester.[137] These findings suggest that hyperglycosylated hCG plays a role in regulating trophoblastic invasion; it is suggested that hyperglycosyalted hCG is mainly autocrine in its activity, whereas regular hCG functions as a classic hormone in maintaining the corpus luteum. Measurement of hyperglycosylated hCG in the first weeks of pregnancy may have a role in screening for Down syndrome, but clinical uses for assays that are specific for the many isoforms of hCG have yet to emerge.[138, 139] Some of the inaccuracy associated with routine pregnancy testing, especially home pregnancy tests, can be attributed to the variability in detecting hyperglycosylated hCG.

We are just beginning to appreciate the complex heterogeneity of hCG, expressed by the many isoforms that are present in biological fluids.[139] It is likely that a specific form of hCG can eventually be linked to a specific condition, offering the possibility of clinical application. For example, an assay specific for hyperglycosylated hCG may be of clinical value in assessing implantation and the early weeks of pregnancy; a low level predicts a failing pregnancy.[140] General clinical use awaits improvements in available assays, requiring the development of pure standards and specific antibodies. Meanwhile, clinicians should keep in mind that current assays measure a pool of hCG and its various molecules.

Virtually all ectopic pregnancies are associated with detectable hCG. The hCG level increases at different rates in normal and ectopic pregnancies, and the quantitative measurement of hCG combined with pelvic ultrasonography has had an enormous impact on the diagnosis and management of ectopic pregnancy. This important clinical problem is discussed fully in Chapter 33. The contributions of hCG measurement can be summarized as follows:

1. The quantitative measurement of hCG can assess pregnancy viability. A normal rate of rise (at least a 50% increase every 2 days) usually indicates a normal pregnancy.

2. When the hCG titer exceeds 1,500–3,000 IU/L, vaginal ultrasonography should identify the presence of an intrauterine gestation.

3. Declining hCG levels are consistent with effective treatment, and persistent or rising levels indicate the presence of viable trophoblastic tissue.

With the use of modern sensitive assays, it is now appreciated that virtually all normal human tissues produce the intact hCG molecule. hCG can be detected in the blood of normal men and women, where it is secreted in a pulsatile fashion in parallel with LH; the source of this circulating hCG is the pituitary gland, perhaps a consequence of evolution when hCG was derived from LH.[141–144] The concentration of this pituitary hCG normally approximates the sensitivity of the usual modern assay, and for this reason, many laboratories will not report the presence of hCG unless the level is 10 IU/L or higher. hCG produced in sites other than the placenta has little or no carbohydrate; therefore, it has a very short half-life and is rapidly cleared from the circulation. Significant levels of free α subunit are also present in the circulation of healthy individuals; however, the levels of the β subunit are extremely low.

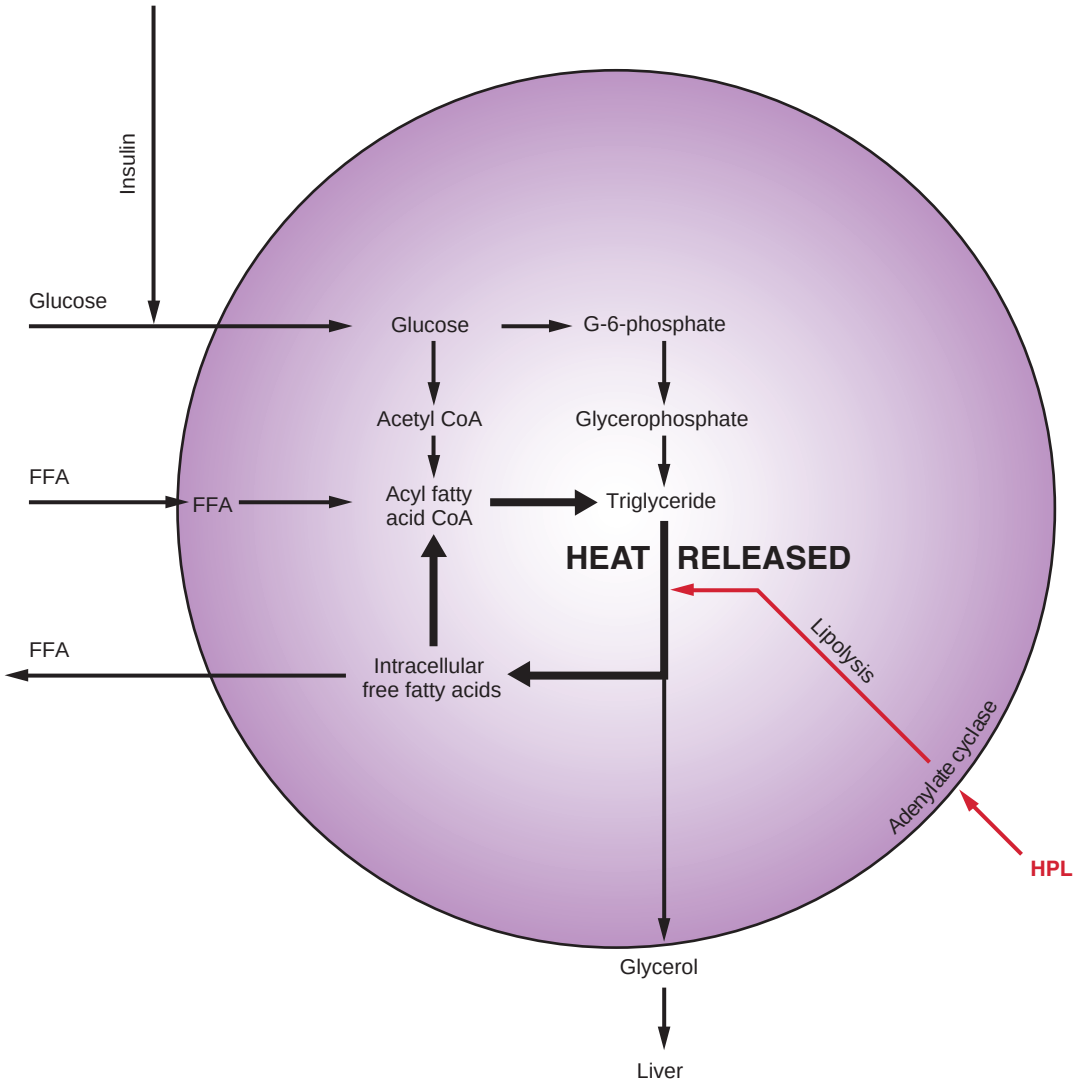
False-positive Tests for hCG. False-positive results with the hCG assay occasionally occur and have been long-recognized, resulting in inappropriate surgical or medical treatment. The level is relatively low, usually less than 150 IU/L. There are many causes, including the hCG secreted by the pituitary, but this clinical problem is usually due to interference in the assay by other substances, especially antibodies to LH or anti-animal immunoglobulins.[145] In addition, nontrophoblastic tumors can secrete detectable hCG. A false-positive result usually remains at the same level over time, neither increasing nor decreasing. When the clinical presentation is uncertain or not consistent with laboratory results (especially an absence of trophoblastic tissue), a positive hCG can be confirmed by several procedures:

1. Obtaining a similar result with a different assay method.
2. Demonstrating the presence of hCG in the urine.
3. Demonstrating parallel results with serial dilutions of the hCG standard and the patient's serum sample.

Human Placental Lactogen (hPL)

Human placental lactogen (sometimes called human chorionic somatomammotropin), also secreted by the syncytiotrophoblast, is a single-chain polypeptide of 191 amino acids held together by 2 disulfide bonds. It is very similar in structure to human growth hormone (hGH), but has only 3% of hGH somatotropin activity. The growth hormone-hPL gene family consists of 5 genes on chromosome 17q22–q24. Two genes encode for hGH, one in the pituitary and one in the placenta, and three for hPL; however, only two of the hPL genes are abundantly active in the placenta, each producing the same hPL hormone.[146] The third hPL gene does generate a protein in the placenta, but its activity is limited.[147]

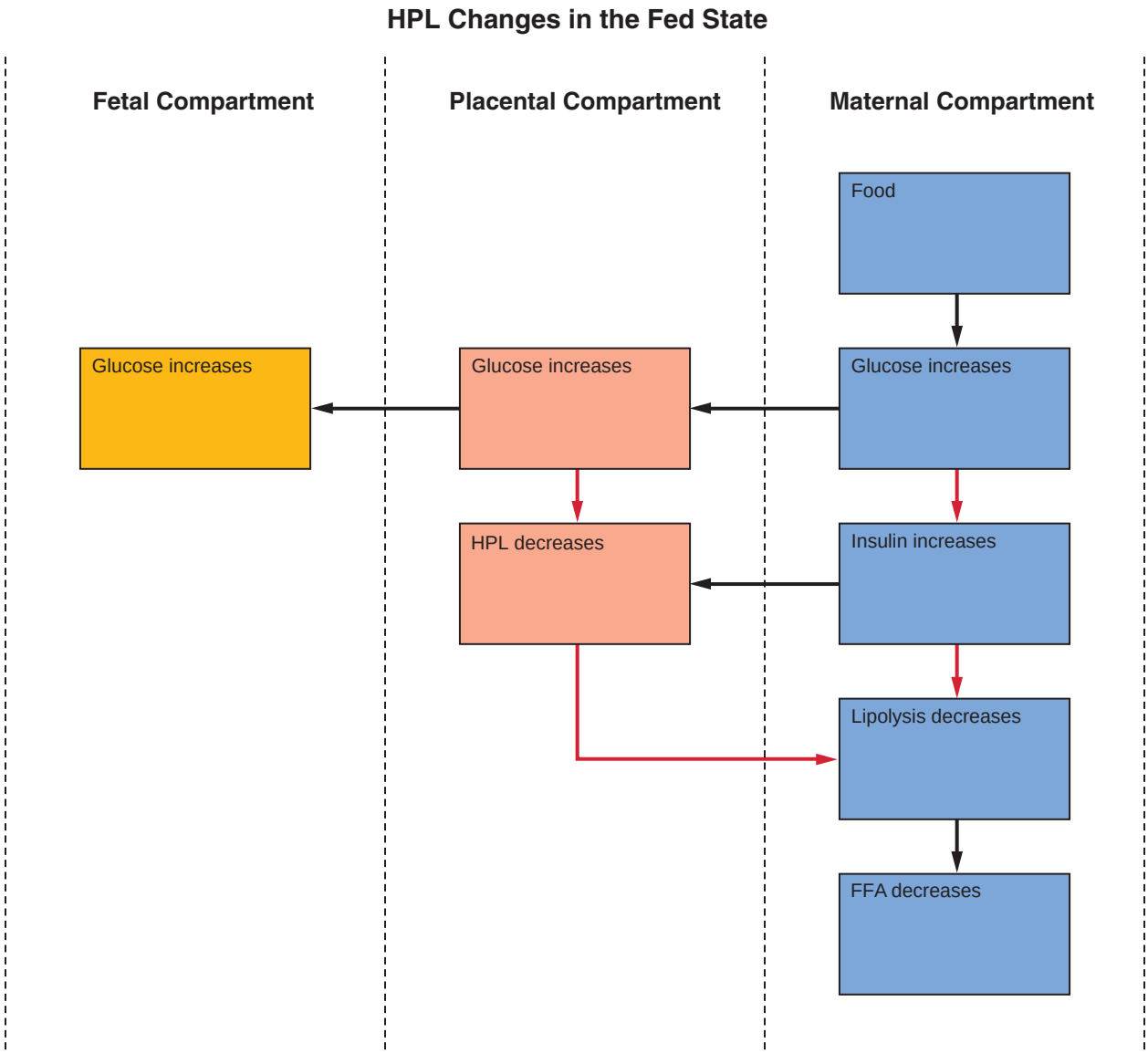
The half-life of hPL is short, about 15 minutes; hence its appeal as an index of placental problems. The level of hPL in the maternal circulation is correlated with fetal and placental weight, steadily increasing until plateauing in the last 4 weeks of pregnancy (5–10 μg/mL). There is no circadian variation, and only minute amounts of hPL enter the fetal circulation. Very high maternal levels are found in association with multiple gestations; levels up to 40 μg/mL have been found with quadruplets and quintuplets. An abnormally low level is anything less than 4 μg/mL in the last trimester.



Physiologic Function

Although hPL is similar in structure to growth hormone, neither growth hormone-releasing hormone nor somatostatin influence placental hPL secretion. One would expect the regulatory mechanism to involve placental growth factors and cytokines, as is the case with other placental steroids and peptides. Although hPL has about 50% of the lactogenic activity of sheep prolactin in certain bioassays, its lactogenic contribution in human pregnancy is uncertain.

In the mother, hPL stimulates insulin secretion and IGF-I production and induces insulin resistance and carbohydrate intolerance. However, the well-recognized insulin resistance in pregnancy is not solely an effect of hPL; for example, placental cytokines (especially TNF-α) influence this metabolic state.[148] Experimentally, the maternal level of hPL can be altered by changing the circulating level (chronically, not acutely) of glucose. hPL is elevated with hypoglycemia and depressed with hyperglycemia. This response in placental hPL may be secondary to the glucose-mediated changes in insulin levels; in vitro experiments with placental tissue indicated a decrease in hPL with a decrease in glucose, followed by an increase



in hPL after exposure to insulin.[149] This information and studies in fasted pregnant women have led to the following formulation for the physiologic function of hPL.[150–156]

The metabolic role of hPL is to mobilize lipids as free fatty acids. In the fed state, there is abundant glucose available, leading to increased insulin levels, lipogenesis, and glucose utilization. This is associated with decreased gluconeogenesis and a decrease in the circulating free fatty acid levels, because the free fatty acids are utilized in the process of lipogenesis to deposit storage packets of triglycerides (see Chapter 19, Obesity).

Pregnancy has been likened to a state of "accelerated starvation," characterized by a relative hypoglycemia in the fasting state.[153] This state is due to two major influences:

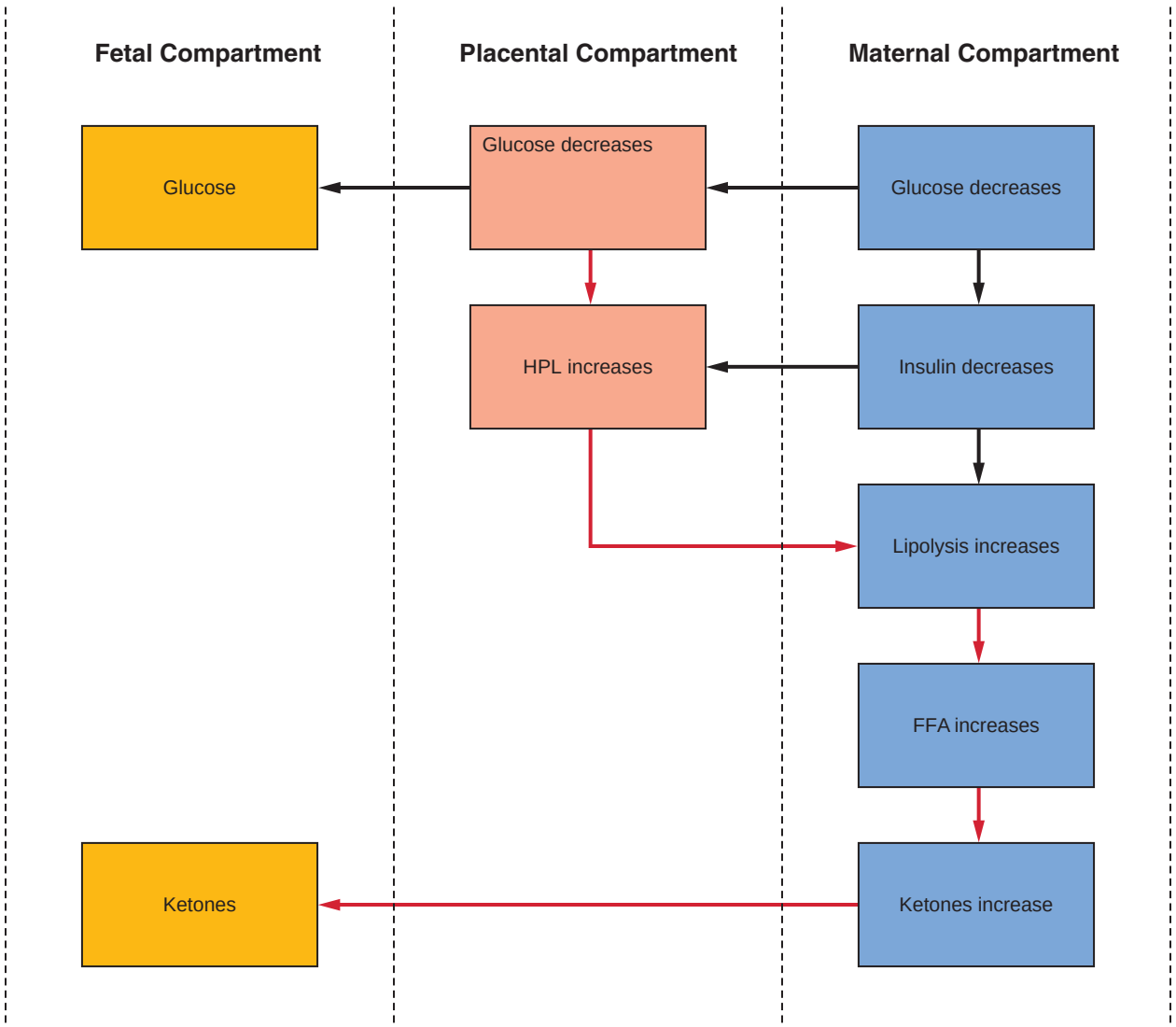
1. Glucose provides the major, although not the entire, fuel requirement for the fetus. A difference in gradient causes a constant transfer of glucose from the mother to the fetus.

2. Placental hormones, specifically estrogen and progesterone, and especially hPL, interfere with the action of maternal insulin. In the second half of pregnancy when

hPL levels rise approximately 10-fold, hPL is a major force in the diabetogenic effects of pregnancy. The latter is characterized by increased levels of insulin associated with decreased cellular response (hyperinsulinemia and peripheral insulin resistance).

As glucose decreases in the fasting state, hPL levels rise. This stimulates lipolysis leading to an increase in circulating free fatty acids. Thus, a different fuel is provided for the mother so that glucose and amino acids can be conserved for the fetus. With sustained fasting, maternal fat is utilized for fuel to such an extent that maternal ketone levels rise. There is limited transport of free fatty acids across the placenta. Therefore, when glucose becomes scarce for the fetus, fetal tissues utilize the ketones that do cross the placenta. Thus, decreased glucose levels lead to decreased insulin and increased hPL, increasing lipolysis and ketone levels. hPL also may enhance the fetal uptake of ketones and amino acids. The mechanism for the insulin antagonism by hPL may be the hPL-stimulated increase in free fatty acid levels, which, in turn, directly interfere with insulin-directed entry of glucose into cells. These interactions significantly involve growth factors, particularly insulin-like growth factor, at the cellular level.

HPL Changes in the Fasting State



This mechanism can be viewed as an important means to provide fuel for the fetus between maternal meals. However, with a sustained state of inadequate glucose intake, the subsequent ketosis may impair fetal brain development and function. Pregnancy is not the time to severely restrict caloric intake. Indeed, impaired fetal growth and development are now recognized to correlate with adverse cardiovascular risk factors and disease in adult life as well as diabetes mellitus.[47, 157, 158]

The lipid, lipoprotein, and apolipoprotein changes during pregnancy are positively correlated with changes in estradiol, progesterone, and hPL.[159] The lipolytic activity of hPL is an important factor because hPL is also linked to the maternal blood levels of cholesterol, triglycerides, phospholipids, and insulin-like growth factor-I.

When glucose is abundant, as in pregnant women with diabetes mellitus, the flow of nutritional substrates (in this case, glucose and amino acids) is in the direction of the fetus. The subsequent hyperinsulinemia in the fetus becomes a strong stimulus to growth, perhaps compounded by maternal hyperinsulinemia caused by obesity as well as the hyperinsulinemia due to the peripheral resistance produced by the hormones of pregnancy.[160] Fetal undernutrition lowers fetal IGF-I levels, and this is associated with a high prevalence of insulin resistance later as adults.[161] In vitro studies indicate that hPL, despite its lower levels in the fetus, directly affects fetal tissue metabolism, including synergistic actions with insulin, especially on glycogen synthesis in the liver. The failure of fetal growth hormone to affect fetal growth (e.g., normal growth in anencephalics) further indicates that hPL may be the fetal growth hormone.

hPL Clinical Uses

Blood levels of hPL are related to placental function. Some studies indicated that hPL was valuable in screening patients for potential fetal complications, but others did not support the use of hPL measurements. Although use of the hPL assay can have an impact on perinatal care, fetal heart rate-monitoring techniques are more reliably predictive and sensitive for assessing fetal well-being. Furthermore, totally uneventful pregnancies have been reported, despite undetectable hPL.[162, 163]

Previous suggestions that a low or declining level of hPL and a high level of hCG are characteristic of trophoblastic disease were not accurate. Because of the rapid clearance of hPL (half-life of about 20 minutes), aborting molar pregnancies are likely to have low levels of hPL, whereas the level of hCG is still high. However, intact molar pregnancies can have elevated levels of both hPL and hCG.[164]

Human Chorionic Thyrotropin (hCT)

The human placenta contains two thyrotropic substances. One is called human chorionic thyrotropin (hCT), which is similar in size and action to pituitary TSH. The content in the normal placenta is very small, and it is unlikely that it has any physiologic importance. hCT differs from the other glycoproteins in that it does not appear to share the common α subunit. Antiserum generated to α-hCG does not neutralize the biologic activities of hCT, but it does neutralize that of hCG and pituitary TSH.

Rarely, patients with trophoblastic disease have hyperthyroidism, and even more rarely, thyroid storm.[165] hCG has intrinsic thyrotropic activity, indicating that hCG is the second placental thyrotropic substance.[166–168] It has been calculated that hCG contains approximately

1/4,000th of the thyrotropic activity of human TSH. In conditions with very elevated hCG levels, such as hydatidiform mole, the thyrotropic activity can be sufficient to produce hyperthyroidism (with elevated free thyroxine, FT4, but suppressed levels of thyroid-stimulating hormone, TSH), and this can even be encountered in normal pregnancy.[169] Another very rare cause of hyperthyroidism in pregnancy is an inherited mutation of the thyrotropin receptor that makes it hypersensitive to hCG.[170]

There is a correlation between elevated thyroid function and hyperemesis gravidarum.[171, 172] Hyperemesis gravidarum is usually associated with very high hCG levels, and some of these patients develop hyperthyroidism as well.[173] Although free T4 will be elevated and TSH suppressed, patients with gestational hyperthyroidism do not develop the clinical signs of Graves' disease and TSH-receptor antibodies, TRAb, will not be detectable. These hyperthyroid manifestations in normal pregnancies may be linked to a specific subpopulation of hCG molecules with greater thyrotropic bioactivity (because highly purified, standard hCG has only trivial TSH-like activity).[174] Specifically, hCG with reduced sialic acid content is increased in pregnant patients with hyperemesis and hyperthyroidism.[175] The thyroid hormone changes in pregnancy and the role of hCG as a thyroid stimulator are also discussed in Chapter 20.

Human Chorionic Adrenocorticotropin

The rise in maternal free cortisol and aldosterone that takes place throughout pregnancy is due to placental ACTH and corticotropin-releasing hormone (CRH) production and secretion into the maternal circulation and due to the effects of estrogen and progesterone on the maternal hypothalamic-pituitary axis.[176–178] The placental content of ACTH is higher than can be accounted for by the contribution of sequestered blood. In addition, cortisol levels in pregnant women are resistant to dexamethasone suppression, indicating that there is a component of maternal ACTH and CRH that does not originate in the maternal hypothalamus and pituitary gland. The placental production of ACTH in the syncytiotrophoblast (and the increase in maternal ACTH levels) is probably due to stimulation by the locally produced CRH in the cytotrophoblast.[179] Placental pro-opiomelanocortin (POMC) gene expression and ACTH content are present throughout pregnancy and increase in the weeks before term.[180] One can speculate that placental ACTH and CRH raise maternal adrenal activity in order to provide the basic building blocks (cholesterol and pregnenolone) for placental steroidogenesis. The increased activity of the maternal adrenal gland is also necessary for the expansion of maternal blood volume during pregnancy.

The maternal ACTH response to the administration of CRH during pregnancy is blunted, reflecting a high level of endogenous CRH and ACTH activity. Vasopressin stimulates ACTH secretion in the pituitary, both directly and indirectly by potentiating the action of CRH. In contrast to the blunted response to CRH during pregnancy, the ACTH response to vasopressin is increased.[181] This is further evidence that placental CRH produces a state of chronic stimulation for the maternal pituitary-adrenal axis. Thus, in contrast to nonpregnant women, CRH levels in maternal plasma are relatively high, rising in the second trimester to peak values at term.[182, 183] In contrast to the hypothalamic-pituitary axis, placental CRH and ACTH are not suppressed by glucocorticoids, and, therefore, maternal ACTH levels are little affected by corticosteroid administration to the mother. Oxytocin is a potent stimulator of CRH and ACTH placental production, a logical mechanism to meet the stress of labor and delivery. Binding of CRH with the CRH-binding protein blunts physiologic response, but the binding protein capacity is reached late in pregnancy, increasing the biological activity of CRH and further increasing cortisol availability during labor and delivery.[184]

Both maternal and fetal levels of CRH are further elevated in pathologic states such as premature labor, hypertension, fetal asphyxia, and intrauterine growth retardation.[185] Because

CRH also stimulates prostaglandin synthesis in the placenta and fetal membranes, it is implicated in the premature labor that accompanies pathologic conditions.[186]

Growth Hormone, Growth Hormone-Releasing Hormone, and Somatostatin

One of the two growth hormone genes on chromosome 17 is expressed only in the syncytiotrophoblast of the placenta, the other is expressed in the pituitary.[147, 187] The placental growth hormone is not identical to pituitary growth hormone, differing in 13 amino acids, and after 15–20 weeks of pregnancy, placental growth hormone gradually replaces pituitary growth hormone in the maternal circulation.[147, 188] Indeed, by term, maternal pituitary growth hormone is undetectable. Placental growth hormone is not present in fetal blood. The changes in maternal levels of insulin-like growth factors and insulin-like growth factor binding proteins reflect regulation by this placental growth hormone.[189] Maternal IGF-I levels in the circulation increase during pregnancy in a pattern similar to that of placental growth hormone. Placental growth hormone is not regulated by placental growth hormone-releasing hormone but responds inversely to maternal glucose and insulin levels, protecting glucose availability for the fetus.[147, 190] Placental growth hormone can also stimulate gluconeogenesis and lipolysis in maternal organs. ***It is believed, therefore, that placental growth hormone influences fetal growth by affecting maternal metabolism***. Placental growth hormone and maternal IGF-1 levels are lower in pregnancies with intrauterine growth retardation and higher in women with female fetuses.[189, 191] Maternal circulating levels of placental growth hormone are higher at midgestation in pregnancies with fetal Down syndrome.[192]

Alpha-Fetoprotein

Alpha-fetoprotein (AFP) is a relatively unique glycoprotein (590 amino acids and 4% carbohydrate) derived largely from fetal liver and partially from the yolk sac until it degenerates at about 12 weeks. In early pregnancy (5–12 weeks), amniotic fluid AFP is mainly from yolk sac origin, whereas maternal circulating AFP is mainly from the fetal liver.[193] Its function is unknown, but it is comparable in size to albumin and contains 39% sequence homology; it may serve as a protein carrier of steroid hormones in fetal blood. AFP may also be a modulator of cell proliferation, synergizing with various growth factors.[194]

Peak levels of AFP in the fetal blood are reached at the end of the first trimester; then levels decrease gradually until a rapid decrease begins at 32 weeks. Maternal blood levels are much lower than fetal levels, rising until week 32 (probably because of the great increase in trophoblast villous surface area during this time period) and then declining. Because AFP is highly concentrated in the fetal central nervous system, abnormal direct contact of CNS with the amniotic fluid (as with neural tube and abdominal wall defects) results in elevated amniotic fluid and maternal blood levels. Other fetal abnormalities, such as intestinal obstruction, omphalocele, and congenital nephrosis, are also associated with high levels of AFP in the amniotic fluid. Besides indicating a variety of fetal anomalies, elevated maternal AFP levels are also present with multiple pregnancies and associated with an increased risk of spontaneous miscarriage, stillbirth, preterm birth, preeclampsia, neonatal death, and low birth weight (probably reflecting an increase in villous surface area in response to an adverse intrauterine environment).[195–197] Conversely, very low maternal AFP levels are associated with large birth weight infants, miscarriage, and stillbirth.[197, 198]

Multiple Marker Screening

Down syndrome is a very common genetic cause of abnormal development. The majority of cases are due to trisomy 21, an extra chromosome usually due to nondisjunction in maternal meiosis. A low maternal level of AFP is associated with trisomy 21. However, there is extensive overlap between normal and affected pregnancies responsible for a significant false-positive rate. Several placental products are secreted in increased amounts in pregnancies with trisomy 21, including hCG and hPL, whereas the maternal circulating level of unconjugated estriol is lower in affected pregnancies. The free β subunit of hCG usually circulates in low concentrations, but in the presence of a fetus with Down syndrome, the levels are high. With trisomy 18, all markers are decreased. Modern screening for fetal aneuploidy combines three markers: AFP, β-hCG, and unconjugated estriol.[199–201] This protocol will detect 85% of open neural tube defects and 80% of Down syndrome, if gestational age is determined by ultrasonography.[202] However, Down syndrome represents only about 50% of the chromosomal abnormalities that can be detected.

The multiple marker screening protocol measures AFP, unconjugated estriol, and hCG in maternal serum at 16–18 weeks gestation, the optimal time for neural tube defect detection. Using the patient's age and the laboratory results, patients are provided a statistical estimation of risks for both neural tube defects and Down syndrome. Corrections are applied for race and weight. A pattern similar to that of Down syndrome has also been reported to be associated with hydropic fetal Turner syndrome.[203]

Triple Test Values

	AFP	*Estriol*	*hCG*
Down Syndrome	Low	Low	High
Trisomy 13	Normal	No Data	Low
Trisomy 18	Low	Low	Low
Open Neural Tube Defects	High	Normal	Normal
IUGR, preterm birth, stillbirth	High	No Data	No Data
Multiple Gestation	High	High	High

The most critical factor for correct risk assessment is accurate gestational dating. A 2-week error in dating can change the calculated risk for Down syndrome 10-fold. Therefore, ultrasound confirmation of gestational dating is essential. In addition, ultrasonography will indicate fetal number (multiple pregnancies are associated with higher marker values) and assess the fetus and placenta for anomalies. Indeed, protocols are in current operation that include ultrasonography for biometric measurements (nuchal translucency, absence of a nose bone) combined with hormone markers as well as substances such as inhibin-A and pregnancy-associated plasma protein A.[204, 205] Protocols are also being developed that use these markers to predict fetal loss.[206] The combination of hormonal measurements with ultrasonography allows earlier antenatal screening, even in the first-trimester, and multiple tests reduce the false-positive rate.[207, 208] Assessment of uterine artery pulsatility by Doppler ultrasonography further adds to evaluation accuracy. First-trimester screening is increasingly being emphasized.

The multiple marker protocol is for screening a low-risk population regardless of age, and amniocentesis or chorionic villus sampling is necessary for final diagnosis. Genetic amniocentesis or chorionic villus sampling has been the standard recommendation for older women; however, although multiple marker screening does not detect all chromosomal abnormalities, it is now strongly argued that the detection rate is so high that multiple

marker screening with ultrasonography should be offered even to younger women, and a decision for amniocentesis or chorionic villus sampling is then based on the estimated risk for an abnormal fetus.[200, 209]

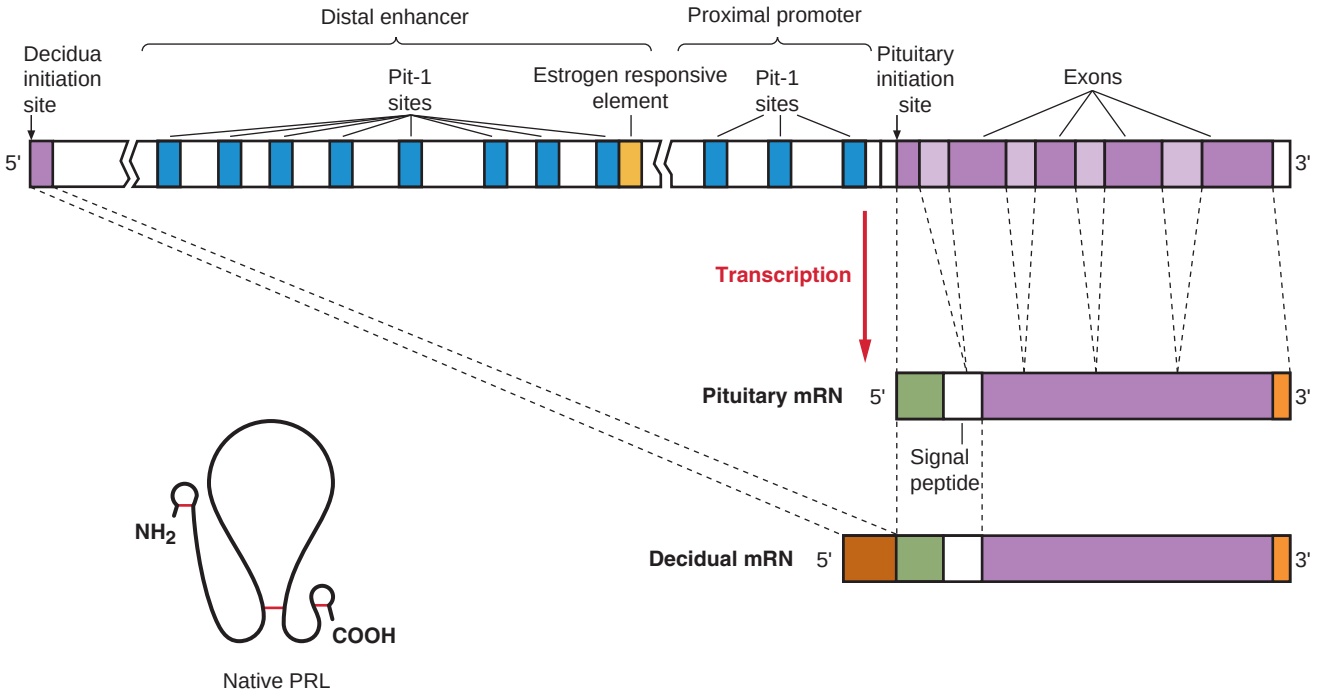
Relaxin

Relaxin is a peptide hormone produced by the corpus luteum of pregnancy, which is not detected in the circulation of men or nonpregnant women. A collection of related peptides compose the relaxin family, encoded by three relaxin genes and four insulin-like peptide genes. These peptides are similar in structure to insulin, composed of two short peptide chains (24 and 29 amino acids, respectively) linked by disulfide bridges. Although it has been argued that the human corpus luteum is the sole source of relaxin in pregnancy, it has also been identified in human placenta, decidua, and chorion.[210–212] The maternal serum concentration rises during the first trimester when the corpus luteum is dominant and declines in the second trimester.[213] This suggests a role in maintaining early pregnancy, but its function is not really known. In animals, relaxin softens the cervix (ripening), inhibits uterine contractions, and relaxes the pubic symphysis; however, relaxin levels do not correlate with changes in peripheral joint laxity in human pregnancy.[214] The animal cervical changes are comparable to those seen with human labor, and in in vitro studies of human cervical stromal cells, relaxin induces changes consistent with clinical ripening.[215, 216] Human relaxin binds primarily to relaxin receptors in the decidua and chorionic cytotrophoblast.[217] Expression of the relaxin receptors is greatest before term, and reduced after labor.[218, 219] Relaxin, originating in the decidua and binding to its receptors in the fetal membranes, increases cytokine levels that can activate matrix metalloproteinases and lead to rupture of the membranes and labor.[220] A search for important roles for relaxin in human pregnancy continues, but these roles appear to be facilitatory, not mandatory.

To examine the contribution of the corpus luteum, normally pregnant women were compared with women pregnant with donated oocytes (and therefore without corpora lutea).[221] Circulating relaxin was undetectable in the women without functioning ovaries, confirming that its major source is the corpus luteum. No effect on prolactin secretion was observed, but it did appear that relaxin enhanced growth hormone secretion by the pituitary. Obviously, relaxin is not necessary for the maintenance of pregnancy and labor because the rest of pregnancy and the outcomes did not differ between those women with circulating levels of relaxin and those with undetectable levels. However, recombinant relaxin and drugs targeted to the relaxin receptors are being tested for clinical applications.[222, 223] Thus far, potential uses include a decrease in collagen formation, an increase in vasodilation, an increase in vascular endothelial growth factor, and the release of histamine. In monkey studies, relaxin treatment combined with estrogen is effective for cervical ripening.[224]

Prolactin

Following ovulation, the endometrium becomes a secretory organ and remains so throughout pregnancy. Decidualized endometrium secretes renin, which may be involved in the regulation of water and electroytes in the amniotic fluid, and relaxin, which may influence prostaglandin production in the membranes. One of the best studied special endocrine functions of the decidual endometrium is the secretion of prolactin. Prolactin is synthesized by endometrium during a normal menstrual cycle, but this synthesis is not initiated until histologic decidualization begins about day 23.[225, 226] The control of prolactin secretion by decidual tissue has not been definitively established. Some argue that once decidualization



is established, prolactin secretion continues in the absence of either progesterone or estradiol, although there is evidence for an inhibitory feedback by decidual proteins (perhaps prolactin itself).[225, 227] Others indicate that endometrial prolactin production requires the combined effects of progestin and estrogen hormones plus the presence of other placental and decidual factors, including relaxin, IGF-I, and specific stimulatory and inhibitory proteins.[228] Indeed, human decidual cells express a prolactin-releasing peptide that stimulates prolactin secretion.[229] It is recognized, however, that transcriptional regulation of the prolactin gene in the decidua is not identical to that in the pituitary and that unique transcription factors are involved.[230] Prolactin, hPL, and growth hormone bind to the same receptor that activates the JAK/Stat signaling pathway resulting in tyrosine phosphorylation and activation of transcription factors.

During pregnancy, prolactin secretion is limited to the fetal pituitary, the maternal pituitary, and the uterus. Neither trophoblast nor fetal membranes synthesize prolactin, but both the myometrium and endometrium can produce prolactin. The endometrium requires the presence of progesterone to initiate prolactin, whereas progesterone suppresses prolactin synthesis in the myometrium. Prolactin derived from the decidua is the source of prolactin found in the amniotic fluid.[231] The prolactin in the fetal circulation is derived from the fetal pituitary. Decidual prolactin is transcribed by a gene with an additional exon compared with the pituitary, accounting for a different system of regulation.[232]

During pregnancy, prolactin maternal blood levels rise from the normal level of 10–25 ng/mL to high concentrations, beginning about 8 weeks and reaching a peak of 200–400 ng/mL at term.[233, 234] The increase in prolactin parallels the increase in estrogen beginning at 7–8 weeks gestation, and the mechanism for increasing prolactin secretion is believed to be estrogen suppression of the hypothalamic prolactin-inhibiting factor, dopamine, and direct stimulation of prolactin gene transcription in the pituitary.[235, 236] There is marked variability in maternal prolactin levels in pregnancy, with a diurnal variation similar to that found in nonpregnant persons. The increase in maternal levels of prolactin represents maternal pituitary secretion in response to estrogen as the fetus prepares the mother for breastfeeding. The mechanisms for pituitary secretion of prolactin are discussed in Chapters 2, 5, and 16.

Amniotic fluid concentrations of prolactin parallel maternal serum concentrations until the tenth week of pregnancy, rise markedly until the 20th week, and then undergo a decrease until delivery. The maternal and fetal blood levels of prolactin are derived from the respective pituitary glands, and, therefore, dopamine agonist suppression of pituitary secretion of prolactin throughout pregnancy produces low maternal and fetal blood levels, yet there is normal fetal growth and development, and amniotic fluid levels are unchanged.[237] Fortunately, decidual secretion of prolactin is unaffected by dopamine agonist treatment because decidual prolactin is important for fluid and electrolyte regulation of the amniotic fluid. This decidual prolactin is transported across the membranes in a process that requires the intact state of amnion and chorion with adherent decidua. The prolactin receptor is expressed in fetal and maternal tissues in the following descending order of intensity: chorionic cytotrophoblast, decidua, amnion, and syncytiotrophoblast.[238] This molecular expression is consistent with local actions.

No clinical significance can be attached to maternal and fetal blood levels of prolactin in abnormal pregnancies. Decidual and amniotic fluid prolactin levels are lower, however, in hypertensive pregnancies and in patients with polyhydramnios.[239, 240] Prolactin receptors are present in the chorion laeve, and their concentration is lower in patients with polyhydramnios.[241] Prolactin reduces the permeability of the human amnion in the fetal to maternal direction. This receptor-mediated action takes place on the epithelium lining the fetal surface.[242] There is also evidence that prolactin derived from the fetal pituitary contributes to the regulation of fetal water and electrolyte balance by acting as an antidiuretic hormone.[243]

Cytokines and Growth Factors

The placenta synthesizes many proteins that are part of the normal composition of cells throughout the body. Local placental cytokine production is believed to be important for embryonic growth and in the maternal immune response essential for survival of the pregnancy.[244] Interleukin-1β is produced in the decidualized endometrium during pregnancy, and colony-stimulating factor-1 (CSF-1) is produced by both decidua and placenta. CSF-1 gene expression in response to interleukin-1β has been localized to mesenchymal fibroblasts from the core of placental villi.[245] Thus, a system of communication is present between maternal decidual and fetal tissue to provide growth factor support for the placenta that would include fetal hematopoiesis, a known response to CSF-1. The placenta also produces interleukin-6, and both interleukins stimulate hCG release by activation of the interleukin-6 receptor.[246] Thus, the interleukin-1 influence on hCG secretion is mediated by the interleukin-6 system. Both trophoblast-derived interleukin-1 and tumor necrosis factor-α (TNF-α) synergistically release interleukin-6 and activate the interleukin-6 system to secrete hCG.[247] Interferons and their receptors are present in virtually all cells, and thus, it is not surprising that they are found in the tissues of pregnancy.

The insulin-like growth factors, IGF-I and IGF-II, are involved in prenatal and postnatal growth and development. These growth factors do not cross the placenta into the fetal circulation; however, they may be involved in placental growth.[248] The maternal levels of IGF-I are significantly regulated by growth hormone-dependent liver synthesis. The fetus can influence maternal IGF-I levels by means of the placental secretion of hPL. An increase in maternal IGF-I levels during pregnancy with a rapid decrease after delivery indicates a significant placental influence. There is no major change in maternal IGF-II levels throughout pregnancy.

The six IGF binding proteins transport IGFs in the circulation, protect IGFs against metabolism and clearance, and, importantly, affect the biologic activity of IGFs by modulating IGF availability at the cellular level. Pregnancy is marked by a rise in maternal levels of

insulin-like growth factor binding protein-1 (IGFBP-1), beginning at the end of the first trimester and reaching a peak at term.[249, 250] IGFBP-1 is now recognized to be the same as placental protein-12, a decidual protein. Thus, IGFBP-1 originates in the decidua, regulated by progesterone, as well as in the liver. The prominence of IGFBP-1 in the pregnant state is in contrast to the nonpregnant state when IGFBP-3 is the main circulating IGFBP. During pregnancy, the levels of IGFBP-3 and IGFBP-2 decrease, apparently due to the activity of a pregnancy-associated serum protease (IGFBP-3 protease).[249] These changes would promote the bioavailability of IGF-I in maternal tissues, and this may be important in enhancing nutrient transfer from the mother to the placenta. There is evidence to indicate that the mother can alter IGFBP-3 proteolytic activity according to her nutritional state, thus increased proteolysis would decrease IGFBP-3 levels increasing the bioavailability of maternal IGF-I.[251]

In the pregnant ewe and fetal lamb, glucose and other nutritional factors regulate the gene expression and, therefore, the circulating levels of IGF binding proteins.[252] Fasting and feeding increased and decreased, respectively, the IGFBP concentrations, perhaps partly a response to insulin levels and the effect of insulin on liver synthesis of IGFBPs. These changes are consistent with IGF and IGFBP involvement in the responses to nutrition and stress. Because IGFBP-1 appears to be the principal binding protein in pregnancy, attention is focused on the changes in IGF-I and IGFBP-1. IGF-I, produced in the placenta, regulates transfer of nutrients across the placenta to the fetus and, thus, enhances fetal growth; IGFBP-1, produced in the decidua, interferes with IGF-I action and inhibits fetal growth.[253] Thus, newborn birth weight correlates directly with maternal levels of IGF-I and inversely with levels of IGFBP-1.

Intrauterine growth retardation is associated with reduced fetal blood levels of IGF-I and IGFBP-3 and increased levels of IGFBP-1 and IGFBP-2.[254] In view of the strong relationship between the IGF system and fetal nutrition, it is logical that fetal glucose availability and insulin are the principal regulating agents. In experimental animals, an increase in fetal insulin or glucose elevates IGF-I levels, whereas nutritional restriction causes an increase in IGFBP-1 and IGFBP-2 and a decrease in IGFBP-3.[255] Insulin is believed to influence growth by promoting cellular uptake of nutrients and by increasing IGF-I production. The fetal blood levels of IGF-II parallel those of IGF-I, and IGF-II promotes fetal growth by means of the IGF-I receptor. IGF-II appears to be important early in embryonic growth, and then after organ development is complete, IGF-I becomes the dominant factor.

Epidermal growth factor (EGF) is present in both cytotrophoblast and syncytiotrophoblast, but more intensely in syncytiotrophoblast, and probably is involved in the differentiation of cytotrophoblast into syncytiotrophoblast. EGF is well known as a mitogen. Other growth factors isolated from human placenta include platelet-derived growth factor, nerve growth factor, fibroblast growth factor, and transforming growth factors. These factors are probably all involved in the proliferation and growth associated with pregnancy.

Inhibin, Activin, and Follistatin

The placenta produces inhibin, which is responsible for the marked increase in maternal inhibin levels throughout pregnancy.[256, 257] Inhibin-A is the principal bioactive inhibin secreted during pregnancy, rising in the maternal circulation at the time of the emergence of placental function, peaking at 8 weeks gestation, and then decreasing before increasing again in the third trimester to reach a level at term that is 100 times greater than that during the normal menstrual cycle.[258–261] Undoubtedly, the high levels of inhibin and estrogen during pregnancy account for the profound suppression of maternal gonadotropins. Trophoblastic inhibin synthesis is inhibited by activin-A and stimulated by hCG, GnRH, epidermal growth factor, transforming growth factor-α, and PGE_2 and $PGF_{2\alpha}$, the

major placental prostaglandins.[257] Activin-A, the major trophoblastic activin product, also increases in the maternal circulation, with elevated but stable levels from 8 to 24 weeks, and then increasing to reach a level at term that is also 100 times greater than that during the normal menstrual cycle.[262]

Similar to their action in the ovarian follicle, inhibin and activin are regulators within the placenta for the production of GnRH, hCG, and steroids; as expected, activin is stimulatory, and inhibin is inhibitory.[118] GnRH and the subunits for inhibin and activin can be found in the same placental cells, in both cytotrophoblast and syncytiotrophoblast.[263] The maternal levels of inhibin-B are very low throughout pregnancy; however, inhibin-B is significantly expressed in the amnion where it is believed to influence prostaglandin synthesis.[264] Trophoblast synthesis and release of inhibin and activin are part of the complex placental story, involving many hormones and locally produced factors. The placental and decidual appearance of inhibin and activin occurs early in pregnancy in time for possible roles in embryogenesis and local immune responses. Higher levels of activin-A are found at mid-gestation in women who subsequently develop preeclampsia.[197, 265]

Follistatin is the activin-binding protein expressed in placenta, membranes, and decidua.[266] Because follistatin binds activin, it antagonizes the stimulatory effects of activin on placental steroid and peptide production.

Endogenous Opiates

Fetal and maternal endogenous opiates originate from the pituitary glands and are secreted in parallel with ACTH, in response to corticotropin-releasing hormone, which is, in part, derived from the placenta.[267] There is reason to believe that in pregnancy the intermediate lobe of the maternal pituitary gland is a major source of elevated circulating endorphin levels. However, the syncytiotrophoblast in response to CRH produces all of the products of proopiomelanocortin (POMC) metabolism, including β-endorphin, enkephalins, and dynorphins. The placenta and membranes are richly endowed with G protein opioid receptors.[268] The presence of CRH in the placenta and placental opiate production in response to CRH and oxytocin indicate an interaction similar to that in the hypothalamic-pituitary axis.[269]

It is not certain whether maternal blood levels of endogenous opiates increase with advancing gestation.[87] However, a marked increase in maternal values is reached during labor, coinciding with full cervical dilation. The maternal levels also correlate with the degree of pain perception and use of analgesia. On the fetal side, hypoxia is a potent stimulus for endorphin release.

There are many hypotheses surrounding the function of endogenous opiates in pregnancy. These include roles related to stress; inhibition of oxytocin, vasopressin and gonadotropins; the promotion of prolactin secretion; and, of course, a natural analgesic agent during labor and delivery.

The Renin-Angiotensin System

The maternal circulating levels of prorenin, the inactive precursor of renin, increase 10-fold during early pregnancy, the result of ovarian stimulation by hCG.[270, 271] This increase in prorenin from the ovary is not associated with any significant change in the blood levels of the active form, renin. Possible roles for this ovarian prorenin-renin-angiotensin system

include the following: stimulation of steroidogenesis to provide androgen substrate for estrogen production, regulation of calcium and prostaglandin metabolism, and stimulation of angiogenesis. This system may affect vascular and tissue functions both in and outside the ovary. Prorenin also originates in chorionic tissues and is highly concentrated in the amniotic fluid. The highest biologic levels of prorenin are found in gestational sacs in early pregnancy; its possible roles in embryonic growth and development remain speculative.[271] Renin and angiotensinogen (the renin substrate) are expressed by the following fetal tissues: chorion, amnion, and placenta.[272] This system responds to a variety of factors, affecting vascular resistance and blood volume.[273] Maternal renin activity is increased 4-fold by mid-gestation, partly a response to an estrogen-induced increase in angiotensinogen but largely a compensatory response to maintain blood pressure in the presence of vasodilation.[274] There is no evidence that fetal or uterine prorenin or renin contribute to the maternal circulation.

Atrial Natriuretic Peptide

Atrial natriuretic peptide (ANP) is derived from human atrial tissue and the placenta.[275] It is a potent natriuretic, diuretic, and smooth muscle-relaxant peptide that circulates as a hormone. Maternal ANP increases in the third trimester and during labor, and cord levels on the arterial side suggest that ANP is a circulating hormone in the fetus.[276] In the mother, ANP release is stimulated by atrial stretch, and this is another mechanism for regulating the volume and electrolyte changes associated with pregnancy and delivery.[277] ANP regulates water and electrolyte balance in the fetus as well, and increased amniotic fluid and maternal blood second-trimester levels of ANP have been reported in the presence of fetal cardiac malformations.[278] In knockout mice without the natriuretic peptide receptor, hearts are enlarged, and surviving adults have hypertension and cardiac hypertrophy.[279] ANP belongs to a family of natriuretic peptides that have been found in the human uterus. ANP is secreted by myometrial cells and exerts a suppressive effect on myometrial contractions; it is speculated that the expanding uterus may release ANP just as the heart does when the atrium is stretched.[280]

Other Proteins

The mother responds to a pregnancy even before implantation. Remarkably, early pregnancy factor (EPF) can be detected in the maternal circulation within 1–2 days after coitus results in a pregnancy.[281] It remains throughout pregnancy but, interestingly, disappears before parturition. EPF prior to implantation is apparently produced by the ovary in response to a signal from the embryo. After implantation, EPF is no longer secreted by the ovary but now is derived from the embryo. EPF is a protein associated with cell proliferation and growth and, therefore, is present in many nonpregnant tissues such as neoplasms. EPF has immunosuppressive properties and is abundant in platelets.

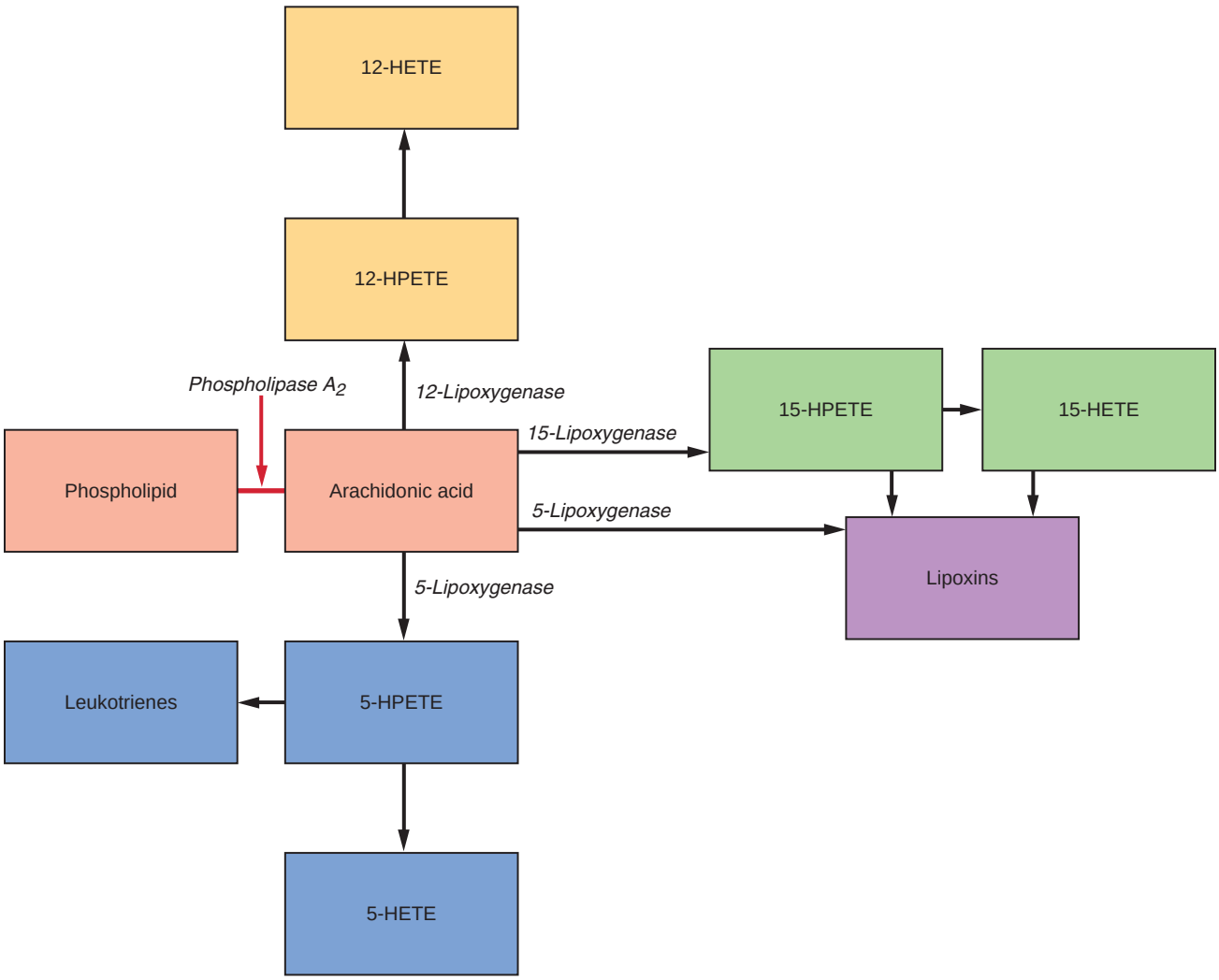
Pregnancy-specific γ_1-glycoprotein (PSG) was previously known as Schwangerschaftsprotein 1. The physiologic function of PSG produced by the placenta is unknown, but it has been used as a test for pregnancy and a marker for malignancies, including choriocarcinoma. Molecular studies have revealed that PSG consists of a family of glycoproteins encoded by genes on chromosome 19.[282] The PSG family is closely related to the carcinoembryonic antigen (CEA) proteins. Pregnancy-associated plasma protein-A (PAPP-A) is a placental protein that is similar to a macroglobulin in the serum, and investigators are still in search of specific functions. Low levels of PAPP-A in the first trimester are associated with adverse obstetrical outcomes.[197] Progesterone-associated endometrial protein, previously called placental protein 14, is now recognized to originate in secretory endometrium and

decidua. No role for this protein has been described thus far. Neuropeptide Y, a peptide extensively distributed in the brain, is found in trophoblast, membranes, and decidua, with higher but nonchanging maternal blood levels during pregnancy.[87]

Prostaglandins

Prostaglandin Biosynthesis

Prostaglandins are autocrine and paracrine factors produced in almost all cells in the human body. The family of prostaglandins with the greatest biologic activity is that having two double bonds, derived from arachidonic acid.[283, 284] Arachidonic acid can be obtained from two sources, directly from the diet (from meats) or by formation from its precursor linoleic acid, which is found in vegetables. In the plasma, 1–2% of the total free fatty acid content is free arachidonic acid. The majority of arachidonic acid is covalently bound in esterified form as a significant proportion of the fatty acids in phospholipids and in esterified cholesterol. Arachidonic acid is only a minor fatty acid in the triglycerides packaged in adipose tissue.



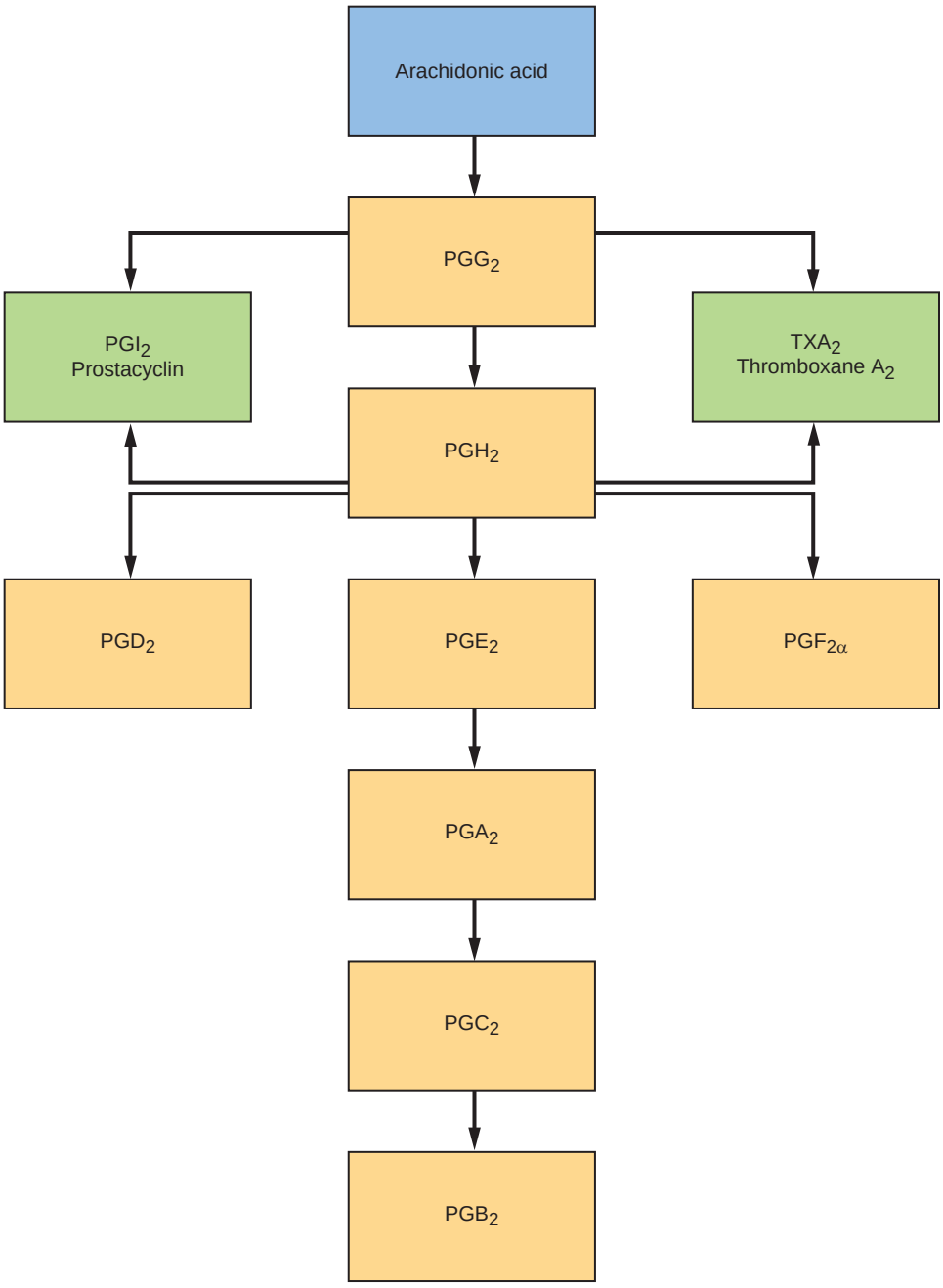
The rate-limiting step in the formation of the prostaglandin family is the release of free arachidonic acid. A variety of hydrolases may be involved in arachidonic acid release, but phospholipase A_2 activation is an important initiator of prostaglandin synthesis because of the abundance of arachidonate in the 2 position of phospholipids. In addition, phospholipase C activity can provide arachidonic acid. Types of stimuli that activate such lipases include burns, infusions of hypertonic and hypotonic solutions, thrombi and small particles, endotoxin, snake venom, mechanical stretching, catecholamines, bradykinin, angiotensin, and the sex steroids.

"Eicosanoids" refer to all the 20-carbon derivatives, whereas "prostanoids" indicate only those containing a structural ring. After the release of arachidonic acid, the synthetic path can go in two different directions: the lipoxygenase pathway or the cyclooxygenase (prostaglandin endoperoxide H synthase) pathway, depending on the local cellular context. There are three lipoxygenase enzymes that lead to active compounds, predominantly in inflammatory white blood cells. Arachidonic acid is first converted to hydroperoxyeicosatetraenoic acids (HPETEs) and then to hydroxyeicosatetraenoic acids (HETEs), lipoxins, or leukotrienes. The leukotrienes are formed by 5-lipoxygenase oxygenation of arachidonic acid at C-5, forming an unstable intermediate, LTA_4.[285] LTB_4 is formed by hydration and LTC_4 by the addition of glutathione. The remaining leukotrienes are metabolites of LTC_4. The previously known "slow reacting substance of anaphylaxis" consists of a mixture of LTC_4, LTD_4, and LTE_4. The leukotrienes are involved in the defense reactions of white cells and participate in hypersensitivity and inflammatory responses. LTB_4 acts primarily on leukocytes (stimulation of leukocyte emigration from the bloodstream), whereas LTC_4, LTD_4, and LTE_4 affect smooth muscle cells (bronchoconstriction in the lungs and reduced contractility in the heart). All leukotrienes increase microvascular permeability. Thus, the leukotrienes are major agonists, synthesized in response to antigens provoking asthma and airway obstruction. Leukotrienes are 100–1,000 times more potent than histamine in the pulmonary airway. Asthma is now treated with specific leukotriene receptor antagonists.

The 12-lipoxygenase pathway leads to 12-hydroxyeicosatetraenoic acid (12-HETE). Little is known about 12-HETE other than its function as a leukostatic agent. The lipoxins (LXA and LXB), products of the 5- and 15-lipoxygenase pathways, inhibit natural killer cell cytotoxicity and are vasodilators.[285]

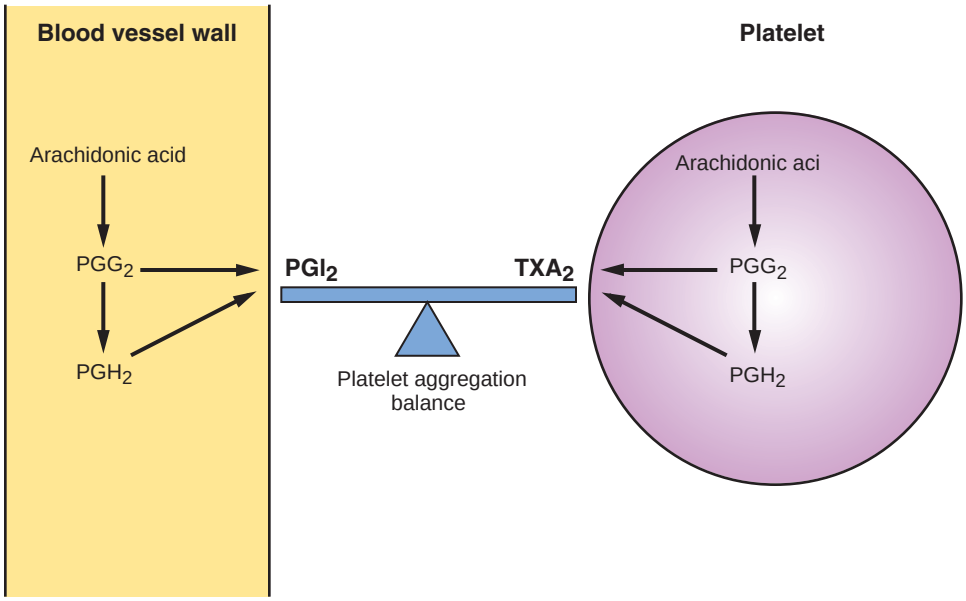
The cyclooxygenase pathway leads to the prostaglandins. The first true prostaglandin (PG) compounds formed are PGG_2 and PGH_2 (half-life of about 5 minutes), the mothers of all other prostaglandins. The numerical subscript refers to the number of double bonds. This number depends on which of the three precursor fatty acids has been utilized. Besides arachidonic acid, the other two precursor fatty acids are linoleic acid, which gives rise to the PG_1 series, and pentanoic acid, the PG_3 series. The latter two series are of less importance in physiology, hence, the significance of the arachidonic acid family. The prostaglandins of original and continuing relevance to reproduction are PGE_2 and $PGF_{2\alpha}$ and possibly PGD_2. The α in $PGF_{2\alpha}$ indicates the α steric configuration of the hydroxyl group at the C-9 position. The A, B, and C prostaglandins either have little biologic activity or do not exist in significant concentrations in biologic tissues. In the original work, the prostaglandin more soluble in ether was named PGE, and the one more soluble in phosphate (spelled with an F in Swedish) buffer was named PGF. Later, naming became alphabetical.

The cyclooxygenase enzyme (prostaglandin synthase) exists in two forms, COX-1 and COX-2, products of separate genes.[286–288] Prostacyclin is produced by COX-1, the constitutive form of the enzyme found in virtually all tissues, whereas COX-2 is induced in responses to inflammatory stimuli. COX-2 is expressed only after stimulation by various growth factors, cytokines, hormones, and endotoxins; therefore, it is called the inducible form. Thus, selective inhibition of COX-2 would possibly be therapeutically advantageous, avoiding the side effects associated with inhibition of COX-1.



Thromboxane and Prostacyclin

Thromboxanes are not true prostaglandins because of the absence of the pentane ring, but prostacyclin (PGI_2) is a legitimate prostaglandin. Thromboxane (TX) (half-life about 30 seconds) and PGI_2 (half-life about 2–3 minutes) can be viewed as opponents, each having powerful biologic activity that counters or balances the other. TXA_2 is the most powerful vasoconstrictor known, whereas PGI_2 is a potent vasodilator. These two agents also have opposing effects on platelet function. Platelets, lungs, and the spleen predominately synthesize TXA_2, whereas the heart, stomach, and blood vessels throughout the body synthesize PGI_2. The lungs are a major source of prostacyclin. Normal pulmonary endothelium makes prostacyclin whereas TXA_2 appears in response to pathologic stimuli.[289] The pulmonary release of prostacyclin may contribute to the body's defense against platelet aggregation.



Let us take a closer look at platelets. The primary function of platelets is the preservation of the vascular system. Blood platelets stick to foreign surfaces or other tissues, a process called adhesion. They also stick to each other and form clumps; this process is called aggregation. Because platelets synthesize TXA_2, a potent stimulator of platelet aggregation, the natural tendency of platelets is to clump and plug defects and damaged spots. The endothelium, on the other hand, produces PGI_2 and its constant presence inhibits platelet aggregation and adherence, keeping blood vessels free of platelet aggregates and ultimately clots. Thus, prostacyclin has a defensive role in the body. It is 4 to 8 times more potent a vasodilator than the E prostaglandins, and it prevents the adherence of platelets to healthy vascular endothelium. However, when the endothelium is damaged, platelets gather, beginning the process of thrombus formation. Even in this abnormal situation, prostacyclin strives to fulfill its protective role because increased PGI_2 can be measured in injured endothelium, thrombosed vessels, and in the vascular tissues of hypertensive animals.

It is believed that endothelial production of prostacyclin plays an important role in the impressive vasodilation that is associated with pregnancy. The placenta is a major source of thromboxane, and preeclampsia may, in part, reflect an imbalance between the vasodilator, prostacyclin, and the vasoconstrictor, thromboxane.[290]

Conditions associated with vascular disease can be understood through the prostacyclin-thromboxane mechanism. For example, atheromatous plaques and nicotine inhibit prostacyclin synthesis. Increasing the cholesterol content of human platelets increases the sensitivity to stimuli that cause platelet aggregation due to increased thromboxane production. The well-known association between low-density and high-density lipoproteins (LDL-cholesterol and HDL-cholesterol) and cardiovascular disease may also be partly explained in terms of PGI_2. LDL from men and postmenopausal women inhibits and HDL stimulates prostacyclin production.[291] Platelets from diabetic pregnant women make more TXA_2 than platelets from normal pregnant women. Smokers who use oral contraceptives have increased platelet aggregation and an inhibition of prostacyclin formation.[292] Incidentally, onion and garlic inhibit platelet aggregation and TXA_2 synthesis.[293] Perhaps the perfect contraceptive pill is a combination of progestin, estrogen, and some onion or garlic.

In some areas of the world, there is a low incidence of cardiovascular disease. This can be directly attributed to diet and the protective action of prostacyclin.[294] The diet of Eskimos

and Japanese has a high content of pentanoic acid and low levels of linoleic and arachidonic acids. Pentanoic acid is the precursor of prostaglandin products with three double bonds, and, as it happens, PGI_3 is an active agent whereas TXA_3 is either not formed, or it is inactive. The fat content of most common fish is 8–12% pentanoic acid, and more than 20% in the more exotic (and expensive) seafoods such as scallops, oysters, and caviar.

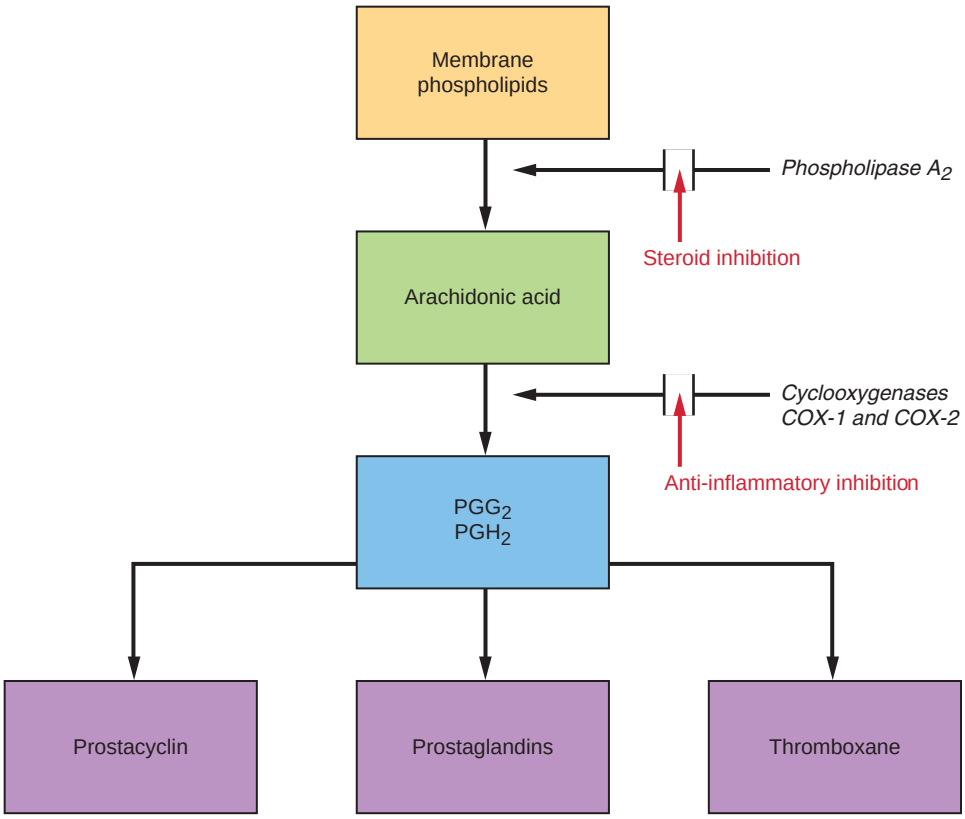
Metabolism

Prostaglandin metabolism is initiated by 15-hydroxyprostaglandin dehydrogenase. The metabolism of prostaglandins occurs primarily in the lungs, kidneys, and liver. The lungs are important in the metabolism of E and F prostaglandins. Indeed, there is an active transport mechanism that specifically carries E and F prostaglandins from the circulation into the lungs. Nearly all active prostaglandins in the circulation are metabolized during one passage through the lungs. Therefore, members of the prostaglandin family have a short half-life and, in most instances, exert autocrine/paracrine actions at the site of their synthesis. Because of the rapid half-lives, studies are often performed by measuring the inactive end products, for example, 6-keto-$PGF_{1\alpha}$, the metabolite of prostacyclin, and TXB_2, the metabolite of thromboxane A_2.

Prostaglandin Inhibition

A review of prostaglandin biochemistry is not complete without a look at the inhibition of the biosynthetic cascade of products. Corticosteroids were thought to inhibit the prostaglandin family by stabilizing membranes and preventing the release of phospholipase. It is now proposed that corticosteroids induce the synthesis of proteins called lipocortins (or annexins) which block the action of phospholipase.[295] Thus far, corticosteroids and some local anesthetic agents are the only substances known to work at this step. Because corticosteroids reduce the availability of arachidonic acid for both the lipoxygenase and cyclooxygenase pathways, they are very effective anti-inflammatory agents and anti-hypersensitivity agents, especially for the treatment of asthma.

Aspirin is an irreversible inhibitor, selectively acetylating the cyclooxygenase involved in prostaglandin synthesis. The other inhibiting agents, nonsteroidal anti-inflammatory drugs (NSAIDS) such as indomethacin and naproxen, are reversible agents, forming a reversible bond with the active site of the enzyme. Acetaminophen inhibits cyclooxygenase in the central nervous system, accounting for its analgesic and antipyretic properties, but it has no anti-inflammatory properties and does not affect platelets. However, acetaminophen does reduce prostacyclin synthesis; the reason for this preferential effect is unknown.[296] The analgesic, antipyretic, and anti-inflammatory actions of these agents are mediated by inhibition of the cyclooxygenase enzymes, COX-1 and COX-2. Aspirin, indomethacin, and ibuprofen are more potent inhibitors of COX-1 than COX-2.[297] Diclofenac, acetaminophen, and naproxen inhibit both enzymes equally. The side effects associated with each agent are a reflection of the degree of selectivity toward the two enzymes; inhibition of COX-1, the constitutive form, is associated with significant side effects, and inhibition of COX-2, the inducible form, is potentially therapeutic for pain and inflammation. Part of the anti-inflammatory activity of glucocorticoids is due to inhibition of COX-2 formation. The well-known gastric ulcerogenic side effect of anti-inflammatory drugs is due to the fact that PGE_2 protects the gastric mucosa by inhibiting gastric acid secretion, and COX-1 is the predominant enzyme in the gastric mucosa. The specific inhibitors of COX-2 are effective analgesics with a better gastrointestinal side effect profile.



Efficacy in treating dysmenorrhea is similar comparing the older agents with the newer specific COX-2 inhibitors. Theoretically, the COX-2 inhibitors should avoid the unwanted inhibition of prostaglandin activity in the process of ovulation; however, in the mouse, it is the COX-2 enzyme that is involved in ovulation, and rofecoxib delayed ovulation in a small trial in women.[298, 299]

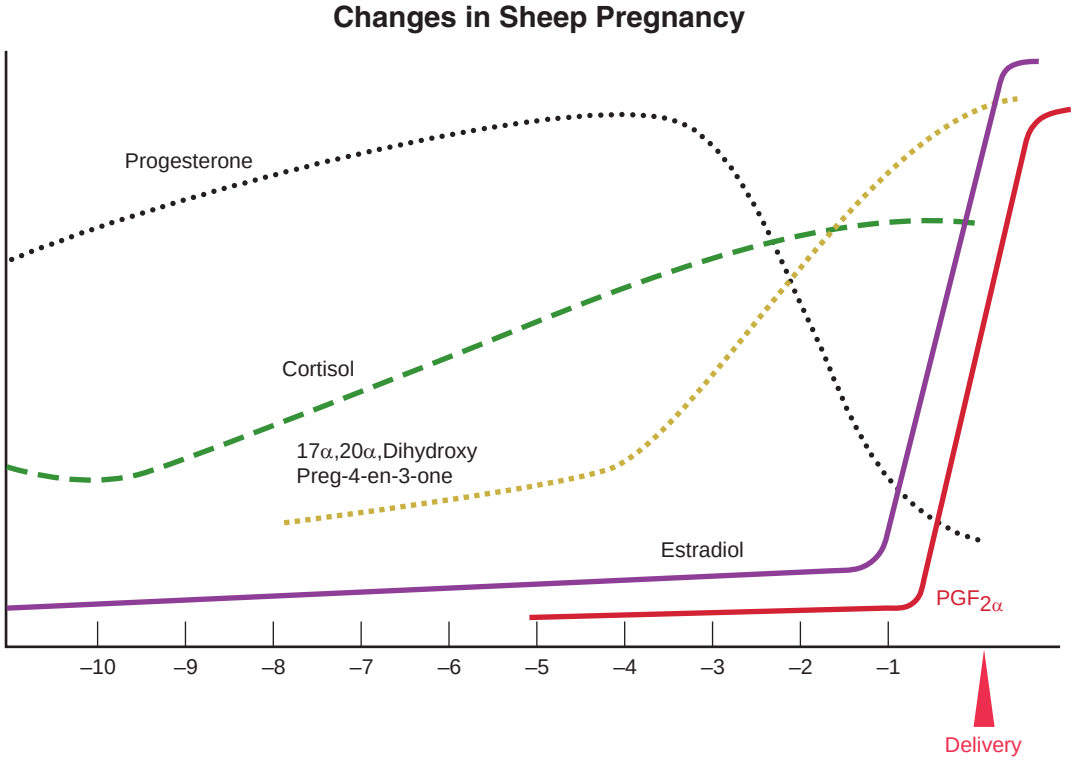
Because of the irreversible nature of the inhibition by aspirin, aspirin exerts a long-lasting effect on platelets, maintaining inhibition in the platelet for its lifespan (8–10 days). Prostacyclin synthesis in the endothelium recovers more quickly because the endothelial cells can resynthesize new cyclooxygenase. Platelets, lacking nuclei, cannot produce new enzyme, probably exclusively COX-1. The sensitivity of the platelets to aspirin may explain the puzzling results in the early studies in which aspirin was given to prevent subsequent morbidity and mortality following thrombotic events. It takes only a little aspirin to effectively inhibit thromboxane synthesis in platelets. Going beyond this dose will not only inhibit thromboxane synthesis in platelets, but also inhibit the protective prostacyclin production in blood vessel walls. Some suggest that a dose of 3.5 mg/kg (about half an aspirin tablet) given at 3-day intervals effectively induces maximal inhibition of platelet aggregation without affecting prostacyclin production by the vessel walls.[300] Others indicate that the dose that effectively and selectively inhibits platelet cyclooxygenase is 20–40 mg daily.[301, 302] The major handicap with the use of inhibitors of PG synthesis is that they strike blindly and with variable effect from tissue to tissue. Obviously, drugs that selectively inhibit TXA_2 synthesis would be superior to aspirin in terms of antithrombotic effects.

A concern with the specific COX-2 inhibitors is their inhibition of prostacyclin formation, whereas COX-1 generation of TXA_2 is unaffected. Unfortunately, arterial thrombotic events, including myocardial infarction and stroke, are about 2-fold increased in users of COX-2 inhibitors, raising appropriate caution, especially in individuals at high risk for cardiovascular disease.[303] It is not certain that non-selective, traditional non-steroidal anti-inflammatory drugs (NSAIDS) don't share in this cardiovascular risk.

The Endocrinology of Parturition

Perhaps the best example of the interplay among fetus, placenta, and mother is the initiation and maintenance of parturition. Hormonal changes in the uteroplacental environment are the principal governing factors accounting for the eventual development of uterine contractions. The sequence of events has been repeatedly reviewed in detail, where references to the original work are available.[304–310]

Extensive work in sheep has implicated the fetal pituitary-adrenal axis in normal parturition. The sequence of events in the ewe begins about 10 days prior to labor with elevation of fetal cortisol in response to fetal pituitary ACTH, in turn a response to increased release of hypothalamic CRH. Fetal adrenalectomy or hypophysectomy prolongs pregnancy, whereas infusion of ACTH or glucocorticoids into the sheep fetus stimulates premature labor. Maternal stimulation of the fetal adrenal is not a factor because in sheep (and in women) there is little or no placental transfer of maternal ACTH into the fetal circulation. Thus, parturition in the ewe is initiated by a signal in the fetal brain activating ACTH secretion.



Increased cortisol secretion by the fetal adrenal gland starts a chain of events associated with labor. The sequence of events continues in the ewe with a decline in progesterone. This change is brought about by the induction of 17α-hydroxylase, 17,20-lyase enzyme activity (P450c17) in the placenta. The up-regulation of P450c17 may be mediated by PGE_2. COX-2 activity is stimulated by cortisol, while at the same time, cortisol inhibits the activity of 15-hydroxyprostaglandin dehydrogenase. An increase in PGE_2 correlates with the increasing activity of P450c17.

Glucocorticoid treatment of sheep placental tissue specifically increases the rate of production of 17α,20α-dihydroxypregn-4-en-3-one. This dihydroxyprogesterone compound

also has been identified in sheep placental tissue obtained after spontaneous labor. Thus, direct synthesis of progesterone does not decline, but increased metabolism to a 17α-hydroxylated product results in less available progesterone. Progesterone withdrawal is associated with a decrease in the resting potential of myometrium; i.e., an increased response to electric and oxytocic stimuli. Conduction of action potential through the muscle is increased, and the myometrial excitability is increased.

Dihydroxyprogesterone also serves as a precursor for the rise in estrogen levels, which occurs a few days prior to parturition. Estrogens enhance rhythmic contractions, as well as increasing vascularity and permeability and the oxytocin response. Thus, progesterone withdrawal and estrogen increase lead to an enhancement of conduction and excitation.

The final event in the ewe is a rise in $PGF_{2\alpha}$ production hours before the onset of uterine activity. A cause-and-effect relationship between the rise in estrogen and the appearance of $PGF_{2\alpha}$ has been demonstrated in sheep. These events indicate that the decline in progesterone, the rise in estrogen, and the increase in $PGF_{2\alpha}$ are all secondary to direct induction of a placental enzyme by fetal cortisol.

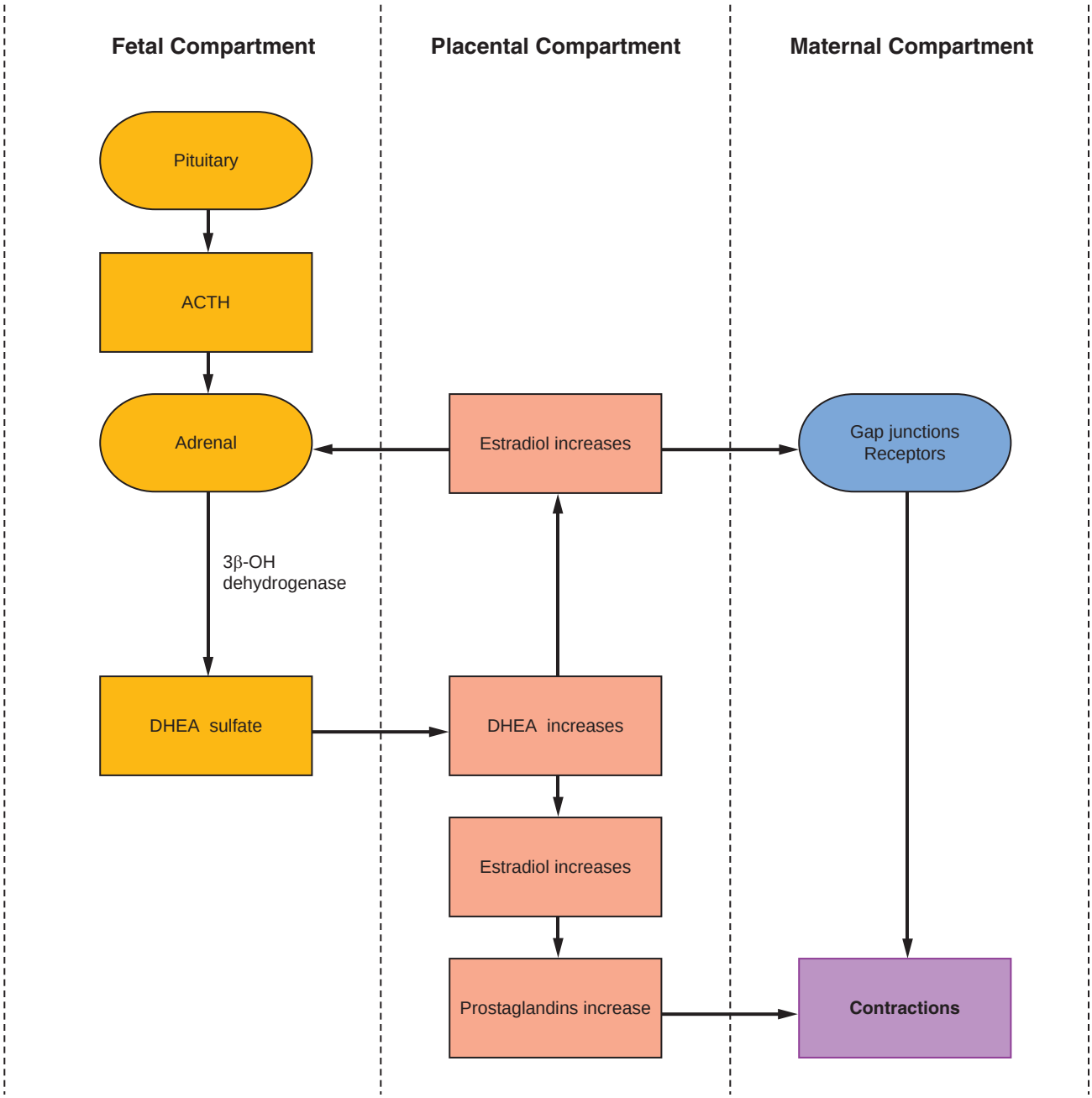
Human Parturition

The steroid events in human pregnancy are not identical to events in the ewe, chiefly because placental P450c17 enzyme activity is absent. In addition, there is a more extended time scale. Steroid changes in the ewe occur over the course of several days, whereas in human pregnancy the changes begin at approximately 34–36 weeks and occur over the last 5 weeks of pregnancy. However, if the time course is expressed as a percentage of gestational length, the percentages in sheep and primates are impressively comparable.

Cortisol rises dramatically in amniotic fluid, beginning at 34–36 weeks, and correlates with pulmonary maturation. Cord blood cortisol concentrations are high in infants born vaginally or by cesarean section following spontaneous onset of labor. In contrast, cord blood cortisol levels are lower in infants born without spontaneous labor, whether delivery is vaginal (induced labor) or by cesarean section (elective repeat section). In keeping with the extended time scale of events, administration of glucocorticoids is not followed acutely by the onset of labor in pregnant women (unless the pregnancy is past due).

It is unlikely that the cortisol increments in the fetus represent changes due to increased adrenal activity in the mother in response to stress. Although maternal cortisol crosses the placenta readily, it is largely (85%) metabolized to cortisone in the process. This, in fact, may be the mechanism by which suppression of the fetal adrenal gland by maternal cortisol is avoided. In contrast to the maternal liver, the fetal liver has a limited capacity for transforming the biologically inactive cortisone to the active cortisol. On the other hand, the fetal lung does possess the capability of changing cortisone to cortisol, and this may be an important source of cortisol for lung maturation. Cortisol itself induces this conversion in lung tissue. Increased fetal adrenal activity is followed by changes in steroid levels, as well as important developmental accomplishments (e.g., increased pulmonary surfactant production and the accumulation of liver glycogen). In human parturition an important contribution of the fetal adrenal, in addition to cortisol, is its effect on placental estrogen production. The common theme in human pregnancies associated with failure to begin labor on time is decreased estrogen production; e.g., delayed parturition in anencephaly or placental sulfatase deficiency.[311]

Progesterone maintenance of uterine quiescence and increased myometrial excitability associated with progesterone withdrawal are firmly established as mechanisms of



parturition in lower species. In primates, the role of progesterone has been less certain, largely because of the inability to demonstrate a definite decline in peripheral blood levels of progesterone prior to parturition.[312] Nevertheless, pharmacologic treatment with progesterone or synthetic progestational agents has some effect in preventing premature labor, although not labor at term.[313–316] There is also reason to believe that progesterone concentration is regulated locally, especially in the fetal membranes, and progesterone withdrawal can be accomplished by a combination of binding, metabolism, and changes in receptor isoform levels, as well as changes in coactivating and corepressing proteins.[317]

In the myometrium, during advancing gestation and with parturition, overall progesterone receptor concentration does not change with the onset of labor; however, a shift in receptor isoforms occurs with a dominance of progesterone receptor-A and other isoforms of the progesterone receptor other than receptor-B (progesterone receptor-C, another truncated

isoform, is expressed in myometrium, decidua, and the membranes).[318–322] ***Because progesterone receptor-A mainly suppresses progesterone receptor-B activity (the principal mediator of genomic progestational action), this change is consistent with a local withdrawal of progesterone in the myometrium. Therefore, there is growing reason to believe that a functional progesterone withdrawal occurs in primates, keeping the mechanism of parturition evolutionarily consistent. This withdrawal of progesterone occurs not only in the myometrium, but also in the decidua and the fetal membranes.***

Because progesterone receptor-B activity suppresses estrogen receptor expression, a shift to progesterone receptor-A would simultaneously allow both progesterone withdrawal and an increase in estrogen activity.[319] This further indicates that the process of parturition begins before the onset of contractions. Prostaglandin involvement continues to be an integral part of this process; prostaglandin $F_{2\alpha}$ stimulates progesterone receptor-A expression in myometrial cells studied in vitro, a response consistent with the emerging dominance of progesterone receptor-A before parturition.[323]

Progesterone receptor concentrations in the monkey amnion change in the presence of labor, consistent with activation of prostaglandin and cytokine production in the membranes following functional progesterone withdrawal.[318] In a fashion similar to human myometrium, progesterone receptor-A levels in monkey myometrium increase in late gestation and during labor. Interruption of exposure to progesterone (e.g., with the antiprogesterone, RU-486) leads to uterine contractions.[324] Furthermore, inhibition of progesterone production in the second trimester of human or the third trimester of monkey pregnancies is followed by a decrease in maternal, fetal, and amniotic fluid progesterone concentrations and preterm labor and delivery.[325, 326] Perhaps multiple mechanisms exist, which affect in a subtle fashion the local concentration and actions of progesterone and the production of progesterone in fetal membranes allowing reduntant pathways to compensate when a specific pathway is compromised.[327] Coactivator and corepressor proteins are known to modulate the responsiveness of steroid hormone target tissues. Appropriate changes in intracellular regulatory proteins would be another potential method to modulate progesterone activity in pregnant tissues.[328]

An increase in estrogen levels in maternal blood begins at 34–35 weeks of gestation, but a late increase just before parturition (as occurs in the ewe) has not been observed in human pregnancy. Perhaps a critical concentration is the signal in human pregnancy rather than a triggering increase. Or the changes are taking place at a local level and are not reflected in the maternal circulation.[329] Although it has not been definitely demonstrated, increased or elevated estrogen levels, as well as a local withdrawal of in progesterone production, are thought to play a key role in increasing prostaglandin. As with progesterone, the change in estrogen may be a *functional* increase at the local tissue level, such as an increase in estrogen receptor concentration and/or activity. Given the central role in parturition for progesterone withdrawal in all species, the local estrogen change could be secondary to the functional progesterone withdrawal, such as the change in progesterone receptor-A expression, allowing estrogen receptor gene expression to escape progesterone inhibition.[319]

The concept of a functional progesterone withdrawal in primate parturition and the obvious importance of progesterone withdrawal in other species prompted extensive study of the administration of progestational drugs to prevent preterm labor. There is some effect with progesterone itself, 100 mg/daily administered vaginally, but there is a one-third reduction in preterm birth with 17α-hydroxyprogesterone caproate given as a weekly 250 mg injection.[330, 331] These results have been obtained in women at high risk for preterm birth by virtue of a previous preterm birth. Progestational therapy is ineffective once labor has begun, and the impact on pregnancies complicated by conditions associated with premature labor is unknown, although studies have indicated no reduction in preterm birth in women with twin or triplet gestations.[332–334] Multiple clinical trials are ongoing.

Evidence for a role of prostaglandin in parturition includes the following:

1. Prostaglandin levels in maternal blood and amniotic fluid increase in association with labor.

2. Arachidonic acid levels in the amniotic fluid rise in labor, and arachidonate injected into the amniotic sac initiates parturition.

3. Patients taking high doses of aspirin have a highly significant increase in the average length of gestation, incidence of postmaturity, and duration of labor.

4. Indomethacin prevents the normal onset of labor in monkeys and stops premature labor in human pregnancies.

5. Stimuli known to cause the release of prostaglandins (cervical manipulation, stripping of membranes, and rupture of membranes) augment or induce uterine contractions.

6. The process of cervical ripening and softening is mediated by prostaglandins.

7. Exogenously administered prostaglandins induce labor.

The precursor fatty acid for prostaglandin production in part may be derived from storage pools in the fetal membranes, the decidua, or both.[295] Phospholipase A_2 has been demonstrated in both human chorioamnion and uterine decidua. The availability of arachidonic acid for prostaglandin production during parturition follows the stimulation of hydrolysis of phosphatidylethanolamine and phosphatidylinositol in decidual, amnion, and chorion laeve tissues.[335–337] Microsomes from amnion, chorion laeve, and decidua vera tissues contain lipases that hydrolyze fatty acids esterified in the 2 position. Specific phospholipase activity (phospholipase A_2 acting on phosphatidylethanolamine and phospholipase C acting on phosphatidylinositol) combined with a diacylglycerol lipase that also has a specificity for arachidonic acid provides a mechanism for the release of arachidonic acid. The activity of these enzymes in fetal membranes and decidua vera tissue increases with increasing length of gestation.

The key may be the increasing levels of estrogen (both estradiol and estriol) in the maternal circulation as well as in the amniotic fluid or, more importantly, locally within the uterus. The marked rise in estrogen near term may affect the activity of the lipase enzymes, leading to the liberation of arachidonic acid. The activity of these phospholipases is increased by increasing concentrations of calcium; and, therefore, the regulation of intracellular calcium is an important mechanism. Nevertheless, a role for local progesterone withdrawal in the activation of prostaglandin production remains a likely mechanism.[326]

Cervical ripening is the process by which the cervix becomes soft and distensible, easily dilated. This change is associated with a decrease in collagen and proteoglycans, and an increase in water, brought about by enzymes and cytokines in response to prostaglandins. Progesterone is believed to exert a stabilizing influence on the cervix during pregnancy, a state that is antagonized by estrogen. Studies of enzyme activity within human cervical tissue indicate that prior to the initiation of labor, progesterone levels are maintained in the cervix whereas estrogen becomes inactivated.[338] With the onset of parturition, 17β-hydroxysteroid dehydrogenase is decreased, resulting in an increase in the local concentrations of estradiol and a metabolite of progesterone, 20α-hydroxyprogesterone (in effect, a local cervical withdrawal of progesterone). These changes are consistent with local responses to increasing estrogen levels, and support the general mechanism of progesterone withdrawal occurring at localized tissue sites.

The human fetal membranes and decidua are incredibly active. Human chorion and decidua produce estrogen utilizing a variety of substrates, especially estrone sulfate and dehydroepiandrosterone sulfate (DHEAS), and this activity is increased around the time of parturition.[339, 340] In addition, the human fetal membranes synthesize and metabolize progesterone.[14] The membranes contain a 17,20-hydroxysteroid dehydrogenase system. One active site converts 20α-dihydroxyprogesterone to progesterone, while another active site on this enzyme converts estrone to estradiol. Thus, this enzyme can play an important role in altering the estrogen/progesterone ratio. The membranes and the decidua contain distinct cell populations with different biochemical activities (which change with labor).[341] Steroidogenic and prostaglandin interactions among these cells could produce the changes necessary for parturition without affecting the concentrations of circulating hormones. In addition, relaxin derived from decidua and/or chorion may exert a paracrine action on amnion prostaglandin production.[212] Throughout most of pregnancy, the amnion and chorion may exert an inhibitory influence over the myometrium by suppressing calcium channel activity.[342] Finally, the fetus may take a very direct role in this scenario by secreting substances into the amniotic fluid, which interact with the fetal membranes to signal the initiation of parturition.

The following observations support an important role for placental corticotropin-releasing hormone (CRH):

1. CRH is produced in trophoblast, the fetal membranes, and decidua.[87]

2. During pregnancy, CRH levels in the amniotic fluid and the maternal circulation progressively increase, and, although amniotic fluid levels do not further increase with labor, the highest maternal levels are found at labor and delivery.

3. Levels of the CRH-binding protein are decreased in trophoblast, membranes, decidua, the amniotic fluid, and maternal circulation prior to labor.[341, 343, 344] This decrease in the CRH-binding protein would allow an increase in CRH activity.

4. CRH directly stimulates DHEA and DHEAS biosynthesis in cells derived from the fetal zone of the adrenal.[54]

5. CRH stimulates prostaglandin release in fetal membranes, decidua, and myometrium.[101, 345]

6. Increased CRH and decreased CRH-binding protein have been measured in women with preterm labor and in women with threatened preterm labor who subsequently deliver within 24 hours.[346–349]

7. Cortisol, in the presence of progesterone, stimulates (probably by blocking progesterone inhibition) trophoblastic CRH synthesis.[52, 350]

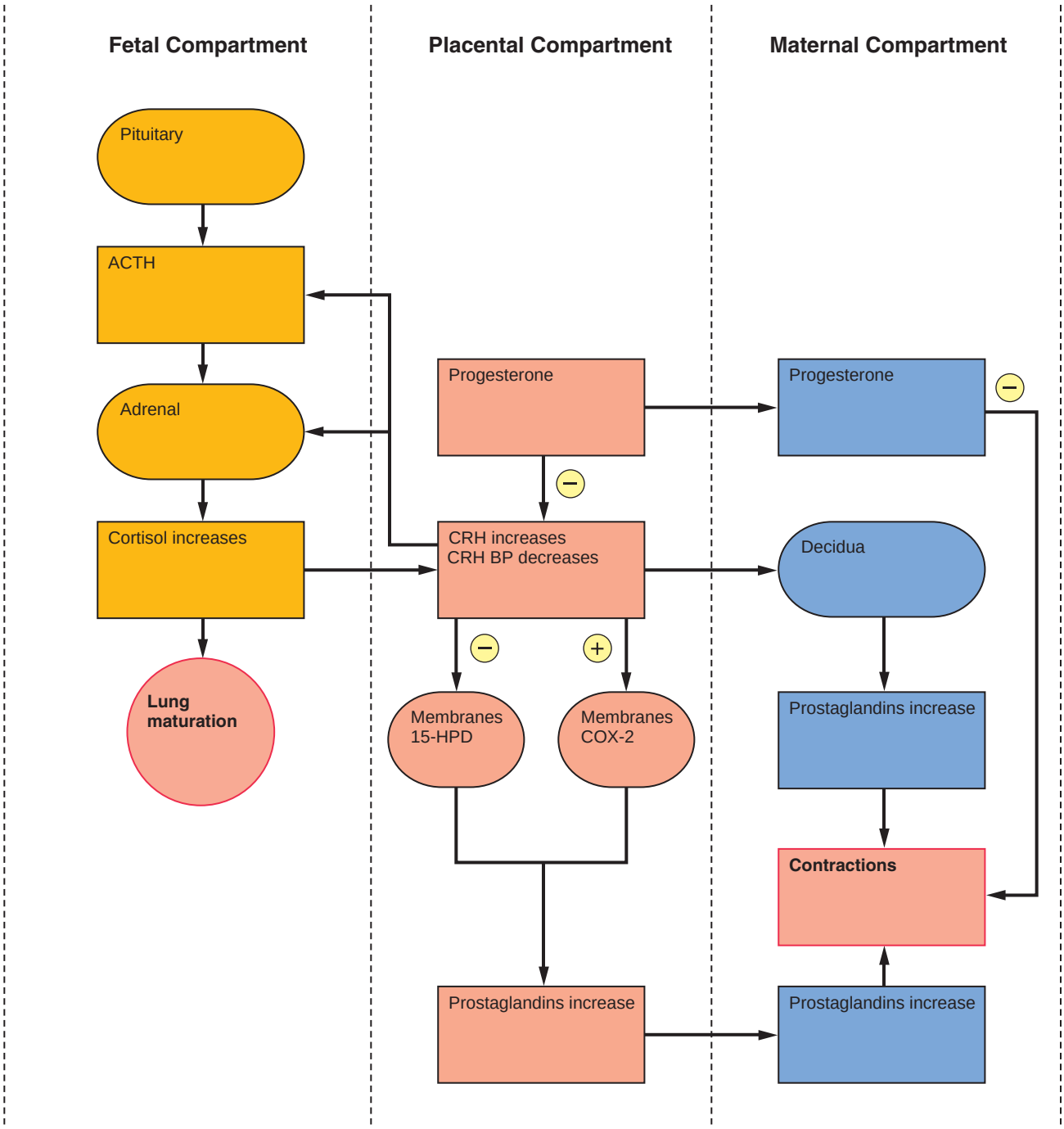
8. CRH, activin A, vasopressin, and prostaglandin $F_{2\alpha}$ stimulate oxytocin release from placental tissues, to augment myometrial contractions.[351]

9. CRH increases the secretion of matrix metalloproteinases in placental cells and fetal membranes, a prelude to the rupture of membranes.[103]

These observations are consistent with a key mechanism involving CRH activity in the initial triggering events of parturition. Although in the ewe the CRH signal begins in the fetal brain, in women, it appears to begin in the uterus. Indeed, placental CRH is expressed in primate placentas.[352] Progesterone and estrogen are major inhibiting factors for CRH production in placental tissue.[353] It has been hypothesized that rising fetal cortisol levels (e.g., in response to stress, especially hypoxia) compete with progesterone for the

glucocorticoid receptor in the placenta, thus blocking the inhibitory action of progesterone on CRH synthesis, leading to an increase in CRH.[52]

Cortisol directly stimulates CRH gene expression in the placenta, providing a mechanism for a specific link between cortisol and CRH.[354] Because CRH directly stimulates steroidogenesis in the fetal zone of the adrenal, the increase in CRH would increase DHEAS to serve as precursor for the increase in estrogen that occurs prior to parturition. The sequence of events could be started by an increase in CRH or a decrease in CRH-binding protein, or both, associated with the estrogen and progesterone changes in late pregnancy (estrogen and progesterone receptor-A repress and progesterone receptor-B increases CRH gene expression).[355, 356] On the other hand, consistent with the sheep studies, the initiating step



in this sequence of events could be an increase in fetal ACTH secretion; e.g., in response to stress and relative hypoxemia and an increase in placental CRH. Although CRH plays a central role, various pathways can lead to its increase, another example of multiple pathways to parturition. Regardless of the specific triggering event, it is increasingly clear that the fetus plays a pivotal, if not controlling, role in parturition.

Regulation of Prostaglandins

With labor, the arachidonic acid pathway in the fetal membranes shifts toward the cyclooxygenase direction with a large increase in the production of PGE_2 due to the induction of COX-2 activity. This COX-2 activity is a response to the increase in cortisol that in turn is a response to CRH. In addition, CRH can directly stimulate prostaglandin production in the membranes.

Specific protein inhibitors of prostaglandin synthase have been demonstrated in placenta, amnion, and chorion, and these proteins cannot be found in tissue from patients who have established labor.[295, 357] The link between infection and the onset of labor (especially preterm labor) may be due to the conversion by bacterial medium (with inflammatory factors such as the interleukins) of arachidonic metabolism in the membranes and decidua to a condition associated with labor marked by the production of PGE_2.[295, 358, 359] In this case, prostaglandin production may be a consequence of inflammatory induction of the second cyclooxygenase enzyme COX-2.[360] In addition, intra-amniotic infection is associated with a loss of the chorionic high concentration of 15-hydroxyprostaglandin dehydrogenase that inactivates prostaglandins, resulting in a shift that favors biosynthesis and activity.[361] These changes are modulated by the cytokines involved in the inflammatory response.

Prostaglandin production during pregnancy reflects the usual complex interaction of a host of autocrine/paracrine factors. Platelet-activating factor, epidermal growth factor, and transforming growth factor-α stimulate prostaglandin production by the fetal membranes apparently by regulating intracellular calcium concentrations.[362, 363] Secretory products of the fetal membranes themselves are active stimulators of membrane prostaglandin production, including renin derived from chorion prorenin.[364] Decidual $PGF_{2\alpha}$ production is enhanced by bradykinin, epidermal growth factor and transforming growth factor-α, and these responses are further increased by interleukin-1β.[365, 366] Prostaglandin production by amnion, chorion, and decidual cells is stimulated by CRH and modulated by progesterone.[100] The ubiquitous substances, activin and inhibin, are involved here as well. Amnion and chorion produce the activin and inhibin subunits, and activin stimulates prostaglandin PGE_2 release from amnion cells.[264]

During labor the maternal circulating levels of PGE_2, $PGF_{2\alpha}$, and the $PGF_{2\alpha}$-metabolite are increased, a change that can be directly attributed to uterine production because the gradient across the uterus for these substances is also increased. This increase in production of prostaglandins within the uterus must be the key factor, because the concentration and affinity of prostaglandin receptors do not change at parturition.[367] Prostacyclin is produced (at least in vitro) by a variety of tissues involved in pregnancy: endometrium, myometrium, placenta, amnion, chorion, and decidua. Prostacyclin and thromboxane are probably more important in the vascular responses of mother and fetus, and in all likelihood do not play a role in initiating or maintaining uterine contractions; however prostacyclin does inhibit myometrial contractility.[368] PGI synthase expression appropriately decreases in myometrium with increasing gestational age.[369]

Decidua produces both PGE_2 and $PGF_{2\alpha}$, but the amnion and chorion produce primarily PGE_2.[370] The inducible cyclooxygenase, COX-2, is expressed at a high level at term in the amnion and chorion.[371] As in sheep, prostaglandin synthesis in membranes and decidua is probably stimulated by cortisol; glucocorticoid receptors are present in the same cells that contain cyclooxygenase.[372]

There is evidence for the transfer of prostaglandin E_2 across the membranes to the decidua and possibly the myometrium.[373] The paradox of PGE_2 production in the amnion being matched not by a PGE-metabolite in the maternal circulation but by a $PGF_{2\alpha}$-metabolite was explained by transfer across the membranes and conversion of PGE_2 to $PGF_{2\alpha}$ in the decidua.[374] However, continued study of this issue strongly indicates that prostaglandins produced on one side of the membranes do not contribute to the prostaglandins on the other side, arguing that uterine contractions must be primarily influenced by decidual or myometrial prostaglandins.[375] Indeed, COX-2 expression in the myometrium increases at term before the onset of labor and is correlated with estrogen receptor-α activity.[319] There is reason to believe that the myometrial exposure to prostaglandins is also influenced by the activity of a catabolic enzyme in the chorion.

At term, prostaglandin synthesis occurs in the amnion and decidua, and throughout pregnancy the chorion forms a barrier preventing passage of bioactive prostaglandins to the myometrium because of a large capacity to catabolize prostaglandins via 15-hydroxyprostaglandin dehydrogenase.[361, 376] The activity of this enzyme is decreased in the presence of labor, including preterm labor, and after premature rupture of membranes or when infection is present.[377] [361, 378]

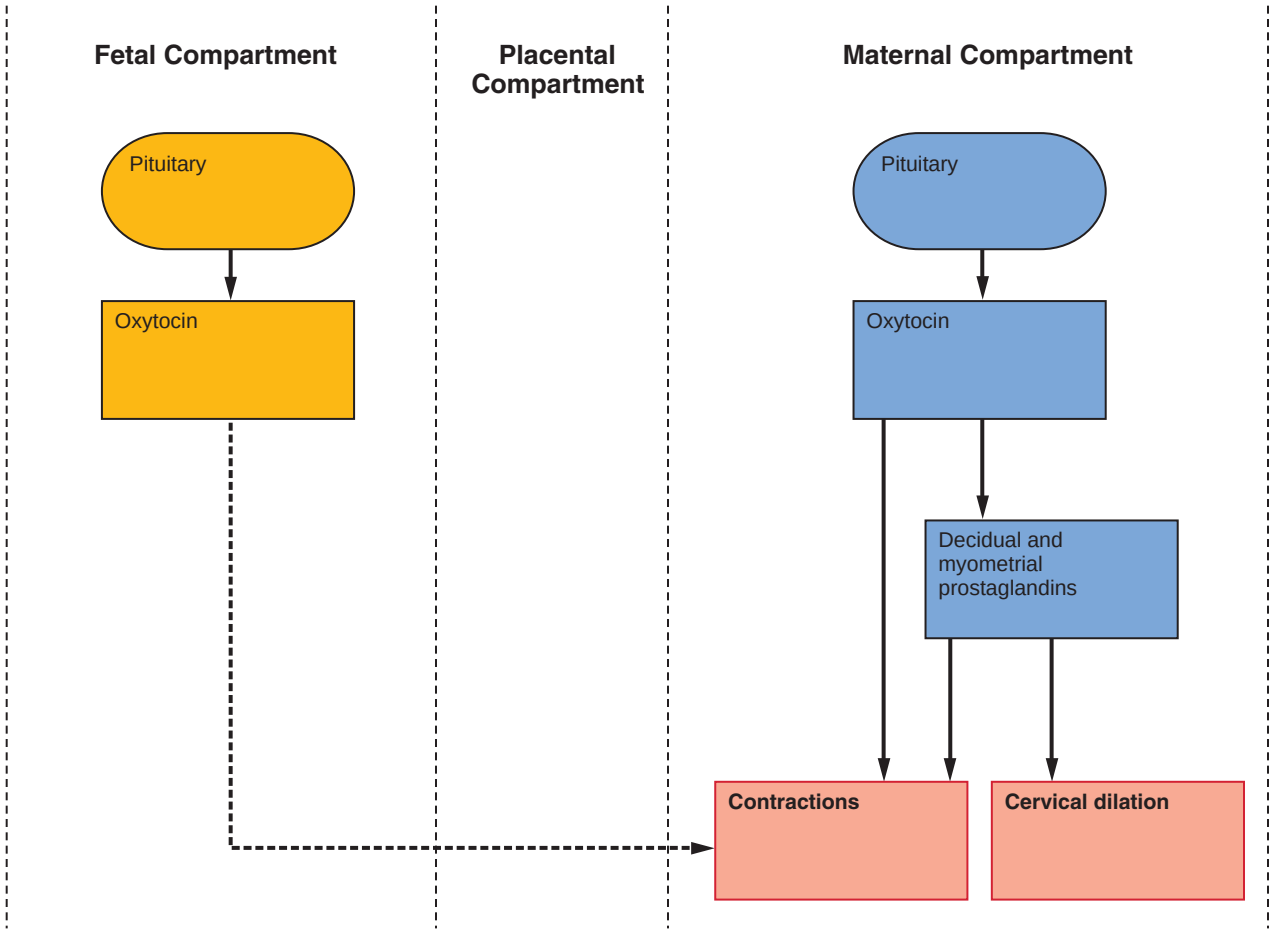
Because the activity of 15-hydroxyprostaglandin dehydrogenase decreases in the myometrium and the chorion during labor, a combination of increased biosynthesis of prostaglandins and a decrease in 15-hydroxyprostaglandin dehydrogenase achieve the increase in prostaglandins associated with parturition, probably mediated by the local changes in estrogen and progesterone bioavailability and activity, with key roles also played by CRH and cortisol. Cortisol decreases and progesterone increases 15-hydroxyprostaglandin dehydrogenase activity in placental tissues.[379, 380] A functional withdrawal of progesterone would allow a greater effect of cortisol resulting in an increase in prostaglandins. Regulation of intracellular calcium ions contributes to this mechanism; an influx of calcium ions increases prostaglandin synthase expression, whereas prostaglandin dehydrogenase expression is suppressed.[381] This is a potential pathway for the input of locally produced peptides involved in parturition.

Metalloproteinases

Disruption and remodeling of the extracellular matrix is part of the process of parturition just as it is in implantation and placentation. Cervical ripening, rupture of the fetal membranes, and detachment of the placenta all involve activity of the matrix metalloproteinases in the decidua and the membranes.[382] A balance between these enzymes and their inhibitors is necessary to maintain the integrity of fetal membranes and uterine structure and function. With the change in function associated with parturition, one would expect this balance to shift towards metalloproteinase expression and activity, and, indeed, this is the case.[383] An early event in premature labor and premature rupture of membranes is the activation of the metalloproteinases. There is evidence to indicate that metalloproteinases are activated by prostaglandins and cytokines, and inhibition is maintained by down-regulation of the conversion of plasminogen to plasmin by progesterone.[384]

Oxytocin and Myometrial Responses

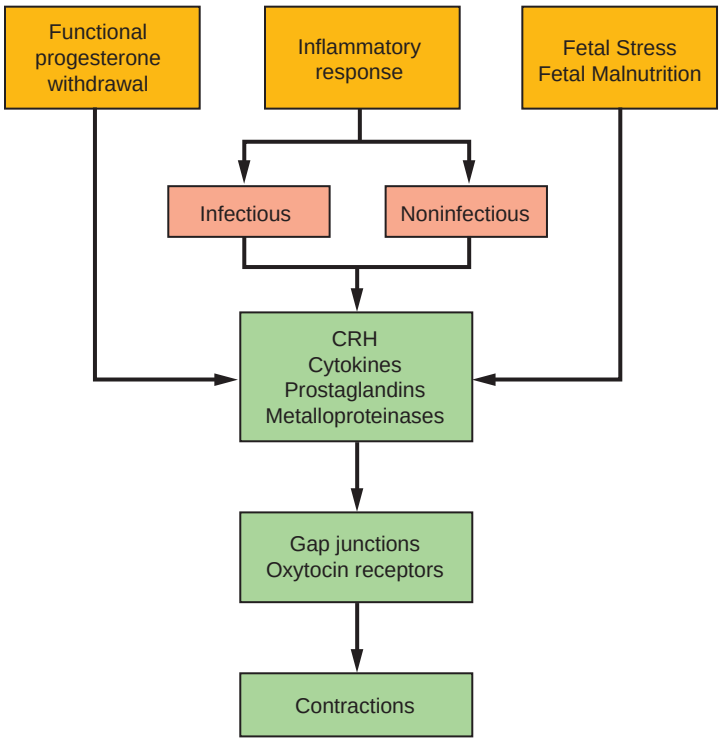
Using sensitive assays, an increase in maternal levels of oxytocin can be detected prior to parturition, occurring at first only at night.[385, 386] Activation of oxytocin secretion is linked to progesterone withdrawal because brainstem oxytocin neurons are suppressed by brain metabolites of progesterone.[387] Once labor has begun, oxytocin levels rise significantly, especially



during the second stage. Thus, maternal oxytocin may be most important for developing the later, more intense uterine contractions. Extremely high concentrations of oxytocin can be measured in the cord blood at delivery, and release of oxytocin from the fetal pituitary may also be involved in labor. However, this is controversial, and studies in monkeys fail to indicate a role for fetal oxytocin in parturition.[386] Part of the contribution of oxytocin to parturition is the stimulation of prostaglandin synthesis in decidua and myometrium.[388] Cervical dilation appears to be dependent on oxytocin stimulation of prostaglandin production, probably in the decidua. The greater frequency of labor and delivery at night may be due to greater nocturnal oxytocin secretion. In addition, oxytocin is synthesized in the amnion, in the chorion, and, significantly, in the decidua.[385, 389, 390] This locally-produced oxytocin, in response to CRH, may be a significant stimulus for myometrial and membrane production of prostaglandins.

It is likely that oxytocin action during the initial stages of labor may depend on myometrial sensitivity to oxytocin in addition to the levels of oxytocin in the blood. The concentration of oxytocin receptors in the myometrium is low in the nonpregnant state and increases steadily throughout gestation (an 80-fold increase), and, during labor, the concentration doubles. This receptor concentration correlates with the uterine sensitivity to oxytocin.[391] The mechanism for the increase is unknown, but it likely is due to a change in the prostaglandin and hormonal milieu of the uterus, especially functional progesterone withdrawal. The local production and effects of oxytocin, estrogen, and progesterone combine in a complicated process of autocrine, paracrine, and endocrine actions to result in parturition.

Animal studies have implicated the formation of low-resistance pathways in the myometrium, called ***gap junctions,*** as an important action of steroids and prostaglandins during labor.[392] In the gap junction, a pore forms, which allows communication from cytoplasm to



cytoplasm between two cells. The pore is a cylinder-shaped channel formed of six special proteins called ***connexins.*** Either substances or electrical current (ions) can follow this pathway without leakage into extracellular space. Thus, gap junctions provide a means of communication between myometrial cells, allowing enhancement of electrical conductivity and synchronization of activity. Gap junction formation is related to the estrogen/progesterone ratio (estrogen up-regulates connexin-43, the gap junction protein, which is down-regulated by progesterone) and to the presence of the stimulating prostaglandins PGE_2 and $PGF_{2\alpha}$. Therefore, it is not surprising that the number of gap junctions increases in the final weeks of pregnancy, especially just before labor. The modulation of the number and the permeability of gap junctions is another contributing factor in the control of uterine contractility.

The final contraction of uterine muscle results from increased free calcium concentrations in the myofibril, the result of prostaglandin action and functional progesterone withdrawal, an effect opposed to the promotion of calcium binding in the sarcoplasmic reticulum by the presence of progesterone.[393, 394] Thus, prostaglandins and oxytocin increase while progesterone decreases intracellular calcium levels. The intracellular calcium concentration is affected by cellular entry and exit of calcium as well as binding in the sarcoplasmic reticulum. It is the intracellular concentration of calcium that determines the rate of myosin phosphorylation and the contractile state of the myometrium. Tocolytic therapy (the use of beta-adrenergic agents) stimulates adenylate cyclase activity, which increases the levels of cellular cyclic AMP, which, in turn, decreases intracellular calcium concentration and inhibits actin-myosin interaction by modulating kinase phosphorylation.

Ducsay and colleagues propose that the coordination of this complex relationship of physiologic, endocrine, and molecular mechanisms is expressed in rhythms.[72, 395] Both mother and fetus experience 24-hour rhythms in hormone secretions, and uterine activity is correlated with day and night (photoperiod regulation). The coordination and enhancement of this rhythmicity play a role in parturition. Improved detection and measurement of this activity could contribute to better prevention and treatment of preterm labor.

Molecular biology is now assessing the activity of genes in the uterus and fetal membranes.[396] Some of this activity is predictable, correlating with autocrine and paracrine

substances known to be involved in parturition. The identification of other genes differentially regulated during parturition will open new areas for research. Ultimately we will come to understand the mechanisms of normal parturition and to be more effective in managing premature and abnormal labor.

Concluding Thought

Imagine yourself as a fetus within a pregnant uterus. Your growth, development, and survival require keeping the uterus quiescent for most of pregnancy. This is accomplished by maintaining progesterone inhibitory dominance of the myometrium. When ready to begin extrauterine life or when your environment becomes inhospitable, you are able to prepare or "activate" the parturition mechanisms by means of hormonal and autocrine/paracrine messengers. Ultimately, uterine contractions and cervical ripening are stimulated, and amazingly, even if you are incapable of initiating these events, the sequence will eventually begin, and delivery will ensue. The extraordinary experience and wonder of labor and birth, as perceived by parents and birth assistants, are matched by your ability and the complexity of the systems you influence.

Treatment of Labor with Prostaglandin Inhibition

The key role for prostaglandins in parturition raises the potential for treatment of premature labor with inhibitors of prostaglandin synthesis. The concern has been that such treatment would result in intrauterine closure of the ductus arteriosus causing pulmonary hypertension. Clinical studies, however, indicate that use of the nonsteroidal anti-inflammatory agents for short periods of time (3 days) yields good results and does not result in this complication.[397] Beyond 34 weeks, the fetus is more sensitive to this pulmonary action, and treatment should be limited to pregnancies less than 32 weeks and with caution from 32 to 34 weeks. If the drug is failing, it should not be maintained because increased blood loss can occur at delivery. Because indomethacin inhibits the synthesis of all members of the prostaglandin family, including the vasodilating prostacyclin, it should be used with caution in hypertensive patients.[398] Sulindac is just as effective as a tocolytic but does not affect urine output and amniotic fluid, and it has a lesser impact on the fetal ductus arteriosus.[399, 400] A specific COX-2 inhibitor, celecoxib, was as effective as indomethacin in treating preterm labor, but importantly, there was no adverse impact on the ductus arteriosus and there was a lesser transient decrease in amniotic fluid volume.[401, 402]

Treatment of pregnant women with indomethacin reduces the amniotic fluid volume due to a decrease in fetal urine output. This is reversible with a decrease in dose. This treatment has been used for polyhydramnios with good response and no effect on the newborn despite treatment for 2 to 11 weeks.[403–405]

Induction of Labor and Cervical Ripening

Pharmacologically and physiologically, prostaglandins have two direct actions associated with labor: ripening of the cervix and myometrial stimulation. Successful parturition requires organized changes in both the upper uterus and in the cervix. The cervical changes are in response to the estrogen/progesterone ratio and the local release of prostaglandins. Whether relaxin plays a role in human parturition is not established; however, recombinant relaxin is being tested for cervical ripening.

Ripening of the cervix is the result of a change that includes an increase in hyaluronic acid and water and a decrease in dermatan sulfate and chondroitin sulfate (these compounds hold the collagen fibers in a rigid structure). How prostaglandins operate in this change is unknown, but enzyme activation must be involved. For ripening of the cervix, PGE_2 is very effective, whereas $PGF_{2\alpha}$ has little effect. The purpose of pharmacologically achieving ripening of the cervix is to increase the success rate with induction of labor and lower the proportion of cesarean sections. Intravaginal prostaglandin E_2 (dinoprostone) administered as tablets, suppositories, and mixed in gels has been very effective for cervical ripening. A synthetic PGE_1 analogue, misoprostol, is also effective when used intravaginally or orally for cervical ripening and labor induction, although there can be a problem of uterine tachysystole (rapid contractions).[406–409]

A major clinical application for the induction of labor in the United States is the use of intravaginal PGE_2 in cases of fetal demise and anencephalic fetuses. The patient should be well hydrated with an electrolyte solution to counteract the induced vasodilation and decreased peripheral resistance. If satisfactory uterine activity is established, the next application should be withheld. And, finally, because there is a synergistic effect when oxytocin is used shortly after prostaglandin administration, there should be a minimum of 6 hours between the last prostaglandin dose and beginning oxytocin augmentation.

Prostaglandins are used to induce term labor. Intravenous prostaglandins are not an acceptable method due to the side effects achieved by the high dosage necessary to reach the uterus. The intravaginal and oral administration of PGE_2 is as effective as intravenous oxytocin, with good results initially reported even in patients with previous cesarean sections.[410, 411] Later, concern was raised that uterine rupture may be more frequent with prostaglandin use in women with previous cesareans.[412] The intravaginal administration of misoprostol, the synthetic prostaglandin E_1 analogue, is safe, effective, and relatively inexpensive for the routine induction of labor.[413] These methods, plus intracervical administration, are in routine use in many parts of the world.

Induced Abortion

Prostaglandins are effective for postcoital contraception and first-trimester abortion but impractical because of the high incidence of side effects, including an unacceptable rate of incomplete abortions. For midtrimester abortions, intraamniotic prostaglandin, intramuscular methyl esters, and vaginal PGE suppositories are available. Again, the major clinical problems have been the efficacy in accomplishing complete expulsion and the high level of systemic side effects. Overall, there is a higher risk of hemorrhage, fever, infection, antibiotic administration, readmission to the hospital, and more operative procedures when compared with saline abortions.

The combination of prostaglandin's oxytocic action with the antiprogesterone effect of RU 486 (mifepristone) has proved to be a safe and effective medical treatment for the induction of therapeutic abortion in both the first and second trimesters.[414–417] Combining a prostaglandin analogue, misoprostol, with mifepristone inexpensively and safely achieves greater than 95% efficacy (Chapter 21).

Prostaglandins and Postpartum Hemorrhage

When routine methods of management for postpartum hemorrhage caused by uterine atony have failed, an analogue of prostaglandin $F_{2\alpha}$ gives excellent results (80–90% successful).[418] Prostin 15 M is (15-S)-15-methyl prostaglandin $F_{2\alpha}$-tromethamine. The dose is 0.25–0.5 mg,

repeated up to 4 times and given with equal efficacy either intramuscularly or directly into the myometrium. It can also be used after the replacement of an inverted uterus. Failures are usually associated with infections or magnesium sulfate therapy. However, clinical trials with injectable prostaglandins have indicated that regular methods are as effective and that the modest reduction in blood loss does not warrant routine use to prevent postpartum hemorrhage.[419] When used after delivery for the prevention of postpartum hemorrhage, misoprostol, the PGE_1 analogue, 600 mg given orally, is less effective with more side effects than the standard use of oxytocin.[419, 420] However, misoprostol can be a life saver when delivery and hemorrhage occur in parts of the world where parenteral drugs are not available.

Prostaglandins and the Fetal Circulation

The predominant effect of prostaglandins on the fetal and maternal cardiovascular system is to maintain the ductus arteriosus, and the renal, mesenteric, uterine, placental, and probably the cerebral and coronary arteries in a relaxed or dilated state. The importance of the ductus arteriosus can be appreciated by considering that 59% of the cardiac output flows through this connection between the pulmonary artery and the descending aorta.

Control of ductal patency and closure is mediated through prostaglandins. The arterial concentration of oxygen is the key to the caliber of the ductus. With increasing gestational age, the ductus becomes increasingly responsive to increased oxygen. In this area, too, attention has turned to PGI_2 and TXA_2.

Fetal lamb ductus homogenates produce mainly PGI_2 when incubated with arachidonic acid. PGE_2 and $PGF_{2\alpha}$ are formed in small amounts and TXA_2 not at all. Although PGE_2 is less abundant than PGI_2 in the ductus, it is a more potent vasodilator of the ductus and is more responsive to oxygen (decreasing vasodilation with increasing oxygen).[421] Thus, PGE_2 appears to be the most important prostaglandin in the ductus from a functional point of view, whereas PGI_2, the major product in the main pulmonary artery, appears to be the major factor in maintaining vasodilation in the pulmonary bed. The ductus is dilated maximally in utero by production of prostaglandins, and a positive vasoconstrictor process is required to close it. The source of the vasoconstrictor is probably the lung. With increasing maturation, the lung shifts to TXA_2 formation. This fits with the association of ductal patency with prematurity. With the onset of pulmonary ventilation at birth leading to vascular changes that deliver blood to the duct directly from the lungs, TXA_2 can now serve as the vasoconstrictor stimulus. The major drawback to this hypothesis is the failure of inhibitors to affect the constriction response to oxygen.

Administration of vasodilating prostaglandins can maintain ductal patency after birth, while preparing an infant for surgery to correct a congenital lesion causing pulmonary hypertension.[422] Infants with persistent ductus patency may be spared thoracotomy by treatment with an inhibitor of prostaglandin synthesis. The use of indomethacin to close a persistent ductus in the premature infant is successful about 40% of the time.[421, 423] Ibuprofen is equally effective and reduces blood flow to critical organs less than that observed with indomethacin.[424] An important factor is early diagnosis and treatment because with increasing postnatal age the ductus becomes less sensitive to prostaglandin inhibitors, probably because of more efficient clearance of the drug.[425] The highest incidence of successful ductus closure has been with infants younger than 30 weeks gestation and younger than 10 days old.

This aspect of the use of prostaglandin inhibitors is of concern in considering the use of agents to inhibit premature labor. The drug half-life in the fetus and newborn is prolonged because the metabolic pathways are limited, and there is reduced drug clearance because of immature renal function. In utero constriction of the ductus can cause congestive heart failure and fetal pulmonary hypertension.[426] Prolonged ductus constriction leads to subendocardial ischemia and fibrotic lesions in the tricuspid valve muscles. Infants with persistent pulmonary hypertension have hypoxemia, cardiomegaly, and right-to-left shunting through the foramen ovale or the ductus. Infants of mothers given either indomethacin or salicylates chronically have been reported to have this syndrome. Duration of exposure and dosage are critical. It takes occlusion of the ductus for more than 2 weeks to produce fetal pulmonary hypertension and cardiac hypertrophy. This side effect is rare in pregnancies less than 27 weeks gestation; the ductus arteriosus usually begins to respond at 27–30 weeks, and, after 30 weeks, this is an important side effect that can be minimized if long-term use is avoided.[427]

Prostaglandins and Fetal Breathing

Prior to parturition, fetal breathing is very shallow. It is proposed that placental PGE_2 suppresses breathing by acting in the fetal brain.[428] Occlusion of the umbilical cord is rapidly followed by a loss of this PGE_2 influence and the onset of air breathing. The administration of indomethacin to fetal sheep increases, whereas infusion of PGE suppresses, fetal breathing movements. This may be the explanation for the decrease in fetal breathing movements observed during human labor (associated with an increase in prostaglandin levels).

Fetal Lung Maturation

The pulmonary alveoli are lined with a surface-active phospholipid-protein complex called pulmonary surfactant, which is synthesized in the type II pneumocyte of mature lungs. It is this surfactant that decreases surface tension, thereby facilitating lung expansion and preventing atelectasis. In full-term fetuses, surfactant is present at birth in sufficient amounts to permit adequate lung expansion and normal breathing. In premature fetuses, however, surfactant is present in lesser amounts, and, when insufficient, postnatal lung expansion and ventilation are frequently impaired, resulting in progressive atelectasis, the clinical syndrome of respiratory distress.

Phosphatidylcholine (lecithin) has been identified as the most active and most abundant lipid of the surfactant complex. The second most active and abundant material is phosphatidylglycerol (PG), which significantly enhances surfactant function. Both are present in only small concentrations until the last 5 weeks of pregnancy. Beginning at 20–22 weeks of pregnancy, a less stable and less active lecithin, palmitoylmyristoyl lecithin, is formed. Hence, a premature infant does not always develop respiratory distress syndrome; however, in addition to being less active, synthesis of this lecithin is decreased by stress and acidosis, making the premature infant more susceptible to respiratory distress. At about the 35th week of gestation, there is a sudden surge of dipalmitoyl lecithin, the major surfactant lecithin, which is stable and very active. Because secretion by the fetal lungs contributes to the formation of amniotic fluid and the sphingomyelin concentration of amniotic fluid changes relatively little throughout pregnancy, assessment of the lecithin/sphingomyelin

(L/S) ratio in amniotic fluid at approximately 34–36 weeks of pregnancy can determine the amount of dipalmitoyl lecithin available and thus the degree to which the lungs will adapt to newborn life.

Gluck and colleagues, in 1971, were the first to demonstrate that the L/S ratio correlates with pulmonary maturity of the fetal lung.[429] In normal development, sphingomyelin concentrations are greater than those of lecithin until about gestational week 26. Prior to 34 weeks, the L/S ratio is approximately 1:1. At 34–36 weeks, with the sudden increase in lecithin, the ratio rises acutely. In general, a ratio of 2.0 or greater indicates pulmonary maturity and that respiratory distress syndrome will not develop in the newborn.[430] Respiratory distress syndrome associated with a ratio greater than 2.0 usually follows a difficult delivery with a low 5-minute Apgar score, suggesting that severe acidosis can inhibit surfactant production. A ratio in the transitional range (1.0–1.9) indicates that respiratory distress syndrome may develop but that the fetal lung has entered the period of lecithin production, and a repeat amniocentesis in 1 or 2 weeks usually reveals a mature L/S ratio. The rise from low to high ratios can occur within 3–4 days.

An increase in the surfactant content of phosphatidylglycerol (PG) at 34–36 weeks marks the final maturation of the fetal lung. When the L/S ratio is greater than 2.0 and PG is present, the incidence of respiratory distress syndrome is virtually zero. The assessment of PG is especially helpful when the amniotic fluid is contaminated because the analysis is not affected by meconium, blood, or vaginal secretions. The L/S ratio has been replaced in many centers by a method that uses fluorescence polarization with a fluorescent probe that binds to surfactant. The fluorescent method is simple, automated, rapid, and less costly.

Abnormalities of pregnancy may affect the rate of maturation of the fetal lung, resulting either in an early mature L/S ratio or a delayed rise in the ratio. Accelerated maturation of the ratio is associated with hypertension, advanced diabetes, hemoglobinopathies, heroin addiction, and poor maternal nutrition. Delayed maturation is seen with diabetes (without hypertension) and Rh sensitization. In general, accelerated maturation is associated with reductions in uteroplacental blood flow (and presumably increased fetal stress). With vigorous and effective control of maternal diabetes, the risk of respiratory distress syndrome in the newborns is not significantly different from infants born to nondiabetics.

Since Graham Liggins observed survival of premature lambs following the administration of cortisol to the fetus,[431] it has become recognized that fetal cortisol is the principal requisite for surfactant biosynthesis. This is true despite the fact that no increase in fetal cortisol can be demonstrated to correlate with the increases in fetal lung maturation. For this reason, fetal lung maturation can be best viewed as the result of not only cortisol but also the synergistic action of prolactin, thyroxine, estrogens, prostaglandins, growth factors, and perhaps other yet unidentified agents.[432] Insulin directly inhibits surfactant protein expression in fetal lung tissue, which explains the increase in respiratory distress syndrome associated with hyperglycemia in pregnancy (although this effect can be overcome by the stress associated with advanced diabetes).[433]

Corticosteroid therapy of pregnant women threatened with preterm delivery reduces neonatal mortality, respiratory distress syndrome, and intraventricular hemorrhage.[434, 435] In general, maximal benefit in terms of enhanced fetal pulmonic maturity has been demonstrated with glucocorticoid administration at 24–32 weeks of gestational age, with some benefit between 32–34 weeks, and little benefit beyond 34 weeks unless there is evidence of pulmonary immaturity. The optimal effect requires that 48 hours elapse after initiation of therapy although some benefit is achieved within hours after administration. The current recommendation in the United States is to administer two doses of betamethasone, 12 mg intramuscularly 24 hours apart or four doses of dexamethasone, 6 mg intramuscularly every 12 hours.[435] There is some support for multiple weekly treatments, but re-treatment remains somewhat controversial and requires consultation with a maternal-fetal medicine

specialist. Although every case of respiratory distress syndrome and subsequent chronic lung disease cannot be prevented, a significant impact can be achieved on infant mortality, and the incidence and severity of respiratory distress syndrome. Additional treatment with thyrotropin-releasing hormone (TRH) was initially believed to be beneficial; however, clinical trials indicate that TRH does not further reduce the incidence of chronic lung disease in glucocorticoid-treated very low birth weight infants.[436–438]

The Postpartum Period

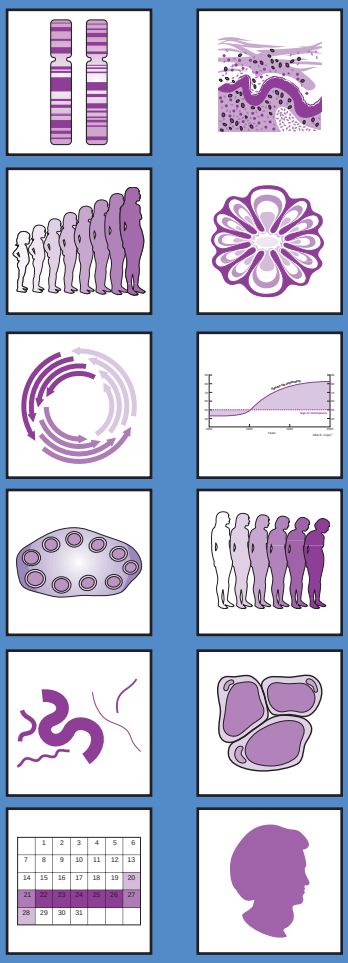
The immediate postpartum period is a time of rapid readjustment to the nonpregnant endocrine state. About 10–15% of women become clinically depressed during this time, and an endocrine mechanism has been suggested.[439] The clinician should always have a high index of suspicion for thyroid dysfunction because of the 5–10% incidence of postpartum thyroiditis in the 3–6 months after delivery. Because of the relative hypercortisolism in the last trimester of pregnancy, it has been suggested that persistent suppression of hypothalamic CRH secretion (and thus the pituitary-adrenal axis) in the postpartum period is a characteristic finding in women with postpartum depression and that this suppression also contributes to a greater vulnerability to autoimmune diseases, such as thyroiditis.[440]

All references are available online at: http://www.clinicalgynendoandinfertility.com

SECTION

II

CLINICAL ENDOCRINOLOGY



9 Normal and Abnormal Sexual Development

Abnormalities of sexual differentiation are seen infrequently in an individual clinician's practice. However, few physicians have not been challenged at least once by a newborn with ambiguous genitalia or by a young woman with primary amenorrhea. Traditional classifications for disorders of sexual differentiation have been confusing, but advances in reproductive science have helped to define their causes and to provide the foundation for a logical and efficient approach to diagnosis.

This chapter first considers the processes involved in normal sexual differentiation, to provide a basis for understanding the various types and causes of abnormal development. Some subjects are discussed in other chapters, but also are included here, for clarity and completeness. The fundamental theme is that disorders of sexual development result primarily from abnormalities in the amount or action of androgens—from excess androgen in females and from too little androgen in males.

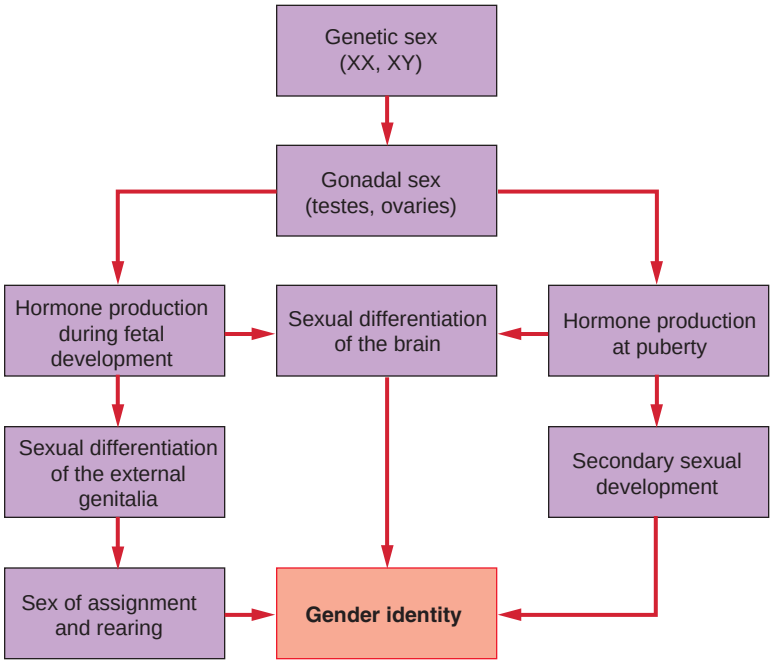
Normal Sexual Differentiation

The gender identity of a person (whether an individual identifies as a male or a female) is determined by their genetic, gonadal, and phenotypic sex and also is influenced by their

environment. Genetic or chromosomal sex is defined by the sex chromosomes, typically XX or XY. Gonadal sex is defined by the direction of gonadal differentiation, into ovaries or testes. Phenotypic sex is defined primarily by the appearance of the external genitalia and the secondary sexual characteristics that develop at puberty. Gender identity includes all behavior having any sexual connotation, such as body gestures and mannerisms, habits of speech, recreational preferences, and content of dreams. Sexual expression, both homosexual and heterosexual, reflects the sum of all sexual influences on the individual, both prenatal and postnatal, the latter referring to the role assigned by society in accordance with the individual's phenotype and behavior.

Normal sexual differentiation involves a sequence of related processes that begins with genetic or chromosomal sex, as established at the time of fertilization.[1] Gonadal sex is determined next; directed by the genetic sex, the indifferent gonads differentiate into ovaries or testes. In turn, gonadal sex controls the hormonal environment of the embryo, which directs the development of the internal and external genitalia. The processes involved in sexual differentiation of the embryonic brain are less clear, but may involve mechanisms similar to those controlling differentiation of the external genitalia. The inductive influences of hormones on the developing central nervous system (CNS) ultimately may determine the patterns of hormone secretion and sexual behavior in the adult.[2–7]

Although the mechanisms that govern sex differentiation are not yet entirely clear, our understanding of the molecular processes involved has advanced significantly in recent years. Current concepts are summarized here, beginning with the genetics of sex determination, followed by germ cell sex differentiation, gonadal differentiation, and development of the internal and external genitalia.



Genetics of Sex Determination

Both the X and the Y chromosomes appear to have evolved from autosomal ancestors over a period of 300 million years.[8] Most of the ancestral genes on the Y chromosome have been lost in the process, leaving only a limited number of currently active genes. A great many

genes are involved in translating the sex chromosome composition of the embryo and in directing the differentiation of the gonadal somatic cells,[9–11] but sex determination depends primarily on the presence or absence of a Y chromosome.

In females, the identical pair of X chromosomes aligns and recombines along its entire length during meiosis, like the autosomes. In males, homology between the X and Y chromosomes is limited to two small regions located at the very distal ends of the short and long arms of the Y. The "pseudoautosomal" region comprises only approximately 5% of the entire Y chromosome and is the only region that normally pairs and recombines during meiosis.[10, 12] Most of the remaining 95% of the Y chromosome is unique to the male, containing multiple copies of genes expressed specifically in the testis and encoding proteins with specialized functions.[8] A single copy of the one gene most critical to testis differentiation, *SRY* (Sex-determining Region on Y), is located on the distal short arm of the Y (Yp11.3), immediately adjacent to the pseudoautosomal region.[13]

Most of what is known about the genetic basis for sexual differentiation derives from studies of mutations in the mouse and human associated with varying degrees of "sex reversal," conditions in which the chromosomal sex does not correlate with the gonadal or phenotypic sex. ***In humans, 46,XX male sex reversal occurs when pairing between the X and Y chromosomes during male meiosis extends abnormally into adjacent non-homologous regions, allowing inappropriate recombination and transfer of Y-specific DNA onto the X chromosome.*** Careful analysis of four XX males having a very small piece of translocated Y DNA (60 kb)[14] prompted a search for highly conserved sequences within that region, which led to discovery of the *SRY* gene.[13] The identification of *SRY* mutations in three XY females supported the hypothesis that *SRY* was the critical and long sought "testis determining factor,"[15, 16] but proof derived ultimately from studies in the mouse. First, a deletion in *Sry* (by convention, mouse genes are designated by small case letters) was identified in a line of XY female mice.[17] Second, *Sry* gene expression in the genital ridge was observed just at the time of testis differentiation.[18] Third, transgenic XX mice carrying *Sry* develop as males.[19] ***SRY now is generally established as the primary genetic signal determining the direction of gonadal differentiation in mammals***.[10, 20] However, XX hermaphrodites having ovotestes but not *SRY* have been described and only a small proportion of phenotypic females with XY gonadal dysgenesis (Swyer syndrome) harbor *SRY* mutations. These observations indicate clearly that sex determination and sex reversal involve genes other than *SRY*.[21]

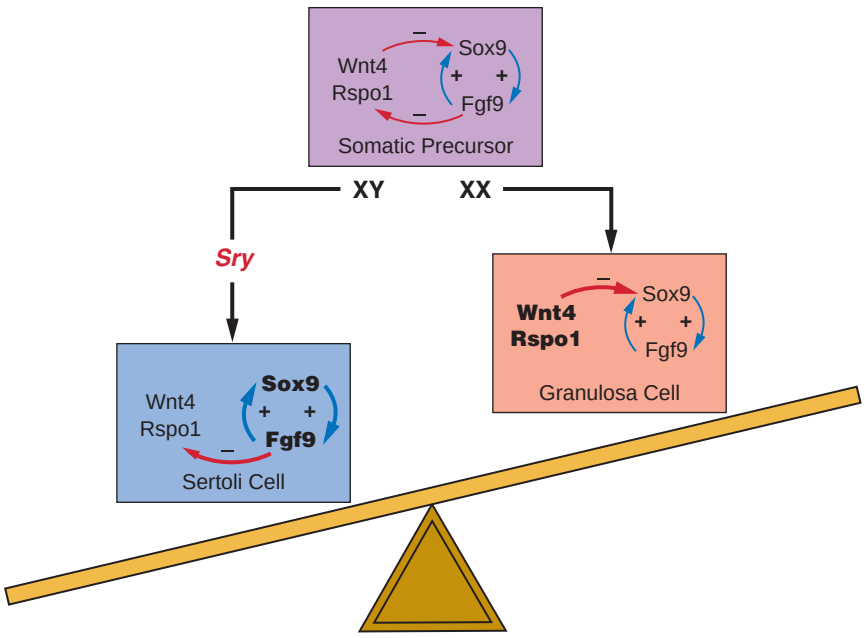
Although the mechanisms that regulate *SRY* expression are still unclear, the nuclear receptor SF1 (Steroidogenic Factor 1) has emerged as a likely and important activator. In the mouse, Sf1 binds to and activates the *Sry* promoter,[22] and heterozygous mutations in the *Sf1* gene (resulting in haploinsufficiency) produce XY female sex reversal.[23–25] In humans, *SF1* haploinsufficiency is a known cause of XY female sex reversal,[26] and an *SF1* polymorphism that reduces transactivation function by approximately 20% is recognized as a susceptibility factor for the development of micropenis and cryptorchidism.[27, 28] Evidence indicates that splice variants of Wt1 (Wilms tumor 1) and GATA4 (GATA binding protein 4) also may be involved in the regulation of *Sry* expression; both are transcription factors containing zinc-finger motifs that can interact and synergistically activate the promoter of human *SRY*.[29] WT1 mutations are associated with gonadal dysgenesis and ambiguous genitalia in males.[30]

The sequence of molecular events involved in testis differentiation is not completely understood, but SRY appears to activate a number of other genes that promote testis development.[31] The 204 amino acid protein product of *SRY* (SRY) contains a 79 amino acid domain very similar to that in a recognized family of transcription factors known as the high mobility group (HMG), which bind to DNA and regulate gene transcription. Members of the related SRY HMG box (SOX) protein family of transcription factors play a crucial role in the cascade of events that drives testis differentiation, and most of the *SRY* point mutations identified in sex-reversed patients translate to abnormalities in the amino acid sequence of SOX proteins.[32]

Substantial evidence now indicates that SOX9 is the most likely SRY target gene. In mice, *Sox9* expression is dramatically up-regulated soon after *Sry* expression begins in XY gonads but down-regulated in XX gonads,[33] and cell-fate mapping experiments have found that Sox9-positive Sertoli cells derive exclusively from Sry-positive gonadal somatic cells.[34] XY mouse embryos having a targeted deletion of *Sox9* develop ovaries,[35, 36] and transgenic activation of *Sox9* expression induces male development in XX embryos.[10] In humans, heterozygous mutations in *SOX9* (resulting in haploinsufficiency) cause a skeletal malformation syndrome (campomelic dysplasia) in which most affected XY patients exhibit female sex reversal, and *SOX9* duplication (resulting in overexpression) is the only known autosomal cause of XX male sex reversal.[32]

The developmental consequences of activating and inactivating mutations in Sox9 resemble those of similar mutations in Sry, implying not only that Sox9 is required for testis differentiation, but also that Sry activation of Sox9 may be all that is necessary to activate other genes important to testis development, such as Fgf9 (fibroblast growth factor 9), and to repress genes that induce ovary development, such as Wnt4 (a member of the wingless family of genes), Rspo1 (R-spondin 1), Dax1 (dosage-sensitive sex reversal, adrenal hypoplasia critical region, on chromosome X, gene 1), and Foxl2 (forkhead box L2).[32] DAX1 is a nuclear transcription factor normally upregulated in the ovary and repressed by SOX9, but *DAX1* duplication (resulting in overexpression) can repress *SRY* (directly, or indirectly by inhibiting SF1) and cause XY female sex reversal.[37, 38] SOX9 probably is the one most important factor regulating the activity of genes involved in Sertoli cell differentiation, and evidence suggests that SOX9 drives the process via feed forward loops that up-regulate its own expression. Sox9 stimulates *Sf1* expression, binds to the same enhancer as Sry (after Sry expression has ended), and also stimulates *Fgf9* expression in nascent Sertoli cells, all of which up-regulate *Sox9* expression and combine to maintain high levels of Sox9 activity.[10, 31, 32] ***Although a great many genes are involved in testis differentiation, virtually all male-to-female sex reversal in mice and in humans can be explained ultimately, directly or indirectly, by the failure to generate sufficient levels of SOX9 to promote the positive-feedback loops that maintain its expression.***

Fgf9 appears particularly critical for maintaining the levels of *Sox9* expression required to induce testis differentiation. Both *Fgf9* and *Sox9* are expressed at low levels in bipotential



XX and XY gonads, but *Fgf9* expression is lost in XX and amplified in XY gonads soon after *Sry* is expressed.[39] Deletion of *Fgf9* does not prevent initial expression of *Sry* or *Sox9* in Sertoli cell precursors, but *Sox9* expression is a prerequisite for *Fgf9* expression, and without it, *Sox9* expression cannot be sustained.[40] *Fgf9* also appears to actively repress genes that promote ovary differentiation, such as *Wnt4*.[39]

Whereas ovarian differentiation has long been considered the "default" pathway of sex determination—the automatic result in the absence of a testis determining factor—recent evidence challenges that traditional concept. In mice, inactivating mutations in genes such as *Wnt4*,[39, 41] *Rspo1*,[42–44] and *Foxl2*[45–47] result in partial or complete XX male sex reversal, and activating mutations in *β-catenin* or *Dax1* result in XY female sex reversal.[32, 48, 49] Rspo1 is required for *Wnt4* expression and activates β-catenin, which, like Foxl2, down-regulates *Sox9* expression.[21] Dax1 acts as a dominant-negative regulator of transcription of other nuclear receptors, including SF1, and thus may repress Sry expression.[32] Taken together, these observations suggest strongly that ovarian development results from the active repression of one or more genes in the testis pathway, rather than from a developmental default mechanism.

It now appears that both testis and ovary differentiation require dominantly acting genes, with SRY inducing testis development via up-regulation of SOX9, and with other genes, primarily WNT4 and RSPO1, teaming to promote ovary development via repression of SOX9. The new concept views the fate of the bipotential gonad as balanced between opposing forces and SRY as the key factor. In XY gonads, SRY induces SOX9 and tips differentiation toward testis development, and in XX gonads lacking SRY, other genes combine to repress SOX9 and promote ovary development.[21, 50]

Germ Cell Sex Differentiation

In human embryos, gonadal development begins during the fifth week of gestation as a protuberance overlying the mesonephric ducts, known as the genital or gonadal ridge. The primordial germ cells do not arise within but migrate into the developing gonads between 4 and 6 weeks gestation, proliferating as they go. At least in the mouse, their survival during migration appears to depend on an interaction between the cell surface tyrosine kinase receptor, *c*-KIT, and a ligand produced by surrounding tissues, called stem cell factor.[51] At this stage of development, the gonads are identical in males and females, indifferent and bipotential, capable of differentiating into either testes or ovaries in response to inductive signals. Although germ cells do not induce gonadal development, they play a more active role in females than in males. In the genetic or pharmacologically induced absence of germ cells, testis cords (the embryonic precursor to seminiferous tubules in the adult testis) can develop, but in females, ovary differentiation fails altogether;[52, 53] somatic cells aggregate but deteriorate, leaving only stromal tissue, and ultimately, a fibrous streak. ***After arrival in the nascent gonads, germ cell differentiation into male (prospermatogonia) or female (oogonia) depends on the sex of the gonadal somatic cells and on signals in the surrounding environment rather than on the chromosomal sex of the germ cells themselves.*** In XY/XX mouse chimeras, XY primordial germ cells can develop as oogonia in female embryos, and XX germ cells as prospermatogonia in male embryos.[54]

It is not yet clear whether the signaling molecules that mediate germ cell sex determination act in the developing testis to inhibit meiosis or in the developing ovary to induce meiosis, what those signaling molecules may be, and whether they act directly on the germ cells themselves, or indirectly via actions on gonadal somatic cells.[31] Recent studies in mice aimed at identifying molecular candidates for the putative meiosis inducing or inhibiting factors have focused attention on retinoic acid, which is produced in the mesonephros.

Whereas retinoic acid treatment induces primordial germ cells in male gonadal explant cultures to express *Stra8*, *Scp3*, *and Dmc1* (meiosis marker genes), germ cells in female gonadal explants treated with a retinoic acid inhibitor continue to express *Oct4* (a marker for pluripotent cells).[55] Moreover, Sertoli cells, which surround the germ cells in the developing testis cords, express *Cyp26B1*, a gene encoding an enzyme (CYP26B1) that metabolizes retinoic acid.[56] ***Taken together, these observations suggest that local levels of retinoic acid may regulate germ cell differentiation in the developing gonad, with retinoic acid diffusing from the adjacent mesonephros acting as the functional meiosis inducing factor in female germ cells, and with CYP26B1 produced by Sertoli cells in the developing testis cords acting as the functional meiosis inhibiting factor in male germ cells.***[10] Alternatively, or in addition, Sertoli cells may secrete a specific meiosis inhibiting factor, with one likely downstream target being *Nanos2*, a gene expressed exclusively in male germ cells.[31, 57]

In the male, the primordial germ cells become incorporated into the developing testis cords and enter mitotic arrest as prospermatogonia, resuming proliferation soon after birth. In the female, the primordial germ cells (oogonia) continue to proliferate by mitosis somewhat longer, reaching a peak of 5–7 million by 20 weeks of gestation. However, only some enter meiosis and become primary oocytes, arresting in diplotene of the first meiotic prophase, and become surrounded by a single layer of flattened pregranulosa cells, forming primordial follicles. Those that are not incorporated into primordial follicles degenerate via apoptosis and, by birth, only approximately 1–2 million germ cells remain. The signals for programmed cell death are unknown but seem likely to involve some form of intercellular communication between the primary oocyte and surrounding pregranulosa cells.

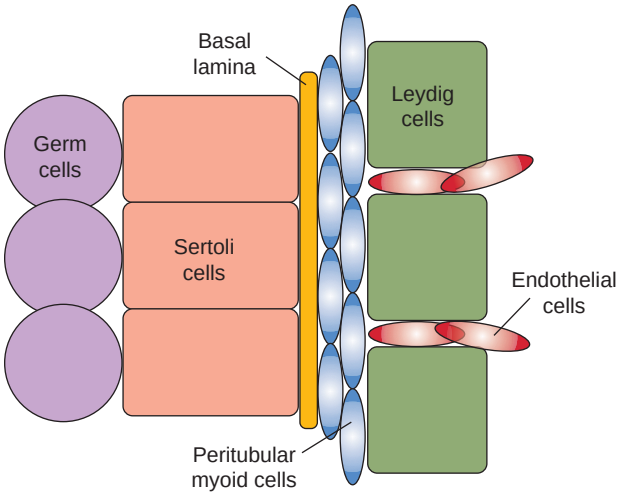
Whereas male germ cells proliferate continuously, the traditional dogma has held that female germ cells proliferate only during embryogenesis and, therefore, that females are born with a finite number of primordial follicles that are steadily depleted and cannot be replenished. However, that dogma has been challenged by studies suggesting that germ line stem cells reside within the bone marrow and may replenish the ovary with new oocytes,[58, 59] stimulating a vigorous scientific debate,[60–66] which continues. Whether or not it occurs normally, the demonstration that mice sterilized by chemotherapy can produce offspring derived from intra-ovarian transplants of germ line stem cells isolated from neonatal or adult ovaries argues that germ line stem cells reside in the ovary and that postnatal oogenesis is possible.[67]

Testis Differentiation and Development

The current model for testis differentiation and development, based primarily on studies in mice, envisions a sequence of events that begins with the formation of the genital ridge, first recognized as a thickening underlying the coelomic epithelium adjacent to the mesonephros. Primordial germ cells migrate into the genital ridge, along with proliferating coelomic epithelial cells, which express *Sf1*. A portion of the epithelial daughter cells expresses *Sry* to become Sertoli cell precursors, the first cell type to differentiate and the only cell type in the developing testis that expresses *Sry*. The subset of somatic cells expressing *Sry* immediately also begins to express *Sox9*, a reliable marker for developing Sertoli cells. In turn, Sox9-positive Sertoli cell precursors secrete other paracrine signaling molecules such as Fgf9 and prostaglandin D_2 (PGD_2), which also play important roles in testis differentiation. Ffg9 reinforces *Sox9* expression and induces neighboring cells to proliferate, thereby increasing the generation of supporting cell precursors that are able to express *Sry*. PGD_2 can induce even Sry-negative cells to express Sox9 and to differentiate into Sertoli cells.[34] Together, Fgf9 and PGD_2 help to maintain Sox9 levels and to ensure a

sufficient number of Sertoli cells to form a testis. Once the number of Sox9-positive cells reaches a critical threshold, Sox9 represses *Sry* expression.

Under the control of *Sry*, Sertoli cells also secrete a factor that induces a migration of cells from the adjacent mesonephros. The developing testis enlarges rapidly with the influx of migrating cells, which differentiate into endothelial cells and Leydig cells upon their arrival in the developing gonad.[10] Male-specific peritubular myoid cells appear to differentiate from cells already within the gonad, flattening and surrounding aggregates of Sertoli cells that organize in layers around clusters of primordial germ cells.[50] The peritubular myoid cells thus help to form the testis cords, later serving to promote the movement of sperm through the seminiferous tubules in the adult testis. Together, the Sertoli cells and peritubular myoid cells induce the development of a basal lamina between them, separating the testis cords from the interstitial tissue. The steroidogenic Leydig cells differentiate within the interstitium, in close proximity to developing blood vessels that derive from endothelial cell precursors. Endothelial cell migration from the mesonephros is specific to the male and required for development of an arterial network that extends throughout the interstitium but not into the testis cords.[50]



Ovary Differentiation and Development

In females lacking a Y chromosome and *SRY*, the bipotential gonad begins to differentiate into an ovary about 2 weeks later than testis development begins in the male. ***Normal ovarian differentiation requires the presence of germ cells; in their absence, the gonadal somatic cells fail to differentiate, indicating some form of communication between germ cells and somatic cells.***[53] *Wnt4* and *Rspo1* are two genes that play an important role in ovarian differentiation; XX mice with targeted deletions of either gene develop ovotestes containing sex cords and functional Leydig cells.[43] *Wnt4* expression is female-specific, suppresses the migration of mesonephric cells as occurs in the developing testis, and is dependent on *Rspo1*.[41, 43] *Rspo1* is specifically upregulated in XX somatic cells from the earliest stages of gonadal differentiation and encodes a secreted protein that, like Wnt4, activates the β-catenin signaling pathway in somatic cells, resulting in a loss of cell-cell adhesion between female germ cells, which is a prerequisite for their entry into meiosis.[43] Consequently, directly or indirectly, Rspo1 regulates female germ cell and ovarian differentiation, by promoting events required for initiation of meiosis, inhibiting migration of mesonephric cells via Wnt4 expression, and by down-regulating Sox9, which drives testis

differentiation. ***Thus, whereas testis differentiation is directed by somatic cells, ovary differentiation requires communication between somatic cells and germ cells.***[68]

Gradually, the developing ovary becomes organized into an outer cortex and an inner medullary region, which ultimately regresses, leaving behind a compressed nest of vestigial tubules and Leydig cells in the hilar region known as the rete ovarii. By 20 weeks of gestation, the ovary achieves mature compartmentalization, consisting of an active cortex containing follicles exhibiting early stages of maturation and atresia, and a developing stroma. Within the cortex, primordial follicles are separated from the somatic cells by a surrounding basement membrane. In some primordial follicles, the pregranulosa cells become cuboidal and proliferate, the ooycte enlarges and produces a zona pellucida (an extracellular glycoprotein matrix deposited between the ooycte and the granulosa cells), and a surrounding layer of thecal cells develops. The remainder stay quiescent until sometime later.

The molecular events that regulate primordial follicle formation and that stimulate or inhibit the initiation of follicular development are understood poorly but appear to involve a variety of factors, all locally produced and regulated, including members of the transforming growth factor β (TGF-β) superfamily of proteins and another family of trophic factors called neurotrophins. Activins, inhibins, antimüllerian hormone (AMH) and bone morphogenetic proteins (BMPs) are members of the TGF-β family of proteins. Activins promote and inhibins retard primordial follicle development, and their relative local concentrations in the fetal ovary during the time of follicle assembly may determine the size of the ovarian follicular pool.[69] AMH appears to be an important inhibitor of primordial follicle growth, and BMPs exert the opposite effect.[69] Neurotrophins and their receptors are essential for the differentiation and survival of various neuronal populations in the central and peripheral nervous systems, but their presence in the developing ovary suggests they also play a role in ovarian development. Four mammalian neurotrophins have been identified, including nerve growth factor (NGF), brain-derived neurotropic factor (BDNF), neurotrophin-3 (NT-3), and neurotrophin 4/5 (NT-4/5), all of which exert their actions via binding to high-affinity trans-membrane tyrosine kinase receptors encoded by members of the *trk* proto-oncogene family (NGF to TrkA, BDNF and NT-4/5 to TrkB, and NT-3 to TrkC).[70] Observations in NGF- and TrkA-null mice indicate that NGF stimulates the proliferation of ovarian mesenchymal cells during the early stages of follicular assembly and promotes differentiation and synthesis of FSH receptors in granulosa cells. Similar experiments with TrkB-null mice suggest that TrkB signaling is required for oocyte survival after follicular assembly and for preantral follicular development.[70] The specific signaling mechanisms that mediate the effects of activins, inhibins, BMPs and neurotrophins remain to be established.

Other paracrine factors mediate a bi-directional communication between oocytes and their surrounding granulosa cells. Oocytes are linked to their investment of granulosa cells via gap junctions which allow passage of small molecules such as ions (e.g., calcium), metabolites (e.g., pyruvate, nucleic acids, inositol), amino acids (e.g., L-alanine), cholesterol, and intracellular signaling molecules (e.g., cyclic adenosine monophophate, cAMP) between granulosa cells and oocytes. In mice, targeted deletions of gap junction proteins (known as connexins), disrupt follicular and oocyte development.[68] Oocytes are unable to use glucose as an energy source to support meiotic maturation, cannot transport certain amino acids, and lack both the enzymes necessary for cholesterol synthesis and the receptors for its uptake from carrier-borne sources. Consequently, they are dependent on adjacent granulosa cells to metabolize glucose into a usable energy substrate, such as pyruvate, for transport of essential amino acids, such as L-alanine, and for synthesis and transfer of cholesterol.[71] ***To meet their needs, oocytes stimulate glycolysis, amino acid transport, and cholesterol synthesis in granulosa cells via paracrine and juxtacrine signals that promote expression of transcripts involved in these metabolic processes, at least in some species.***[71] Candidate signaling molecules include closely related members of the TGF-β family, growth differentiation factor 9 (GDF9) and BMP15; both are expressed robustly in oocytes and appear crucial for normal ovarian follicle development in mammalian species.[72]

Genital Duct Differentiation and Development

Caspar Wolff described the mesonephros in 1759 in his doctoral dissertation, at the age of 26.[73] The paired structures were named wolffian bodies by the 19th century embryologist, Rathke, in recognition of Wolff's initial discovery and description. Johannes Müller, a German physiologist, described the embyrology of the genitalia in 1830. The paramesonephric ducts received his name, not because of his original contributions, but in recognition of his ability to synthesize the existing literature into a coherent concept.

The mesonephric (wolffian) and paramesonephric (müllerian) ducts are discrete primordia that coexist in all embryos during the ambisexual period of development (up to 8 weeks). Thereafter, one duct system persists, giving rise to specialized ducts and glands, and the other regresses, leaving behind only nonfunctional vestiges. ***The wolffian duct develops first, differentiates into the epididymis, vas deferens, and seminal vesicles in males, and regresses in females. The müllerian duct develops later, even after the beginning of sex determination, differentiates into the fallopian tubes, uterus, and upper portion of the vagina in females, and regresses in males.***

The hormonal control of genital duct differentiation and development was established by the classic experiments of Alfred Jost.[74] His landmark studies demonstrated that hormones produced by the testis direct the sexual differentiation of both the internal and external genitalia in the male. Whereas testosterone stabilizes and promotes development of the wolffian ducts, AMH directs the regression of the müllerian system. In females, the wolffian ducts regress, in the absence of testosterone, and the müllerian ducts develop fully, in the absence of AMH. Although not yet clearly defined, our knowledge of the molecular mechanisms involved is growing steadily.

Mesonephric (Wolffian) Duct Development

Testosterone is secreted by the fetal testes soon after Leydig cell formation (at 8 weeks gestation) and rises rapidly to peak concentrations at 15–18 weeks. Fetal testosterone stimulates development of the wolffian duct system, from which the epididymis, vas deferens, and the seminal vesicles derive. Testosterone levels in the male fetus correlate with Leydig cell development, overall gonadal weight, 3β-hydroxysteroid dehydrogenase activity, and chorionic gonadotropin (hCG) concentrations. As maternal hCG levels decline, beginning at approximately 20 weeks gestation, Leydig cell testosterone secretion comes under the control of fetal pituitary luteinizing hormone (LH). In the absence of LH, as in males with anencephaly and other forms of congenital hypopituitarism, Leydig cells all but disappear and the internal and external genitalia do not develop fully.[75]

Testosterone can reach the developing wolffian duct system via the systemic fetal circulation, but the paracrine actions of testosterone produced in nearby Leydig cells are more important for the stabilization and differentiation of the wolffian duct. ***High local concentrations of testosterone stimulate the ipsilateral wolffian duct to differentiate into the epididymis, vas deferens, and seminal vesicle. Duct system differentiation proceeds, therefore, according to the nature of the adjacent gonad.*** High concentrations of testosterone are required because the duct does not have the ability to convert testosterone to dihydrotestosterone (DHT).[76] In rodents, wolffian development can be induced in female embryos by treatment with exogenous androgens, but only to a limited extent,[77] because exogenous androgen treatment cannot achieve and maintain the high local concentrations required to induce duct differentiation. For the same reason, the wolffian ducts do not develop in female fetuses exposed to excess endogenous adrenal androgens, as in

classical congenital adrenal hyperplasia, or to excess maternally-derived androgens, as occurs in women with pregnancy luteoma. Testosterone acts via binding to androgen receptors in the wolffian duct, which are detectable in both males and females, but androgen production in females does not approach the levels required to promote wolffian duct differentiation.[77]

The paired wolffian ducts arise within the urogenital ridge during embryogenesis, running its length and terminating in the cloaca. The ducts form by a rearrangement of mesenchymal cells rather than by cell proliferation.[78] The regulatory signals involved have not been established, but evidence from studies in mice having targeted deletions of candidate genes has implicated a number of transcription factors, including Pax2, Lim1, and Emx2. All are expressed in mesenchymal condensations before duct formation and respond to opposing signals from adjacent mesoderm and overlying ectoderm, which appear to restrict their expression to the specific area in the mesoderm from which the ducts arise.[78] Along the axis of the forming wolffian ducts, a series of smaller tubules develop. The most anterior tubules fuse with the wolffian duct to become the precursors of the efferent ducts, ultimately connecting the testis to the epididymis; the more posterior or caudal tubules regress. In the human, parallel efferent ductules form multiple connections with the head (caput) of the epididymis.

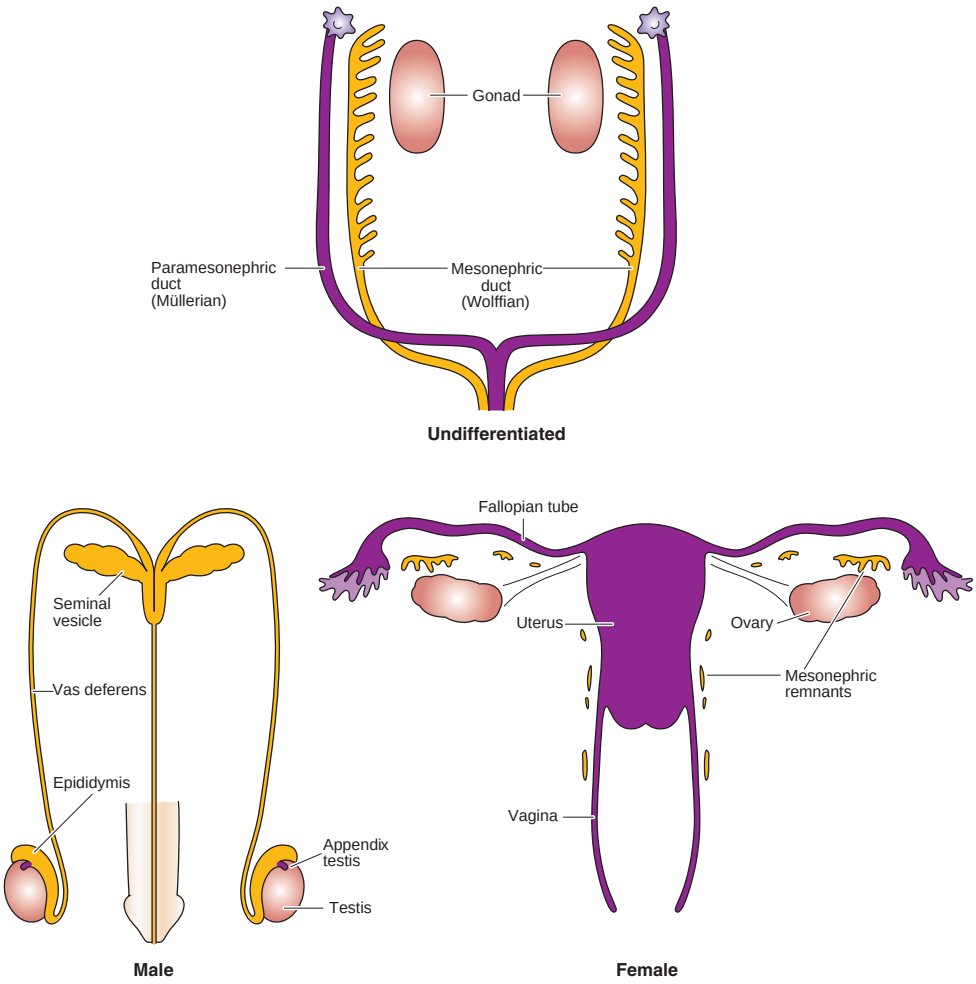
Gradually, the straight wolffian ducts elongate and coil as a result of epithelial cell proliferation, stimulated by testosterone transported from the testes via the lumen of the duct, as well as by growth factors (e.g., epidermal growth factor, EGF; basic fibroblast growth factor, bFGF), which also are found in high concentrations in luminal fluid.[79, 80] The structure of the developing epididymis becomes increasingly complex. Elongation and three-dimensional coiling begin at the end nearest the testis (the caput), and progress distally, except at the most caudal end of the duct, which remains straight and ultimately gives rise to the vas deferens. The factors that stimulate or control coiling of the duct are uncertain but may involve a combination of regional signals from the surrounding mesenchyme, focal "hot spots" of epithelial cell growth, and physical space limitations.[78] Region-specific expression of homeobox (HOX) genes, which are transcriptional regulators of patterning, appears important for the differentiation of the duct into its morphologically and functionally distinct segments (caput, corpus, and caudal regions). For example, *Hoxa10* and *Hoxa11* appear to act distally to define the boundary between the epididymis and vas deferens.[81] Others HOX genes appear to direct differentiation of the seminal vesicle (derived from the posterior wolffian duct) and the prostate (derived from the urogenital sinus).[82] Evidence suggests that HOX genes may act by controlling the expression of other morphogenic factors such as inhibin beta A, which is expressed most highly in the greatly coiled caput region and to a progressively lesser extent in the mesenchyme surrounding more distal regions of the duct.[83] Growth factors in the testicular fluid also appear to play an important role in cellular differentiation along the length of the epididymis.[84]

The extraordinary length of the epididymis—approximately 6 meters in the human—reflects its functional importance. ***As sperm leave the testis, they are functionally immature, having neither full motility nor the ability to recognize and fertilize an oocyte.*** They mature and acquire those functions as they pass through the epididymis, undergoing both biochemical and physical changes in a changing luminal environment regulated by a region-specific epididymal epithelium. The vas deferens is distinguished from the epididymis by its structure and by its function. It originates at the caudal end of the epididymis, where sperm are stored, and ends in the ejaculatory duct, which joins with the urethra. The vas deferens is surrounded by layers of smooth muscle that contract in response to sympathetic nerve stimulation, moving sperm through the vas deferens, into the ejaculatory duct (formed by the union of the vas with the duct of the seminal vesicle), and into the urethra.

Paramesonephric (Müllerian) Duct Development and Regression

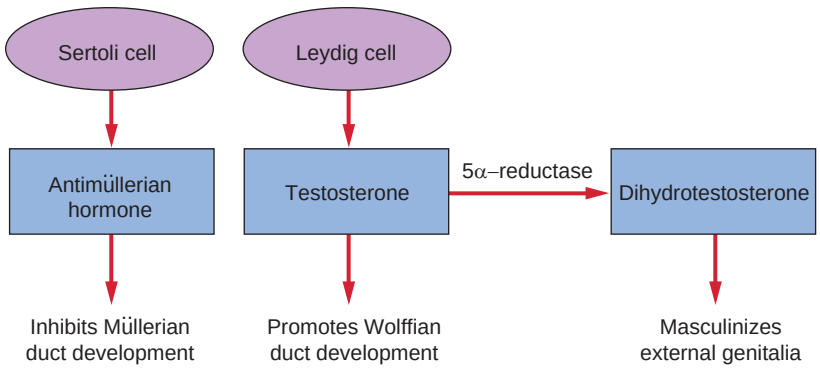
The müllerian ducts begin by invagination of the coelomic epithelium, which progresses until reaching the wolffian ducts, then elongate, by cellular proliferation, along the length of the wolffian ducts until reaching and fusing with the urogenital sinus.[85] The wolffian ducts make no direct contribution to the müllerian ducts, but are essential for normal müllerian development, serving as a guide or migrational template.[86] ***If the wolffian ducts do not form, müllerian duct development also fails. Consequently, abnormalities in the renal system are highly associated with abnormalities in development of the fallopian tubes, uterus, and upper vagina.***

Müllerian duct development can be separated into three phases, each controlled by different genes, as demonstrated by careful analyses of mutant mice. Selection of the cells in the coelomic epithelium that will become the müllerian ducts is controlled by *Lim1*, which encodes a protein also involved in formation of the wolffian ducts.[87] Expression of *Wnt4* and other genes in the *Wnt* family (*Wnt7a*, *Wnt9b*) appears necessary for epithelial invagination.[85] *Pax2* is required for duct elongation[88] and, together with *Pax8*, also for differentiation of the duct into a uterus and vagina.[89] Directly or indirectly, müllerian duct development also involves other genes such as those encoding retinoic acid receptors; mice having targeted deletions of the retinoic acid receptors fail to develop müllerian ducts or to differentiate specific portions of the duct.[90]



AMH is a member of the TGF-β superfamily family of growth and differentiation factors that includes inhibin and activin.[91, 92] The gene encoding AMH is located on the short arm of chromosome 19 (19p13.3). Like other members of the TGF-β superfamily, AMH signaling is mediated via a heterodimeric receptor consisting of a type I and a type II serine/threonine kinase receptor; the type II part of the receptor mediates ligand specificity and the type I receptor activates a downstream signaling cascade. The specific type II receptor that binds AMH, called AMHR2, has been isolated in several mammalian species; in the human, the gene encoding AMHR2 is located on chromosome 12 (12q13). Three different type I receptors have been linked to AMH signaling—ALK2, ALK3, and ALK6; ALK2 and ALK3 appear particularly important, because decreased expression or deletion of either disrupts müllerian duct regression.[85] ***AMH gene expression is induced by SOX9 in Sertoli cells soon after testicular differentiation and results in the ipsilateral regression of the müllerian ducts by 8 weeks of gestation, before the emergence of testosterone and stimulation of the wolffian ducts.***[93] Inactivating mutations of AMH or AMHR2 result in persistent müllerian ducts in males.[94]

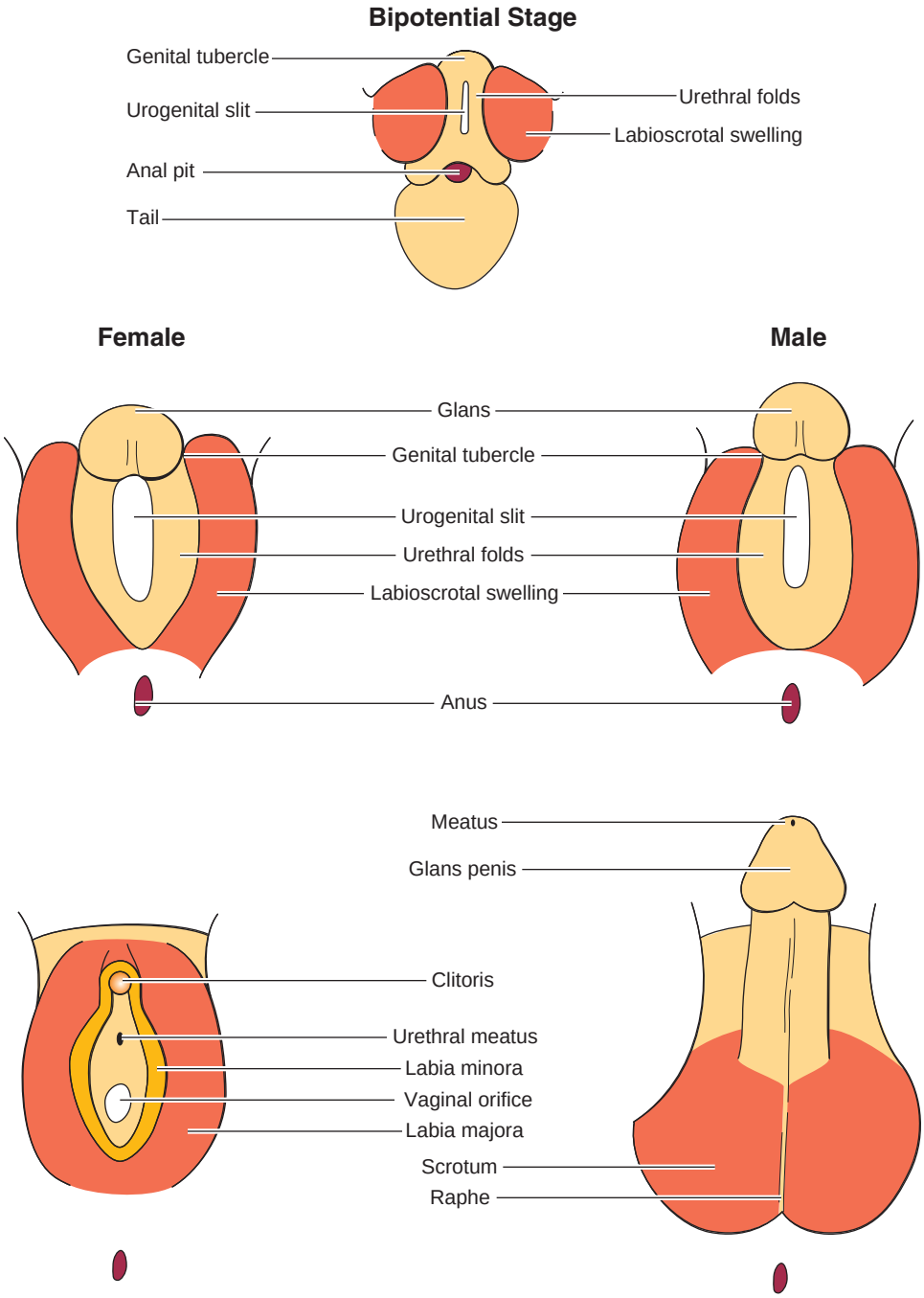
The process of müllerian duct regression involves a number of genes, but studies in mice indicate that *Wt1* and *Wnt7a* play key roles. AMH signaling induces coelomic epithelial cells expressing *Wt1*, *Amhr2*, and *Alk3* to migrate and surround the müllerian duct, transforming to mesenchymal cells in the process.[95, 96] *Wnt7a* expression in the müllerian duct mesoepithelium promotes secretion of a signaling molecule (Wnt7a) that activates Amhr2 in the neighboring mesenchymal cells via Wt1, which binds and activates the *Amhr2* promoter.[85] At the same time, β-catenin gene expression increases in the mesenchymal cells surrounding the duct, and accumulation of β-catenin is accompanied by increased apoptosis in the müllerian ductal epithelium.[95, 96] Whether Wnt-dependent β-catenin activity is required to induce Amhr2 expression or functions downstream of AMH signaling, or both, is not yet clear. Regardless, the process of müllerian duct regression appears to involve both apoptosis and the transition of ductal epithelial cells to mesenchymal cells.[85] The matrix metalloproteinase MMP2 also plays a role, by mediating destruction of the extracellular matrix; available evidence indicates that MMP2 activity also is AMH-dependent, although the mechanism involved has not been established.[97]



Development of the External Genitalia

In the bipotential state, which persists until 9 weeks of gestation, the external genitalia consist of a genital tubercle, a urogenital sinus, and lateral labioscrotal folds or swellings. Unlike the internal genitalia where both duct systems initially coexist, the external genitalia are neutral primordia able to develop into either male or female structures, depending on gonadal steroid hormone signals.

In the male, the Leydig cells of the fetal testis begin to secrete testosterone at 8–9 weeks of gestation and masculinization of the external genitalia begins one week later, at approximately 10 weeks. ***The genital tubercle grows, forming the penis, the edges of the urogenital sinus fuse to form the penile urethra, and the labioscrotal folds fuse to form a scrotum.*** The process typically is completed by 12 to 14 weeks of gestation. Thereafter, the principal change is in the growth and length of the penis. Complete development of the male external genitalia and differentiation of the prostate requires the conversion of testosterone to dihydrotestosterone (DHT), via the action of the intracellular enzyme 5α-reductase. The genital tubercle and the labioscrotal swellings are highly sensitive to DHT, being rich in both androgen receptors and 5α-reductase activity.



In the female, and in males with defects in androgen synthesis or action, the external genital primordia do not masculinize. ***The genital tubercle remains small and becomes the clitoris, the margins of the urogenital sinus remain separate and form the labia minora, the labioscrotal folds form the labia majora, and the urogenital sinus develops into the lower vagina and urethra.***

In females, abnormal androgen exposure between 9 and 14 weeks of gestation results in varying degrees of masculinization, such as clitoral hypertrophy and labial fusion. In males, the external genitalia will not masculinize completely if androgen action is deficient during the same critical time interval, yielding a small phallus, hypospadias, or scrotal defects. ***In both sexes, because the external genitalia share a common origin, genital ambiguity results from abnormalities in androgen action—in females from too much, and in males from too little.***

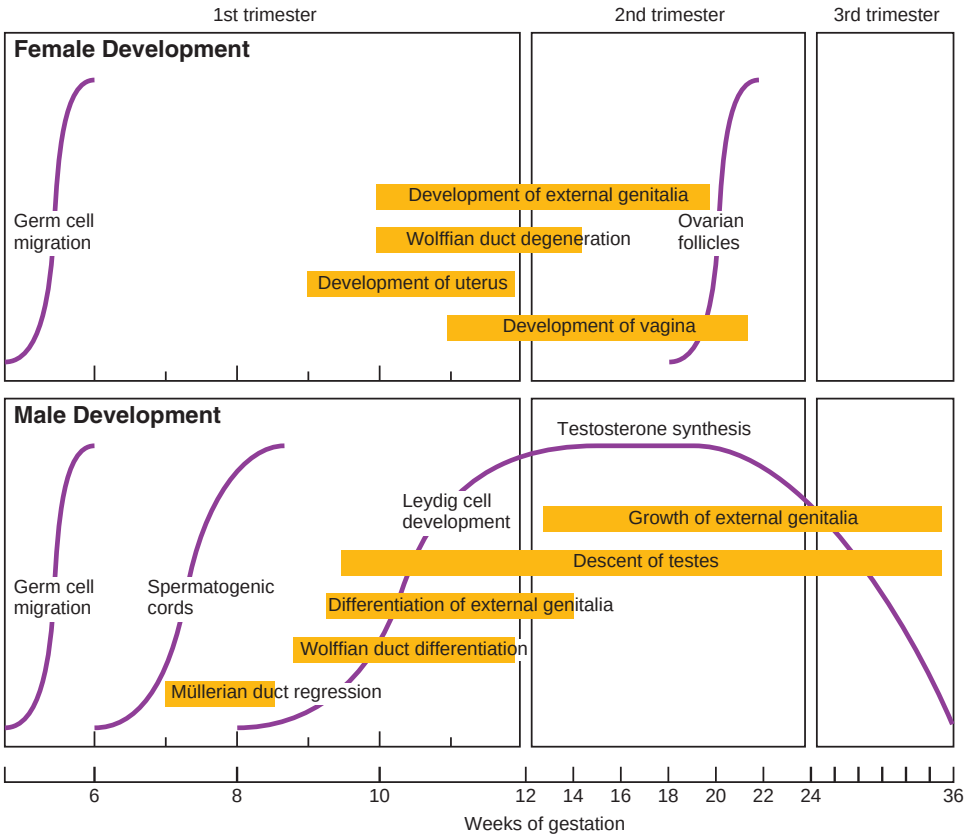
Sexual Differentiation of the Central Nervous System

Experimental evidence from studies in rodents and nonhuman primates suggests strongly that the fetal hormonal environment directs sexual differentiation of not only the genitalia, but also the central nervous system (CNS). Treatment with testosterone during early development increases reproductive and other behaviors more common in males and decreases behaviors more common in females. These observations suggest that testosterone and its metabolites play a role in brain development and neuronal organization.[98, 99]

Most of our knowledge about the early influence of testosterone on the brain and behavior in humans derives from clinical disorders associated with abnormal hormone production in early life, such as congenital adrenal hyperplasia (CAH). In male fetuses with classical CAH, sexual development progresses normally, but in female fetuses, testosterone is markedly elevated and causes masculinization of the external genitalia (clitoral enlargement and labial fusion). ***Studies in girls with classical CAH indicate that increased prenatal androgen exposure also affects their brains and behavior.*** Compared to unrelated age- and sex-matched controls or to unaffected female relatives of similar age, their toy preferences (vehicles, weapons) and play behaviors (rough, active play) are more typical of boys than of girls, to an extent that correlates with the severity of their disorder.[4, 100, 101] Girls with classical CAH also display more physical aggression and greater spatial abilities.[102–104] Although less well studied, there also is evidence to suggest that antenatal androgen exposure may influence sexual orientation. Whereas most females with classical CAH are heterosexual, as a group they are more likely to exhibit a bisexual or homosexual orientation; the effect is more pronounced in women with the severe salt-wasting form of CAH than in those with the milder, simple virilizing CAH.[105] Other studies observing a significant linear relationship between childhood behaviors and maternal serum or amniotic fluid testosterone concentrations during pregnancy suggest that even normal variations in prenatal androgen exposure may influence behavior, in both males and females.[106, 107]

Presumably, the behavioral consequences of variations in prenatal androgen exposure reflect changes in neuronal development and organization. In rodents, an area of the anterior hypothalamic/preoptic region, called the sexually dimorphic nucleus of the preoptic area, is substantially larger in males than in females and treatment with androgens increases its size in females.[98] Whereas no comparable, specific, sexually dimorphic region has been identified in the human brain, there is some evidence from studies in females with classical CAH using functional MRI to suggest that prenatal androgen exposure may

"masculinize" certain regions of the brain such as the amygdala, which is involved with regulation of emotion and aggression.[108] Variations in fetal hormonal programming may contribute, therefore, to the spectrum of psychosexual behavior observed in humans. In addition, gender role is influenced heavily by the sex of rearing and by social interactions based upon genital appearance and secondary sexual characteristics.



Disorders of Sexual Development

Disorders of sexual development (DSD) are congenital conditions characterized by atypical development of chromosomal, gonadal, or phenotypic sex. Traditionally, they have been classified according to gonadal sex. A ***true hermaphrodite*** has both ovarian and testicular tissue. A ***male pseudohermaphrodite*** has testes, but a female genital phenotype, and a ***female pseudohermaphrodite*** has ovaries, but masculine genital characteristics. However, recent advances in molecular genetic diagnosis and increasing awareness of ethical issues and patient advocacy concerns suggested the need to re-examine the traditional classification scheme and to retire gender-based terms that many now consider pejorative.

Ideally, a classification system must be flexible, to allow incorporation of new information, logical, to maintain a consistent structure, reflect genetic cause when that is known, and accommodate the spectrum of phenotypic variation. The classification and nomenclature used here, organized by chromosomal composition and causation, conforms with recommendations arising from a 2006 consensus conference involving experts in pediatric endocrinology and other specialties involved in the management of patients with disorders of sexual development.[109]

46,XX Disorders of Sexual Development

- Disorders of gonadal (ovarian) development
- Ovotesticular disorder of sexual development (true hermaphroditism)
- Testicular disorder of sexual development (46,XX male sex reversal)
- Gonadal dysgenesis
- Androgen excess—Fetal origin (congenital adrenal hyperplasia)
  - 21-Hydroxylase (P450c21) deficiency
  - 11β-Hydroxylase (P450c11β) deficiency
  - 3β-Hydroxysteroid dehydrogenase deficiency
- Androgen excess—Fetoplacental origin
  - Aromatase (P450arom) deficiency
  - P450 oxidoreductase deficiency
- Androgen excess—Maternal origin (gestational hyperandrogenism)
  - Drug ingestion
  - Excess androgen production
    - Pregnancy luteoma
    - Theca-lutein cysts
- Other disorders of genital development
  - Cloacal extrophy
  - Müllerian agenesis (Mayer-Rokitansky-Küster-Hauser syndrome)
  - Müllerian, renal, and cervicothoracic somite dysplasia (MURCS association)

46,XY Disorders of Sexual Development

- Disorders of gonadal (testicular) development
  - Complete gonadal dysgenesis (Swyer syndrome)
  - Partial gonadal dysgenesis
  - Testicular regression syndrome
  - Ovotesticular disorder of sexual development
- Disorders of androgen synthesis
  - Steroid 5α-reductase deficiency
  - 17α-Hydroxylase (P450c17) deficiency
  - 3β-Hydroxysteroid dehydrogenase deficiency
  - 17β-Hydroxysteroid dehydrogenase deficiency
  - P450 oxidoreductase deficiency
  - Steroid acute regulatory (StAR) protein deficiency
- Disorders of androgen action
  - Complete androgen insensitivity syndrome
  - Incomplete (partial) androgen insensitivity syndromes
- LH receptor defects
  - Leydig cell hypoplasia
- Disorders of antimüllerian hormone (AMH) and its receptor
  - Hernia uterine inguinale syndrome

Sex Chromosome Disorders of Sexual Development

- 45,X (Turner syndrome and variants)
- 47,XXY (Klinefelter syndrome and variants)
- 45,X/46,XY (mixed gonadal dysgenesis, ovotesticular disorder of sexual development)
- 46,XX/46,XY (chimerism, ovotesticular disorder of sexual development)

46,XX Disorders of Sexual Development

Disorders of sexual development in chromosomal females can result from abnormalities in gonadal development, but most are caused by androgen excess, which may be of fetal, fetoplacental, or maternal origin. Excess fetal androgen production results from steroidogenic enzyme deficiencies causing congenital adrenal hyperplasia. Androgen excess of fetoplacental origin results from enzyme deficiencies involving both the fetal adrenal and the placenta. Maternal androgen excess can result from the ingestion of drugs having androgenic properties and from disorders causing gestational hyperandrogenism.

Disorders of Gonadal (Ovarian) Development

Rarely, 46,XX disorders of sexual development (DSD) can result from abnormalities of gonadal development, which include ovotesticular DSD (true hermaphroditism), testicular DSD (46,XX sex reversal), and gonadal dysgenesis.

Ovotesticular Disorder of Sexual Development (True Hermaphroditism)

Ovotesticular DSD previously was called true hermaphroditism.[109] Hermaphroditus, the Greek god with bisexual attributes, was the child of Hermes—the god of athletics, secrets, and occult philosophy—and Aphrodite, the goddess of love. The bisexual theme was immortalized in Greek and Roman sculptures depicting a woman with male external genitalia. Pliny (23–79 A.D.) was the first to apply the term hermaphrodite to humans, offering a description in his massive work, *Historia Naturalis*.

Ovotesticular DSD is a rare condition characterized by mixed ovarian and testicular tissue, which may include bilateral ovotestes or an ovotestis and a contralateral ovary or testis. The disorder is described here because the majority of patients have a 46,XX karyotype. However, because 7% of patients with ovotesticular DSD have a 46,XY karyotype and 10–40% exhibit chromosomal mosaicism,[110] the disorder also must be listed among the causes of 46,XY- and Sex Chromosome Disorders of Sexual Development.

Whereas gonads containing testicular tissue are observed most frequently on the right, normal ovaries are observed most often on the left.[110] Usually, both müllerian and wolffian internal genital structures are present and, as could be predicted, internal genital structures correspond with the adjacent gonad. Whereas most have a vagina, the uterus can be normal and functional, hypoplastic, vestigial, or altogether absent.[110, 111] External genital development reflects the level of androgen production and exposure and the phenotype can range widely, from ambiguous genitalia to isolated hypospadias. Most are virilized sufficiently to allow male sex assignment, but three-fourths develop gynecomastia and half menstruate after puberty.

The genetics and pathophysiology of ovotesticular DSD are not well established. Mechanisms that might explain the testicular development include the translocation of testis-determining genes from the Y to the X chromosome or an autosome, and autosomal dominant mutations that promote testis development in the absence of a Y chromosome.[112]

In one individual, the condition has been associated with an inactivating mutation in the *RSPO1* gene,[113] which is located on chromosome 1p34.2–3

Testicular Disorder of Sexual Development (46,XX Sex Reversal)

Testicular DSD is a rare "sex reversal" syndrome in which the chromosomal sex (46,XX) is not consistent with the gonadal sex (testes). The disorder was first described by de la Chapelle in 1964,[114] and can be divided into two types, *SRY*-positive and *SRY*-negative. ***Approximately 90% of cases result from abnormal recombination between the distal portions of the short arms of the X and Y chromosomes and transfer of SRY from the Y to the X chromosome during male meiosis; in 10% of cases, SRY cannot be detected.***[115] In most *SRY*-negative patients, the mechanism causing testis development cannot be determined.[115–118]

Although some patients with *SRY*-positive testicular DSD have ambiguous genitalia, which may result from preferential inactivation of the *SRY*-bearing X chromosome,[119] the large majority are sterile males with normal genital development, a normal male hair pattern, and short stature. Consequently, unless they have cryptorchid testes, most are not recognized until after puberty, when they may present with hypogonadism, gynecomastia, and/or infertility.[115] In contrast, *SRY*-negative XX males usually have ambiguous genitalia and often develop gynecomastia or fail to masculinize fully after puberty.[115–118] Rarely, they may exhibit occult gonadal mosaicism for *SRY*.[120] In some, the phenotype has been linked to a duplication of sequences on chromosome 17q, including the *SOX9* gene, which acts downstream of *SRY* in the testis-determining pathway.[32, 121] However, in most patients with *SRY*-negative testicular DSD, the cause remains unclear. In theory, XX male sex reversal might result from an inactivating mutation or deletion in genes encoding factors that inhibit testis development, but there is no direct evidence they are a cause of testicular DSD.[116]

Gonadal Dysgenesis

Some individuals with primary amenorrhea, hypergonadotropic hypogonadism, and gonadal dysgenesis (streak gonads) have a normal 46,XX karyotype, providing indirect evidence that autosomal genes also play a critical role in ovarian differentiation. Affected women are normal in stature and, in most cases, have no apparent somatic anomalies. A wide variety of candidate genes have been identified, primarily via experiments involving murine knock-out models, including several that encode DNA and RNA binding proteins and transcription factors expressed during oogenesis.[122]

Androgen Excess—Fetal Origin (Congenital Adrenal Hyperplasia)

Virilizing CAH is a genetic disorder caused by enzyme defects in adrenal cortisol biosynthesis. More than 90% of cases result from a deficiency in the enzyme 21-hydroxylase.[123–125] Deficiencies of 11β-hydroxylase and 3β-HSD are less common causes of CAH. ***In all, the pathophysiology relates primarily to decreased cortisol production, which stimulates a compensatory increase in pituitary adrenocorticotropic hormone (ACTH) secretion, causing adrenal hyperplasia; increased levels of steroid hormones proximal to the***

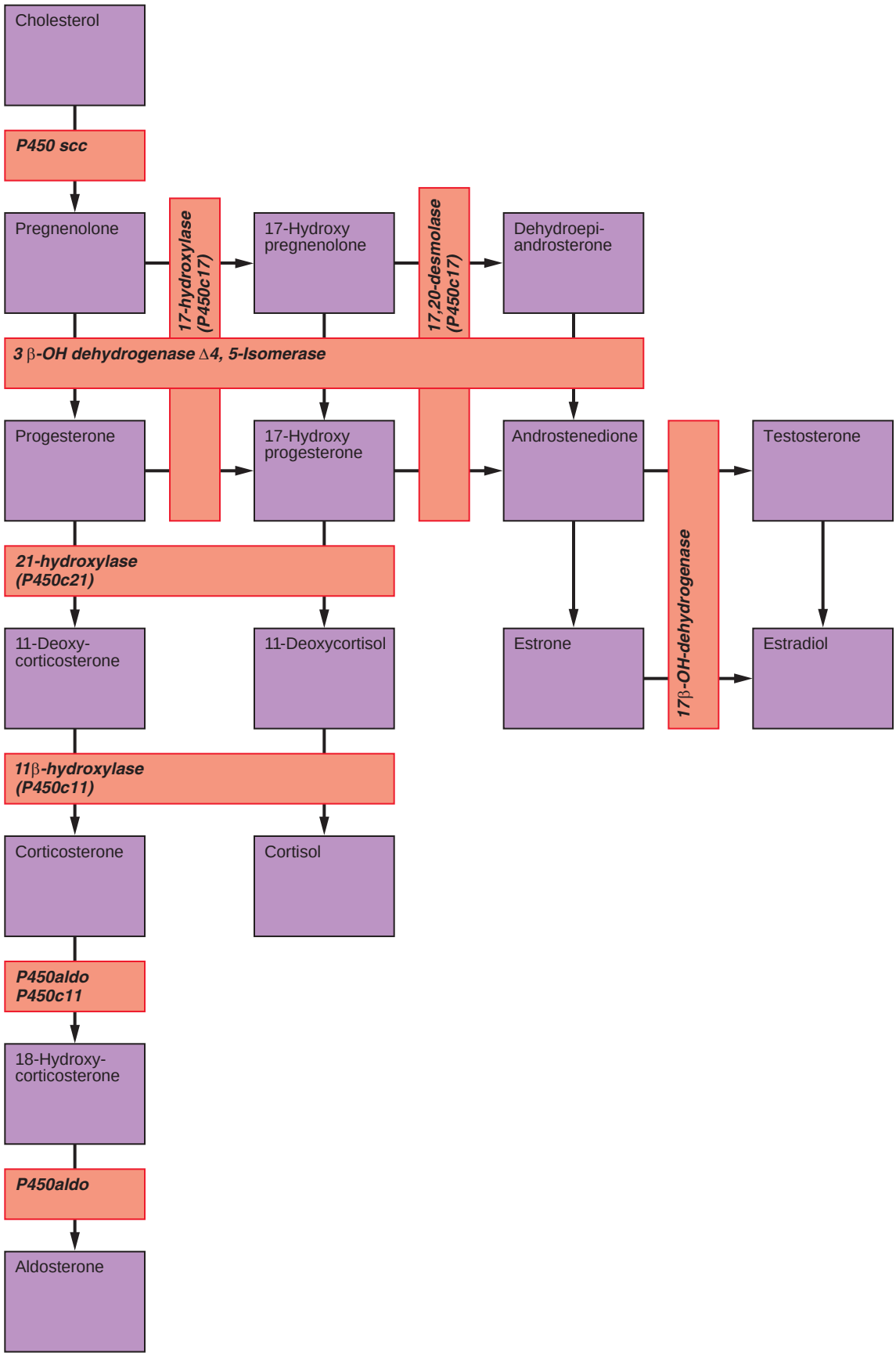
enzyme block seek an alternative metabolic pathway, resulting in increased production of androgens.

In females, the classic forms of CAH (with and without salt-wasting) are characterized by genital ambiguity. ***Depending on the time, duration, and level of exposure, abnormally high androgen concentrations in utero result in varying degrees of clitoral enlargement and labial fusion and abnormalities of the urethra and vagina; generally, the urethra and vagina share a urogenital sinus that opens at the base of the clitoris.*** The fetal adrenal cortex does not achieve a significant level of function before 10 weeks gestation and, by that time, the vagina and urethra normally have become separated. However, between 10 and 12 weeks, rising androgen levels can promote progressive clitoral enlargement, labial fusion, and even partial closure of the urethra. At birth, the genital anatomy is similar to that in males with hypospadias and bilateral cryotochidism and can result in incorrect sex assignment. The effects of elevated adrenal androgen levels arising after 12–14 weeks of gestation are more limited. ***Female external genital development normally is not completed until approximately 20 weeks of gestation and the size of the clitoris depends more on the level than on the timing of excess androgen exposure.*** Development of the internal genitalia is normal in females with classical CAH because the excess androgen derives from the adrenals and the normal ovaries produce neither AMH nor significant amounts of androgen. ***Absent AMH and the high local androgen concentrations required to promote wolffian duct development, the fallopian tubes, uterus, and upper vagina develop normally.***

21-Hydroxylase (P450c21) Deficiency

The enzyme 21-hydroxylase (also designated P450c21 and CYP21A2) mediates the conversion of 17α-hydroxyprogesterone (17OHP) to 11-deoxycortisol (the immediate precursor of cortisol) and of progesterone to 11-deoxycorticosterone (an intermediate steroid in aldosterone synthesis). ***CAH due to 21-hydroxylase deficiency is the most frequent cause of sexual ambiguity and the most common endocrine cause of neonatal death.*** The more serious "salt-wasting" variety of classical 21-hydroxylase deficiency is characterized by severe deficiencies of both cortisol and aldosterone, resulting in salt-wasting and dehydration, in addition to virilization. In the less severe "simple virilizing" form of the disorder, elevated levels of ACTH are able to drive sufficient glucocorticoid and mineralocorticoid production to prevent circulatory collapse, but excess androgen production *in utero* results in masculinization of the external genitalia. The third and least severe "nonclassical" form of 21-hydroxylase deficiency generally does not become apparent until adolescence or early adulthood, when abnormally high androgen levels cause hirsutism and menstrual irregularities.

Data derived from neonatal screening programs for detection of classical CAH indicate that prevalence varies widely with ethnicity. Whereas the overall prevalence is approximately 1 in 15,000 live births,[126] prevalence ranges from 1 in 28,000 Chinese[127] and between 1 in 5,000 and 1 in 23,000 Caucasians,[128, 129] to as high as 1 in 280 Yupic Eskimos.[130] In the United States, the prevalence of classical CAH is lower in African-Americans (1 in 42,000) than in Caucasians (1 in 15,500).[131] Approximately two-thirds exhibit salt-wasting and one-third has the simple virilizing form of the disorder.


Cholesterol
P450 scc
Pregnenolone
17-hydroxylase (P450c17)
17-Hydroxy pregnenolone
17,20-desmolase (P450c17)
Dehydroepi-androsterone
3 β-OH dehydrogenase Δ4, 5-Isomerase
Progesterone
17-Hydroxy progesterone
Androstenedione
Testosterone
21-hydroxylase (P450c21)
11-Deoxy-corticosterone
11-Deoxycortisol
Estrone
17β-OH-dehydrogenase
Estradiol
11β-hydroxylase (P450c11)
Corticosterone
Cortisol
P450aldo P450c11
18-Hydroxy-corticosterone
P450aldo
Aldosterone

Nonclassical 21-hydroxylase deficiency is one of the most common autosomal recessive diseases and, as in the classical form of the disorder, prevalence varies with ethnicity. Nonclassical 21-hydroxylase deficiency affects between 1 in 100 and 1 in 1,000 Caucasians,[130–132] and may be even more common among those of Mediterranean, Hispanic, Slavic, and Eastern European Jewish descent.[133] Most affected individuals are not identified in neonatal screening programs because their serum levels of 17OHP are not sufficiently elevated.[134] Estimates of the carrier frequency (heterozygotes) for nonclassical 21-hydroxylase deficiency generally have ranged between 1 in 60 and 1 in 80 individuals,[127, 130] but have been as high as 1 in 10 in a European population.[135]

All forms of CAH, including 21-hydroxylase deficiency, are transmitted as autosomal recessive disorders. Humans have 2 *CYP21A* genes; one is a nonfunctional pseudogene (*CYP21A1*, also designated *CYP21P*, encoding an inactive form of the enzyme), and the other is the active gene (*CYP21A2*). The 2 genes have greater than 90% homology and reside in the same region within the HLA histocompatibility complex on the short arm of chromosome 6 (6p21.3), which provides ample opportunity for recombination during meiosis.[136–139] Most *CYP21A2* mutations (approximately 75%) result from non-reciprocal gene conversions in which a segment of the *CPY21A1* pseudogene is inserted into the active *CYP21A2* gene, altering its sequence and resulting in point mutations that yield a defective enzyme.[138–142] Approximately 20% of *CYP21A2* mutations result from unequal cross-over exchanges between the two genes, yielding a larger fusion gene that produces an enzyme having reduced or no activity.[128, 133, 142, 143] About 20 gene conversion mutations account for almost all of the affected alleles observed among various ethnic groups.[141, 144–151] The remaining 5% of patients with *CYP21A2* mutations have 1 or 2 of the more than 60 different point mutations that have been identified.[141, 144–146]

Women who carry a classic mutation are at risk for having a child with the severe form of the disorder. They may be asymptomatic, having one classic mutation and one normal allele, or exhibit the nonclassical form of CAH, having one classic mutation and a variant allele associated with mild enzyme deficiency (compound heterzygote). Compound heterozygotes having two variant alleles can exhibit the features of nonclassical CAH but are not at risk for having a child with classical CAH.

Although phenotype does not reliably predict genotype, the effect of a given mutation generally can be predicted by site-directed mutagenesis and expression, and by analysis of enzyme activity *in vitro*.[132, 141, 147–150, 152–159]

- The salt-wasting form of classical 21-hydroxylase deficiency usually is associated with large gene deletions or a mutation that affects splicing and results in no enzyme activity.
- Patients with the simple virilizing form of classical 21-hydroxylase deficiency most often have point mutations that result in low but detectable enzyme activity (e.g., 1–2% of normal) that supports adequate aldosterone and cortisol production.
- Those with the nonclassical form of classical 21-hydroxylase deficiency usually are compound heterozygotes, having one classic mutation and one variant allele or two variant alleles; the phenotype of compound heterozygotes usually correlates with the less severe of the two mutations.[144]
- Heterozygotes may exhibit biochemical abnormalities but typically have no clinically significant endocrinopathy.[160, 161]

Females with classical 21-hydroxylase deficiency (both salt-wasting and simple virilizing forms) present at birth with ambiguous genitalia (adrenogenital syndrome).[162–164] Boys with salt-wasting CAH typically present as neonates or during early infancy with symptoms of adrenal insufficiency (failure to thrive, dehydration, hyponatremia, hyperkalemia), and those with simple virilizing CAH not identified by neonatal screening generally present as young children with early virilization. ***Females with the nonclassical "late-onset" form***

of 21-hydroxlyase deficiency have normal external genitalia and present later, during childhood or early adolescence with precocious puberty, or as young adults with other signs of hyperandrogenism such as hirsutism.

As discussed earlier in this chapter in reference to the sexual differentiation of the CNS, females with classical CAH tend to exhibit greater interest in male-typical toys and play and more cross-gender and aggressive behavior than unaffected healthy women.[4, 100–105] Studies of cognitive function in women with classical CAH have yielded inconsistent results. Whereas some have suggested that such women exhibit lower[165, 166] or higher intelligence[167] and differences in verbal learning and memory,[168, 169] compared to unaffected women, others have found no evidence to indicate that prenatal androgen exposure has a consistent or predictable effect on cognition in women with CAH.[170]

Fertility in women with classical CAH is lower than in normal women,[3, 105] primarily due to chronic anovulation relating to excess production of adrenal androgens and progestogens (progesterone, 17OHP) and disordered patterns of gonadotropin secretion;[145] abnormalities of genital anatomy and psychological factors, such as delayed psychosexual development and decreased sexual activity, also contribute.[171] In one study of quality of life in women with classical CAH, half reported that their disease adversely affected their sexual life and most were less than satisfied with their genital anatomy and function, regardless whether they had received reconstructive surgery; vaginal stenosis or narrowing were commonly observed.[172] Women with classical CAH also had a later sexual debut and fewer pregnancies and children. Fertility rates correlate with the severity of the disorder and are significantly lower in women with salt-wasting than in those with the simple virilizing form of classical CAH.[173] However, outcomes of pregnancies among women with classical CAH who conceive are normal except for an increased incidence of gestational diabetes.[171] Children born to mothers with classical CAH have normal birthweight, no increased incidence of malformations, and exhibit normal intellectual and social development.[171, 174] ***Although maternal serum androgen concentrations can increase significantly during pregnancy and should be monitored, the high capacity of placental aromatase activity effectively protects the female fetus from the masculinizing effects of maternal hyperandrogenism.[174]***

Diagnosis of 21-hydroxylase deficiency is based on a high serum concentration of 17OHP, the primary substrate for the enzyme. ***In neonates with either salt-wasting or simple virilizing CAH, 17OHP levels typically are greater than 3,500 ng/dL;[123, 175] levels in normal newborns generally are below 100 ng/dL.[141]*** To distinguish 21-hydroxylase deficiency from other causes of CAH (11β-hydroxylase and 3βHSD deficiencies), serum concentrations of 11-deoxycortisol and 17α-hydroxypregnenolone also should be measured. When the diagnosis is suspected but uncertain, it can be confirmed by performing an ACTH stimulation test, obtaining blood samples before and 60 minutes after administering cosyntropin (synthetic ACTH 1–24; 1 μg/m^2 or 0.25 mg);[176] in affected infants, stimulated 17OHP levels typically exceed 10,000 ng/dL.[162] Diagnosis also can be confirmed by genotyping, which can detect approximately 95% of mutations.[177]

In couples known to be at risk for having an affected child, (affected sibling, both partners carriers for a classic mutation), prenatal diagnosis is possible by genotyping amniocytes or, preferably, cells obtained by chorionic villus sampling (CVS).[133, 146] Early prenatal diagnosis offers the option for intervention, before the most critical period of fetal genital differentiation, in efforts to avoid severe masculinization of the external genitalia in affected female fetuses.

Neonatal screening programs measure 17OHP in blood samples dried on filter paper, comparing results to established reference values that vary with weight and gestational age.[178, 179] Antenatal corticosteroid treatment can decrease 17OHP levels and increase the risk for

a false negative result, particularly when administered repeatedly[180]; screening can be repeated at 1–2 weeks of age, with careful monitoring in the interim, or genotyping can be performed on the dried blood sample.[181]

In the late-onset nonclassical form of 21-hydroxylase deficiency, serum 17OHP concentrations often are only slightly elevated, especially late in the day, and the serum dehydroepiandrosterone sulfate (DHEAS) concentration usually is normal. In children, morning values greater than 82 ng/dL suggest the diagnosis, which can be confirmed by performing an ACTH stimulation test. ***In adult women, morning values less than 200 ng/dL (obtained during the early follicular phase of the cycle) exclude the diagnosis, levels over 800 ng/dL are virtually diagnostic, and intermediate results require additional evaluation with an ACTH stimulation test; in most patients with nonclassical 21-hydroxylase deficiency, the stimulated 17OHP level will exceed 1,500 ng/dL.***[133, 175, 182] A 21-hydroxylase deficiency can be distinguished from 11β-hydroxylase and 3βHSD deficiencies by also measuring 11-deoxycortisol and 17α-hydroxypregnenolone, but the distinction in patients with late-onset CAH has little or no clinical relevance and generally is unnecessary.

11β-Hydroxylase (P450c11) Deficiency

The enzyme 11β-hydroxylase (also designated P450c11 and CYP11B1) mediates the conversion of 11-deoxycortisol to cortisol and of 11-deoxycorticosterone to corticosterone (an intermediate steroid in aldosterone synthesis). The clinical features of 11β-hydroxylase deficiency result from the excess production of adrenal androgens and the mineralocorticoid action of 11-deoxycorticosterone; 11-deoxycortisol has no significant biological activity.

Although 11β-hydroxylase deficiency is the second most common cause of CAH, it accounts for only about 5–8% of adrenal steroid enzyme defects.[162, 163, 183] Like 21-hydroxylase deficiency, 11β-hydroxylase deficiency has severe salt-wasting and simple virilizing forms, and a milder late-onset form. In females, 11β-hydroxylase deficiency can result in virilization of the external genitalia, but also may present later, in children with sexual precocity or in adolescent or young women with hirsutism and menstrual irregularity.[184–186] In most affected individuals, the disorder has unique clinical features that help to distinguish it from 21-hydroxylase deficiency. ***Whereas both 21-hydroxylase deficiency and 11β-hydroxylase deficiency may result in salt-wasting, approximately two-thirds of patients with 11β-hydroxylase deficiency exhibit hypertension due to an increased production of mineralocorticoids.***[183, 187–189] Hypokalemia also may be observed and plasma rennin activity often is low. These effects generally have been attributed to excess production of 11-deoxycorticosterone, which has significant mineralocorticoid activity, although blood pressure and serum 11-deoxycorticosterone concentrations do not correlate closely.[184, 190] The explanation for the wide variation in the clinical manifestations of 11β-hydroxylase deficiency is not clear.

The overall incidence of 11β-hydroxylase deficiency is approximately 1 in 100,000 live births, but like 21-hyroxylase deficiency, incidence varies with ethnicity. In Israel, the incidence of 11β-hydroxylase deficiency is as high as 1 in 5,000 births among Jews of Moroccan ancestry.[191] The enzyme deficiency is an autosomal recessive disorder caused by mutations in the *CYP11B1* gene, which is located on the long arm of chromosome 8 (8q21–q22). The known mutations include missense mutations that result in production of an inactive enzyme,[159, 192–194] frameshift and nonsense mutations that prevent enzyme synthesis,[195–197] and others resulting from unequal recombination between the *CYP11B1* and *CYP11B2* genes.[198, 199] The *CYP11B2* gene is located in the same region on chromosome 8 and encodes an enzyme having both 11β-hydroxylase and 18-hydroxylase

(also designated P450c18 or P450aldo) activity, mediating the conversion of corticosterone to 18-hydroxycorticosterone and, subsequently, aldosterone. There are no specific correlations between genotype and phenotype in patients with 11β-hydroxylase deficiency.[200] Although the late-onset form of 11β-hydroxylase deficiency may be caused by mutations yielding an enzyme with reduced but still significant activity, none has yet been identified.

Diagnosis of 11β-hydroxylase deficiency is based on demonstrating high serum concentrations of 11-deoxycortisol and 11-deoxycorticosterone, as well as testosterone; both basal and ACTH-stimulated levels generally are elevated in affected neonates.[183, 201, 202] In adolescents and young adults, basal 11-deoxycortisol and 11-deoxycorticosterone levels may be normal and ACTH stimulation often is required to make the diagnosis; results must be compared to established age and sex-specific normal values.

3β-Hydroxysteroid Dehydrogenase Deficiency

The enzyme 3β-hydroxysteroid dehydrogenase/Δ^5-Δ^4 isomerase (3β-HSD) catalyzes the oxidation and isomerization of Δ^5-3β-hydroxysteroid precursors into Δ^4-ketosteroids, an essential step in the formation of all classes of steroid hormones (glucocorticoids, mineralocorticoids, progestogens, androgens and estrogens). There are two 3β-HSD isoenzymes, designated type I and type II. The type I 3β-HSD gene (*HSD3B1*) mediates 3β-HSD activity in the placenta and peripheral tissues (skin, breasts, prostate) and the type II 3β-HSD gene (*HSD3B2*) is active in the adrenal, ovary and testis. Deficiency of type II 3β-HSD causes an uncommon form of CAH, accounting for less than 5% of cases.[203] The type I isoenzyme is normal in patients with 3β-HSD deficiency. Consequently, serum concentrations of Δ^4 steroids, such as 17OHP and androstendione, can be normal or even sometimes modestly elevated in affected patients. Serum levels of the substrates for the type I enzyme (pregnenolone, 17α-hydroxypregnenolone, DHEA) are increased due to the defect in the type II enzyme in the adrenals and gonads.

The clinical presentation of patients with 3β-HSD deficiency varies significantly, but can be divided into salt-wasting and non-salt-wasting forms. The salt-wasting form has been associated with nonsense mutations introducing stop codons,[204] frameshift mutations,[204–206] and a variety of point mutations in the *HSD3B2* gene.[207–211] Those with the non-salt-wasting form have had missense mutations causing single amino acid substitutions that dramatically decrease the enzyme's affinity for substrates or cofactors.[211–214]

The external genitalia of females with 3β-HSD deficiency can be mildly virilized, presumably because DHEA levels are high and some is converted to androstenedione and, subsequently, to testosterone in the periphery. Whereas the salt-wasting form of classical 3β-HSD deficiency (analogous to those of 21-hydroxylase and 11β-hydroxylase deficiencies) usually is diagnosed during the first few months of life, the non-salt-wasting form of the disorder generally presents later. In females, because the external genitalia often are normal at birth, diagnosis of the non-salt-wasting form of 3β-HSD deficiency typically is delayed, presenting in childhood with premature puberarche, or in young women with signs of hyperandrogenism.[203]

Although basal levels of Δ^5-3β-hydroxy steroids (pregnenolone, 17α-hydroxypregnenolone, DHEA and DHEAS) generally are elevated in affected individuals, an increased ratio of Δ^5/Δ^4 steroids is a better indication of a possible 3β-HSD deficiency. ***The most reliable diagnostic criterion is the serum 17α-hydroxypregnenolone concentration after ACTH stimulation.*** Proposed threshold values are based on observations in patients with documented mutations (neonates, ≥ 12,600 ng/dL; Tanner stage I children ≥ 5,490 ng/dL, children with premature pubarche, ≥ 9,790 ng/dL; adults ≥ 9,620 ng/dL). Some

have argued that many women with a clinical diagnosis of polycystic ovary syndrome actually may have a late-onset form of 3β-HSD deficiency that may be as or more common than the late-onset form of 21-hydroxylase deficiency.[215] An exaggerated 17α-hydroxypregnenolone response to ACTH stimulation is relatively common in women with hyperandrogenism, but levels rarely approach those observed in women with proven mutations, suggesting that the response likely reflects only adrenal hyperactivity and not an enzyme deficiency.[216] Furthermore, molecular studies have only rarely identified any mutations in *HSD3B2* in patients suspected of having a mild form of 3β-HSD deficiency.[217–219]

Treatment of Congenital Adrenal Hyperplasia

Treatment for classical forms of CAH is aimed at providing sufficient amounts of the deficient hormone, cortisol, to reduce excessive ACTH secretion and to prevent the consequences of excessive androgen production. In mothers at risk for having an affected child, treatment can reduce or prevent masculinization of a female fetus. In neonates with classical CAH, treatment can be life-saving and prevents further virilization. In children, treatment permits normal growth and sexual maturation. In adults with classical or nonclassical CAH, treatment helps in the management of hirsutism, menstrual abnormalities, and infertility.

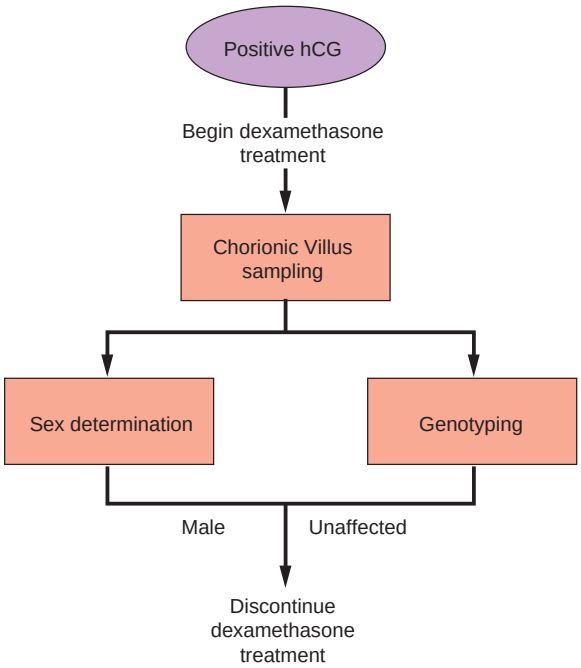
Preimplantation Genetic Diagnosis in Couples at Risk for Having an Affected Child

Polymerase chain reaction (PCR)-based genotyping has greatly improved genetic counseling of families with CAH. In couples at risk for conceiving an affected child, the technology also can be applied in preimplantation genetic diagnosis (PGD) to detect affected embryos resulting from in vitro fertilization (IVF).[220, 221] Typically, a single cell is removed from each embryo reaching the 6–8 cell stage on the third day after oocyte retrieval and fertilization by intracytoplasmic sperm injection (ICSI), with transfer of an unaffected embryo(s) 2 days later at the blastocyst stage. Although PGD requires IVF that otherwise would be unnecessary in fertile couples, some may prefer this option over others based on early prenatal diagnosis, as described below.

Prenatal Treatment of Mothers at Risk for Having an Affected Child

Prenatal maternal treatment with dexamethasone (up to 1.5 mg daily in divided doses) can greatly decrease or prevent fetal female genital virilization.[222] Dexamethasone is not metabolized by the placenta and crosses effectively into the fetal circulation. ***For maximum effectiveness, treatment should begin at 4 to 5 weeks of gestation, and not later than 9 weeks.***[142, 222–225] Prenatal maternal treatment poses some potential risks for the fetus, such as postnatal failure to thrive and psychomotor developmental delay, and also can have significant maternal side effects, including severe abdominal striae, hyperglycemia, hypertension, gastrointestinal symptoms, and emotional lability.[225, 226]

Given that only one in eight fetuses will benefit from maternal treatment (one in four affected, half of which will be males), the best approach involves early prenatal diagnosis by CVS with rapid sex determination (fluorescence in situ hybridization for the X and Y chromosomes, or karyotype) and genotyping, continuing or beginning treatment only in those mothers having an affected female fetus. However, because even short-term prenatal treatment with dexamethasone may adversely affect postnatal physical, cognitive, and emotional development, careful pre-treatment counseling, monitoring, and long-term follow-up is required and best provided in a research setting.[227, 228]



Neonatal Treatment

Newborn infants with classical CAH may be identified by prenatal diagnosis or neonatal screening, or because they have genital ambiguity (females), or an adrenal crisis (males). Infants who exhibit signs of adrenal crisis (hypotension, hyponatremia, hyperkalemia, hypoglycemia, vomiting and diarrhea, weight loss, anorexia) require urgent medical treatment, focusing first on administration of fluids (10–20 mL/kg 0.9% saline) and correction of any significant hypoglycemia (2–4 mg/kg 10% dextrose); hyperkalemia should be corrected by administering glucose and insulin, if necessary. After a blood sample is obtained for measurement of steroid hormones (17OHP primarily), a stress dose of hydrocortisone should be administered (50–100 mg/m^2 intravenously, typically 25 mg), followed by 50–100 mg/m^2 daily in divided doses (every four hours). Additional stress doses of hydrocortisone are administered until the infant is stable and feeding normally. Immediate mineralocorticoid replacement is not necessary but will be required if a diagnosis of salt-wasting CAH is confirmed. Initially, doses of fludrocortisone up to 0.3 mg daily and sodium chloride supplementation (1–3 g daily; 17–51 mEq daily) are required.

In infants having a positive neonatal screening test for CAH, the diagnosis should be confirmed with a second blood sample for measurement of 17OHP and electrolytes. While awaiting the results, electrolytes should be monitored closely if the infant is not treated empirically with glucocorticoids and mineralocorticoids. ***Again, the urgent need is to identify infants with salt-wasting CAH before they develop adrenal crisis, which can occur anytime within the first few days or weeks after birth without treatment.***[141, 222]

Treatment in Children

Ideally, the medical, surgical, and psychological management of children with CAH should be guided by a multidisciplinary team, including pediatric endocrinologists, surgeons, urologists, geneticists, and psychologists.[222]

Children with classical or symptomatic nonclassical 21-hydroxylase deficiency require treatment with glucocorticoids.[141, 222, 229] The goal of treatment is to promote normal growth and development by providing sufficient hormone to minimize adrenal sex steroid production while avoiding the consequences of glucocorticoid excess. Usually, that can be achieved by treatment with hydrocortisone (cortisol) in a dose of 12–18 mg/m^{2}daily,[133, 222, 229] which still exceeds normal daily cortisol secretion in children and adolescents (6–9 mg/m^2/day).[230–232] Whereas long-acting glucocorticoids (e.g., prednisone, dexamethasone) also can be used, their longer duration of action and greater potency also increase the risk of over-treatment, which can adversely affect growth before closure of the epiphyses.[142, 233, 234] Normal growth has been observed in some studies of children treated with prednisone (approximately 1 mg/m^2/day)[235] or dexamethasone (approximately 0.27 mg/m^2/day),[236] but hydrocortisone remains the treatment of choice during childhood.[222, 229]

Mineralocorticoid treatment with fludrocortisone is required for children having classical 21-hydroxylase deficiency, regardless whether they have the salt-wasting or simple virilizing form of the disorder. The goal of treatment is to maintain normal serum sodium and potassium concentrations while avoiding the consequences of over-treatment or under-treatment. Excessive mineralocorticoid treatment can cause hypertension, hypokalemia, and may impair growth.[237] Inadequate treatment can result in poor growth because it increases the glucocorticoid requirement,[237, 238] and may increase adrenal androgen production, because chronic volume depletion causes increased production of renin and angiotensin II, which can stimulate steroidogenesis.[239] In children, fludrocortisone is administered in a dose ranging between 0.05 and 0.2 mg daily.[222] Salt supplementation can be discontinued as the child begins to eat table food, but may be needed during hot weather or strenuous exercise.

The effectiveness of treatment generally should be monitored approximately every 3 months in infants and every 4–12 months in children,[222] by measuring the serum concentrations of 17OHP, androstenedione, plasma rennin activity, growth velocity, and skeletal maturation, comparing results to normative data for age and sexual maturation.[133] Ideally, serum hormone measurements should be obtained in the morning when results will reflect peak concentrations.[141, 222, 229] ***Serum 17OHP levels generally should be maintained in a range between 400 and 1200 ng/dL, but care must be taken to avoid undertreatment of hyperandrogenism and the consequences of iatrogenic hypercortisolism.***[240] Plasma rennin activity should be kept within the normal range for age by adjusting treatment with fludrocortisone and salt supplementation, before adjusting the level of glucocorticoid treatment. When necessary, a brief 7–10 day course of treatment with dexamethasone can effectively suppress high androstenedione levels that may result from poor compliance. Bone age and growth rate should be monitored every 6 months, with the goal of avoiding a decrease in growth and advanced bone age.[241, 242] Patients with classical CAH have an increased risk for developing central precocious puberty due to poor control of adrenal androgen production; in those who do, treatment with a long-acting GnRH agonist may be needed.[243]

Illness can precipitate adrenal crisis in children with classical CAH unless they receive adequate glucocorticoid treatment. Signs and symptoms suggesting the possibility include hypotension, electrolyte imbalance (hyponatremia, hyperkalemia, hypoglycemia), and vomiting and diarrhea that sometimes can be accompanied by abdominal pain, fever, loss of appetite, and weight loss. In children with mild illness, the maintenance dose of glucocorticoid generally should be increased by 2- to 3-fold. When illness is associated with diarrhea or vomiting and reduced oral intake, intravenous glucocorticoids, saline, and glucose may be required. In children with severe illness or who require major surgery, intravenous hydrocortisone should be administered in a dose appropriate for age; for those 12 years of age or older, a one-time dose of 100 mg should be administered, followed by

100 mg/day. During recovery, stress doses of hydrocortisone can be gradually decreased, by approximately 50% per day.[222]

Children with classical CAH are at increased risk for early puberty and short stature because high levels of sex steroids promote premature epiphyseal closure. In treated patients with classical CAH, adult height usually is lower than in reference populations, by an average of approximately 10 cm, independent of the level of control of adrenal androgen concentrations, which suggests that treatment with exogenous glucocorticoids also suppresses growth.[244, 245] The effectiveness of treatment during the first 2 years of life and during puberty appears to have the most important influence on final height.[246–248] Treatment with growth hormone and a long-acting GnRH agonist can help to maximize growth and adult height.[249, 250] Obesity is a common complication of glucocorticoid treatment in children with classical CAH; body mass index correlates with the dose of prescribed medication.[251] In obese children, the incidence of hypertension also is increased.[252]

The surgical management of the genital abnormalities in virilized female children with classical CAH is quite complicated. Traditionally, surgery has been performed in the first few years of life, when the child is still too young to remember the procedure, and to avoid any psychological problems associated with having abnormal external genitalia. ***However, the wisdom and outcomes achieved with early surgery recently have been challenged and many now advocate delaying unnecessary surgery until the child is older and can participate in the decision.***[109] The controversy is discussed in a later section of this chapter devoted to the management of ambiguous genitalia. If clitoroplasty is performed, the clitoral recession procedure, conserving the glans and its innervation, should be employed. It is important to know that women who undergo clitoroplasty and even total clitoral amputation generally do not have an impaired erotic response or decreased capacity for orgasm. When necessary, vaginal reconstruction is best postponed until after puberty when mature compliance is possible. In patients with severe classical CAH, bilateral adrenalectomy offers the potential advantage of preventing adrenal hyperandrogenism, but also increases the risk for developing adrenal crisis.[253–255]

Treatment in Adults

For older adolescent and adult women with classical CAH, the goal of treatment is to lower and maintain serum concentrations of adrenal precursors (17OHP) and androgens to the upper limits for normal women. ***After epiphyseal closure is complete, treatment with long-acting glucocorticoids (e.g., dexamethasone, prednisone) generally is preferred.*** When administered at bedtime in a dose ranging between 0.25 and 0.75 mg, ACTH is effectively suppressed for most or all of the following day. Bedtime treatment effectively inhibits the peak of ACTH secretion, which occurs between 2:00 A.M. and 10:00 A.M.[233] ***To avoid the risks of osteoporosis and developing Cushing's syndrome, dosage must be adjusted to the needs of the individual patient.*** Alternative treatment regimens include prednisone (median dose 7 mg/day; range 4–10 mg/day) or single or divided doses of hydrocortisone (median 30 mg/day; range 15–40 mg/day).[256] Supplemental doses of glucocorticoid, generally involving a 2–3 fold increase in the usual daily dose, are indicated during times of stress such as febrile illness, surgery, and trauma; normal exercise does not require stress doses of glucocorticoids.[229]

As in children with classical CAH, mineralocorticoid treatment in adults is provided with fludrocortisone, in the dose required to maintain normal serum sodium and potassium concentrations and plasma rennin activity, usually ranging between 0.1 and

0.2 mg/day. When mineralocorticoid treatment is optimized, the dose of glucocorticoids can be minimized.[237, 238] Inadequate treatment can result in chronic volume depletion that promotes excess production of renin and angiotensin II, which, in turn, can stimulate increased adrenal androgen synthesis.[239] Patients with the simple virilizing form of classical CAH who exhibit increased plasma renin activity and aldosterone concentrations can benefit from mineralocorticoid treatment, which helps in controlling 17OHP levels.[257, 258]

Treatment should be monitored by periodic measurements of bone density, and serum 17OHP, DHEAS, androstenedione, and testosterone concentrations, remaining alert to the development of signs or symptoms of Cushing's syndrome. In those who require mineralocorticoid treatment, plasma renin activity should be monitored and maintained near the upper limit of normal.

Many women with classical CAH who underwent reconstructive surgery during childhood later require further reconstructive surgery during late adolescence or early adulthood, generally involving clitoroplasty and vaginoplasty. Approximately half of procedures performed during infancy will require later revision.[133, 259]

Psychological counseling, ideally beginning soon after the diagnosis is established, is an important part of the treatment of classical CAH. Although data are limited and conflicting, the incidence of adult psychiatric disorders may be increased in women with classical CAH.[260, 261] Sexual relationships may develop somewhat later than usual and sexual function may not be completely normal, even in those having had reconstructive surgery.[5]

Treatment During Pregnancy

Although normal reproduction is possible with effective treatment, fertility generally is decreased in women with classical CAH, particularly in those with the salt-wasting variety of the disorder, due to chronic anovulation and, in some cases, to poor surgical results.[3] ***In those who do conceive, serum concentrations of androstenedione, testosterone and 17OHP should be carefully monitored and the dosage of glucocorticoids increased as needed to maintain normal levels for gestational age. Treatment with long-acting glucocorticoids should be discontinued in favor of treatment with hydrocortisone, which is metabolized by the placenta and thereby avoids the risk of suppressing the fetal hypothalamic-pituitary-adrenal axis.*** In general, term pregnancies, delivery of healthy female infants with normal external genitalia, and normal growth and development in both girls and boys can be achieved.[174, 262] Even when maternal androgen levels cannot be suppressed to normal, the high capacity of placental aromatase activity effectively protects the fetal female genitalia.[174]

The incidence of cesarean delivery is increased, primarily because of concerns that vaginal delivery may disrupt a previous surgical reconstruction of perineal anatomy. An android pelvis is no more common than usual, because the form and size of the adult pelvis are determined during the pubertal growth spurt. However, a small pelvis might result if bone age is advanced to age 13–14 before treatment started. ***The need for stress doses of glucocorticoids during labor and delivery is obvious and does not increase the risk for infection or poor wound healing.***

Androgen Excess—Fetoplacental Origin

Two rare enzyme deficiencies associated with androgen excess—aromatase deficiency and P450 oxidoreductase deficiency—are distinct from those causing classical forms of CAH because they involve both the fetal adrenal and the placenta.

Aromatase (P450arom) Deficiency

The enzyme aromatase (also designated P450arom and CYP19A1) catalyzes the conversion of 19-carbon androgens (androstenedione, testosterone, 16α-hydroxy DHEA) to aromatic 18-carbon estrogens (estrone, estradiol, and estriol, respectively) and is encoded by the *CYP19A1* gene, located on chromosome 15 (15p21.1). The enzyme is active in the gonads, the placenta, the brain, and in adipose; tissue-specific regulation is controlled, in part, by alternative tissue-specific promoters. Aromatase deficiency is a rare autosomal recessive disorder caused by mutations in the *CYP19A1* gene. As a consequence, fetal androgens are not converted to estrogens in the placenta, resulting in female fetal virilization (due to the accumulation of fetal androgens), low maternal serum estrogen levels, and maternal hirsutism, which typically develops during the second half of pregnancy and regresses after delivery. ***Affected females classically present with ambiguous genitalia at birth and, at puberty, exhibit signs of hyperandrogenism, absent breast development, primary amenorrhea associated with hypergonadotropic hypogonadism, and multicystic ovaries.***[263–266] Aromatase mutations also can produce variable or nonclassic phenotypes characterized by varying degrees of breast development.[267]

P450 Oxidoreductase Deficiency

The classical forms of CAH all are caused by mutations in genes encoding steroidogenic enzymes, resulting in reduced or absent enzyme activity, and in clinical signs and symptoms caused by the accumulation of steroid precursors and/or decreased production of the principal steroid end product. Another newly described form of CAH results from a deficiency in the P450 oxidoreductase (POR) enzyme. Although not a steroidogenic enzyme *per se*, POR nonetheless affects several steroidogenic pathways and now is recognized as a cause of both 46,XX disorders of sexual development (female virilization) and 46,XY disorders of sexual development (incomplete male virilization), which is discussed below.[268]

First described in 2004,[269] POR deficiency is perhaps the most complex form of CAH because it affects the activity of all of the P450 enzymes involved in steroidogenesis, to varying degrees, resulting in varying patterns of abnormal steroid hormone production and a spectrum of clinical manifestations, and has other "non-endocrine" effects on skeletal development and drug metabolism. POR is a flavoprotein associated with the endoplasmic reticulum and is encoded by the *POR* gene, located on chromosome 7 (7q11.2). POR serves as the electron donor in the activation of *all* microsomal P450 enzymes, including P450c21 (the adrenal 21-hydroxylase, CYP21A2), P450c17 (CYP17A1, which catalyzes both 17α-hydroxylase and 17,20-lyase activities), and P450arom (aromatase, CYP19A1, which mediates the conversion of androgens to estrogens). POR deficiency is an autosomal recessive disorder and more than 25 different POR mutations already have been identified, most being missense mutations in the central electron transfer domain of the protein.[268]

Patients with the same mutation, even siblings, can exhibit phenotypic differences, but the hormonal profile of all patients with POR mutations reflects partial deficiencies of 21-hydroxylase and 17α-hydroxylase/17,20-lyase. Because 21-hydroxylase and 17,20-lyase activities are impaired to a greater degree than 17α-hydroxylase activity, basal serum 17OHP concentrations are elevated and exhibit an exaggerated response to ACTH stimulation (due to impaired 21-hydroxylase activity), and levels of DHEA/DHEAS and androstenedione are low (due to impaired 17,20-lyase activity). Basal cortisol levels usually are normal or near normal, but do not rise normally with ACTH stimulation, revealing a chronically compensated adrenal insufficiency.[268]

Hormone Profile Associated with P450 Oxidoreductase Deficiency

Hormone	*Basal Level*	*ACTH-stimulated Response*
17OHP	High	Exaggerated
DHEA/DHEAS	Low	Low
Androstenedione	Low	Low
Cortisol	Normal	Low

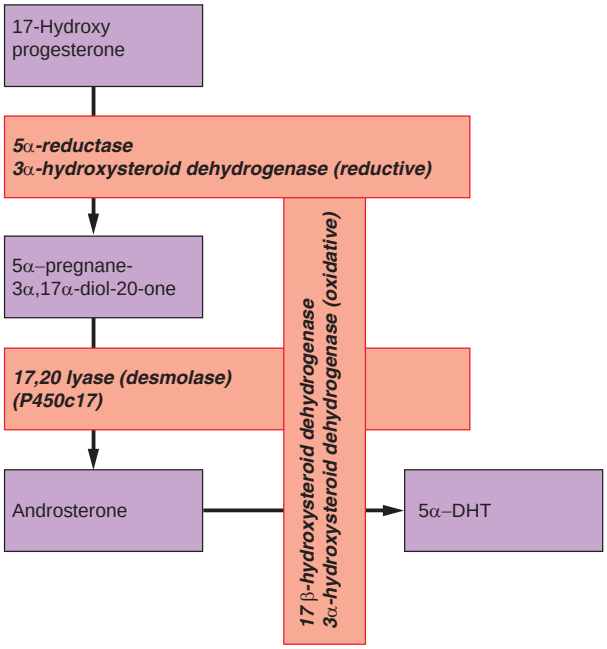
Surprisingly, females with POR deficiency frequently become virilized *in utero*, something not expected, given that fetal adrenal androgen production should be decreased, not increased. There are two hypotheses concerning the source of androgen excess that might explain the apparent dichotomy, but neither has been established conclusively.[270–272] The first envisions that even the modest amount of androgens produced could accumulate due to the deficiency of placental P450arom in patients with POR deficiency. The second invokes an alternative "backdoor pathway" to androgen production in which elevated levels of 17OHP, which cannot be efficiently metabolized via P450c21 or P450c17 activities, seek alternative metabolism via 5α-reduction and ultimately are converted to dihydrotestosterone (DHT), circumventing the usual pathway via androstenedione and testosterone.[271–273] Although the backdoor pathway still involves P450c17, the enzyme's affinity for its substrate in the alternative pathway (5α-pregnane-3α,17α-diol-20one) is much higher than for 17OHP. Therefore, the backdoor pathway likely functions better than the conventional metabolic pathway in patients with POR deficiency.[273]

The phenotype of POR deficiency varies widely. Whereas some exhibit a characteristic spectrum of skeletal abnormalities known as the Antley-Bixler Syndrome (craniosynostosis, mid-face hypoplasia, choanal atresia or stenosis, radio-humeral and/or radio-ulnar synostosis, femoral bowing and fractures, and joint contractures), indistinguishable from that observed in patients with mutations in the fibroblast growth factor receptor-2 gene (*FGFR2*), bony abnormalities are subtle or altogether absent in others.[268] The phenotypic spectrum in patients with proven POR deficiency has included asymptomatic patients identified by neonatal screening for 21-hydroxylase deficiency, asymptomatic patients whose mothers virilized during pregnancy, virilized female infants, and an adult female with primary amenorrhea and muticystic ovaries.[269, 270, 273, 274] The widely varying phenotype has led to speculation that POR deficiency may be relatively common and frequently goes unrecognized or is misdiagnosed.

Diagnosis of POR deficiency is not straightforward. The diagnosis should be considered in the evaluation of children with sexual ambiguity and when prenatal screening for

trisomy 21 reveals low maternal estriol levels. Mutation analysis is indicated for patients who exhibit compatible steroid hormone profiles.

The "Backdoor Pathway" to Androgen Production



Androgen Excess—Maternal Origin (Gestational Hyperandrogenism)

Maternal gestational hyperandrogenism is another, albeit very uncommon, cause of fetal virilization and may result from maternal ingestion of androgens or drugs having androgenic actions, or from excess maternal androgen production. The possibility should be considered when a pregnant woman exhibits a rapid onset of masculinizing signs, including hirsutism, temporal balding, clitoromegaly and deepening of the voice. It also should be considered after delivery of a virilized female infant, remembering that luteomas and theca-lutein cysts regress after delivery.

The possible or probable extent of fetal virilization relates to the time of exposure to maternal androgens. Whereas exposure during early pregnancy can cause labioscrotal fusion and clitoromegaly, exposure after 12 weeks of gestation causes only clitoral hypertrophy.

Drug Ingestion

Most cases of female fetal virilization resulting from maternal drug ingestion have involved treatment with danazol for endometriosis,[275] or with progestins for threatened or recurrent abortion.[276, 277] ***The risk appears limited to progestins that bind to the androgen receptor.***[278] Given their potential risks, progestational agents other than progesterone or 17-hydroxyprogesterone are no longer administered to pregnant women. ***However, virilization of female infants has not been observed in women exposed to oral contraceptives after conception.***[279] Maternal ingestion of androgens also can cause fetal virilization, but often does not, probably because fetal exposure is limited by the high capacity of placental aromatase activity.

Excess Androgen Production

Women who develop gestational hyperandrogenism merit a thorough evaluation.[280] Luteomas and theca-lutein cysts are the most common causes; virilizing ovarian or adrenal tumors are rarely encountered during pregnancy.[281–287] However, all diagnostic possibilities warrant consideration.

Pelvic ultrasonography is helpful for distinguishing between adrenal and ovarian tumors, solid and cystic ovarian masses, and unilateral from bilateral ovarian disease. Approximately one-half of luteomas and almost all theca-lutein cysts are bilateral;[288] other ovarian tumors usually are unilateral. Malignant tumors are most likely to be solid and unilateral. Serum hormone measurements have limited diagnostic value, given that maternal serum testosterone concentrations normally rise progressively during pregnancy, primarily due to the marked increase in SHBG levels. ***Surgery is rarely needed for women with luteomas or theca-lutein cysts because both characteristically regress after delivery.*** However, when a tumor is highly suspected and cannot be excluded, laparoscopy or laparotomy may be required to establish the correct diagnosis.

Pregnancy Luteoma

Pregnancy luteomas are hyperplastic masses of luteinized cells and not true tumors. Their true incidence is unknown; most likely go unrecognized because they produce little androgen or have little or no significant androgenic effect. Luteomas may be discovered incidentally at the time of cesarean delivery or other abdominal surgery during pregnancy or the early puerperium, or when they are large or cause maternal virilization. Typically, luteomas are solid masses ranging between 6 and 10 cm in size; in approximately half of cases, they are bilateral.[289, 290]

In women with luteomas, serum concentrations of androstenedione, testosterone, and dihydrotestosterone are increased, sometimes dramatically.[280, 289, 291] However, only approximately one-third of reported pregnancy luteomas have been associated with maternal hirsutism or virilization,[280, 292] probably because any increase in serum free testosterone is limited by the large increase in sex hormone-binding globulin (SHBG) levels that occurs during pregnancy. ***There is essentially no risk that the female fetus of a woman with a luteoma will become virilized if the mother herself does not.*** However, approximately 80% of female infants born to virilized mothers also are virilized, to an extent that correlates with the severity, duration, and most importantly, the stage of pregnancy at the time of androgen exposure.[293, 294]

The typically prompt regression of luteomas after delivery suggests that hCG may play some role in stimulating or perpetuating androgen production in luteomas.[295] However, most luteomas are identified late in gestation, long after the peak in maternal serum hCG concentrations. Moreover, theca-lutein cysts, rather than luteomas, are more commonly associated with excessively high hCG levels, as observed in women with gestational trophoblastic disease. Consequently, it seems likely that some other mechanism is responsible for the growth and androgen production of luteomas in late gestation.

Theca-Lutein Cysts

Clinically apparent theca-lutein cysts, also known as hyperreactio-luteinalis, develop most frequently in women with multiple pregnancies, isoimmunized mothers, those with molar pregnancies or gestational trophoblastic disease, and women with diabetes mellitus, all of which are associated with increased maternal serum hCG concentrations; the highest incidence (10–20%) is observed in women with trophoblastic disease.[296] However, not all women with such conditions develop theca-lutein cysts, which also may persist long after evacuation of molar pregnancies despite the rapid decrease in serum hCG levels.[296–298] Rarely, mothers with pre-existing hirsutism related to polycystic ovary

syndrome or ovarian stromal hyperthecosis may develop theca-lutein cysts and gestational hyperandrogenism.[299, 300]

Ovaries containing theca-lutein cysts can become quite enlarged, reaching 10–15 cm in diameter. Histologically, the ovarian cortex usually exhibits focal hyalinization. ***Approximately 30% of pregnant women with clinically apparent theca-lutein cysts become hirsute or virilize.***[280, 301–303] In most of those who exhibit virilization, serum concentrations of testosterone and androstenedione are elevated; cord serum testosterone levels also may be elevated in their infants,[301, 302, 304, 305] but no cases of virilized female infants have been reported.

Other Disorders of Genital Development

There is a last group of 46,XX disorders of sexual development that cannot yet be classified by cause, because their causes are unknown. The category includes cloacal extrophy, müllerian agenesis, and the syndrome of Müllerian, renal, and cervicothoracic somite dysplasia known as the MURCS association.

Cloacal Extrophy

Cloacal extrophy is a rare and complex anorectal and genitorurinary malformation in which the rectum, vagina and urinary tract share a common everted orifice, accompanied by an omphaalocele, and an imperforate anus. Typically, the bladder and genitalia are divided into two halves on either side of an exposed segment of bowel; a number of variants have been described.[306] The disorder is believed to result from the failed migration of the lateral mesodermal folds of the infraumbilical anterior abdominal wall, leading to an enlarged cloacal membrane that ruptures prematurely, before descent of the urorectal septum, sometime prior to 8 weeks of gestation.[306, 307]

Müllerian Agenesis (Mayer-Rokitansky-Küster-Hauser Syndrome)

Müllerian agenesis is a disorder of genital development characterized by absence of the vagina, an absent or hypoplastic uterus, and normal or hypoplastic fallopian tubes.[308] The disorder is a relatively common cause of primary amenorrhea and is described in detail in Chapter 11 (Amenorrhea). Typically, the ovaries are entirely normal, although one or both also may be undescended, hypoplastic, or associated with an inguinal hernia. Affected patients often also have urologic anomalies (unilateral renal agenesis, ectopic or horseshoe kidney, and duplication of the collecting systems) and skeletal malformations (e.g., hemiverterbrae and scoliosis, or the Klippel-Feil syndrome, which includes a short neck, low hairline, limited range of motion, and neurologic symptoms, resulting from one or more fused vertebrae).[309, 310] The cause is unknown, although some cases are associated with chromosomal translocations or occur in familial aggregates, suggesting a genetic basis. Logically, müllerian agenesis might be attributed to an activating mutation in the gene encoding AMH or its receptor, causing excess AMH activity, but none have been identified.[311]

Patients with müllerian agenesis typically present in late adolescence or as young adults with primary amenorrhea, exhibiting normal breast and pubic hair development and no visible vagina. A few may have functional islands of endometrium, resulting in obstructed menses and symptoms of cyclic pain.[309, 310] Evaluation should include a karyotype, renal ultrasonography, spinal X-rays, and pelvic ultrasonography or MRI when there is reason to suspect a functional uterine remnant.[312, 313] Surgery generally is indicated only for those with symptoms relating to hematometra, endometriosis, or an inguinal hernia. When the time is appropriate, a functional vagina can be created by progressive vaginal dilation,[314–316]

traditional vaginoplasty,[317] or the modified Vecchietti operation, which is performed laparoscopically.[318, 319] Women with müllerian agenesis are infertile, but can expect normal sexual function and have their own genetic offspring via IVF using oocytes retrieved from their own normal ovaries and their partner's sperm, with subsequent transfer of embryos to a gestational surrogate.[320, 321]

Müllerian Renal Cervicothoracic Somite Dysplasia (The MURCS Association)

The MURCS association is a syndrome characterized by müllerian (MU) aplasia or hypoplasia, unilateral renal (R) agenesis or ectopy and cervicothoracic somite (CS) dysplasia, which results in vertebral defects (e.g., Klippel-Feil anomaly, scoliosis), and abnormalities of the ribs, upper limbs, and scapula.[322] Other associated anomalies have included cleft lip and palate, ovarian agenesis, abnormal pulmonary fissures, tetralogy of Fallot, anorectal malformations, and transmissive deafness.[323–326] The pathophysiology involved is unclear but logically may involve an event occurring very early in development when the blastemas of the pronephric buds and cervicothoracic buds are closely located. The disorder has similarities to the 22q11 deletion syndrome (aortic arch anomalies, facial deformities, nasal voice, mild learning difficulties, renal agenesis, autoimmune disease and cervical spine anomalies) and to the Mayer-Rokitansky-Küster-Hauser syndrome, suggesting a similar pathophysiology.[322]

46,XY Disorders of Sexual Development

Disorders of sexual development occurring in chromosomal males (46,XY) can result from abnormalities in gonadal development, from decreased fetal androgen synthesis relating to deficiencies of steroidogenic enzymes or regulatory proteins, from androgen receptor defects that prevent normal androgen action, from LH receptor defects causing Leydig cell hypoplasia, or from mutations affecting AMH or its receptor.

Disorders of Gonadal (Testicular) Development

Normal gonadal development requires normal germ cells and normal gonadal somatic cells. Disorders of testicular development include complete gonadal dysgenesis (Swyer syndrome), partial gonadal dysgenesis (a variety of single gene disorders and chromosomal abnormalities involving key genes), and the loss of otherwise normally developed testes during fetal life (testicular regression syndrome). In addition, a small proportion of patients with ovotesticular DSD (discussed in an earlier section of this chapter) has a 46,XY karyotype.

Complete Gonadal Dysgenesis (Swyer Syndrome)

Swyer syndrome is an uncommon form of gonadal dysgenesis, characterized by a 46,XY karyotype.[327] ***Despite the presence of a Y chromosome, the phenotype is female because the dysgenetic (streak) gonads produce neither AMH nor androgens.*** Consequently, the vagina, cervix, uterus, and fallopian tubes develop normally and the internal and external genitalia do not masculinize.[328] ***In approximately 10–15% of patients, the disorder results from an inactivating mutation in the SRY gene, but in most, no cause can be identified.***[329] Mutations in other genes involved in the regulation of *SRY* expression or encoding important downstream elements in the testis-determining pathway have been implicated.[330–332]

Patients with Swyer syndrome generally present after the expected time of puberty with delayed sexual maturation, primary amenorrhea, normal pubic hair, and normal female internal and external genital anatomy. Evaluation reveals hypergonadotropic hypogonadism, prompting a karyotype that establishes the diagnosis. ***Gonadectomy is indicated soon after diagnosis due the significant risk for development of germ cell tumors in occult testicular elements (20–30%).***[333]

Sex of assignment and rearing and gender identify is unequivocally female and no specific treatment is required other than estrogen therapy to induce breast development and, subsequently, estrogen and progestin therapy (cyclic or combined) to maintain sexual maturation. Pregnancy can be achieved with IVF using donor oocytes and has not been associated with any specific risks or complications.[334]

Partial Gonadal Dysgenesis

Partial gonadal dysgenesis describes a group of disorders resulting from a wide assortment of genetic mutations causing abnormal gonadal development and function. In affected patients, müllerian structures may be present or absent, the external genitalia may be female, ambiguous, or male, and the phenotype can include developmental abnormalities outside of the reproductive tract. The wide variations in phenotype reflect the many different actions of the gene products, which are involved in the regulation of SRY expression, müllerian regression, testis differentiation, and developmental patterning. Examples include single gene disorders involving *WT1*, *SF1*, *SRY*, *SOX9*, *DHH* (an intercellular signaling molecule having an important role in morphogenesis and testis development), *ATRX* (a transcriptional regulator expressed during development) and *ARX* (a homeobox-containing gene expressed during development), as well as chromosomal aberrations involving key genes such as a *DMRT1* (hemizygosity), *DAX1* (duplication), and *WNT4* (duplication).[109]

Genes Associated with 46,XY Partial Gonadal Dysgenesis[109]

Gene	*Locus*	*Inheritance*	*Gonad*	*Mullerian Structures*	*External Genitalia*	*Associated Features*
Single Gene Disorders						
WT1	11p13	Autosomal dominant	Dysgenetic testis	+/–	Female or ambiguous	Wilms' tumor, renal anomalies, gonadal tumors
SF1	9q33	Autosomal dominant/ recessive	Dysgenetic testis	+/–	Female or ambiguous	Adrenal failure (some)
SRY	Yp11.3	Y	Dysgenetic testis/ ovotestis	+/–	Female or ambiguous	
SOX9	17q24-5	Autosomal dominant	Dysgenetic testis/ ovotestis	+/–	Female or ambiguous	Camptomelic dysplasia
DHH	12q13.1	Autosomal recessive	Dysgenetic testis	+	Female	
ATRX	Xq13.3	X	Dysgenetic testis	–	Female, ambiguous, or male	α-thalassemia, mental retardation
ARX	Xp22.13	X	Dysgenetic testis	–	Ambiguous	Lissencephaly, epilepsy, temperature instability
Chromosomal Aberrations						
DMRT1	9p24.3	Hemizygosity	Dysgenetic testis	+/–	Female or ambiguous	Mental retardation
DAX1	Xp21.3	Duplication Xp21	Dysgenetic testis or ovary	+/–	Female or ambiguous	
WNT4	1p35	Duplication 1p35	Dysgenetic testis	+	Ambiguous	Mental retardation

Testicular Regression Syndrome

Testicular regression syndrome is a condition in which a developmentally normal testis existed during fetal life but subsequently regressed or was lost. ***The disorder can be unilateral or bilateral and is characterized by partial or complete absence of testicular tissue in the presence of normal male external genitalia.***[335] Typically, the vas and associated vessels end blindly with a varying amount of testicular tissue remaining. The natural history of the disorder is poorly understood. Current concepts presume normal early embryonic development and testicular descent, followed by a catastrophic event such as torsion. If the process occurs relatively late in pregnancy, the internal and external genitalia virilize and the müllerian ducts regress normally, but the testes are absent at birth (anorchia). However, earlier loss of both testes can result in a small phallus or incomplete masculinization.[336]

In histopathologic studies of atretic nodular testicular remnants obtained from affected neonates or young boys, only 10% of specimens have contained any identifiable seminiferous tubules, which consisted mainly of Sertoli cells enveloped in fibrous strands with no visible germ cells.[335] Hemosiderin-laden macrophages, generally a late cellular response to tissue damage, are present in approximately two-thirds of cases and, occasionally, a vas and epididymis can be observed.[335, 337, 338]

Disorders of Androgen Synthesis

A number of steroidogenic enzymes and regulatory proteins are involved in androgen synthesis, and a deficiency in any one of them can result in decreased fetal androgen production, and its consequences. All are rare, but each is a recognized cause of 46,XY disorders of sexual development; together they account for less than 5% of cases. Disorders of testosterone synthesis usually impair virilization of the external genitalia to a greater extent than the internal genitalia.

Steroid 5α-Reductase Deficiency

Steroid 5α-reductase (type 2) deficiency is an autosomal recessive disorder characterized by a 46,XY karyotype and severe perineal hypospadias (describing a genital configuration consisting of a phallus midway in size between penis and clitoris, a chordee tethering the phallus to the perineum, a urethral opening usually on the perineum, and an incompletely closed urogenital opening resembling a small and shallow vagina), which results from impaired virilization during embryogenesis due to defective conversion of testosterone to dihydrotestosterone (DHT).[339–341] In the classical presentation, the external genitalia are predominantly female at birth, exhibiting failed fusion of the labioscrotal folds and a urogenital sinus or separate urethral and vaginal openings, with or without clitoromegaly. The internal genitalia are male; the epididymides, vasa deferentia, seminal vesicles, and ejaculatory ducts form, but empty into a shortened, blind vagina. In some, the wolffian duct derivatives end on the perineum, on either side of the urethra. The testes are located in the inguinal canals, the labia majora, or in the scrotum and exhibit impaired spermatogenesis. ***The distinguishing feature of the disorder is that affected individuals virilize, to varying degrees, at the time of puberty.*** Unlike in disorders relating to abnormalities in the androgen receptor, breast development in men with steroid 5α-reductase is like that in normal males. Although a few individuals with 5α-reductase deficiency are sufficiently virilized to be assigned a male gender at birth,[342] most have been reared as females and assumed a male gender and behavior at the time of puberty.[343]

The clinical features of steroid 5α-reductase deficiency again illustrate the mechanisms involved in phenotypic sexual differentiation.[344] The wolffian duct derivatives (the ejaculatory ducts, epididymides, vasa deferentia, and seminal vesicles) form normally, in response to normal fetal testosterone levels, but the genital structures that derive from the urogenital sinus and genital tubercle (the external genitalia, urethra, and prostate) do not virilize normally, because they are dependent on the intracellular conversion of testosterone to DHT. Affected men develop a normal muscle mass, libido, and deepening of the voice, which result from the actions of testosterone, but have less body hair and less temporal hairline recession, and no problems with acne, all of which result primarily from the actions of DHT. The unique importance of DHT during fetal development is demonstrated by the significant genital virilization that occurs after puberty.[345]

There are two types of steroid 5α-reductase, designated types 1 and 2,[346] encoded by 2 separate genes; the gene encoding the type I enzyme (*RD5A1*) is located on chromosome 5 (5p15) and that encoding the type 2 enzyme (*RD5A2*) is located on chromosome 2 (2p23). In those with the disorder recognized as steroid 5α-reductase deficiency, the type 2 enzyme is defective and the type 1 enzyme is normal.[347] ***The resulting impaired conversion of testosterone to DHT prevents normal virilization of the male external genitalia during fetal development.*** Affected individuals typically have very low but measurable serum DHT concentrations, which could reflect limited activity of the abnormal enzyme but more likely results from the actions of the type 1 enzyme.[342] The virilization that occurs at puberty may be driven by serum DHT or by testosterone itself. Transient gynecomastia may develop at puberty but does not persist, because androgen and estrogen production is like that in normal adult men.

A wide variety of *RD5A2* mutations has been described, most being point mutations that yield a low concentration of enzyme, an unstable enzyme having reduced activity, or an enzyme with decreased affinity for testosterone and/or essential cofactors. Approximately 40% of affected individuals are homozygous for the same mutation, the remainder being compound heterozygotes.[347] Nearly half have similarly affected family members, likely reflecting consanguinity and a founder effect.[348] In women, mutations are essentially silent; although body hair may be reduced and menarche delayed, even those with homozygous mutations are phenotypically normal and have normal menstrual function and fertility.[349, 350]

The diagnosis of 5α-reductase deficiency should be suspected in infants with genital ambiguity and in adolescents or young adults having the characteristic phenotype and serum hormone profile (a normal male serum testosterone concentration and an increased testosterone/DHT ratio), which typically exceeds 10 in infants and often exceeds 20 in older children and adults.[341, 351] In infants and prepubertal children, basal testosterone and DHT levels may not be sufficient for diagnosis and are best evaluated by performing an hCG stimulation test, measuring testosterone and DHT before (basal, day 1) and after (stimulated, days 3 and 6) administering exogenous hCG (1,500 IU/m^2 on days 1 and 3).[351–353] Individuals with 5α-reductase deficiency can be distinguished from those having defects of testosterone synthesis by their normal or elevated serum testosterone levels, and from those with incomplete androgen insensitivity by demonstrating normal ratios of 5β- to 5α-reduced glucocorticoid metabolites, as indicators of hepatic steroid metabolism.[354] A definitive diagnosis can be established by analyzing DNA extracted from blood or tissue.[347]

The management of patients with steroid 5α-reductase deficiency is complicated because half or more initially assigned as females undergo a change in gender identity and behavior in later life.[343] The sex of rearing, the age of the subject, and the gender identity influence the choice of management. The decision to raise an individual as female should be made only after thorough psychological evaluation to confirm a female gender identity, but once made, management is relatively straightforward. Gonadectomy should be performed to prevent later virilization and tumor development in cryptorchid testes.[355] Any clitormegaly can be surgically corrected, taking care to maintain the glans clitoris. If required, a functional

vagina can be established by progressive vaginal dilation,[315, 316, 356] or by surgical vaginoplasty.[357] Estrogen treatment to induce and maintain feminine characteristics should be started at the usual time of puberty or immediately after gonadectomy in adults.

The decision to raise an individual as male is no less complex and also will involve surgical and medical treatment. The timing of surgery to correct the hypospadias and cryptorchidism depends on the degree of hypospadias and the size of the phallus. Unfortunately, the extent of virilization at puberty usually is less than desired,[351, 358] prompting efforts to improve results by treatment with exogenous testosterone or DHT. Treatment with testosterone before puberty can help to enlarge the phallus.[351] Treatment with DHT can raise serum DHT concentrations, but must be specially prepared as there is no commercially available preparation.[359, 360] Although sperm counts are quite low in most patients,[361] fertility can be achieved via intrauterine insemination[362] or IVF and ICSI.[363]

The change from a female to a male gender identity can be extremely traumatic psychologically, but some have managed the transition quite successfully.[343] In one such case, the patient effectively conducted a "double-life," functioning in all public respects as a female, while having numerous clandestine heterosexual relationships. Aware of his male sexual identity since puberty, he nonetheless delayed seeking medical assistance for fear that exposure would bring shame and guilt to his religiously devout elderly "old world" mother. Although he had planned to keep his secret until his mother died, he finally sought diagnostic help at age 65, because his mother, then age 93, continued to enjoy good health. The transition from a female to a male gender in an individual with steroid 5α-reductase deficiency was chronicled in the Pulitzer Prize-winning novel, *Middlesex*, authored by Jeffrey Eugenides, which features the heroine Calliope Stephanides who becomes the hero, Cal.[364]

17α-Hydroxylase Deficiency

The *CYP17A1* gene encodes an enzyme having both 17α-hydroxylase and 17,20-lyase activities, which are required for synthesis of cortisol, androgens and estrogens. Deficiency of 17α-hydroxylase is a rare cause of CAH, with little more than 100 cases reported.[365, 366] Human 17α-hydroxylase deficiency syndromes involving a loss of only 17α-hydroxylase or 17,20-lyase have been observed,[367, 368] but in most affected patients, both enzymes are deficient.[369]

The compensatory increase in ACTH stimulation that accompanies decreased cortisol synthesis stimulates increased production of 11-deoxysteroids (via 21-hydroxylase), including corticosterone and the mineralocorticoids 11-deoxycorticosterone and 18-hydroxy-deoxycorticosterone.[370, 371] In turn, mineralocorticoid excess leads to volume expansion, which inhibits renin release and the synthesis of aldosterone.[372] Production of androgens (dependent on 17,20-lyase activity) and, subsequently, estrogens is decreased in both the adrenals and the gonads. Serum concentrations of progesterone are increased, but those of 17OHP, cortisol, DHEA, DHEA-S, androstenedione, testosterone, and estradiol are low.

Like other forms of CAH, 17α-hydroxylase deficiency is an autosomal recessive disorder. The *CYP17A1* gene is located on chromosome 10 (10q24.3) and numerous different mutations have been described,[373] including small insertions that disrupt the normal reading frame of the gene (resulting in early termination),[374] deletions of a single or several codons,[375, 376] large deletions with insertion of foreign DNA,[377] and nonsense or missense mutations that yield stop codons or an enzyme with decreased activity.[376, 378–385] ***Females with 17α-hydroxylase deficiency typically present with delayed puberty, primary amenorrhea, and hypergonadotropic hypogonadism; most are hypertensive (due to hypernatremia and hypervolemia) and some also have hypokalemia.***[386, 387] ***Affected males usually have female external gentalia (male pseudohermaphroditism), a blind vagina,***

and intra-abdominal testes; most have been raised as girls, with the underlying disorder recognized only later during evaluation for delayed puberty.[388]

The treatment of 17α-hydroxylase deficiency involves giving sufficient glucocorticoids to suppress excess production of ACTH and mineralocorticoids, while avoiding glucocorticoid excess. Because almost all affected patients are raised as females, estrogen therapy also should be provided, at the time of diagnosis at puberty or the expected time of puberty. In genetic females having a uterus, progestational treatment also must be provided.

3β-Hydroxysteroid Dehydrogenase Deficiency

Defects in the enzyme 3β-hydroxysteroid dehydrogenase/Δ^5-Δ^4 isomerase (3β-HSD) and its endocrine and developmental consequences in genetic females are discussed in detail in the earlier section of this chapter devoted to causes of female virilization (46,XX disorders of sexual development). The enzyme defect is considered here again, briefly, because it can cause incomplete masculinization of males as well as virilization in females.

Type II 3β-HSD catalyzes the oxidation and isomerization of Δ^5-3β-hydroxysteroid precursors into Δ^4-ketosteroids in the adrenals and gonads. A 3β-HSD deficiency results in the accumulation of excessive amounts of Δ^5-3β-hydroxysteroids, including pregnenolone, 17α-hydroxypregnenolone, DHEA, and DHEA-S, and in low levels of Δ^4-ketosteroids such as androstenedione and testosterone and, subsequently, dihydrotestosterone (DHT). Consequently, affected males exhibit varying degrees of incomplete masculinization, ranging from hypospadias to nearly normal female external genitalia.[213, 389, 390]

17β-Hydroxysteroid Dehydrogenase Deficiency

The 17β-hydroxysteroid dehydrogenase (17β-HSD) family of enzymes includes the type 3 isoenzyme, which catalyzes the conversion of androstenedione into the biologically active androgen, testosterone, in the Leydig cells of the testis. Mutations in the *HSD17B3* gene, located on chromosome 9 (9q22), can result in 17β-HSD deficiency, an autosomal recessive disorder caused by impaired testicular testosterone production.[391, 392] Although rare, 17β-HSD deficiency is the most common hereditary defect in testosterone synthesis.

Males with homozygous or compound heterozygous mutations have testes and normally developed internal genitalia but have severely under-virilized external genitalia, which typically appear female and include a short, blind vagina, much like patients with incomplete androgen insensitivity.[392, 393] Consequently, most are assigned a female gender at birth and are raised as females. ***Alternatively, they can exhibit genital ambiguity, with varying degrees of clitoromegaly and labial fusion, or have male genitalia with micropenis or hypospadias.***[391, 394] The testes may be located in the abdomen, in the inguinal canals, or in the labia majora. ***Virilization occurs at puberty, probably due to extra-testicular conversion of androstenedione to testosterone by unaffected 17β-HSD isoenzymes in peripheral tissues (e.g., liver, skin, adipose).***[392, 395–397] The phallus enlarges, muscle mass increases, a male body habitus and hair pattern develops, and the voice may deepen. Gender role reversal has been observed in one- to two-thirds of those affected and raised as girls.[398] Ideally, therefore, the diagnosis is best made before puberty and followed with gonadectomy and estrogen therapy in those with female genitalia. In those born with ambiguous genitalia, early diagnosis may allow male sex assignment because androgen treatment can promote development of a nearly normal adult male phenotype.[399, 400]

When suspected, an elevated basal serum androstenedione level and low serum testosterone/androstenedione ratio (<0.8–0.9) after exogenous hCG stimulation suggests

the diagnosis of 17β-HSD deficiency but the method lacks specificity, because normal values have not been firmly established in age-matched controls and because similar results can be observed in individuals with other defects in testosterone biosynthesis or Leydig cell hypoplasia. Genotyping permits definitive diagnosis. A variety of different mutations has been described, most yielding an enzyme having little or no significant activity.[392]

P450 Oxidoreductase Deficiency

The genetics and pathophysiology of P450 oxidoreductase (POR) deficiency are discussed in detail in the earlier section of this chapter devoted to causes of 46,XX disorders of sexual development (female virilization). The disorder is included again here because, like 3β-HSD deficiency, P450 oxidoreductase deficiency also is among the causes of 46,XY disorders of sexual development (incomplete masculinization in males).

As described earlier, POR is not a steroidogenic enzyme but a flavoprotein that serves as the electron donor in the activation of all microsomal P450 enzymes, including P450c21 (the adrenal 21-hydroxylase, CYP21A2), P450c17 (CYP17A1, which catalyzes both 17,20-lyase and 17α-hydroxylase activities), and P450arom (aromatase, CYP19A1, which mediates the conversion of androgens to estrogens).[268] The developmental consequences of POR deficiency in males result primarily from partial deficiencies of 21-hydroxylase and 17,20-lyase, and to a lesser extent, 17α-hydroxylase activity. Serum concentrations of 17OHP are elevated and androgen levels are low. Not surprisingly, affected boys often are undervirilized, because decreased 17,20-lyase activity prevents generation of 19-carbon androgens, including testosterone.

Steroid Acute Regulatory (StAR) Protein Deficiency

The rarest and most severe form of CAH is known as congenital lipoid adrenal hyperplasia. The disorder is characterized by a deficiency of *all* adrenal and gonadal steroid hormones, increased ACTH secretion, and marked adrenal hyperplasia associated with progressive accumulation of cholesterol esters.

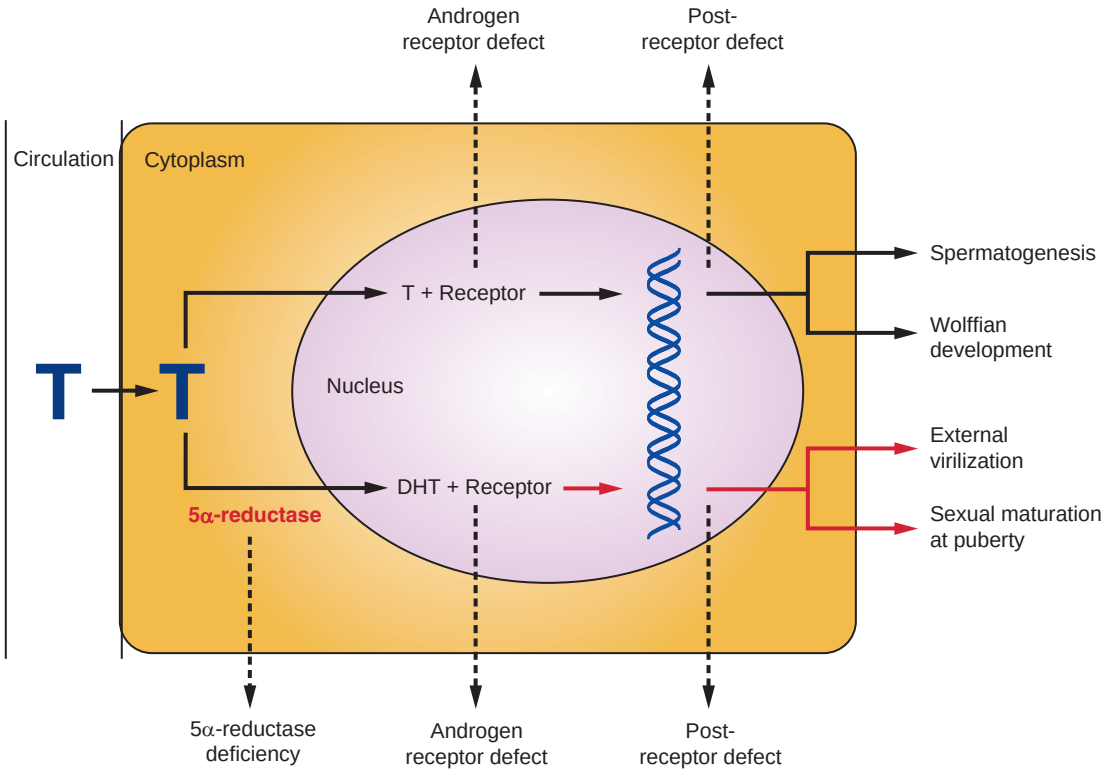
Congenital lipoid adrenal hyperplasia is an autosomal recessive disorder that results from mutations in the gene encoding the steroidogenic acute regulatory (StAR) protein.[401, 402] StAR mediates the acute response to steroidogenic stimuli by facilitating transport of cholesterol from the outer to the inner mitochondrial membrane, the rate-limiting step in steroidogenesis,[402, 403] and is expressed in the adrenal cortex and gonads, but not in the placenta. A wide variety of mutations in the *StAR* gene (located at 8p11.2) have been described, most resulting in reduced activity of the protein.[401, 404–406] The disorder also can result from a heterozygous mutation in the *CYP11A1* gene encoding the cholesterol side chain cleavage enzyme, which converts cholesterol to pregnenolone.[407] In either case, all steroidogenic pathways are affected, resulting in a global deficiency of steroid hormones that reflects both the intrinsic defect in steroidogenesis and the progressive cellular damage that results from the accumulation of cholesterol.[404] Serum cortisol and aldosterone concentrations are very low, ACTH and plasma renin activity levels are very high, sex steroid concentrations are low and serum gonadotropin levels are elevated, even in young children.[407, 408]

Patients with congenital lipoid adrenal hyperplasia typically present very soon after birth or in early infancy with symptoms of severe adrenal insufficiency (vomiting, diarrhea, volume depletion, hyponatremia, hyperkalemia).[409] Male infants usually have female external genitalia due to the severe androgen deficiency. Females are normally developed at birth and even may undergo spontaneous puberty,[410] possibly because the prepubertal ovary, unlike the adrenals and testes, is relatively dormant and thus may escape cellular

damage from cholesterol accumulation. Although two-thirds of reported patients died in infancy,[408] some treated with glucocorticoids and mineralocorticoids have survived to reach puberty.[411]

Disorders of Androgen Action

Mutations in the gene that encodes the androgen receptor (*AR*) can produce a variety of phenotypes in males having normal testes and testosterone production; more than 400 different *AR* mutations have been identified.[412] Collectively, these 46,XY disorders of sexual development are known as androgen insensitivity syndromes. The genetics, pathophysiology and endocrinology of androgen receptor disorders are quite similar. The phenotype depends on whether androgen receptors are absent entirely,[413, 414] present but functionally abnormal,[415–417] or normal but decreased in quantity.[415, 418]



Complete Androgen Insensitivity Syndrome

Complete androgen insensitivity was first described in detail by Morris, at Yale, who coined the name "testicular feminization."[419] However, complete androgen insensitivity syndrome (AIS) now is the preferred term.[420] Complete AIS can result from a wide variety of inactivating mutations in the *AR* gene, including major gene deletions, premature stop codons, splicing abnormalities, and missense mutations that result in amino acid substitutions in the androgen receptor. The *AR* gene is located on the X chromosome (Xq12) and complete AIS therefore follows an X-linked recessive pattern of inheritance. One in three phenotypic sisters of an affected individual and one in six female offspring of a normal sister will

have an XY karyotype. Consequently, careful investigation to identify other affected family members is warranted. Approximately 40% of patients with complete AIS have no family history of the disorder,[421] presumably representing *de novo* mutations.

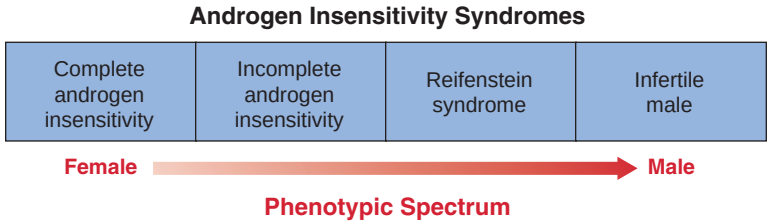
The pathophysiology, and the phenotype, of complete AIS are logical and predictable. The receptor defect results in insensitivity to androgen. Consequently, androgen-induced wolffian duct development cannot proceed normally; the presence or absence of wolffian duct derivatives (epididymides, vasa deferentia) varies with the type of mutation. Whereas remnants can be observed (adjacent to the testes) in those having point mutations in the ligand-binding domain of otherwise normally expressed receptors, which may permit a very limited response to high local androgen concentrations *in utero*, wolffian structures are completely absent in those having mutations that yield premature stop codons or frameshift mutations that prevent androgen receptor expression.[422] The normal testes produce normal amounts of AMH, which effectively suppress müllerian duct development. Therefore, the uterus and fallopian tubes typically are absent, although vestiges also have been observed.[423] The testes may be found in the abdomen but, more commonly, are located in the inguinal canals or in the labia majora, probably because AMH normally mediates their descent; their histology resembles that of undescended testes, with a normal or increased number of Leydig cells and absent spermatogenesis. The external genitalia are clearly female, due to androgen insensitivity, although the labia and clitoris may be slightly underdeveloped. The vagina is either blind and short or altogether absent, reflecting only the developmental contribution of the urogenital sinus. Axillary and pubic hair is scant or absent, again due to androgen insensitivity. Breast development is female and may be enhanced, probably due to the action of estrogen unopposed by the actions of androgens. The overall body habitus also is female, although the average height and weight of women with complete AIS is greater than that of normal women.[424, 425] Women with complete AIS also exhibit normal female sexual orientation and maternal instincts.[426, 427]

In women with complete AIS, serum testosterone concentrations are normal or moderately increased, LH levels are increased, and the serum FSH usually is in the normal range. The increase in LH levels results from resistance to the negative feedback effects of androgens at the hypothalamic-pituitary level; both the frequency and amplitude of pulsatile LH secretion are increased in patients with complete AIS.[428, 429] Whereas estrogen production (estrone and estradiol) also is increased by approximately 70% over that in normal men, dihydrotestosterone (DHT) production is decreased due to the absence of male urogenital tissues, which are the primary site of DHT production.[354]

The diagnosis of complete AIS should be suspected in girls with inguinal hernias or labial masses,[430] and in women with primary amenorrhea. In the adolescent or adult, diagnosis usually is not difficult. ***Most patients with complete AIS present with primary amenorrhea, normal breast development, absent or scant pubic and axillary hair, a short vagina, and an absent cervix and uterus; a serum testosterone in the normal male range and a 46,XY karyotype establish the diagnosis.*** Whereas males with defects in testosterone biosynthesis can have a female phenotype, breast development does not occur. ***Patients with complete AIS generally are easily differentiated from those with müllerian agenesis who have normal amounts of pubic and axillary hair, normal female serum testosterone concentrations, and a 46,XX karyotype.*** The location of the testes usually can be defined by ultrasonography or MRI.[431]

The clinical management of complete AIS includes appropriate hormone therapy, creation of a functional vagina, gonadectomy to prevent tumorigenesis in cryptorchid testes, and psychological support, all discussed in detail in Chapter 11 (Amenorrhea) and summarized here. Estrogen treatment is indicated when gonadectomy is performed after puberty is completed, or at the time of expected puberty if the gonads were removed before puberty. Options for creation of a functional vagina include progressive vaginal dilation and vaginoplasty.

The short but distinct vagina observed in most patients with complete AIS facilitates efforts with vaginal dilation,[315, 316, 356] but surgical treatment also produces good results, when necessary.[357] ***In patients with complete AIS, gonadectomy generally is best delayed until after puberty is completed (approximately age 16–18) because pubertal development generally proceeds more smoothly in response to endogenous hormone production, and because the overall risk for tumor development is quite low (5–10%), particularly before puberty.***[418, 432–434] Psychological support should be directed towards reinforcement of their female gender identity and include truthful education for both patient and parents.



Incomplete Androgen Insensitivity Syndromes

Incomplete AIS describes a variety of disorders that result from defects in androgen action less severe than those associated with complete AIS. ***The spectrum of clinical presentations can vary from phenotypic females with mild virilization to under-virilized males who may be fertile or infertile, even within one affected family.***[435, 436]

Phenotypic women with mild virilization are at one end of the clinical spectrum of incomplete or partial AIS. They resemble women with complete AIS but have normal body hair, external genitalia exhibiting partial fusion of the labioscrotal folds, with or without clitoromegaly, and they both virilize and feminize at puberty. They have no müllerian structures (due to the actions of AMH), underdeveloped male internal genitalia (epididymides, vasa deferentia, seminal vesicles, ejaculatory ducts), and testes similar to those with complete AIS. Axillary and pubic hair is normal. Breast development, the overall body habitus, and gender identity are distinctly female. The phenotype is approximately one-tenth as common as that of complete AIS.[418]

Reifenstein syndrome describes individuals having a predominantly male phenotype who are under-virilized.[437] ***The most common clinical presentation is an infertile man with a bifid scrotum and perineoscrotal hypospadias.*** However, the appearance of the external genitalia can vary widely, from a microphallus with a normal penile urethra to complete failure of scrotal fusion. The internal genitalia are male but not completely developed; müllerian structures are absent and, usually, so too is the prostate. The testes can be cryptorchid or normally descended and small, and exhibit a maturation arrest in spermatogenesis. Men with Reifenstein syndrome have normal axillary and pubic hair but, typically, little or no chest or facial hair. They have a male body habitus but usually develop gynecomastia at the time of puberty. Gender identity corresponds with the sex of rearing and sexual dysfunction is common in those raised as males.[438–440]

Some men with partial androgen insensitivity are only mildly under-masculinized and infertile.[441] Their internal and external genitalia are normal and their testes are normally descended, but exhibit either an absent germinal epithelium or spermatogenic arrest. They have normal amounts of body hair and some have gynecomastia. ***The prevalence of partial androgen insensitivity in men with azoospermia or severe oligospermia is unknown but may be as high as 10%.***[442, 443] Still other men with partial androgen insensitivity are under-virilized but fertile.[444, 445]

Androgen Insensitivity Syndromes				
	Complete	*Incomplete*	*Reifenstein*	*Infertile*
Inheritance	X-linked recessive	X-linked recessive	X-linked recessive	X-linked recessive
Spermatogenesis	Absent	Absent	Absent	Decreased
Müllerian	Absent	Absent	Absent	Absent
Wolffian	Absent	Underdeveloped	Male	Male
External	Female	Female (clitoromegaly)	Male (hypospadias)	Male
Breasts	Female	Female	Gynecomastia	Male (gynecomastia)

Serum hormone levels in phenotypic women with incomplete AIS, in men with Reifenstein syndrome, and in infertile men with partial androgen insensitivity are similar to those in individuals with complete AIS.[437, 441, 446] In under-virilized fertile men with partial androgen insensitivity, testosterone concentrations are elevated but LH levels are in the normal range.[444]

Whereas major *AR* gene deletions and premature termination codons have been identified only in patients with complete AIS, point mutations resulting in amino acid substitutions in the androgen receptor can cause the entire spectrum of phenotypes associated with androgen insensitivity. Approximately 80% of amino acid substitutions are located in the hormone-binding domain of the receptor. Most of the remainder is in the DNA-binding domain, permitting normal androgen binding but preventing activation of androgen-responsive genes.[447] How or why different mutations or even the same mutation can result in varying degrees of androgen insensitivity is poorly understood, but may relate to differences in the timing of receptor expression, differences in testosterone synthesis or metabolism, differences in transcription factors, or to polymorphisms that influence the effect of a given mutation.[447–450] In some, androgen resistance results not from a mutation in the androgen receptor but from a defective coactivator protein required for normal function of the androgen-androgen receptor complex.[451, 452] In others with somatic mosaicism, the androgen receptor is normal in some, but not all, tissues.[453, 454]

Incomplete AIS often presents in newborns, as ambiguous genitalia. In adults, the diagnosis generally is not difficult. Affected individuals present as phenotypic women with mild virilization of the external genitalia (otherwise appearing the same as women with complete AIS), or as phenotypic men with gynecomastia and perineoscrotal hypospadias. In phenotypic women with incomplete AIS, the family history can be very helpful because others who are affected have a similar appearance, but in under-virilized men, the phenotype of affected family members can vary significantly. At all ages, the differential diagnosis includes steroid 5α-reductase deficiency, defects in testosterone biosynthesis, and mixed gonadal dysgenesis, which is discussed below. The ratio of testosterone to dihydrotestosterone (DHT) in serum helps to differentiate incomplete AIS from steroid 5α-reductase deficiency; the ratio is normal in most with incomplete AIS, but increased in patients with steroid 5α-reductase deficiency. In most cases, androgen receptor defects can be differentiated from defects of testosterone biosynthesis by the serum testosterone concentration, which typically is normal or high in the former and decreased in the latter. Individuals with mixed gonadal dysgenesis (characterized by a unilateral testis, a contralateral streak gonad, and a 46,XY or 45X/46,XY karyotype) often have a single descended gonad or exhibit some of the phenotypic features of Turner syndrome. Androgen insensitivity also should be suspected in under-virilized men with azoospermia or severe oligospermia. Testosterone or LH concentrations may be elevated but are normal in most;[442] some have high serum FSH levels and resemble men with microdeletions in the AZF (azoospermia factor) region of the Y chromosome.[455]

Although only a few families with the under-virilized fertile male syndrome have been reported, the disorder might be suspected in under-virilized men with a normally formed male urethra and gynecomastia, particularly in those having other family members who are similarly affected.

Androgen receptor binding can be evaluated *in vitro*, using cultured fibroblasts derived from genital skin. However, the method is labor-intensive and costly and cannot exclude abnormalities of androgen receptor function unrelated to binding. Techniques involving the insertion of an androgen-responsive reporter gene in fibroblasts that can demonstrate impaired androgen receptor function have been described, but their usefulness for diagnosis of mild androgen insensitivities is unclear.[456] ***The most reliable method for diagnosis of androgen insensitivity is to sequence the AR gene using DNA derived from blood or tissue, with reference to a database that lists all of the mutations that have been identified in patients with androgen insensitivity.*** Once characterized, the defect can be identified in fetuses at risk using DNA obtained via CVS or amniocentesis.[457–459]

The clinical management of incomplete AIS syndromes is directed toward appropriate gender assignment in infants with ambiguous genitalia, hormone therapy, psychological support, gonadectomy to prevent tumorigenesis in cryptorchid testes, reconstructive surgery where it is needed, and the treatment of gynecomastia in men.

In determining gender assignment, the size of the phallus and the feasibility of constructing a penile urethra are the most important considerations. For phenotypic females and those who are raised as females, estrogen treatment is indicated, when gonadectomy is performed after puberty, or at the expected time of puberty if gonadectomy was performed earlier. ***Whereas adult women can receive normal doses of estrogen therapy immediately, the age and the dose of treatment in children must consider the growth percentile, growth velocity, bone age, target height, and predicted adult height.*** In boys with Reifenstein syndrome, treatment with high doses of testosterone or DHT can achieve greater phallic growth;[460–462] in adult men, high-dose testosterone treatment can,[358, 463] but does not always,[429] improve masculinization. As in women with complete AIS, psychological support should be aimed at truthful education after consultation with the family and the development of a support network; early disclosure may help to limit emotional trauma.[464] Gonadectomy is performed to eliminate the risk of tumors developing in cryptorchid testes (1–2% of undescended testes, more often in abdominal than in inguinal testes), some of which are malignant.[465] ***Whereas gonadectomy generally is best postponed until after puberty is completed in women with complete AIS, earlier surgery is indicated to prevent the virilization at puberty in those with incomplete AIS.*** In boys with Reifenstein syndrome, early surgery to correct cryptorchidism both decreases the risk of tumor and helps to maximize testicular function. The gynecomastia observed in men with Reifenstein syndrome and in under-virilized fertile males, which results from both increased estrogen production and androgen resistance, can be treated by mastectomy when it is disfiguring or otherwise disturbing to the individual. The incidence of breast cancer may be increased in men with Reifenstein syndrome.[466]

LH Receptor Defects

Leydig cell hypoplasia, describing the absence of mature Leydig cells in the testes, is a rare autosomal recessive 46,XY disorder of sexual development caused by inactivating mutations in the LH/hCG receptor.[467, 468] Testicular Leydig cell testosterone production is stimulated by hCG during fetal life and by LH after birth. In the fetus, the number and differentiation of Leydig cells and levels of androgen production parallel the changes in serum hCG concentrations during pregnancy. Consequently, a decrease in the number or

function of Leydig cells results in decreased fetal testosterone production and in the failure of normal male sexual differentiation. Large deletions or nonsense mutations in the LH/hCG receptor gene (*LHCGR*) yield defective receptors that prevent normal hormone binding; more subtle mutations allow binding but prevent normal signal transduction, or cause misfolding of the receptor that interferes with its normal transport to the cell surface.[468–470]

In affected individuals, müllerian duct derivatives are absent (reflecting the normal action of AMH), wolffian duct development is impaired (reflecting the decreased level of testosterone production), and the testes fail to descend, because normal descent requires the actions of testosterone and insulin-like factor 3, both deriving from Leydig cells.[471, 472] ***The phenotype of patients with Leydig cell hypoplasia otherwise generally correlates with the level of residual LH/hCG receptor activity, ranging from completely female external genital development to nearly normal male genitalia.*** Those having no significant testosterone production appear female at birth and present at puberty with primary amenorrhea and sexual infantilism, lacking both pubic hair development (due to the lack of testosterone) and breast development (due to the absence of aromatizable substrate); the serum LH concentration is elevated and testosterone levels are abnormally low. Others in whom LH receptor function is only partially impaired and testosterone production is decreased but still significant can present with ambiguous genitalia, hypospadias, or micropenis.[468, 470]

Disorders of Antimüllerian Hormone and Its Receptor

The ***hernia uterine inguinale syndrome*** is a rare autosomal recessive disorder that results from a failure of müllerian duct regression, due to mutations in the genes encoding AMH or its receptor.[473, 474] Affected patients appear as normal males having an inguinal hernia containing relatively well-differentiated müllerian duct structures, usually including a uterus and fallopian tubes. They have normal male internal and external genitalia and, usually, cryptorchid testes. In affected families, genetic studies have identified mutations in the *AMH* gene in 45%, and in the *AMHR2* receptor gene in another 39%; in the remaining 15%, no mutation was detected, implicating genes coding for other factors in the AMH transduction cascade.[475]

Sex Chromosome Disorders of Sexual Development

Sex chromosome disorders of sexual development describes a group of disorders associated with an abnormal karyotype, including 45,X (Turner syndrome and variants), 47,XXY (Klinefelter syndrome and variants), 45,X/46,XY mosaicism (mixed gonadal dysgenesis) and 46,XX/46,XY mosaicism (chimerism). Ovotesticular DSD (true hermaphroditism), discussed in the earlier section of this chapter devoted to 46,XX disorders of sexual development, also can be associated with a mosaic 45,X/46,XY or 46,XX/46,XY karyotype.

45,X (Turner Syndrome and Variants)

Turner syndrome was first described in 1938,[476] and now is recognized as important cause of short stature in girls and primary amenorrhea in young women. The disorder also is discussed in Chapter 11, as a cause of amenorrhea. Turner syndrome results from loss of

all or part of an X chromosome and is very common, affecting up to 3% of all conceptions, although only 1 in 1,000 45,X embryos survives to birth; approximately 15% of all spontaneous abortions have a 45,X karyotype.[477, 478] The incidence of Turner syndrome among newborn girls ranges between 1 in 2,000–5,000 live born phenotypic females.[479–481]

Although Turner syndrome classically is associated with a 45,X karyotype, more than half of patients are mosaics (e.g., 45,X/46,XX). The prevalence of mosaicism varies directly with the method used for detection, ranging from 34% with conventional cytogenetics, to 60% with fluorescence in situ hybridization, to nearly 75% when a reverse transcriptase PCR assay is used.[482] Some patients with Turner syndrome lack only part of one X chromosome or exhibit one of a variety of structural abnormalities of the X, including ring chromosomes, isochromosomes, and terminal deletions.

Clinical Features

The sine qua non of Turner syndrome is short stature. It is the only abnormality present in virtually all patients and can be attributed to deletion of the short stature homeobox-containing gene (*SHOX)*, which is located in the pseudo-autosomal region at the distal end of the short arm of the X chromosome (Xp22.33).[483, 484] ***The classical phenotype of Turner syndrome also includes the absence of sexual development, a webbed neck, low set ears and posterior hairline, widely-spaced nipples ("shield chest") short fourth metacarpals, and an increased carrying angle at the elbow ("cubitus valgus").*** The phenotype of patients with Turner syndrome relates, in part, to the parental origin of their X chromosome; most with a 45,X karyotype retain the maternal X.[485]

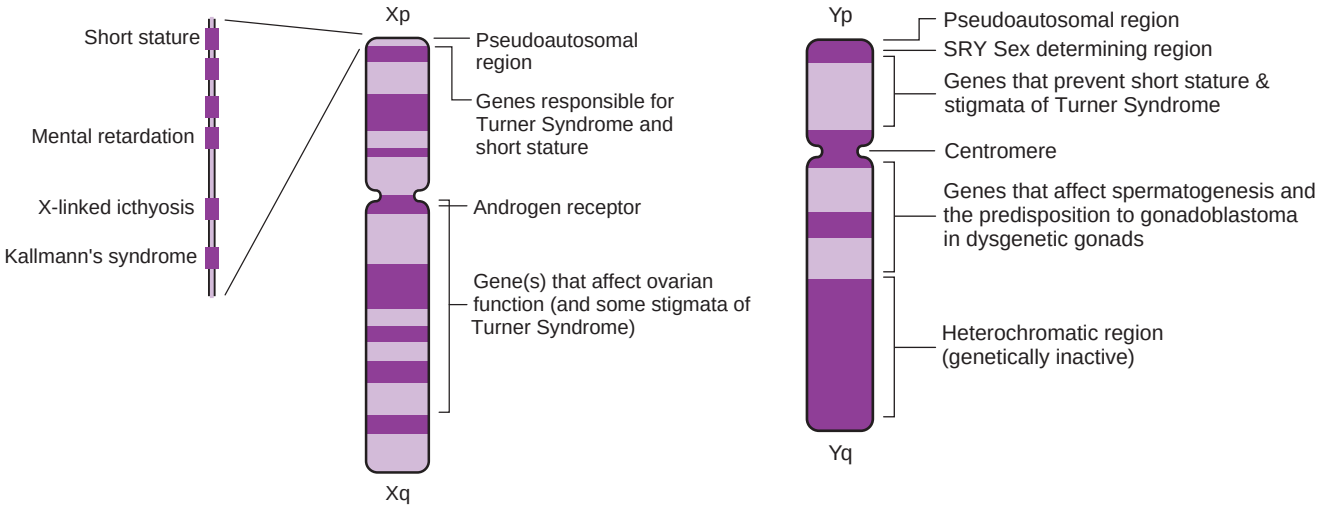
Most women with Turner syndrome have no pubertal development and primary amenorrhea. However, some develop normally and later present with secondary amenorrhea. Approximately 15% of patients with Turner syndrome begin but do not complete pubertal development and approximately 5% complete puberty and begin menstruation.[486] The streak gonads in women with Turner syndrome characteristically are composed of connective tissue with no follicles or only a few atretic follicles. About one-third have ovaries that can be imaged with pelvic ultrasonography, and those who do more commonly exhibit spontaneous breast development at puberty.[487] Although a few women with Turner syndrome conceive naturally, pregnancies are rare and associated with a relatively high risk for sex chromosome aneuploidy and spontaneous abortion.

Approximately 30–50% of patients with Turner syndrome have renal anomalies; horseshoe kidney is the most common and predisposes to hydronephrosis due to obstruction.[488, 489] Cardiac malformations also are common. Approximately 20–30% of patients have aortic valve disease (bicuspid aortic valve, aortic root dilation) and 3–10% have coarctation;[490–493] The prevalence of cardiac anomalies may be higher in those with a 45,X karyotype than in mosaics.[491] Other cardiovascular anomalies include elongation of the transverse aortic arch (49%), persistent left superior vena cava (13%), anomalous pulmonary venous return (13%), and an aberrant right subclavian artery (8%); their prevalence is higher in patients with abnormal neck and chest development.[494] A prolonged QT interval can be observed, even in young children with Turner syndrome.[495, 496] Idiopathic hypertension also is common, even in the absence of any apparent renal or cardiac malformations.[497, 498]

Patients with Turner syndrome are predisposed to developing osteoporosis, primarily due to ovarian failure but possibly also relating to haploinsufficiency for genes affecting bone located on the X chromosome.[499, 500] Many also have ocular abnormalities, including amblyopia, strabismus, ptosis, hypertelorism, epicanthus, farsightedness, and red-green color blindness.[501–503] The prevalence of endocrine diseases such as hypothyroidism and diabetes is increased and, for hypothyroidism, correlates with the karyotype (X iso-chromosome, 38%; 45,X 14%, other, 6%).[504] Celiac disease, hearing loss and liver function abnormalities also are more common in patients with Turner syndrome.[505–508]

Patients with Turner syndrome typically have normal intelligence.[509] The rare patient having a small X-ring chromosome may have severe mental retardation, because the ring chromosome does not undergo X-inactivation.[510, 511] Nonetheless, formal evaluation of intellectual, learning, and motor skills generally is recommended before enrollment in school.[512] Attention-deficit/hyperactivity disorder (ADHD)[513] and problems with visual-spatial organization[514] are more common in girls with Turner syndrome.

Overall mortality is increased approximately 3-fold, relating primarily to circulatory disease, diabetes, liver and renal disease.[515] Although overall cancer risk in women with Turner syndrome is similar to that in the general population, the incidence of CNS tumors, bladder cancer, and endometrial cancer may be increased, and the risk of breast cancer is decreased.[516]



Diagnosis

The diagnosis of Turner syndrome sometimes is incidental, discovered when prenatal CVS or amniocentesis is performed for advanced maternal age or an increased nuchal translucency raises suspicion for chromosomal anomalies.[517] Diagnosis can be made at birth but often is delayed until childhood or adolescence. Neonates frequently exhibit lymphedema of the hands and feet, a webbed neck, nail dysplasia, a high palate, or short fourth metacarpals; short stature and delayed puberty are the later diagnostic keys.[518] ***Therefore, karyotyping is recommended for all girls with unexplained short stature, delayed puberty, a webbed neck, lymphedema, or aortic coarctation, and should be considered for those with 2 or more phenotypic features suggesting the diagnosis.***

Karyotype should include an examination of at least 30 cells, to detect significant mosaicism (e.g., 45,X/46,XX; 45,X/46,XY). For those with a suspicious phenotype but having a normal lymphocyte karyotype, a second tissue (e.g., skin fibroblasts) should be examined to exclude tissue-specific mosaicsm.[519] ***Patients with Turner syndrome having a chromosomal fragment of uncertain origin and those exhibiting any evidence of virilization also should be evaluated specifically using FISH and Y chromosome-specific probes, because those having all or part of a Y (approximately 5%) are at increased risk for developing gonadoblastoma.***[520–523] Routine FISH is not useful because the gene that confers an increased risk for gonadoblastoma has not been identified.[524]

Clinical Management

The spectrum of medical problems in patients with Turner syndrome and their health implications require specific evaluation and periodic monitoring, as described in detail in Chapter 11.[521] In brief summary, evaluation should include periodic echocardiography (or MRI if required), renal ultrasonography, thyroid function studies, a complete blood count,

fasting glucose, lipid profile, and renal and liver function tests, anti-endomysial antibodies (to detect celiac disease), and audiometry. Those with coarctation of the aorta should undergo surgery for its correction and those with other anomalies must be monitored carefully. Hypertension, hypothyroidism, hearing and visual problems also require specific treatment.

Growth hormone therapy should begin as soon as height falls below the 5th percentile for age, usually between 2 and 5 years of age.[525] Early diagnosis and treatment with growth hormone can increase lean body mass[526] and help patients to achieve a normal adult height.[527–529] Combined treatment with growth hormone and low doses of oxandrolone (an anabolic steroid) can help to maximize growth for older girls between 9 and 12 years of age when diagnosis is delayed.[530, 531] Estrogen therapy decreases height velocity and gain in height and therefore generally is not recommended before age 13 or 14 years.[530, 532, 533] ***Estrogen treatment should begin at a low dose (0.25–0.5 mg micronized estradiol, or its equivalent) and increase gradually at 3–6 month intervals until reaching the final dose (2.0 mg micronized estradiol, or its equivalent); the goal is to complete sexual maturation over a period of 2–3 years.*** Once treatment with estrogen begins, linear growth will continue for no longer than another 18–36 months. Treatment with a cyclic progestin (e.g., medroxyprogesterone acetate, 5 mg daily for 12–14 days of each month) should begin after the first episode of menstrual bleeding or after 12–24 months of treatment.[521] Treatment with oral contraceptives offers a convenient alternative for longer-term management.

Oocyte donation offers the possibility of pregnancy to patients with Turner syndrome, but the cardiovascular demands of pregnancy pose unique and potentially serious risks that must be carefully considered. The risk of death during pregnancy is increased as much as 100-fold, primarily due to complications of aortic dissection or rupture. Risk is greatest for those with preexisting abnormalities such as a bicuspid aortic valve or a dilated aortic root, but even those without such findings remain at risk. Consequently, Turner syndrome generally should be regarded as a relative contraindication to pregnancy. Those expressing serious interest in oocyte donation must receive thorough evaluation and counseling, and those having any significant cardiac abnormality should be strongly discouraged.[534]

47,XXY (Klinefelter Syndrome and Variants)

Klinefelter syndrome is the most common congenital cause of hypogonadism in males, affecting approximately 1 in 1,000 male births.[535, 536] The most common karyotype associated with the disorder is 47,XXY, but additional X chromosomes (e.g., 48,XXXY) and mosaics (e.g., 46,XY/47,XXY) also have been described.[537] Klinefelter syndrome results from nondisjunction of the sex chromosomes of either parent during meiosis; mosaics likely result from mitotic nondisjunction.

The phenotype of men with Klinefelter syndrome varies with the number of extra X chromosomes.[537] The gonads are almost always small and firm and sperm production usually is severely decreased. Serum testosterone concentrations generally are low, causing decreased virilization, and gonadotropin levels are elevated.[538] The body habitus exhibits long arms and legs, due both to a long bone abnormality and the influence of testosterone deficiency, and a short trunk. Affected patients also commonly exhibit a variety of psychosocial problems unrelated to hypogonadism.[539–541] The prevalence of pulmonary disease, breast[542] and mediastinal cancers,[543] varicose veins, and diabetes is increased in patients with Klinefelter syndrome.[544] Their mortality from breast cancer is higher than in the general population, and that from prostate cancer is lower.[545]

Diagnosis of Klinefelter syndrome is made by karyotype. Hypogonadism can be treated effectively with testosterone. Fertility is possible via ICSI, even when there are no sperm

in the ejaculate. In a substantial number of azoospermic men with Klinefelter syndrome, sperm can be obtained via testicular sperm extraction (TESE). However, the prevalence of sex chromosome hyperploidy and autosomal aneuploidies is increased in sperm obtained from men with Klinefelter syndrome, compared to normal men, and those chromosomal errors might in some cases be transmitted to their offspring.[546]

45,X/46,XY (Mixed Gonadal Dysgenesis)

Mixed gonadal dysgenesis is a term used to describe asymmetrical gonadal dysgenesis, where one gonad can be identified as a testis and the other is a streak or absent altogether. The most common karyotype associated with the condition is 45,X/46,XY. The dysgenetic testis typically contains immature seminiferous tubules lined by immature Sertoli cells and primitive germ cells. Rarely, the gonad contains primitive sex cord-like structures, with or without germ cells, within an ovary-like stroma.[547]

The phenotype in mixed gonadal dysgenesis can vary widely, probably reflecting the relative proportions of 45,X and 46,XY cells in the gonadal ridge. ***Typically, the genitalia are ambiguous, but also can be female or male.*** Müllerian and wolffian duct development correspond with the character of the ipsilateral gonad.[548, 549] In prepubertal children, basal serum testosterone and gonadotropin concentrations are normal and the testosterone response to exogenous hCG stimulation is highly variable. After puberty, patients exhibit varying degrees of virilization, depending on the level of testosterone production, and serum gonadotropin concentrations are elevated.[549] As might be expected, the incidence of gonadal tumors is relatively high (25%).

45,XX/46,XY (Chimerism)

Chimerism is the term used to describe one body derived from the fusion of cells from both twins of a dizygotic pair. ***All chimeras are, by definition, mosaics, but derive from two distinct zygotes rather than from a single zygote.***[550] Chimeras are not visibly different unless a developmental anomaly in one of the cell lines or sex discordance between the cell lines causes a visibly abnormal phenotype. Most known chimeras are discovered in one of two ways. One is random chance, when people with normal phenotypes are genotyped (e.g., as prospective transplant donors or recipients) and found to carry three or four alleles at multiple loci, instead of one or two.[551–554] Most of the other known chimeras are recognized because of a sex difference between their cell lines, resulting in anomalies of sexual anatomy, maturation, or function that spur the search for an explanation, leading to discovery of mixed cell lines.[555, 556] Consequently, a predominance of abnormal sexual development can be expected in chimeras, due to ascertainment bias. The disturbance in sexual development presumably results from conflicting directions generated by sexually distinct gonadal cells during embryogenesis.

Diagnosis and Management of Ambiguous Genitalia

A newborn infant with ambiguous external genitalia presents a major diagnostic challenge, and a social and medical emergency. The physicians involved must make an important

decision regarding the sex of rearing. The evaluation must be organized and efficient to ensure that the appropriate gender is assigned, that potential life-threatening conditions are recognized, and that the necessary medical, surgical, and psychological interventions begin promptly. A multidisciplinary approach is essential, drawing on the expertise of specialists in neonatology, endocrinology, urology, genetics, psychology, and if available, medical ethics.[557] Diagnostic procedures may delay the decision, but a period of delay is far better than a later reversal of sex assignment. Naming the child should be delayed until a gender is firmly assigned. Open communication with the parents and family is key, emphasizing initially that the team will work with them to make the best possible decisions and that the child can become a well-adjusted, functional member of society. Ideally, one member of the team should be designated to conduct discussions with the family.[558] Parental education, guidance, support, and ongoing communication with the family's primary care physician are essential.[559, 560]

Diagnostic Evaluation

The initial evaluation of an infant with ambiguous genitalia should include a thorough history, complete physical examination, abdominal/pelvic ultrasonography, a karyotype and FISH for *SRY*, and endocrine studies of adrenal and gonadal steroid secretion.

History and Physical Examination

The history should focus on identifying any prenatal exposure to androgens or medications that might act as endocrine disruptors, maternal virilization during pregnancy, and any previously affected relatives, unexplained infant death, or consanguinity.

A careful physical examination cannot establish a diagnosis but can provide useful clues. The presence or absence of dysmorphic features or other anomalies should be noted because the finding usually excludes most forms of CAH and suggests a broader pattern of malformations characteristic of a trisomy (trisomy 21, 18, or 13) or a specific syndrome. Hyperpigmentation suggests high levels of ACTH as in CAH. A systematic examination of the genitalia should answer each of the following questions:

- ***Are gonads palpable?*** Palpation of the genital and inguinal regions is perhaps the most important part of the physical examination. Gonads in the inguinal regions or in scrotal folds are almost certainly testes.[561] Asymmetry of the gonads or genitalia suggests gonadal dysgenesis or the presence of both testis and ovary (ovotesticular DSD). ***Gonads that are not palpable may be in the abdomen and can be ovaries or testes, but virilizing CAH must be excluded specifically.***
- ***What is the length and diameter of the phallus?*** Careful examination of the phallus may help differentiate between a penis and a clitoris. Whereas the penis has a midline ventral frenulum, the clitoris has two folds extending from its lateral aspects to the labia minora. The phallus should be measured from the pubic ramus (compressing any suprapubic fat) to the tip of the glans (excluding any excess foreskin) after stretching to the point of resistance; diameter should be measured at the midpoint. Measurements should be compared to established norms, which are adjusted for gestational age. In a term infant, normal penile length is 2.5 cm or greater and normal diameter is 0.9 cm or greater.[562] In normal neonates, clitoral length ranges between 2 and 6 mm and measurements greater than 9 mm are unusual.[563, 564] Micropenis that is not accompanied by hypospadias can be caused

by decreased testosterone production *in utero* or by growth hormone or gonadotropin deficiency. Clitoromegaly resulting from androgen exposure can be caused by maternal androgen exposure, CAH, or ovotesticular DSD.

- ***What is the position of the urethral meatus?*** The urethral meatus may be found anywhere along the ventral surface of the phallus or on the perineum. Hypospadias is almost always accompanied by chordee, which is a ventral curvature of the phallus resulting from a shortened urethra. A single opening at the base of the phallus can represent either an incompletely fused penile urethra (hypospadias) or a virilized urogenital sinus. In either case, the findings must be confirmed radiologically or by cystoscopy/vaginoscopy.
- ***Are the labioscrotal folds fused?*** The findings can range widely from the unfused labia majora of a normal female, to varying degrees of posterior fusion, to a bifid scrotum, to a fully fused normal-appearing male scrotum. The anogenital ratio is defined by the distance from the anus to the posterior fourchette of the vagina, divided by the distance from the anus to the base of the clitoris.[565] A ratio greater than 0.5 suggests virilization and some degree of labioscrotal fusion. Once the karyotype is known, the extent of virilization can be further defined by comparison to established gender-specific standards[133, 566] to more objectively document the genital phenotype.

Imaging

Imaging with abdominal and pelvic ultrasonography can help to determine the location of the gonads and the presence or absence of a uterus. However, an infantile uterus can be difficult to image, even with MRI. Imaging is informative when it reveals a uterus but inconclusive when it does not. Although a retrograde urethrogram can be used, cystoscopy/vaginoscopy generally is the best method for defining the urethral and vaginal anatomy.[567] Rarely, laparoscopy may be needed to confidently define the reproductive anatomy and to biopsy the gonads.

Initial Laboratory Evaluation

The initial laboratory evaluation should include the following:

- Karyotype, to determine the chromosomal sex
- FISH, to determine the presence or absence of *SRY*
- Measures of adrenal and gonadal steroid secretion, including:
  - 17OHP—rapidly excludes CAH due to 21-hydroxylase deficiency, a common cause of ambiguous genitalia that can be life-threatening.
  - Electrolytes—should be measured immediately and monitored at least daily until salt wasting can be excluded.
  - ACTH, cortisol, DHEA, 17α-hydroxypregnenolone, and 11-desoxycortisol—identifies less common causes of CAH caused by enzyme deficiencies other than 21-hydroxylase deficiency.

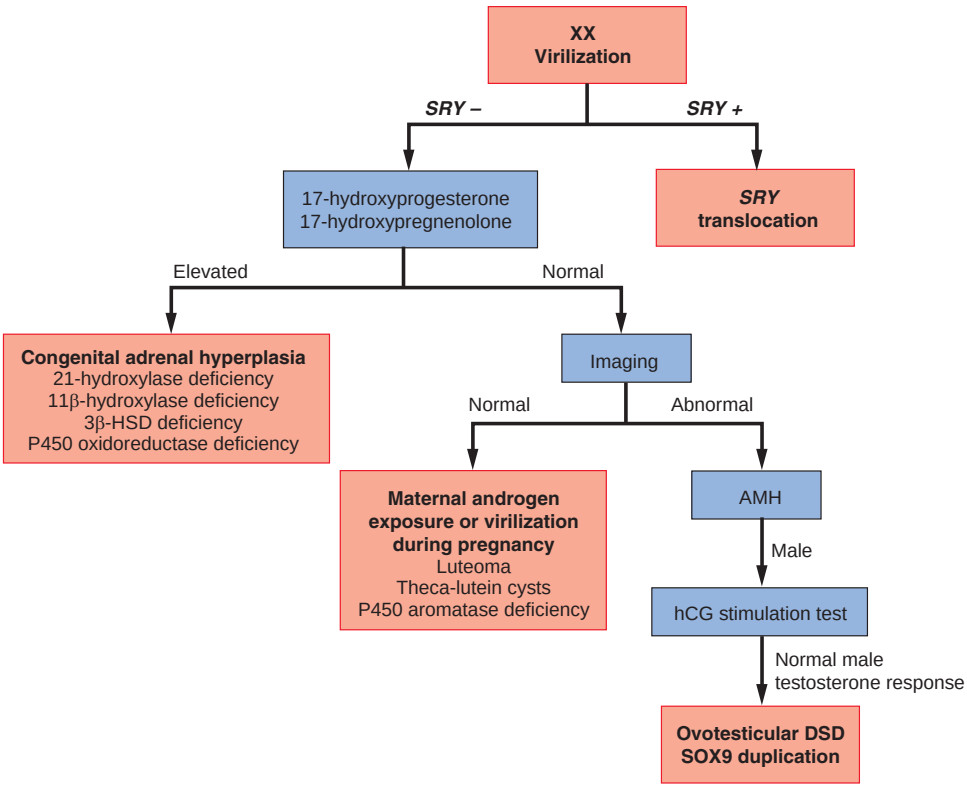
Differential Diagnosis

Based on the karyotype, the infant can be placed in one of three diagnostic categories: (1) XX virilization, (2) XY undervirilization, or (3) mixed sex chromosome pattern.[109]

XX Virilization

The differential diagnosis of XX virilization includes disorders of gonadal (ovarian) development and androgen excess of fetal, fetoplacental, or maternal origin. FISH detection of *SRY* indicates an *SRY* translocation. In the absence of *SRY*, CAH is the most common diagnosis in XX virilized infants. An elevated serum 17OHP or 17α-hydroxypregnenolone concentration clearly suggests CAH. The most common cause, 21-hydroxylase deficiency, can be differentiated from other less common causes of CAH (11β-hydroxylase, 3β-HSD, and P450 oxidoreductase deficiencies) by the serum levels of ACTH, cortisol, DHEA, and 11-desoxycortisol, because each has a characteristic serum steroid hormone pattern (described in earlier sections of this chapter).

When CAH has been excluded, imaging revealing a uterus or a history of maternal androgen exposure or virilization during pregnancy points clearly toward maternal gestational hyperandrogenism (pregnancy luteoma, theca-lutein cysts) or P450 aromatase deficiency. Some of the remaining uncommon causes of XX virilization due to testicular DSD can be differentiated by measuring AMH (or inhibin B) and performing an hCG stimulation test (described below); male levels of AMH and a normal testosterone response to hCG indicate the presence of functional testicular tissue, suggesting the diagnosis of ovotesticular DSD or a *SOX9* duplication.[121, 568]



XY Undervirilization

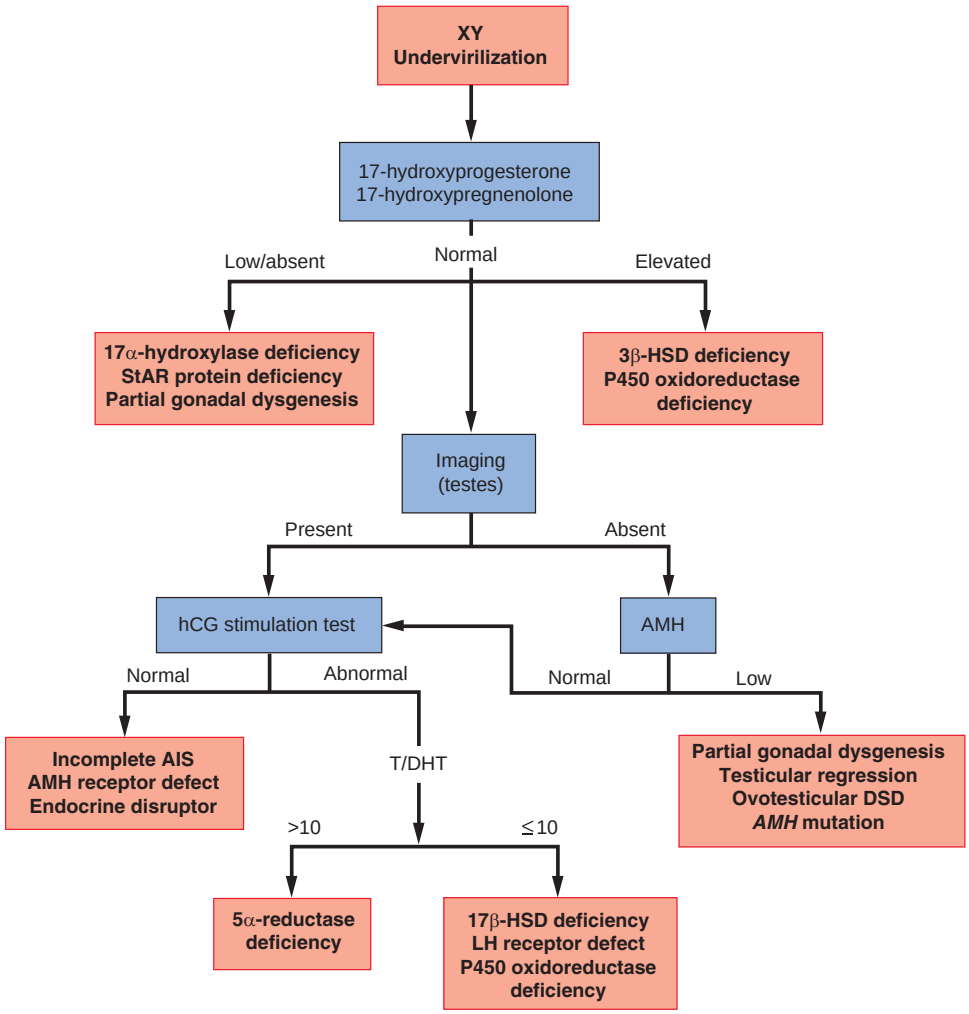
The differential diagnosis of XY undervirilization includes disorders of gonadal (testicular) development, disorders of androgen synthesis or action, LH receptor defects, and disorders of AMH and its receptor. The diagnostic evaluation for *SRY*-positive infants is challenging

due to the large number of possible causes. In addition to the initial laboratory evaluation described above, XY undervirilized infants require further evaluation by measuring the serum LH, FSH, AMH, testosterone, androstenedione, and dihydrotestosterone (DHT) concentrations. Gonadotropin and sex steroid levels should be measured when they normally are detectable, either in the first 24 hours after birth or between 2 and 6 months of age.[569, 570] Provocative tests (ACTH stimulation test, hCG stimulation test) and other specialized tests also may be required to establish a diagnosis.

The serum concentrations of 17OHP and 17α-hydroxypregnenolone help to narrow the spectrum of diagnostic possibilities. Low or absent levels indicate a 17α-hydroxylase or StAR protein deficiency, or partial gonadal dysgenesis (*SF1* mutation, *DAX1* duplication), and elevated levels point to a 3β-HSD or P450 oxidoreductase deficiency; an ACTH stimulation test can better demonstrate the different adrenal enzyme deficiencies,[571] and also excludes adrenal insufficiency. The test involves measurements of serum ACTH, cortisol, progesterone, pregnenolone, 17OHP, 17α-hydroxypregnenolone, DHEA and androstenedione before and 60 minutes after administering cosyntropin (synthetic ACTH 1–24; 1 μg/m^2 or 0.25 mg)[176]; hormone values must be compared to age-adjusted normal ranges. The absence of any significant steroidogenic response suggests partial gonadal dysgenesis due to a StAR deficiency, a *SF1* mutation, or *DAX1* duplication. Normal 17OHP and 17α-hydroxypregnenolone concentrations require additional evaluation.

AMH (or inhibin B) is a marker of Sertoli cell mass and identifies patients having functional testicular tissue, even when the testes cannot be imaged.[572] A low serum AMH level suggests a form of partial gonadal dysgenesis, testicular regression syndrome, ovotesticular DSD, or an *AMH* mutation. Those with *AMH* mutations typically have normal male external genitalia, variable testicular descent, and persistent müllerian ducts. When the serum AMH level is normal or imaging reveals testes, disorders of androgen synthesis (steroid 5α-reductase, 17β-HSD, and P450 oxidoreductase deficiencies), LH receptor defects, incomplete androgen insensitivity, and AMH receptor defects must be considered. In patients with steroid 5α-reductase deficiency, the testosterone/DHT ratio typically is greater than 10. In those with 17β-HSD deficiency, the serum testosterone concentration often is in the lower normal range, but the serum androstenedione level is elevated several-fold and the testosterone/androstenedione ratio usually is less than 0.8.[573] In patients with LH receptor defects, LH levels are high, testosterone concentrations are low, and androstenedione levels are not elevated. An hCG stimulation test helps to better define and distinguish suspected enzyme deficiencies from incomplete androgen insensitivity and LH receptor defects. The test involves measurements of serum hCG, LH, FSH, testosterone, androstenedione, and DHT on days 1 (basal), 3 and 6, with exogenous hCG (1,500 IU/m^2) administered on days 1 and 3. A normal response is a 2-fold increase in the testosterone level on day 3 and a 4-fold rise on day 6, a testosterone/DHT ratio less than 10,[353] and a testosterone/androstenedione ratio greater than 0.8.[573] However, because endocrine evaluation may not distinguish clearly between patients with 17β-HSD deficiency and those with LH receptor defects, genotyping may be required to establish the correct diagnosis.

In patients exhibiting normal basal and stimulated androgen levels, the remaining possibilities include incomplete androgen insensitivity, an AMH receptor defect, and prenatal exposure to an endocrine disruptor. Sequencing of the *AR* gene will identify some, but not all, patients with incomplete AIS; a demonstrable mutation will be found in fewer than half of those in which the diagnosis is suspected,[451] and other possibilities (e.g., *SF1* mutation) must then be considered. Those having an AMH receptor mutation typically have normal male external genitalia and cryptorchid testes. Occasionally, ambiguous genitalia may result from prenatal exposure to phenytoin, phenobarbital, or an environmental exposure.[574]



Mixed Sex Chromosome Pattern

The differential diagnosis of a mixed sex chromosome pattern includes only a few disorders associated with genital ambiguity, such as mixed gonadal dysgenesis, ovotesticular DSD and chimerism. Patients with mixed gonadal dysgenesis typically exhibit external genital asymmetry, and those with ovotesticular DSD generally can be expected to have a low serum AMH level.

Clinical Management of Children with Ambiguous Genitalia

The management of children born with ambiguous genitalia focuses initially on stabilization, averting the possibility of adrenal crisis in infants with salt-wasting forms of CAH being the most urgent medical issue. Thereafter, management decisions regarding the sex of rearing must consider a great many interacting and sometimes conflicting factors. Recent years have witnessed a growing appreciation for the complexities of psychosexual

development, which reflects the influence of genes, societal and cultural norms, and family dynamics, and the effects of prenatal androgen exposure on sexual differentiation of the brain. Consequently, the wisdom of traditional management paradigms, in which sex assignment has been based primarily on the potential for reproduction and traditional sexual function, has come into question.[575–577]

Stabilization

CAH is the most common cause of 46,XX disorders of sexual development, one of the causes of 46,XY disorders of sexual development, and can be life-threatening. Consequently, infants with ambiguous genitalia and no palpable gonads should be assumed to have CAH and treated empirically until the diagnosis is confirmed or excluded. If CAH is not promptly and effectively treated, hypoglycemia and symptoms of salt-wasting (vomiting, diarrhea, hypovolemia, hyperkalemia, and cardiovascular collapse) can occur anytime within the first few days and weeks of life. The treatment of infants with classical forms of CAH is described in detail in an earlier section of this chapter and is only summarized here. Initial treatment includes fluid administration (5% dextrose in 0.9% saline), correction of any electrolyte abnormalities, and administration of stress doses of steroids. Patients with a confirmed diagnosis of CAH will require careful monitoring of electrolytes (hyperkalemia usually is the first indication of impending adrenal crisis), prompt treatment with adequate doses of hydrocortisone (which has some mineralocorticoid activity), and treatment with mineralocorticoids once stable and feeding normally.

Family Counseling

Initial discussions should focus on helping the parents to gain a basic understanding of the causes of ambiguous genitalia, first explaining that the genitalia are not fully formed or are overdeveloped, and providing additional, more detailed information in accordance with the family's intellectual and emotional capacity and considering their cultural and religious background. Until the diagnosis and sex of rearing have been established, birth announcements describing gender should be postponed, requiring careful and sensitive counseling. If the family inquires, matters relating to long-term reproductive function and sexuality should be discussed openly and frankly.

Gender Decisions

In some cases, the decision is relatively straightforward, such as in mildly virilized girls with CAH. However, in many, the influence of sex assignment and rearing on ultimate gender identity cannot be predicted confidently, primarily because there are few data regarding long-term outcomes to guide the decision.[578–580]

Traditional approaches have focused on early gender assignment and reconstructive surgery, but many now advocate postponing surgery until the patient can participate in the decision, when possible.[581] The traditional approach was based on the assumptions that gender identity reflects gender assignment and the sex of rearing and can be imposed, that preservation of fertility (if possible), sexual function, and appearance should be the primary goals, and that a good anatomical result will translate to a healthy adaptation and patient satisfaction. Consequently, virilized female infants were assigned a female sex and underwent staged genital reconstruction during infancy, and undervirilized boys often were

managed similarly, based on judgments regarding the feasibility of constructing a penile urethra.[582] These traditional principles were promulgated by the American Academy of Pediatrics as recently as 2000,[583] but an international (American and European) consensus conference conducted in 2006 has questioned their wisdom,[109] citing the lack of evidence that early genital surgery effectively reinforces gender assignment or influences gender identity or that genital appearance directs gender role decisions in adults. Vocal advocates have argued passionately that reconstructive surgery should be delayed until the patient can participate in the decision, but whether most affected adults agree is unknown.

Available data from long-term studies reflecting the traditional approach to gender assignment must be interpreted cautiously, because many of the most discontented affected adults may have declined to participate. Studies involving patients who had early feminizing surgery indicate that many had a poor cosmetic result and most required further surgery,[584] and that sexual dysfunction is common among women who underwent clitoral surgery as infants or children.[585] Surgical techniques have improved,[586, 587] but long-term outcomes remain uncertain. Two studies of long-term psychosocial outcomes in patients with 46,XY disorders of sexual development found that half were living as men and half as women, and that the two groups did not differ in satisfaction with their appearance, function, or sex of rearing.[438, 588] Most patients assigned a female gender were satisfied with their sex of rearing, but only half had exclusively heterosexual interests. Overall, half felt they did not have adequate information about their medical histories.[438, 588] Whereas these data suggest that early gender assignment based on appearance usually results in a healthy psychosocial adjustment and outcome, the studies focused primarily on patients having limited potential for virilization. In contrast, patients having significant potential for virilization, such as those with steroid 5α-reductase or 17β-hydroxysteroid dehydrogenase deficiencies, do not readily accept female gender assignment.[589]

Currently, there are no universally accepted guidelines for gender assignment. The Consensus Statement on the Management of Intersex Disorders offered numerous specific conclusions and recommendations.[109] Guidelines published by another group separate decisions regarding gender assignment from those regarding genital surgery.[590] The Intersex Society of North American, a peer support, education, and advocacy group founded and operated by and for intersexuals also has published recommendations, emphasizing the importance of avoiding harmful or unnecessary surgery, qualified professional mental health care for the child and family, and empowering patients by helping them to understand their condition and to choose or decline medical interventions.

Although expert opinions and recommendations vary, the guiding principles can be summarized as follows:

- Gender assignment and sex of rearing should be based on the most probable adult gender identity and the potential for adult function.
- Decisions should respect the family's own values and preferences.
- Children should be raised in the predicted and selected gender role, but should also participate actively in longer-term gender decisions.

There is general agreement on some issues. Almost all virilized 46,XX children with CAH should be raised as females, primarily because more than 90% identify as females as adults, and 46,XX infants having essentially normal male external genitalia probably should be raised as males. When female gender is assigned, clitoral surgery probably is best delayed until the child expresses gender identity and can participate in the decision. When male gender is assigned, phallic reconstruction can be performed at a time acceptable to the family and to the surgeon, while still permitting a later change in gender decision by the adult patient. Finally, purely cosmetic surgery should be postponed until the patient can participate in the decision.

Long-Term Care

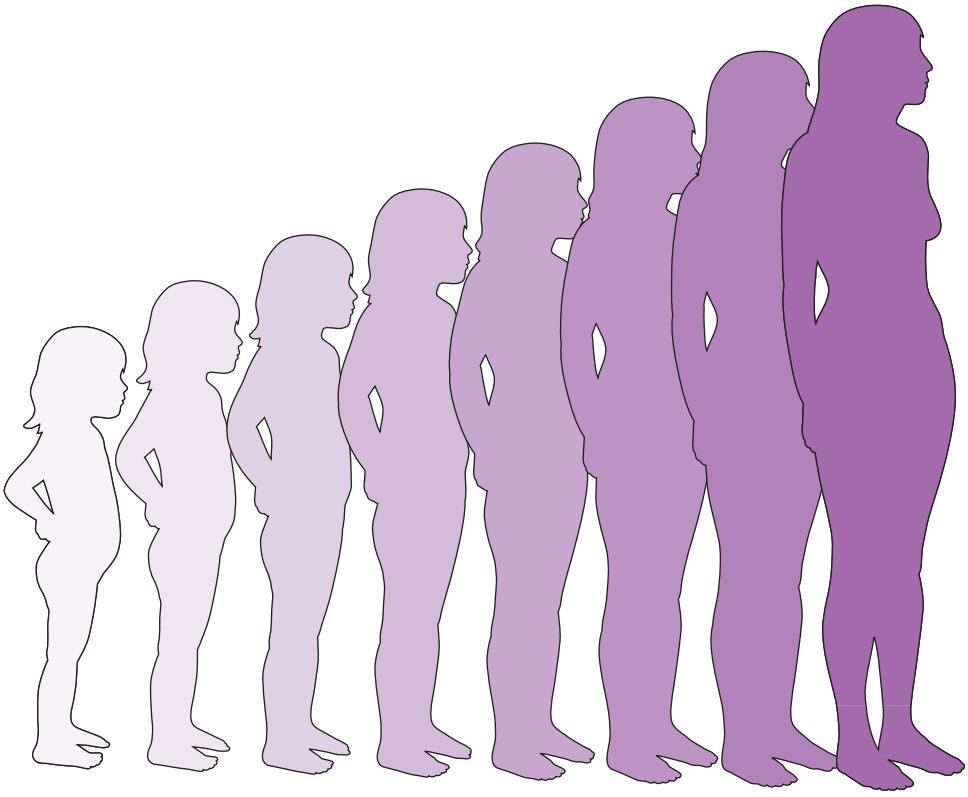
Patients with disorders of sexual development and their families should receive ongoing support to help in their psychosexual development. Support groups offer many affected patients useful information and insights and include the Intersex Society of North America (www.isna.org), the Androgen Insensitivity Syndrome Support Group (www.aissg.org), Bodies Like Ours (www.bodieslikeours.org), and The Magic Foundation (www.magicfoundation.org).

Patients having all or part of a Y chromosome whose gonads are located in the abdomen are at significant risk for developing a gonadal tumor. Consequently, they should be removed, or when possible and appropriate, moved to the scrotum, soon after diagnosis, except in those with complete AIS, in whom surgery generally is best postponed until after puberty.

Long-term care also must consider the potential effects of sex steroid exposure, the possibility and effects of a changing hormone environment at puberty, and the need for timely and effective sex steroid therapy.

All references are available online at: http://www.clinicalgynendoandinfertility.com

10 Normal and Abnormal Growth and Pubertal Development



In many societies throughout history, puberty has been a time of celebration. The changes that accompany puberty announce the transition from childhood to adulthood and the development of fertility. Puberty is the process of cognitive, psychosocial, and biologic maturation. Whereas growth and the development of secondary sexual characteristics are the most visible manifestations of the onset of puberty, changes in body composition and cognitive development are no less significant.[1] Puberty can be a difficult transition for many adolescents, even when it progresses normally, and presents substantially greater challenges when its onset is premature, or delayed. The recent trend towards an earlier pubertal maturation and some of its consequences, notably earlier sexuality and the problem of teen pregnancy, make it all the more important to understand the physiology of normal puberty.

This chapter focuses first on the endocrinology and physiology of normal puberty, to provide the foundation for subsequent discussion of the pathophysiology, diagnosis, and management of abnormalities of growth and pubertal development.

The Endocrinology of Normal Puberty

The hypothalamus, anterior pituitary gland, and gonads of the fetus, neonate, infant, and prepubertal child are all capable of secreting hormones in adult concentrations. The key to understanding the endocrinology of puberty lies in first understanding the mechanisms that govern the hypothalamic-pituitary-gonadal axis.

The Ontogeny of the Hypothalamic-Pituitary-Gonadal Axis

The "hypothalamic pulse generator," the term chosen by Ernst Knobil to describe the rhythmic, pulsatile nature of gonadotropin-releasing hormone (GnRH) secretion,[2] consists of approximately 1,500–2,000 specialized neurosecretory cells in the arcuate nucleus, located in the medial basal hypothalamus. The resident GnRH neurons exhibit spontaneous "autorhythmicity" and function as an oscillator in the pulsatile secretion of GnRH.[3, 4] In response to the pulsatile GnRH signal, pituitary gonadotropes, which contain plasma membrane GnRH receptors, secrete follicle-stimulating hormone (FSH) and luteinizing hormone (LH), also in a pulsatile manner. In turn, the episodic gonadotropin signal stimulates maturation of the germinal elements of the gonads and is transmitted into the pulsatile secretion of sex steroids. Ultimately, the patterns of gonadotropin and gonadal steroid secretion during fetal life, infancy, childhood, adolescence, and adulthood primarily reflect changes in the activity of the hypothalamic pulse generator.

Fetal Life and Infancy

The hypothalamic-pituitary-gonadal axis becomes functional even before birth. Neurons that synthesize GnRH originate in the olfactory placode and migrate to the hypothalamus between 6 and 9 weeks of gestation.[5] By 10 weeks, the hypothalamus contains significant amounts of GnRH.[6] Development of the hypothalamic-pituitary portal venous system begins between 9 and 10 weeks of gestation and is completed by 19–20 weeks.[7] Consequently, FSH and LH concentrations in fetal pituitary glands increase dramatically soon thereafter. The pituitary begins to secrete FSH and LH into the fetal circulation by week 12.[8] Fetal serum gonadotropin levels rise progressively, reaching a peak between 20 and 24 weeks,[9] then decrease steadily over the last 10 weeks of pregnancy, probably due to a developed sensitivity to the negative feedback effects of high circulating estrogen and progesterone concentrations derived from the placenta.[10, 11]

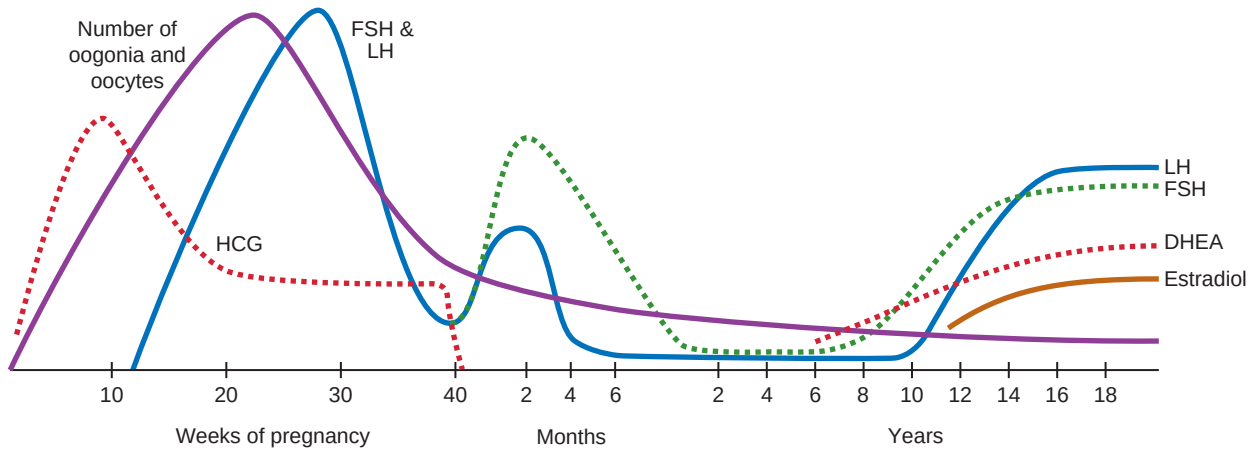
After birth, steroid levels fall precipitously due to the loss of maternal and placental hormones, allowing the newborn's hypothalamic-pituitary-gonadal axis to escape their suppressive effects. The characteristic pulsatile pattern of hypothalamic GnRH secretion emerges,[12, 13] and serum gonadotropin concentrations rise again promptly, with a striking sex difference; FSH rises to a greater extent in females and LH to a greater extent in males.[11] In female infants, FSH concentrations occasionally reach levels even greater than those observed in the normal adult menstrual cycle.[14, 15] Consequently, waves of ovarian follicular development begin and estradiol levels during the first few months of life are comparable to those observed during the midfollicular phase of the menstrual cycle.[16] In male infants, elevated LH levels stimulate increased testosterone secretion from the testes. Gonadotropin and gonadal steroid levels peak at about 3–6 months in boys and 12–18 months in girls

and steadily decline thereafter, presumably because normal negative feedback mechanisms become fully functional. By approximately 9–12 months of age in boys and 24–36 months in girls, gonadotropin concentrations fall to typical prepubertal levels, remaining at very low concentrations until the onset of puberty.[11] Suppression of hypothalamic pulse generator activity is less intense and shorter in duration in females than in males, probably reflecting the influence of testosterone on hypothalamic programming.[17]

Childhood and Early Adolescence

During the interval between infancy and puberty, known as the "juvenile pause" in nonhuman primates, the hypothalamic-pituitary-gonadal axis lies dormant. Normal ovulatory menstrual cycles can be induced in prepubertal female monkeys by administering a higher amplitude pulsatile infusion of exogenous GnRH, indicating that neither the anterior pituitary nor the gonads are the limiting factor.[18] Although the GnRH pulse generator is active, the frequency and amplitude of pulsatile GnRH secretion generally are irregular and very low.[19–22] Low amplitude pulses of gonadotropin secretion can be detected in prepubertal children as young as 5 years of age, primarily during sleep.[22–24] FSH levels rise more than LH, but there is no detectable increase in steroid hormone concentrations.

For a long time, the prevailing theory to explain the juvenile pause that precedes puberty envisioned a hypothalamic "gonadostat" controlling the level of sensitivity to the central negative feedback actions of gonadal steroids. In that context, the changing patterns of gonadotropin secretion were attributed to changes in the "gonadostat" setting. Decreasing gonadotropin levels in late infancy and sustained low concentrations during childhood reflected a rising and ultimately high sensitivity to even very low levels of sex steroid feedback, and increasing gonadotropin concentrations at the onset of puberty reflected a decrease in feedback sensitivity.[25, 26] The "gonadostat" theory prevailed until cross-sectional and longitudinal studies in children with gonadal dysgenesis revealed a similar, but exaggerated, "diphasic" pattern of gonadotropin secretion. Serum gonadotropin levels in girls with Turner syndrome are markedly elevated in infancy, decline to very low levels during childhood, and rise again to grossly high concentrations at pubertal age, all in the absence of any possible change in the level of negative feedback from gonadal steroids.[27] These and similar observations in castrate nonhuman primates demonstrated that steroid hormone feedback affects the amount, but not the pattern, of gonadotropin secretion, contradicting the traditional "gonadostat' theory and establishing a new paradigm. ***The typical "diphasic" pattern of gonadotropin secretion from infancy to puberty results primarily from changing levels of central inhibition of pulsatile GnRH secretion, and to a lesser extent, from a high sensitivity to low levels of gonadal steroid feedback.***



Puberty

About 1 year before breast budding in prepubertal girls, the character of nocturnal pulses of gonadotropin secretion changes with LH levels exceeding those of FSH. Breast budding occurs when the nocturnal pulses of gonadotropin secretion become great enough to generate detectable coincident increases in serum estradiol concentrations. At that time, LH peak amplitude increases about 10-fold, whereas FSH pulse amplitude only doubles, resulting in a marked decrease in the serum FSH/LH ratio.[22, 24] The change reflects an increase in pituitary responsiveness to GnRH, which has a priming effect on pituitary LH secretion and increases the number of GnRH receptors on gonadotropes (up-regulation). Gonadotropes first increase their capacity for response to GnRH by synthesis, and later by secretion of gonadotropins. Pulse frequency also increases, but to a much lesser extent. Gonadotropin pulses become diurnal and the duration of increases in estradiol levels becomes more prolonged. As puberty progresses, the amplitude of pulsatile LH secretion increases further, to levels 20–40 times greater than those detected during prepubertal years, probably reflecting the influence of rising estradiol levels at both the hypothalamic and pituitary levels. Although nocturnal pulse amplitudes are still greatest, significant pulses occur during daytime and basal LH levels become detectable.[22] LH bioactivity also increases, due to changes in glycosylation.[28]

In response to increasing gonadotropin secretion, basal estradiol levels increase progressively.[20] Inhibin B levels, which are low or undetectable in prepubertal girls, increase sharply in mid-puberty, then decline in its later stages, first reflecting increasing ovarian stimulation, then the onset of the menstrual cycle and the appearance of a luteal phase, when levels are low.[29] Inhibin A concentrations, undetectable or very low through early puberty, increase gradually thereafter but reach adult levels only after menarche, consistent with the corpus luteum being the primary source.[29] Menarche occurs in late puberty, after a year-long rise in daily estrogen production,[30] probably when estradiol and inhibin B levels become sufficient to exert significant negative feedback on gonadotropin secretion, resulting in cyclic estrogen production. Cycle length and menstrual characteristics vary until the positive feedback relationship between estradiol and gonadotropin secretion matures and ovulation becomes established, often a year or more after menarche.

Central Control Mechanisms

Some of the factors governing the "neuroendocrine switch" for the GnRH pulse generator that is "on" in early infancy, turns "off" during childhood, and switches back "on" again at puberty have now been identified. Most of the important work on the neuroendocrinology of puberty has been performed in nonhuman primate models. The list of factors that modulate the activity of the hypothalamic-pituitary-gonadal axis includes both inhibitory and excitatory neurotransmitters and peptides.

Gamma-Aminobutyric Acid

Gamma-aminobutyric acid (GABA) is an inhibitory neurotransmitter produced by specialized neurons in the hypothalamus and has an important role in regulating the activity of the GnRH pulse generator. Elegant hypothalamic perfusion studies have revealed that release of GABA into the median eminence decreases as pulsatile GnRH secretion increases at the onset of puberty.[31] Conversely, central perfusion with a $GABA_A$ receptor antagonist (bicuculline) or the antisense oligodeoxynucleotide for the mRNA coding the $GABA_A$ synthesizing enzyme (glutamate acid decarboxylase) stimulates GnRH release.[31, 32] Chronic administration of bicuculline into the third ventricle induces precocious puberty and menarche in prepubertal female monkeys.[33] Evidence suggests that changes in the

subunit composition of $GABA_A$ receptors may contribute to the disinhibition of pulsatile GnRH secretion at the onset of puberty.[34] These observations suggest that central GABA signaling is one of the factors that restrains GnRH neuronal activity during childhood.

Neuropeptide Y

Neuropeptide NPY (NPY) is a hypothalamic peptide involved in the control of food intake behavior and reproductive function in adults. In castrate adult female monkeys, intracerebroventricular administration of NPY inhibits pulsatile GnRH release.[35] In males, the postnatal pattern of GnRH pulse generator activity is inversely related to NPY gene and protein expression in the medial basal hypothalamus and central administration of an NPY receptor antagonist stimulates GnRH release in juveniles.[36] These observations suggest that NPY, like GABA, is an important component of the "neurobiologic brake" that restrains the GnRH pulse generator in prepubertal primates. However, others have observed that NPY levels increase in the median eminence at the onset of puberty,[37] that infusion of NPY into the median eminence increases GnRH release,[38] and that infusion of NPY antiserum into the median eminence did not stimulate GnRH secretion in prepuberal monkeys.[37] The varying effects of NPY appear to relate to the site of infusion within the brain and not on the steroid milieu.[35] Additional work will be required to clarify the role of NPY in regulation of the hypothalamic pulse generator and the onset of puberty.

Glutamate

Glutamate is an excitatory neurotransmitter in the hypothalamus and stimulates GnRH release via N-methyl-D-aspartate (NMDA) receptors both *in vivo* and *in vitro*.[39] An intravenous bolus of NMDA stimulates hypothalamic GnRH release,[40] and treatment with a specific glutamate receptor antagonist blocks the effect in nonhuman primates. Moreover, prolonged (16–30 weeks) intermittent NMDA stimulation (1 minute every 3 hours) activates the hypothalamic-pituitary-gonadal axis and stimulates precocious puberty and the initiation of spermatogenesis in juvenile males.[41] These observations suggest that glutamate signaling may play a role in the resurgence of pulsatile GnRH secretion at the onset of puberty.

Kisspeptins

In just the past few years, kisspeptins have emerged as a critical component of the system that controls the level of GnRH neuronal activity between infancy and puberty. Kisspeptins are neuropeptides (encoded by the *KISS1* gene) that signal via the G-protein coupled receptor, GPR54 (encoded by the *KISS1R* gene).[42] Interestingly, the first evidence of their importance in the regulation of the hypothalamic-pituitary-gonadal axis came from observations in humans. Several members of a large consanguineous family with hypogonadotropic hypogonadism and delayed puberty were found to harbor homozygous inactivating mutations for GPR54.[43, 44] One affected compound heterozygote exhibited an exaggerated pituitary response to exogenous pulsatile GnRH administration, suggesting a hypothalamic locus for the disorder.[44] More importantly, the observation also suggested that kisspeptin signaling via GPR54 might play a major role in the resurgence of pulsatile GnRH secretion at puberty in primates. The results of subsequent studies in nonhuman primates and humans strongly support that interpretation.

Neurons expressing *KISS1* are located exclusively in the arcuate nucleus,[45, 46] where GnRH neurons also express GPR54.[42] In castrate male and intact female monkeys, the pubertal resurgence of pulsatile GnRH secretion is associated with a nearly 5-fold increase in *KISS1* expression and, in females, also with an increase in *KISS1R* expression.[47] Hypothalamic kisspeptin secretion is distinctly pulsatile and highly correlated with that of GnRH.[48] An intermittent kisspeptin infusion can sustain pulsatile LH secretion in castrate juvenile animals after discontinuation of a priming pulsatile infusion of exogenous GnRH, but not in the presence of a GnRH receptor antagonist, indicating that the effect of kisspeptin is mediated via pulsatile GnRH secretion.[47] The observation that patients with inactivating mutations of GPR54 exhibit pulsatile LH secretion with low amplitude and normal

frequency suggested that kisspeptin might only amplify and not stimulate GnRH pulse generator activity directly.[44, 49] However, a continuous kisspeptin infusion, which down-regulates GPR54, suppresses both LH pulse amplitude and frequency, implying that kisspeptin has similar effects on pulsatile GnRH secretion.[50, 51] Finally, an activating *KISS1R* mutation resulting in prolonged activation of the *KISS1R* signal transduction pathway has been described in a young girl with GnRH-dependent (central) precocious puberty.[52] ***Taken together, these observations indicate that hypothalamic kisspeptin-GPR54 signaling is a key component of the neurobiologic mechanism that triggers the onset of puberty. They further suggest that kisspeptin neurons may provide the fuel for the hypothalamic GnRH pulse generator.*** The report of undetectable serum gonadotropins in an infant boy bearing a loss-of-function mutation in the *KISS1R* gene suggests that kisspeptin input to the GnRH neuronal network also is necessary for the increased pulsatile GnRH secretion normally observed during early infancy.[53]

There is increasing evidence from studies in nonhuman primates that kisspeptin neurons also are involved in mediating the feedback actions of both testicular and ovarian hormones. In males, testosterone negative feedback, which regulates LH secretion by slowing the pace of pulsatile GnRH secretion,[54] is associated with a decrease in hypothalamic *KISS1* mRNA levels.[55] The observation suggests that kisspeptin neurons play an important role in the negative feedback loop that regulates LH secretion in the male, which also involves opioid and GABA neuronal input.[56] Gonadal steroids also suppress hypothalamic *KISS1* expression in females. In postmenopausal women and ovariectomized monkeys, the density of neurons expressing *KISS1* mRNA is significantly higher than in premenopausal women and in intact females monkeys, and treatment with estrogen and progesterone markedly decreases *KISS1* expression,[46] suggesting that kisspeptin neurons also participate in mediating the hypothalamic negative feedback actions of ovarian steroid hormones. ***The collective body of evidence thus indicates that kisspeptin neurons are a critical component of the neurobiologic mechanism that regulates the activity of the hypothalamic pulse generator.***

In some way, the system that controls the ontogeny of pulsatile GnRH secretion integrates kisspeptin signaling with that of other neuotransmitters (glutamate, GABA) and neuropeptides (NPY). Whether kisspeptin neurons in the arcuate nucleus function as a "pubertal clock," as a growth-tracking "somatometer," or simply relay information from such centers to the GnRH neuronal network is unknown.[57] Regardless, kisspeptin neurons have emerged as one of the primary transducers of the internal and external environmental cues that regulate the neuroendocrine reproductive axis.

Peripheral Signaling

The age at onset of puberty has been declining steadily as the prevalence of obesity has been increasing, suggesting that a critical body weight[58] or body composition[59] may be an important factor in determining the timing and progression of puberty.[60] Conversely, the suppressive effects of fasting[61, 62] and chronic malnutrition[63] on the neuroendocrine control of reproduction are well known and consistent with the hypothesis. The manner in which such metabolic signals might be communicated and integrated with the reproductive axis in primates is unknown, but studies in ungulates suggest an effect on pulsatile hypothalamic GnRH secretion.[64]

Leptin

Leptin is produced by adipocytes and serum concentrations are strongly associated with body fat and changes in body fat content. Not surprisingly, leptin has been implicated as one way in which metabolic signals might be communicated to the higher centers controlling the activity of the hypothalamic pulse generator at the onset of puberty.

Leptin-deficient mice and rats fail to enter puberty, and treatment with leptin induces the onset of puberty.[65] In humans, serum leptin concentrations in boys and girls diverge at puberty. In males, leptin levels first increase then decrease again to prepubertal concentrations, whereas in females leptin concentrations rise throughout puberty.[66, 67] One study found that serum leptin levels were directly related to the amount of subcutaneous fat and inversely related to androgen levels.[68] Another in girls observed that an increase in the mean serum leptin concentration to 12.2 ng/mL, corresponding to 29.7% body fat and a body mass index (BMI) of 22.3, was associated with a decrease in age at menarche, and that a 1 ng/mL increase in serum leptin lowered the age at menarche by 1 month.[69]

Evidence from studies in children with congenital leptin deficiency has provided insights into the potential importance of leptin as a somatic stimulus for the onset of puberty. In affected pubertal age children, treatment with recombinant leptin has been associated with endocrine changes consistent with the onset of puberty, whereas all adults with congenital leptin or leptin receptor deficiency described have had severe hypogonadotropic hypogonadism.[70] However, similar treatment in younger children has not induced premature puberty.[71] These clinical observations suggest that leptin plays an important, but only permissive, role in the onset of puberty. Nonetheless, they are consistent with the idea that a circulating somatic hormone might have the ability to influence or modulate the activity of the hypothalamic GnRH pulse generator.[17]

Other Candidate Metabolic Signals

Numerous other metabolic signals have been suggested as playing a role in the nutritional regulation of reproduction, such as insulin, ghrelin (the endogenous ligand of the growth hormone secretagogue with a putative role in energy balance)[72] galanin-like peptide (a potential neuronal target of leptin),[73] and free fatty acids.[74] However, the manner in which these signals might interact with inhibitory and excitatory hypothalamic neurotransmitters and peptides and any role they may have in the onset of puberty remain to be established.

The Physiology of Normal Puberty

Although the timing, sequence, and pace of pubertal maturation vary among individuals, the sentinel events of puberty generally follow a predictable pattern. ***Adrenarche*** describes the activation of adrenal androgen secretion that begins before puberty and ultimately stimulates ***pubarche***, the appearance of pubic hair. ***Gonadarche*** describes the activation of the hypothalamic-pituitary-gonadal axis, which facilitates the pubertal growth spurt, stimulates ***thelarche***, the appearance of breast tissue, and finally ***menarche***, the onset of menses.

Adrenarche

Adrenarche is the term used to describe the increase in adrenal androgen production that begins at approximately 6 years of age in both boys and girls.[75, 76] ***Although adrenarche is independent of the maturation of the hypothalamic-pituitary-gonadal axis, the two often are temporally related.***[77] The increase in adrenal androgen production results from a change in the adrenal response to adrenocorticotropic hormone (ACTH) stimulation, characterized by a shift towards increased production of Δ^5-3β-hydroxysteroid intermediates (17α-hydroxpregnenolone; dehydroepiandrostendione, DHEA) and decreased production

of Δ^4-ketosteroids (17α-hydroxprogesterone, 17-OHP; androstendione), with no change in cortisol secretion.[78] Consequently, an increase in serum DHEA-sulfate (DHEA-S) levels heralds the onset of adrenarche. ***Generally, the best indicator of adrenarche is a serum DHEA-S concentration greater than 40 µg/dL, which is higher than that normally seen in children 1–5 years of age (5–35 µg/dL).***

Adrenal androgens derive from the zona reticularis, the innermost layer of the adrenal cortex,[79] which begins to form at approximately 3 years of age and becomes well defined coincident with the increase in DHEA-S production at adrenarche. The molecular mechanisms that govern adrenal cortical differentiation involve a variety of different genes, but little is known about their transcriptional regulation.[80] The zona reticularis exhibits a unique enzymatic profile. The activity of 3β-hydroxysteroid dehydrogenase/Δ^5-Δ^4 isomerase (3β-HSD), which catalyzes the oxidation and isomerization of Δ^5-3β-hydroxysteroid precursors into Δ^4-ketosteroids, is low.[81] In contrast, the activities of P450c17, including both 17α-hydroxylase (catalyzing the conversion of pregnenolone to 17α-hydroxpregnenolone) and 17,20 lyase (catalyzing the conversion of 17α-hydroxpregnenolone to DHEA), are high, as is steroid sulfotransferase activity.[82] Cytochrome b5, which facilitates 17,20 lyase activity, also is preferentially expressed.[83] Taken together, the enzyme profile of the zona reticularis favors the formation of DHEA and DHEA-S.

The primary stimulus for adrenarche is unknown. Although ACTH is an obvious candidate, circulating levels of adrenal androgens change without any corresponding changes in ACTH or cortisol during fetal life, puberty, and with aging. In other conditions such as chronic disease, surgical stress, recovery from secondary adrenal insufficiency, and anorexia nervosa, changes in ACTH-induced cortisol secretion are not accompanied by any change in serum adrenal androgen concentrations.[84] Although derivatives of proopiomelanocortin (POMC, produced by pituitary corticotropes)[85] and other pituitary-dependent factors have been implicated,[86] conclusive evidence for an adrenal androgen stimulating hormone is lacking.[87, 88] Whatever the stimulus might be, it might act to spur the growth and differentiation of the zona reticularis, which may derive from cells originally contained within the "fetal zone" of the adrenal cortex having a unique enzymatic profile (described above). Alternatively, it might act by suppressing 3β-HSD,[81] or by stimulating the 17,20 lyase activity of P450c17.[82] Interleukin-6 has been implicated as a mediator because it is highly expressed in the zona reticularis and can stimulate DHEA secretion.[89] Leptin also has been implicated, because adrenarche coincides with the preadolescent increase in body fat[90] and leptin levels[91] and leptin stimulates 17,20 lyase activity.[92] However, the temporal linkage between adrenarche and increasing body fat also might result from a compensatory hyperinsulinemia or from activation of the growth hormone (GH)/insulin-like growth factor-1 (IGF-1) axis.[93]

The steady increase in adrenal androgen secretion after adrenarche ultimately stimulates ***pubarche***, the appearance of pubic hair, and also the development and activity of the pilosebacious unit, consisting of a hair follicle and sebaceous gland.[94] Adrenal androgen levels correlate with changes in bone density, suggesting they also may contribute to growth in cortical bone.[95] If adrenarche has any more fundamental role in the onset of puberty, the mechanism is unknown. ***Adrenarche generally precedes activation of the hypothalamic-pituitary-gonadal axis, or gonadarche, by approximately 2–3 years.*** The temporal relationship suggests that adrenal androgen secretion might stimulate the pubertal transition, but several lines of evidence indicate otherwise. First, premature adrenarche generally is not associated with an earlier onset of thelarche or menarche. Second, adrenarche occurs in those with congenital hypergonadotropic hypogonadism (e.g., gonadal dysgenesis) or hypogonadotropic hypogonadism (e.g., Kallmann syndrome). Third, gonadarche occurs in children with Addison disease (hypoadrenalism) treated with glucocorticoids. Finally, in children under age 6 with true precocious puberty, gonadarche precedes adrenarche.

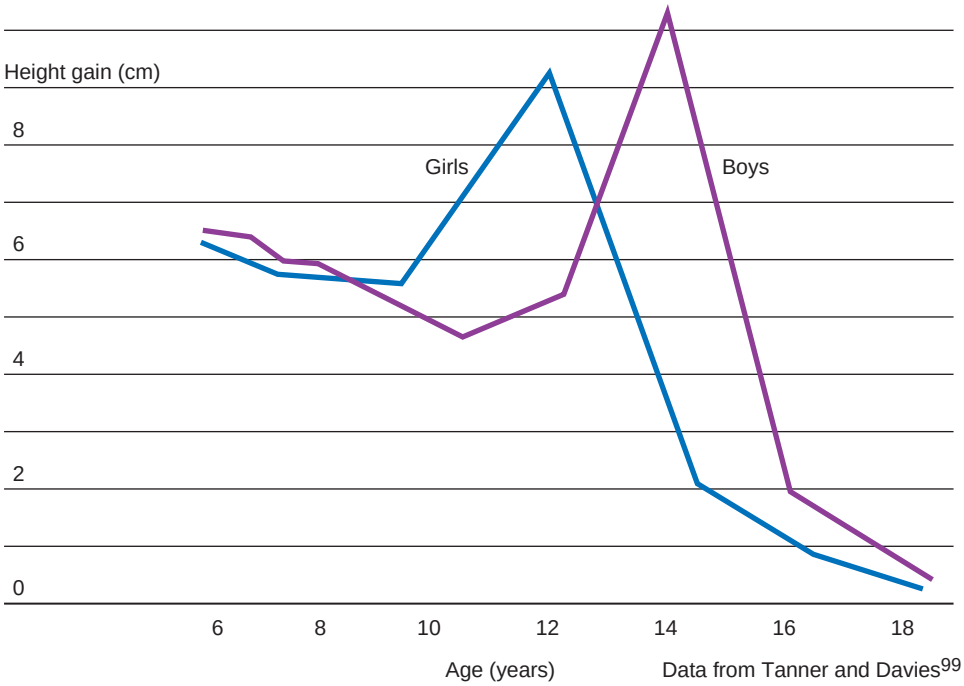
Growth

A substantial proportion of adult height, about 17–18%, is gained during puberty.[96] Growth of the limbs precedes that of the trunk, beginning in the distal portions and later also involving the proximal part of the limbs; growth of the trunk occurs primarily during later puberty.[97] ***The pubertal growth spurt occurs approximately 2 years earlier in girls than in boys and, in general, peak height velocity is reached approximately 6 months before menarche.***[98]

The difference in adult height of men and women relates to the onset and time of the growth spurt in boys and girls. Boys are approximately 2 cm taller than girls when girls reach their peak height velocity, grow 3–4 cm/year for another 1 to 2 years before entering puberty, and reach a greater peak height velocity (10.3 cm/year) than girls (9.0 cm/year).[99]

Bone mass accumulation during puberty is critical to the development of peak bone mass, which is a major determinant of the risk for developing osteoporosis in later life. Although genetics may be the most important determinant of peak bone mass, other factors such as nutrition and hormone exposure during puberty also contribute. About one-half of total body calcium is accrued during puberty in females, and one-half to two-thirds in males.[100, 101] ***In girls, the peak velocity in bone mineral accretion occurs at menarche, approximately 9–12 months after peak height velocity is attained.***[102, 103] The pubertal increase in bone density is greater in black females than in white females.[104] Taken together, these data suggest that the window of opportunity to maximize peak bone mass is relatively narrow.[101]

Weight changes during pubertal maturation reflect changes in body composition and the relative proportions of lean body mass and fat. Skinfold thickness decreases in early puberty and increases after peak height velocity, particularly in girls. Adolescent girls have more body fat than boys, most being deposited in the upper arms, thighs and back; the gender difference increases throughout puberty. The increase in BMI before 16 years of age relates primarily to changes in fat-free mass and thereafter to an increase in fat mass.[105] If desired, BMI can be tracked from adolescence to adulthood using published tables for comparison by age and ethnicity.[106]



Data from Tanner and Davies[99]

Growth Hormone

Growth hormone, produced by somatotropes, is the pituitary hormone produced in greatest abundance. The GH gene family includes five distinct genes, all of which are located on chromosome 17 (17q22).[107] The pituitary GH gene (*GH1*) encodes two alternatively spliced mRNAs, yielding a predominant 22 kDa GH molecule and another 20 kDa molecule that accounts for approximately 10% of circulating GH. Placental syncytiotrophblasts express a GH variant in addition to three other genes encoding human chorionic somatotropin, also known as human placental lactogen (discussed in Chapter 8). The regulation of pituitary GH secretion is highly complex. GH secretion is controlled primarily by hypothalamic GH-releasing hormone (GHRH) and by peripheral factors acting on somatotropes that stimulate (e.g., ghrelin),[108] or inhibit (e.g., somatostatin)[109] GH release.[110–112] Most of the peripheral actions of GH are mediated by insulin-like growth factor I (IGF-I), which inhibits GH release. Nutritional factors also play a role in the regulation of GH secretion; whereas fasting[113] and high protein meals[114] stimulate GH release, hyperglycemia and leptin inhibit GH secretion.[115] Estrogens stimulate, and excess glucocorticoids inhibit GH release. ***GH secretion peaks during puberty and declines with age, by approximately 50% every 7 years.***[113]

Like the gonadotropins, GH is secreted in a pulsatile fashion, and at the onset of puberty, GH pulse amplitude increases, especially during sleep.[116] Consequently, GH concentrations rise progressively. ***The rate of the increase in circulating GH levels is the most important determinant of the pubertal growth rate; slower growing children exhibit fewer and lower amplitude GH pulses and a more gradual increase in serum GH concentrations.***[117]

GH acts via binding to a specific receptor to stimulate hepatic synthesis and secretion of IGF-I, which promotes both growth and differentiation.[107] GH receptor mutations result in GH insensitivity and growth failure (Laron dwarfism);[118] in affected individuals, serum GH concentrations are elevated and levels of IGF-I are low. GH stimulates growth via direct and indirect (via IGF-I) actions on the epiphyseal plates of long bones. GH also has a number of metabolic actions, which include increased lipolysis, stimulation of protein synthesis, insulin antagonism, and water and sodium retention.

Insulin-like Growth Factor I

IGF-I is synthesized and secreted by the liver in response to GH stimulation and circulates in serum bound to high affinity IGF binding proteins (IGFBPs). The genes encoding IGF-I, IGF-II, and insulin all belong to the same family. The *IGF1* gene has several components and yields several different mRNAs, including the 6 kb form that is regulated by GH. IGF-I acts via its own receptor, which is widely distributed in a variety of tissues and organs.[119] IGF-I receptor concentrations are controlled by GH, thyroxine, and other growth factors such as fibroblast growth factor and platelet-derived growth factor. IGF-I acts via a complex signaling cascade to stimulate cell growth and to inhibit apoptosis.

The family of IGFBPs includes six proteins having greater affinities for IGF-I than the IGF-I receptor. IGFBPs are present in all extracellular fluids and serve both to transport IGF-I and to control the amount of IGF-I available to bind to the IGF-I receptor. IGFBP-3 is the most abundant in serum and has the highest affinity for IGF-I, but generally is saturated. Although present in lower concentrations, IGFBP-1 is unsaturated and therefore has greater impact on the levels of free IGF-I. The serum IGFBP-1 concentration is regulated by insulin, increasing during fasting when insulin levels are low, and decreasing after feeding or administration of insulin.[120] IGF-I levels are decreased in diseases associated with malnutrition such as inflammatory bowel disease and in hypothyroidism. IGF-I augments the effects of FSH and LH in the ovary, the effect of ACTH on adrenal steroidogenesis, and

the thyroid response to thyroid-stimulating hormone (TSH). ***IGF-I levels rise 7-fold from very low concentrations at birth to peak values at puberty, fall rapidly by approximately 50% by age 20, then decline slowly with advancing age.***[121]

Gonadal Steroids

The pubertal growth spurt is stimulated primarily by rising levels of GH and IGF-I, but a substantial body of evidence indicates that sex steroids also play an important role. In children with central (gonadotropin-dependent) precocious puberty treated with a long-acting GnRH agonist, mean height velocity and nocturnal serum GH and IGF-I levels, initially above the means for chronological age, decrease significantly after 6–12 months and remain suppressed for the duration of treatment.[122, 123] A study in children with central precocious puberty and GH deficiency (due to an intracranial lesion) observed that bone age was advanced in GH-deficient subjects, but not as much as in control subjects with precocious puberty and normal GH secretion; IGF-I levels were lower in GH-deficient subjects, but greater than in age-matched prepubertal GH-deficient children.[124] In a subset of GH-deficient subjects, treatment with a GnRH analog suppressed gonadal sex steroid levels and decreased height velocity, with no appreciable change in GH or IGF-I levels. In girls with Turner syndrome, treatment with exogenous estrogen increases growth velocity and bone age, compared to those observed in placebo-treated controls.[125, 126] ***Taken together, these observations indicate that the pubertal growth spurt is mediated, at least in part, by a sex steroid-induced increase in GH secretion. Moreover, they demonstrate that precocious puberty can induce a substantial growth spurt, even in the absence of a normal pubertal increase in circulating GH or IGF-I. However, normal pubertal growth requires the combined actions of sex steroids and GH. Ultimately, sex steroids limit adult height by stimulating epiphyseal fusion.***

The Timing of Puberty

What triggers the onset of puberty remains one of the most compelling unanswered questions in reproductive endocrinology. The age at onset of puberty and menarche is influenced by genetics, overall health, social environment, and environmental exposures.

An analysis of two genome-wide association studies including more than 17,000 women from the Nurses' Health Study and the Women's Genome Health Study identified 10 common variants or single nucleotide polymorphisms (SNPs) clustered in the regions of chromosomes 6q21 and 9q31.2 that were associated with age at menarche.[127, 128] Genetic variation in or near the locus (6q21) of the *LIN28B* gene (encoding a developmentally regulated RNA binding protein)[129] has been associated with age at menarche in a number of human populations.[128, 130, 131] Some genetic variants associated with adult height also have been associated with age at menarche, suggesting the association between height and age at menarche has a genetic basis.[128, 130] Other genes associated with age at menarche include *FTO* (fat mass and obesity associated gene) and *NEGR1* (neuronal growth regulator 1), both of which also are associated with childhood obesity.[130] Children with a family history of early puberty are more likely to experience an early puberty themselves; age at menarche correlates relatively well between mothers and daughters and between sisters.[132]

Children who live closer to the equator, at lower altitudes, in urban areas, and mildly obese children generally begin puberty earlier than those who live in northern latitudes, at higher elevations, in rural areas, and those of normal weight. Accumulating evidence suggests that

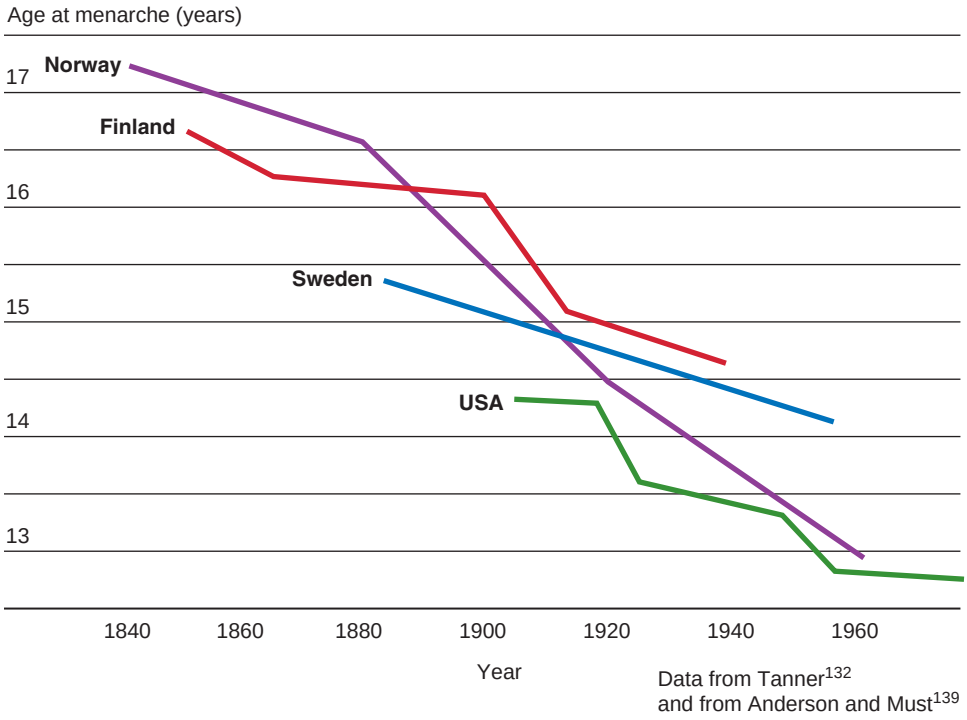
certain environmental toxicants acting as "endocrine disruptors" also may influence the timing of sexual development.[133]

The age at onset of puberty has been declining gradually in the general population of the United States over the past century. Although the rate of decrease has slowed considerably more recently, the trend has continued. Overall, the average age at menarche for American girls decreased from approximately 12.75 years in the 1960s to approximately 12.5 years in the early 1990s.[134, 135] A 1997 study conducted by the Pediatric Research in Office Settings (PROS) network examined the timing of pubertal development in more than 17,000 American girls (90% white, 10% black) and found that the earliest signs of puberty were occurring at ages significantly younger than in the past, with striking racial differences, as follows:[136]

Pubertal Milestone	*Black American Girls*	*White American Girls*
Thelarche		
Mean age	8.9 yr	10.0 yr
Age 6	6.4 %	2.9 %
Age 7	15.4 %	5.0 %
Age 8	37.8 %	10.5 %
Age 9	62.6 %	32.1 %
Age 10	80.2 %	61.5 %
Age 11	96.0 %	85.4 %
Age 12	98.9 %	96.0 %
Pubarche		
Mean age	8.8 yr	10.5 yr
Age 6	9.5 %	1.4 %
Age 7	17.7 %	2.8 %
Age 8	34.3 %	7.7 %
Age 9	62.6 %	20.0 %
Age 10	85.6 %	46.4 %
Age 11	95.2 %	74.3 %
Age 12	98.9 %	92.2 %
Thelarche and/or Pubarche		
Age 6	14.3 %	3.7 %
Age 7	27.2 %	6.7 %
Age 8	48.3 %	14.7 %
Age 9	77.4 %	38.2 %
Age 10	94.6 %	67.9 %
Age 11	98.4 %	88.0 %
Age 12	100.0 %	96.6 %
Menarche		
Mean age	12.2 yr	12.9 yr
Age 9	2.7 %	0.2 %
Age 10	6.3 %	1.8 %
Age 11	27.9 %	13.4 %
Age 12	62.1 %	35.2 %

These data indicate that a substantial proportion of American girls begin pubertal development 6–12 months earlier than previously observed. ***On average, black American girls begin puberty between ages 8 and 9, and white American girls by age 10. However, thelarche and/or pubarche can occur normally in black girls as early as age 6 and in white girls as early as age 7.***

Subsequent studies analyzing data from the National Health and Nutrition Examination Survey (NHANES) have observed a 2.3 month decrease in the average age of menarche between surveys for the years 1988–1994 (12.53 years) and 1999–2002 (12.34 years), and an overall 4.9 month decrease since 1960.[134, 137–139] The decrease in age of menarche has been observed in all ethnic groups, declining from 12.57 to 12.52 years in non-Hispanic white girls, from 12.09 to 12.06 years for non-Hispanic black girls, and from 12.24 to 12.09 for Hispanic American girls.[139] Changes in the population distribution of race and ethnicity over time explain the larger change in overall average age compared to those within groups.



Historically, the trend to an earlier onset of sexual development has been attributed to improved nutrition and less stressful living conditions.[140] The age at menarche has declined as the prevalence of obesity has increased, suggesting that a critical body weight[58] or body composition[59] is an important factor in determining the onset and progression of puberty.[60] ***Indeed, higher weight and body fat mass are associated with an increased likelihood of early menarche.***[60, 134, 141–143] Data from the U.S. NHANES survey indicated that a girl with a BMI at the 85th percentile is more than twice as likely to have reached menarche as a girl of the same age and race/ethnicity having a BMI at the 50th percentile.[139] However, girls reach menarche over a wide range of weight and BMI, and age at menarche cannot be predicted reliably for individuals on that basis.[144] ***Importantly, early pubertal development is associated with slightly decreased adult height and an increased risk for obesity, compared to a late menarche.***[145, 146]

Average Ages of Pubertal Milestones in Different Populations[147]

Country	Year(s)	Subjects	Ages	Thelarche	Pubarche	Menarche
Chile	2000	758	6–16	8.9	10.4	12.7
China	1993	3,749	7–19	9.8	11.6	12.4
Denmark	1991–1993	1,100	6–20	10.9	11.3	13.4
Egypt	2000	1,550	6–18	10.7	10.5	12.4
England	1960–1970	192	3–19	11.2	11.7	13.5
India	1988–1991	9,951	5–18	10.2	—	12.6
Iran	2003–2004	1,420	6–17	9.7	10.5	12.7
Italy	1998–2001	1,642	6–15	10.5	10.6	11.9
Japan	1990–2000	832	6–14	9.7	—	12.2
Korea	1993–1995	4,237	14–20	—	—	12.5
Lithuania	1999–2000	1,231	7–12	—	10.2	11.7
Netherlands	1996–1997	3,028	8–20	10.7	11.0	12.9
Spain	2000	266	8–10	10.7	—	12.4
Thailand	1997–1999	300	9–19	9.4	11.1	11.2
Turkey	2005	1,562	6–16	10.2	10.6	12.4
U.S.	1992–1993	15,439 white	3–12	9.9	10.5	12.9
		1,638 black		8.9	8.8	12.2

Stages of Pubertal Development

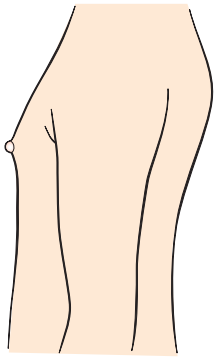
Puberty includes a series of predictable events that vary in timing, sequence, and pace. ***In general, the first sign of puberty in most adolescent girls is an acceleration of growth, followed by breast budding (thelarche), the appearance of pubic hair (pubarche), and finally, the onset of menses (menarche).***

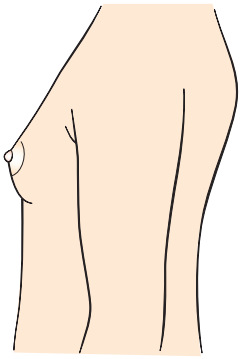
The staging systems used most frequently to describe the physical changes of puberty were first described by Marshall and Tanner in 1969 (girls)[148] and 1970 (boys).[149] The Tanner stages describe secondary sexual characteristics, including breast development in girls, pubic hair growth in both sexes, and genital development in boys. As shown in the diagrams, there are five Tanner stages of breast and pubic hair development in girls, with stage 1 representing the prepubertal state and stage 5 representing adult development.

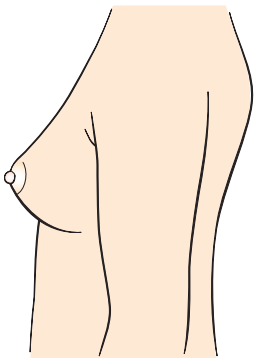
Breast development follows a recognized sequence of events. Breast budding (Tanner stage 2) is distinguished by enlargement and by widening of the areolae. The breast then enlarges, becoming elevated beyond the areolae (Tanner stage 3). The breast enlarges further and the areolae and nipple form secondary mounds (Tanner stage 4), just before the breast achieves an adult contour (Tanner stage 5).

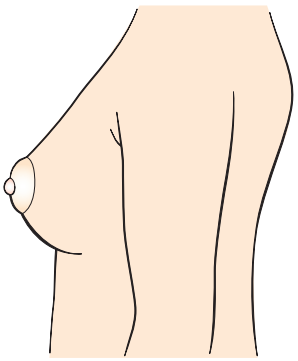
In the majority of adolescents, pubarche closely follows thelarche, but in a substantial minority the sequence is reversed and pubarche precedes thelarche. In either case, the two are closely linked and progress in parallel. Pubarche (Tanner stage 2) is distinguished by the emergence of a small amount of long, relatively straight hair on the labia majora. The hair then becomes curly, coarser, and extends outward (Tanner stage 3). Hair extends further to cover the labia (Tanner stage 4) before assuming an adult pattern with extension onto the medial thigh (Tanner stage 5).

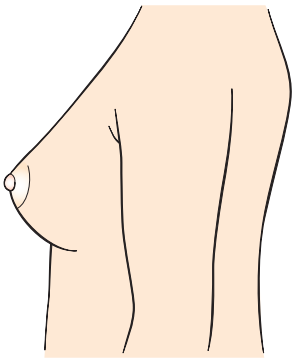
Menarche occurs an average of 2.6 years after the onset of puberty and after the peak of growth has passed.[98, 102, 148] ***On average, the pubertal sequence of accelerated growth, thelarche, pubarche, and menarche requires a period of 4.5 years (range, 1–6 years).*** The relationship between menarche and the growth spurt is relatively fixed. After menarche, growth slows and generally does not increase more than about 6 cm (2.4 inches). The menses immediately following menarche usually are anovulatory, irregular, and occasionally heavy. Anovulatory cycles frequently persist for as much as 12–18 months and are not

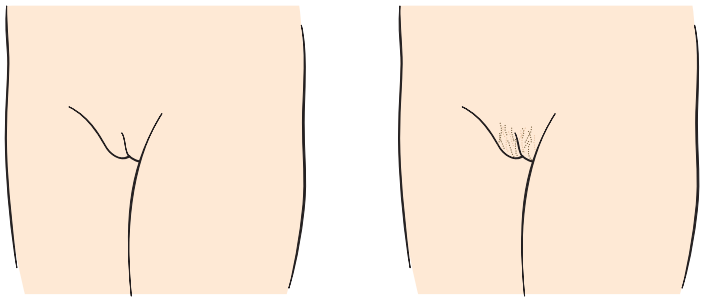

1 Prepubertal


2 Breast bud


3 Breast elevation

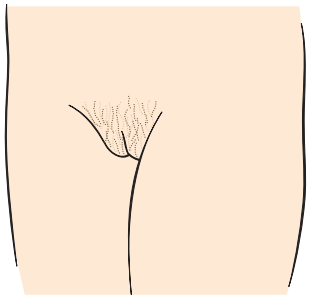

4 Areolar mound


5 Adult contour

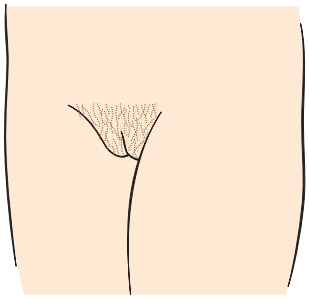


1 Prepubertal

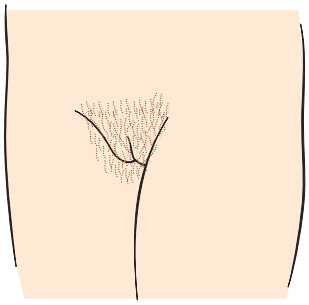
2 Presexual hair



3 Sexual hair

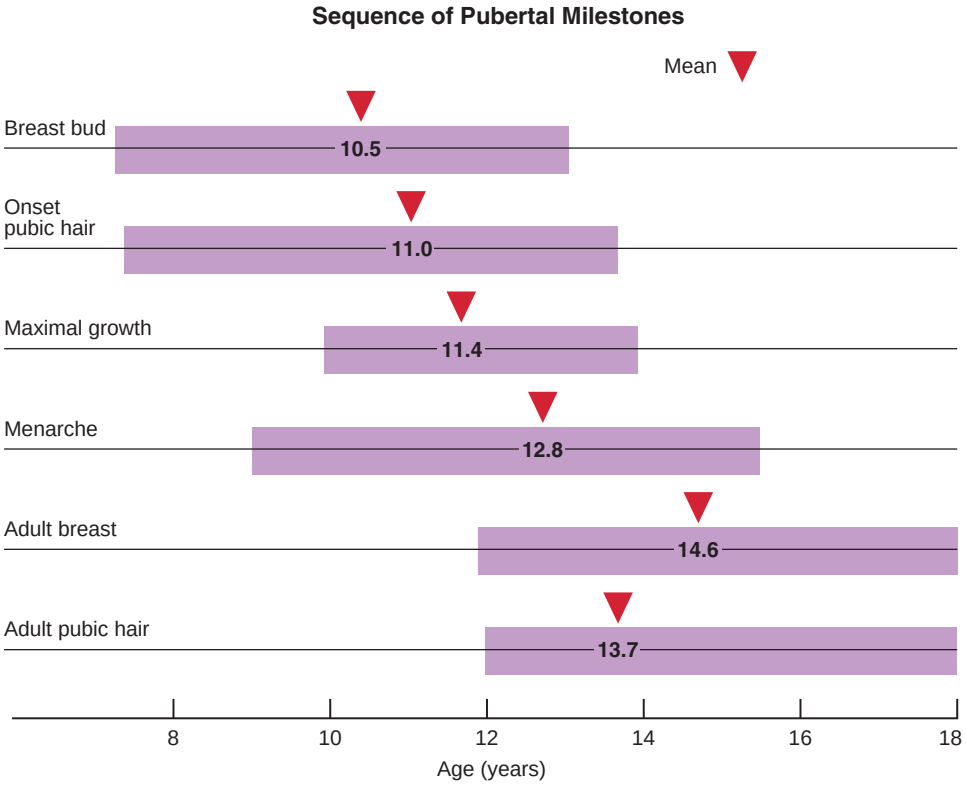


4 Mid-escutcheon



5 Female escutcheon

uncommon even up to 4 years after menarche.[150, 151] However, the frequency of menses generally increases rapidly over the first year after menarche; 65% of adolescent girls report having 10 or more periods per year at the end of the first postmenarcheal year, and 90% after 3 years.[30] The hallmark of the maturation of the hypothalamic-pituitary-ovarian axis and of the completion of puberty is the development of estrogen positive feedback, which stimulates the midcycle LH surge and ovulation. In general, ovulatory cycles become progressively more frequent. The time required to establish ovulatory cycles relates to the age at menarche; when menarche occurs after the age of 13, only one-half have ovulatory cycles within 4.5 years.[152]



Summary of Pubertal Events

The onset of puberty is an evolving sequence of maturational steps. The hypothalamic-pituitary-gonadal axis differentiates and develops during fetal life and becomes fully functional before birth. Beginning in late infancy and continuing through childhood, the axis is inactive because pulsatile secretion of hypothalamic GnRH is suppressed to very low levels of activity by central inhibitory mechanisms and, to a lesser extent, by a high sensitivity to low levels of gonadal steroid feedback.

In late childhood, adrenal androgen production increases (adrenarche), in response to an unknown stimulus, rises steadily thereafter, and ultimately stimulates the growth of pubic and axillary hair (pubarche). After a decade of quiescence, pulsatile GnRH secretion increases and the hypothalamic-pituitary-gonadal axis is reactivated (gonadarche), probably in response to metabolic signals from the periphery. FSH and LH levels rise moderately before age 10, followed by a gradual increase in estradiol concentrations, which stimulates breast development (thelarche). The increase in pulsatile gonadotropin secretion occurs first at night, during sleep, but gradually extends throughout the day.

The rapid increase in skeletal growth (the pubertal growth spurt) that precedes thelarche and pubarche is mediated primarily by a sex steroid-induced increase in GH secretion, which, in turn, stimulates increased production of IGF-I, and, to a lesser extent, by the direct effects of increasing sex steroid concentrations. Ultimately, rising sex steroid levels limit adult height by stimulating epiphyseal fusion.

At midpuberty, gonadal estrogen production increases sufficiently to stimulate endometrial proliferation, ultimately resulting in the onset of menses (menarche). Postmenarcheal cycles are at first anovulatory. Gradually, as the estrogen positive feedback mechanism matures, ovulatory cycles increase in frequency and, in late puberty, become established.

Serum Hormone Concentrations During Female Puberty[153–159]				
Tanner	*FSH*	*LH*	*Estradiol*	*DHEA-S*
Stage	*IU/L*	*IU/L*	*pg/mL*	*ng/dL*
Stage 1	0.9– 5.1	1.8– 9.2	<10	19– 302
Stage 2	1.4– 7.0	2.0–16.6	7– 37	45–1,904
Stage 3	2.4– 7.7	5.6–13.6	9– 59	125–1,730
Stage 4	1.5–11.2	7.0–14.4	10–156	153–1,321
Adult Follicular	3.0–20.0	5.0–25.0	30–100	162–1,620

Common Problems Associated with Puberty

Some of the common physical manifestations of pubertal maturation may be viewed by patients or their parents as abnormal, including anemia, acne, psychosocial problems, myopia, and dysfunctional uterine bleeding. Adolescents also are at high risk for sexually transmitted infections.

Girls tend to eat less of foods high in iron content, such as meat, and a low heme iron intake increases the risk of low iron stores.[160] The third National Health and Nutrition Examination Survey (1988–1994) observed a 9% prevalence of anemia among American girls between the ages of 12 and 15 years.[161]

Acne is a disorder of the pilosebaceous unit caused by androgen stimulation characterized by follicular occlusion and inflammation. During puberty, the number of acneiform lesions increases, at all stages.[162] In girls, acne tends to be more severe in the late stages of puberty, which are associated with higher serum concentrations of DHEA-S.[163]

The psychosocial changes during puberty predispose to a increased incidence of depression, which is twice as common in girls as in boys.[164] Many girls become unhappy with their physical appearance, resulting in a decrease in self-esteem that is more common in white girls than in black girls.[165] The problem is most common when pubertal development is not synchronous with that of peers.[166] Early maturing girls are more likely to develop psychopathology,[167] to have older friends,[168] and to be vulnerable to peer pressures.[169]

The prevalence of myopia (nearsightedness), caused by growth in the axial diameter of the eye, is greatest during puberty. Dysfunctional uterine bleeding is a consequence of anovulatory cycles and is common in adolescent girls during the first year or two after menarche.

Adolescents represent the highest-risk age group for nearly all sexually transmitted infections (STIs).[170] The risk for acquiring an STI relates to age at first intercourse, the number of partners, the perceived risk, and attitudes about acquiring an STI.[171] The persistence of a columnar epithelium on the exocervix (ectropion) also may predispose to infection with Chlamydia[172] and human papillomavirus.[173, 174]

Precocious Puberty

Precocious puberty describes pubertal development that begins at an earlier age than expected, based on established normal standards. Its causes are many, ranging from variants

of normal development, such as premature adrenarche, to serious pathology, including malignant intracranial neoplasms. Children with precocious puberty warrant careful evaluation to define the cause and, when indicated, prompt treatment to avoid the psychosocial and growth consequences of abnormally early sexual development.

Indications for Evaluation

Abnormally early or precocious puberty generally is defined as pubertal development occurring more than 2.5 standard deviations earlier than the average age. Traditionally, using 10 years as the average age of onset of puberty in girls, precocious puberty has been defined as secondary sexual development before the age of 8 years. However, as discussed in an earlier section of this chapter,[134, 135] the age at onset of puberty has been declining over the past few decades, raising questions about when pubertal development should be considered precocious and warrants clinical evaluation.

The 1997 study conducted by the Pediatric Research in Office Settings (PROS) network observed that 6.7% of American white girls and 27.2% of black girls had breast or pubic hair development before the age of 8 years.[136] These observations suggested that continued application of the traditional definition of precocious puberty would result in a large number of potentially normal girls having extensive, costly, and unnecessary testing. Consequently, new guidelines were proposed, lowering the age at which evaluation is indicated to age 7 in white girls and age 6 in black girls,[175] sparking a vigorous debate.

Some authorities questioned the new recommendations because the PROS study population was not a random sample drawn from the general population, because the study focused on premature thelarche and premature adrenarche and not on "true" precocious puberty (typically characterized by early breast *and* pubic hair development), and because no cause was determined in those having precocious puberty. Criticism centered on concerns that the recommended lower ages in the newly proposed guidelines might increase significantly the risk for under-diagnosis of important endocrine pathology.[176–179] Indeed, a subsequent large European study involving 443 girls with central precocious puberty identified 35 with an occult intracranial lesion (8%) and reported that application of the revised American guidelines lowering the age for evaluation for precocious puberty would have missed 4/35 girls (11%) with cranial pathology.[180, 181] Others favoring the new recommendations emphasized that fewer than 2% of girls with precocious puberty over age 6 had an intracranial lesion and that unnecessary imaging has high financial and emotional costs.[182]

In another American study involving 223 patients (white girls ages 7–8 years and black girls ages 6–8 years) referred to a single tertiary center solely for evaluation of precocious puberty over a 5-year period, 105 (47%) exhibited both breast and pubic hair development, 83 (37%) had only pubic hair, 24 (11%) had only breast development, and 11 (5%) had no signs of early sexual development.[179] Ultimately, 186/212 (88%) with signs of early puberty had a diagnosis of idiopathic gonadotropin-dependent precocious puberty and 26 (12%) had a treatable endocrinopathy amenable to early intervention, including acanthosis nigricans/hyperinsulinemia, hypothyroidism, neurofibromatosis, GH deficiency, pituitary adenoma, McCune-Albright syndrome, and congenital adrenal hyperplasia. More importantly, more than one-third of the girls with only breast development or with breast and pubic hair development had bone ages that were advanced significantly and were, therefore, at risk for diminished growth potential.[179]

Clearly, more and larger prospective studies are needed because questions regarding the indications for evaluation for precocious puberty remain unsettled. However, taking all of the available data into consideration, we believe that the following guidelines offer a good balance between safety and cost-effectiveness:

All girls under the age of 6 who have either breast or pubic hair development and girls under age 8 having both breast and pubic hair development merit a thorough evaluation to determine the cause.

Girls under 8 years of age having only early breast development (premature thelarche) or pubic hair growth (premature adrenarche or pubarche) warrant a careful history and physical examination and, at a minimum, an evaluation of bone age and close follow-up to determine their linear growth rate in efforts to identify those who may be at risk for decreased growth potential.

Between the ages of 6 and 8 years, clinicians must make individual judgments regarding the extent of evaluation, based on the results of the initial evaluation and, inevitably, on the level of anxiety in the patient and her parents.[9]

Factors associated with an increased risk for intracranial pathology that clearly warrant complete evaluation and imaging include onset of puberty before age 6, rapid pubertal progression, and associated symptoms of headache, seizures, or focal neurologic deficits.[175, 182]

Classification of Precocious Puberty

Traditionally, precocious puberty has been classified according to the underlying pathophysiology. However, the classification has limited practical utility in clinical practice because it reflects the final diagnosis, after evaluation is completed.

Gonadotropin-dependent precocious puberty, also known as "central precocious puberty" or "true precocious puberty," describes early maturation and activation of the hypothalamic-pituitary-gonadal axis and is characterized by both breast and pubic hair development in girls, and by pubic hair development and testicular enlargement (>4 mL in volume or 2.5 cm in diameter) in boys. The early developing sexual characteristics are "isosexual," meaning they are consistent with the child's gender.

Gonadotropin-independent precocious puberty, also known as "peripheral precocious puberty" or "pseudo-precocious puberty," describes early sexual development that is independent of GnRH and gonadotropins and generally results from exposure to sex steroid hormones that derive from the gonads, the adrenals, or the environment. Gonadotropin-independent precocious puberty is further sub-classified as isosexual when sexual characteristics are consistent with gender, and as "contrasexual" when inconsistent with gender (virilization in girls, or feminization in boys).

Incomplete precocious puberty describes children with isolated premature thelarche or premature adrenarche. Both usually represent variants of normal pubertal development, but some will progress to complete precocious puberty that may be gonadotropin-dependent or independent.

Gonadotropin-Dependent Precocious Puberty

Gonadotropin-dependent precocious puberty results from early maturation of the hypothalamic-pituitary-gonadal axis and is much more common in girls than in boys.[183] Although puberty begins earlier than normal, the sequence of pubertal events generally is normal and proceeds at the normal pace.

Up to 90% of children with gonadotropin-dependent precocious puberty have no identifiable cause (idiopathic), a diagnosis made by exclusion.[184, 185] However, the disorder can be associated with a variety of central nervous system lesions, including tumors, irradiation, hydrocephalus, cysts, trauma, inflammatory diseases, and midline developmental defects such as septo-optic dysplasia. ***Consequently, head magnetic resonance imaging (MRI) is indicated even when there are no neurological signs or symptoms.***[180, 185, 186]

Tumors associated with gonadotropin-dependent precocious puberty include hamartomas, astrocytomas, ependymomas, pineal tumors, and optic and hypothalamic gliomas. Hamartomas are heterotopic neuronal masses containing GnRH neurons that typically attach to the tuber cinereum or floor of the third ventricle where they can function as an ectopic hypothalamic GnRH pulse generator, divorced from the central inhibitory mechanisms that normally restrain activity during childhood; they are the most common tumor associated with precocious puberty and can be associated with gelastic seizures (laughing, giggling)[187, 188] some produce transforming growth factor alpha, which mediates release of GnRH.[189] The precocious puberty that can be observed in children with neurofibromatosis usually relates to an optic glioma.[190]

As described in an earlier section of this chapter, activating mutations in the gene encoding the GPR54 receptor (*KISS1R*), which mediates the actions of kisspeptin (an excitatory neuroregulator of GnRH secretion) can cause gonadotropin-dependent precocious puberty.[52]

Children exposed to high circulating androgen or estrogen concentrations, as may occur with congenital adrenal hyperplasia, virilizing tumors, and the McCune Albright syndrome, often exhibit early maturation of the hypothalamic-pituitary-gonadal axis, which then results in gonadotropin-dependent precocious puberty.[191–193]

Although quite rare, girls with severe primary hypothyroidism can present with precocious puberty, exhibiting breast development, galactorrhea, and episodic menstrual bleeding. In most cases, the very high serum levels of TSH, which has structural similarity to FSH, appear to activate the FSH receptor.[194]

Rarely, gonadotropin-dependent precocious pubertal development has resulted from an autonomous pituitary gonadotropin-secreting tumor rather than from early maturation of the hypothalamic-pituitary-gonadal axis.[195, 196]

Gonadotropin-Independent Precocious Puberty

Gonadotropin-independent precocious puberty can result from excess sex steroids secretion from the gonads or adrenals or from exposure to exogenous estrogens.

Autonomous functional ovarian follicular cysts are the most common cause of gonadotropin-independent precocious puberty in girls. Transient breast development and vaginal bleeding are the most common presentation, which can be an isolated event or recur at unpredictable intervals.[197–199] Serum estrogen levels typically are elevated, but not always (due to regression of the cyst), and both basal and GnRH-stimulated gonadotropin concentrations are low. In most cases, bone age is not advanced. Ovarian ultrasonography usually demonstrates one or more unilateral or bilateral ovarian cysts greater than 15 mm in diameter.[200] The disorder is self-limited in most and requires no treatment. However, recurrent cysts resulting in prolonged or repeated estrogen exposure can precipitate early maturation of the hypothalamic-pituitary-gonadal axis, resulting in gonadotropin-dependent precocious puberty.[198] Autonomous ovarian cysts also can be an early manifestation of McCune-Albright syndrome, arising before emergence of the characteristic skin (“café-au-lait spots”) or bone lesions; affected patients therefore require careful longer-term follow-up.[197, 199]

Ovarian tumors are rare causes of gonadotropin-independent precocious puberty in girls and include granulosa cell tumors, Leydig cell tumors and gonadoblastomas.[201–203]

McCune-Albright syndrome is a rare disorder characterized classically by precocious puberty, café-au-lait skin pigmentation, and polyostotic fibrous dysplasia of bone, all caused by a somatic mutation of the alpha subunit of the G-protein (encoded by the GNAS1 gene), which results in a mosaic distribution of cells bearing constitutively active adenylate cyclase.[204–206] The mutation results in continuous stimulation of endocrine function and, in addition to precocious puberty, also can cause gigantism, Cushing syndrome, adrenal hyperplasia, and thyrotoxicosis, in varying combinations. Although precocious puberty is the most common clinical manifestation,[207] the phenotype varies with the tissues that are affected by the mutation and can include hepatitis, intestinal polylps, and cardiac arrhythmias. As in other forms of gonadotropin-independent precocious puberty, the sequence of pubertal development may be abnormal; for example, vaginal bleeding frequently precedes breast development.[208] The skin and bone lesions can increase over time and may not be present at the initital presentation. Early and repeated exposure to elevated sex steroid levels can result in accelerated growth, advanced bone age, and reduced adult height; it also may induce early maturation of the hypothalamic-pituitary-gonadal axis, resulting in secondary gonadotropin-dependent precocious puberty. McCune-Albright syndrome is more common in girls than in boys. The diagnosis merits consideration in girls presenting with recurrent functional ovarian follicular cysts and episodic menses.[209] Partial forms of the syndrome also have been described.[206]

Adrenal pathology, such as androgen-secreting tumors and congenital adrenal hyperplasia, is another cause of gonadotropin-independent precocious pubertal development.

Exposure to exogenous estrogens or environmental pollutants having estrogenic activity (xenoestrogens) can result in premature sexual development in infants or toddlers.[210–212] Examples include accidental exposure to estrogens, xenoestrogens, or placental extracts contained in cosmetics or personal hair and skin care products and environmental pollutants that can act as endocrine disruptors by mimicking estradiol, such as polychlorinated biphenyls, herbicides, pesticides, and plasticizers, which may be found in water contaminated with industrial products.[213] Serum hormone levels in affected children typically are in the normal range, but can vary widely depending on the nature, time, and frequency of use or exposure. ***Children are extremely sensitive to the effects of estrogen and may respond with increased growth or breast development even at serum levels below the limits of detection.***[214]

Incomplete Precocious Puberty

Incomplete precocious puberty includes premature adrenarche or premature thelarche and usually is a variant of normal puberty. Such cases present a clinical dilemma due to uncertainty regarding whether the condition is entirely benign, as usual, or might be the first indication of true precocious puberty.

Premature Adrenarche

Premature adrenarche is the most common cause of premature pubarche, describing otherwise unexplained early growth of genital hair associated with increased levels of adrenal androgens.[215] ***Generally, the best indicator of adrenarche is a serum DHEA-S concentration greater than 40 µg/dL, which is higher than that normally seen in children 1–5 years of age (5–35 µg/dL).*** In children with premature adrenarche, the growth rate and bone age often are above average but still within normal ranges. ***Exaggerated adrenarche*** is the term used to describe the clinical extreme of premature adrenarche, wherein the serum DHEA-S

level exceeds that typical of adrenarche or for age and usually, but not always, is associated with a somewhat early onset of true puberty.[216]

The cause of premature adrenarche is unknown. The condition traditionally has been considered an early variant of normal development and, as such, generally has no serious consequences. However, up to 20% of girls with premature adrenarche may subsequently develop gonadotropin-dependent precocious puberty and close follow-up therefore is recommended.[77, 217] Other evidence indicates that girls with premature adrenarche also are at increased risk for developing polycystic ovary syndrome, suggesting that premature adrenarche may be an early manifestation of the disorder.[218–223] In many, premature pubarche is preceded by low birth weight and is followed by hyperandrogenism, hirsutism, and oligomenorrhea in adolescence, often accompanied by hyperinsulinemia and dyslipidemia. These observations suggest that insulin resistance may be the underlying metabolic disorder, causing decreased growth during fetal life, premature pubarche, and hyperandrogenism that worsens during late puberty or the early postmenarcheal years.[224] In those affected, metformin treatment can decrease insulin resistance and hyperandrogenism, improve the lipid profile, often restore cyclic menses, and may help to prevent later development of diabetes and cardiovascular disease.[225]

Premature pubarche usually results from a premature adrenarche but also has other causes. Idiopathic premature pubarche, unassociated with any demonstrable increase in adrenal androgen production, probably reflects an increased sensitivity of hair follicles to normal androgen concentrations. Premature pubarche sometimes can be the only clinical manifestation of a mild form of congenital adrenal hyperplasia (CAH).[226, 227] Other rare causes of ACTH-dependent childhood virilization include Cushing syndrome, glucocoticoid resistance, cortisone reductase deficiency, and androgen-producing neoplasms of the adrenal gland or ovary.

The evaluation of premature pubarche should focus first on determining whether the growth of pubic hair is an isolated phenomenon or may be associated with other signs and symptoms suggesting another of the diagnoses mentioned above. ***The single most important and useful test is an x-ray of the left hand and wrist for bone age. If sexual hair is small in amount and slow-growing and bone age is normal, precocious puberty is unlikely and expectant management is appropriate, with re-evaluation at 6 months and periodically thereafter.*** A limited endocrine evaluation should include measurements of serum testosterone and DHEA-S, for comparison to age-adjusted normal values. The presumptive diagnosis of premature adrenarche can be made when both are appropriate for pubarche, bone age is normal, and predicted adult height is within the range expected for the family.[228] More extensive endocrine evaluation can be reserved for those children having other signs suggesting true precocious puberty or a virilizing disorder.

An ACTH stimulation test to exclude the diagnosis of CAH is indicated when bone age is advanced abnormally, the predicted adult height is abnormally low, or when the serum testosterone and DHEA-S concentrations are elevated above the ranges typical of premature adrenarche. The test is performed by obtaining blood samples before and 60 minutes after administering cosyntropin (synthetic ACTH 1–24; 1 $\mu g/m^2$ or 0.25 mg). A stimulated serum 17-OHP concentration greater than 1,000 ng/dL generally indicates 21-hydroxylase deficiency.[229] In children with premature pubarche, diagnosis of the rare 3β-HSD deficiency requires a stimulated 17α-hydroxpregnenolone level greater than 9,790 ng/dL.[227]

Premature adrenarche is a benign condition and requires no specific treatment. Parents can be reassured that the condition is a normal variant relating to increased sensitivity of hair follicles to low levels of androgen, or an early occurring incomplete form of puberty. However, children with a diagnosis of premature adrenarche merit periodic re-evaluation for evidence of progressive virilization.

Premature Thelarche

Premature thelarche generally is defined as isolated breast development in girls before the age of 8 years. In girls, premature thelarche usually is a benign condition considered a variant of normal puberty. Early breast development is particularly common during the first year of life when the hypothalamic-pituitary-gonadal axis is still active.[230, 231] Studies using ultrasensitive bioassays for estrogen have detected higher estrogen levels in many, but not all, girls with premature thelarche than in normal controls.[232] The breast also may be more sensitive to estradiol than normal in some girls.[233] Although most affected children subsequently experience normal puberty and growth,[234–236] a significant proportion experiences an earlier than average menarche.[237]

Physical examination typically reveals a light pink areola with an infantile appearance, with Tanner stage 2 or 3 breast development; frequently, the change may be unilateral or asymmetrical. There is no sign of androgen exposure.

Exaggerated thelarche describes those with premature thelarche who also exhibit increased growth velocity and/or advanced bone age and may represent an intermediate between premature thelarache and precocious puberty.[232, 238] Even girls with exaggerated thelarche exhibit a prepubertal pattern of response to acute stimulation with GnRH or a GnRH agonist; FSH levels rise more than LH, which remains below 5 IU/L.[239] Some cases have been associated with the presence of functional ovarian cysts.[197] Genetic studies in girls with exaggerated thelarche have revealed that some harbor a mutation in the *GNAS1* gene, suggesting that the disorder can be an early or the only sign of McCune-Albright syndrome.[232, 240]

Premature thelarche also has been related to exposure to exogenous estrogen, including environmental chemicals that degrade slowly in the environment and can accumulate in the food chain, but no clear relationship with premature thelarche has been established.

The evaluation of premature thelarche, like that of premature adrenarche, should focus on determining whether breast development is an isolated phenomenon or associated with other signs of precocious puberty; here again, the most important initial test is an evaluation of bone age. ***In children with Tanner stage 2 breast development and normal bone age, precocious puberty is unlikely and expectant management is appropriate, with re-evaluation at 6 months and periodically thereafter.***

Approximately 15–20% of girls with premature thelarche subsequently develop gonadotropin-dependent precocious puberty, at a mean age of 7.1 ± 0.7 years and mean bone age of 9.0 ± 1.1 years.[241, 242] A longitudinal study involving more than 150 girls with premature thelarche observed that 69% had complete regression of breast development (13% of these later developing true precocious puberty), 21% had recurrent episodes of breast development (32% later developing true precocious puberty), and 10% had persistent breast development (57% later developing true precocious puberty).[242]

Evaluation of Precocious Pubertal Development

The evaluation of early sexual development begins with a careful history and physical examination and measurement of bone age to determine whether there is any corresponding increase in linear growth. Subsequent evaluation is limited to those with precocious puberty and is aimed at determining the cause and at directing treatment.

The ***medical history*** should determine when the physical change(s) were first noticed, in the siblings and parents as well as in the patient, seek evidence of growth acceleration, exclude previous history of neurological disease or trauma or exposure to sex steroids, and identify any associated symptoms of headache, seizures, or abdominal pain.

The ***physical examination*** should include height, weight, and calculation of growth velocity (cm/year), which often is an early indication of evolving precocious puberty.[243] A fundoscopic examination should be performed to detect papilledema, a sign of increased intracranial pressure. Evaluation of visual fields may reveal evidence to suggest a sellar mass lesion. A careful examination of the skin should be performed to identify any café-au-lait spots, which suggest the diagnosis of McCune-Albright syndrome.

Tanner staging of pubic hair and/or breast development should be performed. The diameter of the glandular breast tissue should be measured, taking care to distinguish it from adipose. Accurate assessments are important for determining whether additional evaluation is warranted.

A measurement of ***bone age*** is indicated when examination demonstrates signs of early sexual development.

Tanner Staging

	Breast	*Public Hair*
Stage 1(prepubertal)	Elevation of papilla only	No public hair
Stage 2	Elevation of breast and papilla as small mound, increased areola diameter	Sparse, long, pigmented hair, primarily on labia majora
Stage 3	Further enlargement without separation of breast and areola	Dark, coarse, curled hair sparsely distributed over mons
Stage 4	Secondary mound of areola and papilla above the breast	Adult-type hair, abundant but limited to the mons
Stage 5	Recession of areola to contour of the breast	Adult-type hair, extending onto the medial thigh

Endocrine Evaluation and Imaging

Children with advanced bone age and those having normal bone age accompanied by both breast and pubic hair development, or normal bone age with evidence of accelerated growth and breast or pubic hair development, warrant further endocrine evaluation and imaging.

Basal and GnRH-stimulated serum gonadotropin levels differentiate gonadotropin-dependent from gonadotropin-independent precocious puberty, which then guides further evaluation. Serum gonadotropin concentrations should be measured using ultra-sensitive assays having low detection limits for pediatric patients (approximately 0.1 IU/L).[244–246]

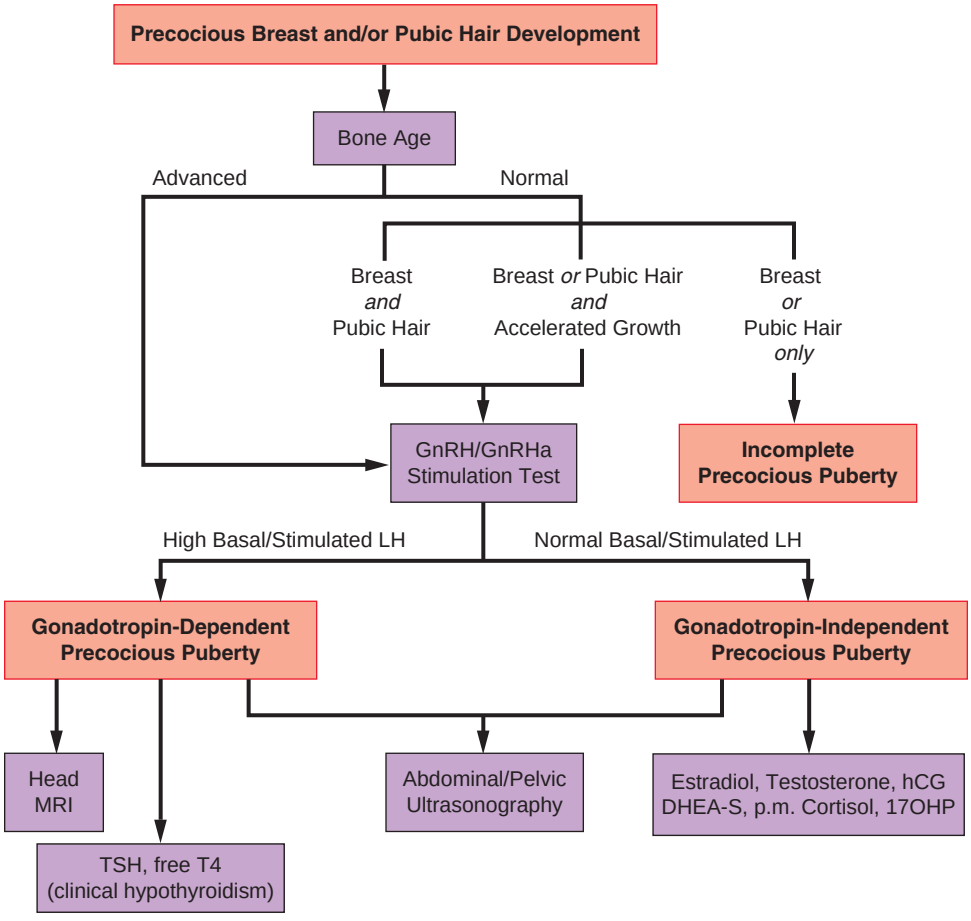
The ***GnRH stimulation test*** is performed by obtaining blood samples before and 30–40 minutes after a single dose of GnRH (100 μg), administered intravenously. Because synthetic GnRH currently is not available in the United States, a GnRH agonist can be used instead,[247–249] obtaining blood samples before and 60 minutes after a single dose of leuprolide acetate (20 μg/kg), administered subcutaneously.[249] ***The stimulated serum LH concentration is the most useful diagnostic parameter; although a normal threshold value has not been firmly established, due to differences in assay methods and the limited amount of normative data, a stimulated LH value of 3.3–5.0 IU/L defines the upper limit of normal for prepubertal children (Tanner stage 1, T1) with most assays.***[249] Both basal and stimulated serum LH concentrations have high specificity and positive predictive value for diagnosis of gonadotropin-dependent precocious puberty. In a study comparing the results of GnRH stimulation tests performed in normal children (T1) with those obtained in

children with gonadotropin-dependent and gonadotropin-independent precocious puberty, the mean basal serum LH concentration was 1.6 IU/L in the group with gonadotropin-dependent precocious puberty and less than 0.6 IU/L in the other two groups. The mean stimulated LH value in the group of children with gonadotropin-dependent precocious puberty was 21.6 IU/L, compared to 3.2 IU/L in normal children (T1) and 1.4 IU/L in the group with gonadotropin-independent precocious puberty.[250]

In children with gonadotropin-dependent precocious puberty (as identified by elevated basal or stimulated serum LH levels), a head MRI is indicated to exclude an intracranial mass lesion.[249,251] Thyroid function tests (TSH and free T4) should be obtained if there is any clinical evidence of hypothyroidism.

In children with gonadotropin-independent precocious puberty (as identified by normal basal and stimulated serum LH levels), serum concentrations of estradiol, testosterone and hCG (functional ovarian cysts and tumors, functional adrenal tumors), late afternoon cortisol (Cushing syndrome), DHEA-S (premature adrenarche), and 17OHP (congenital adrenal hyperplasia) should be obtained to determine the peripheral source of sex steroid production and the cause of early sexual development.

Abdominal and pelvic ultrasonography is indicated in all girls with precocious puberty to identify functional ovarian cysts or tumors. Ultrasonography is indicated even for those with gonadotropin-dependent precocious puberty because early and repeated or sustained exposure to sex steroids from autonomous peripheral sources can induce a secondary premature maturation of the hypothalamic-pituitary-gonadal axis.



Treatment of Precocious Puberty

The treatment of precocious puberty differs according to whether it is gonadotropin-dependent or gonadotropin-independent and on the underlying cause, when that can be determined. ***The principal goals of treatment are to stop or slow development until normal pubertal age, to maximize adult height, and to reduce the risk of psychosocial problems associated with early sexual maturation.***

Treatment of Gonadotropin-Dependent Precocious Puberty

The decision to treat gonadotropin-dependent precocious puberty depends on the underlying pathology and on the speed of sexual development. In those having an identified intracranial lesion, treatment should be directed to the lesion, if that is possible. In those having no intracranial lesion, the decision to treat should be based primarily on the pace of progression and on the estimated adult height.

Treatment for gonadotropin-dependent precocious puberty generally is indicated when sexual maturation progresses to the next stage within 3–6 months, when growth velocity is accelerated to greater than 6 cm/year (unless peak height velocity has already passed), when bone age is advanced by 1 year or more, or when predicted adult height is below the target range or is decreasing on serial determinations.[251] Conversely, those with stable or regressing pubertal signs, normal growth velocity (for age), bone age within 1 year of chronological age, and a predicted adult height within the target range may not require treatment. In most cases, growth velocity should be monitored for 3–6 months before making the decision to treat.[251] Treatment aimed only at avoiding the potential psychosocial consequences of precocious puberty should be carefully considered because there are few data regarding outcomes and effectiveness.

GnRH Agonist Treatment

Long-acting GnRH agonists have proven both safe and effective for the treatment of idiopathic gonadotropin-dependent precocious puberty.[252–259] GnRH agonist treatment causes a brief initial "flare" of gonadotropin release, followed by pituitary desensitization (exhaustion of available stores of releasable gonadotropins), and down-regulation (decrease in GnRH receptors). By suppressing the pituitary-gonadal axis, GnRH agonist therapy can prevent progressive pubertal development, and increase final adult height, compared to pretreatment predictions. ***Young children and those who exhibit rapidly progressive development can be expected to have early epiphyseal fusion, are at greatest risk for compromised adult height, and can benefit most from treatment.***[249]

In girls under 6 years of age with idiopathic gonadotropin-dependent precocious puberty, treatment with a GnRH agonist can be expected to add 9–10 cm to adult height. In older children already past their peak with slowing growth velocity, treatment can be expected to slow it further, to delay epiphyseal fusion, and to yield slow but steady increases in predicted adult height. In girls between 6 and 8 years of age, GnRH agonist treatment typically results in a gain of 4–7 cm in height, less if bone age is significantly advanced.[249] Girls already close to the age of normal puberty, those with slowly progressive maturation, and girls with a predicted height above 150 cm have less to gain and may not benefit significantly from treatment.[260–262]

The choice among the available GnRH agonist formulations depends mostly on physician preference and availability. Depot preparations generally are preferred because of improved compliance. Direct comparisons in randomized trials have not been made,

but any of the following treatment regimens generally can be expected to suppress the pituitary-gonadal axis:[263–265]

- Buserelin 6.3 mg every 2 months
- Goserelin 3.6 mg every month, or 10.8 mg every 3 months
- Histrelin 50 mg implant every year
- Leuprolide 3.75–7.5 mg monthly, or 11.25 mg every 3 months
- Triptorelin 3.0–3.75 mg monthly, or 11.25 mg every 3 months

Nonetheless, the dose of GnRH agonist treatment required can vary significantly.[266] Inadequate treatment can permit progressive sexual development and bone maturation. Conversely, over-treatment can suppress endogenous GH and decrease growth velocity and bone mineral accumulation to levels below those normally expected during the prepubertal years.[267] The adequacy of GnRH agonist treatment can be monitored simply by measuring the serum LH concentration 30–60 minutes after each repeated injection of the agonist; the LH level should be less than 3.0 IU/L, consistent with prepubertal norms after acute GnRH agonist stimulation.[268]

GnRH agonist treatment should be monitored at 3–6 month intervals with serial physical examinations to detect any progressive pubertal development; bone age also should be evaluated periodically.[249] Breast development should cease and growth velocity and the pace of advancing bone age should decrease. Pubic hair development may continue due to normal adrenarche.[269] Although bone density may decline during longer durations of treatment, bone mass is regained after treatment ends and peak bone mass is normal; consequently, there is no reason or need to monitor bone density.[249]

Treatment with GnRH agonists does not appear to have any significant long-term adverse effects on function of the hypothalamic-pituitary-gonadal axis.[270] ***It can be continued until the epiphyses are fused or until the pubertal and chronological ages are appropriately matched. Prompt reactivation of the pituitary-gonadal axis and pubertal development, in a pattern similar to that in normal adolescents, generally follows the discontinuation of treatment.***[271]

GnRH agonist therapy also is recommended for treatment of GnRH-secreting hypothalamic hamartomas[187, 272]; the tumor can be monitored by serial imaging and risky surgery can be avoided. Treatment for other hypothalamic, pituitary, cerebral, or pineal tumors must be individualized. Many that are small and do not extend around or into vital structures can be excised successfully.

Treatment of Gonadotropin-Independent Precocious Puberty

The treatment of gonadotropin-independent precocious puberty is aimed at the underlying pathology. Girls with functional tumors involving the ovaries or adrenals are treated surgically; hCG-secreting tumors also may require adjunctive radiation or chemotherapy, depending on the type and location of the tumor. Solitary unilateral functional ovarian cysts also can be excised surgically. Children with congenital adrenal hyperplasia should receive treatment with glucocorticoids, and those with McCune-Albright syndrome generally are best treated with drugs that inhibit steroidogenesis or hormone action rather than surgery, to preserve fertility.

McCune-Albright Syndrome in Girls

In girls with McCune-Albright syndrome, treatment can be aimed at blocking aromatization and estrogen production, but available evidence indicates that aromatase inhibitors,

such as fadrozole, letrozole, and anastrozole, tend to lose their effectiveness over time.[273–276] The alternative is to block the effects of estrogens by treatment with an antiestrogen such as tamoxifen, which has been used successfully for treatment of associated vaginal bleeding.[277] Bisphosphonate treatment can be useful in treatment of the fibrous dysplasia of bone that causes pain and fractures.[278] Those who develop a gonadotropin-dependent component to their precocious development, due to chronic premature exposure to sex steroids, may benefit from adjunctive treatment with a GnRH agonist, as in children with idiopathic gonadotropin-dependent precocious puberty.[208]

Management of Incomplete Precocious Puberty

Although girls with isolated premature thelarche or premature adrenache do not require treatment, they do merit regular examinations to detect other emerging evidence of precocious sexual development that may signal the need for further evaluation and possible treatment.

Delayed Puberty

Delayed puberty is defined by absent or incomplete sexual maturation by the age at which 95% of children of the same sex has started pubertal development. ***In the United States, breast development, the usual first sign, begins by the age of 12 years in more than 95% of girls.***[136] Delayed puberty results from hypogonadism, which, in turn, can result from an inactive hypothalamic-pituitary axis (hypogonadotropic hypogonadism) or from primary gonadal failure (hypergonadotropic hypogonadism).

The most common cause of hypogonadotropic hypogonadism is a functional GnRH deficiency, reflecting a constitutional delay in the reactivation of the hypothalamic-pituitary-gonadal axis, or the suppressive effects of chronic stress due to illness, malnutirition, or excessive exercise. GnRH deficiency also can result from genetic defects (e.g., Kallmann syndrome) or anatomical abnormalities (e.g., hypothalamic and pituitary tumors). Other causes of hypogonadotropic hypogonadism include pituitary failure, hypothyroidism, and hyperprolactinemia. Hypergonadotropic hypogonadism can result from idiopathic primary gonadal failure, from previous treatment of malignancy (gonadectomy, chemotherapy, gonadal radiation) or from a variety of congenital and genetic abnormalities or syndromes.

The distribution of diagnostic frequencies among girls with delayed puberty is shown in the table below, representing the findings in a series of 326 patients.[279] The series included all girls who were referred for evaluation of delayed pubertal milestones, including some with only primary amenorrhea relating to müllerian or other developmental anomalies or androgen insensitivity syndrome, who did not have true delayed puberty. In a subsequent study involving 74 females 18 years old and younger (mean age 14 ± 1.4 year) referred to a tertiary center solely for the evaluation of delayed puberty, the final diagnosis was constitutional delay in 22 (30%), functional hypogonadotropic hypogonadism (chronic illness, eating disorders, excessive exercise) in 14 (19%), irreversible hypogonadotropic hypogonadism (genetic causes, CNS tumors) in 15 (20%), hypergonadotropic hypogonadism (previous chemotherapy, gonadal radiation, congenital and genetic abnormalities) in 19 (26%), with 4 patients (5%) left unclassified.[280]

Relative Frequency of Delayed Pubertal Abnormalities[279]		
Hypergonadotropic Hypogonadism		**43.0%**
Ovarian failure, abnormal karyotype		26.0%
Ovarian failure, normal karyotype		17.0%
46,XX	15.0%	
46,XY	2.0%	
Hypogonadotropic Hypogonadism		**31.0%**
Reversible		18.0%
Physiologic delay	10.0%	
Weight loss/anorexia	3.0%	
Primary hypothyroidism	1.0%	
Congenital adrenal hyperplasia	1.0%	
Cushing syndrome	0.5%	
Prolactinoma	1.5%	
Irreversible		13.0%
GnRH deficiency	7.0%	
Hypopituitarism	2.0%	
Congenital CNS defects	0.5%	
Other pituitary adenomas	0.5%	
Craniopharyngioma	1.0%	
Malignant pituitary tumor	0.5%	
Eugonadism		**26.0%**
Müllerian agenesis		14.0%
Vaginal septum		3.0%
Imperforate hymen		0.5%
Androgen insensitivity syndrome		1.0%
Inappropriate positive feedback		7.0%

Evaluation of Delayed Pubertal Development

The initial evaluation of delayed puberty begins in the same way as that for precocious puberty, with a careful history, physical examination, and a measurement of bone age.

The ***medical history*** should determine whether pubertal development has not yet started, or began and then stopped. A careful evaluation of the previous growth pattern can provide important clues.[281] ***Those with constitutional delay typically exhibit delayed growth, adrenarche, and sexual development, associated with declining growth velocity and delayed skeletal maturation.*** Other important historical factors include dietary and exercise habits, previous serious illnesses, and medications that might delay the onset or slow the pace of pubertal progression.[282]

Delayed puberty can be among the first clinical indications of an underlying metabolic disorder, such as inflammatory bowel disease or hypothyroidism. Neurologic symptoms, including headache, visual disturbances, anosmia, dyskinesia, seizures, and mental retardation, suggest a CNS disease or disorder. Anosmia suggests strongly a genetic cause,

such as a *KAL1*, *FGF8*, *FGFR1*, *PROK2*, or *PROKR2* gene mutation (all associated with different forms of Kallmann syndrome). ***A complete family history, with emphasis on the age at pubertal milestones in older siblings and parents, also provides useful information; in most patients with constitutional delay, other family members have a similar history.***[280]

The ***physical examination*** should include height, weight, arm span, and Tanner staging of breast and pubic hair development. Height should be compared to norms for age and for bone age, and then carefully monitored for at least 6 months. A eunochoid body habitus (arm span exceeds height by ≥ 5cm) suggests delayed epiphyseal closure due to hypogonadism. In the presence of breast budding (Tanner stage 2), a normal spontaneous puberty generally can be expected and both the patient and family can be reassured. Congenital malformations such as midline defects and skeletal abnormalities (clift lip/palate, scoliosis) suggest congenital GnRH deficiency resulting from genetic mutations involving elements of the fibroblast growth factor signaling pathway. As in patients with precocious puberty, a fundoscopic examination should be performed to detect papilledema and visual fields should be evaluated.

A measurement of ***bone age*** should be obtained for comparison with chronological age and for assessment of the potential for future growth. ***Patients with constitutional delay of puberty typically exhibit a bone age between 12 and 13.5 years, which generally does not progress further without the exposure to gonadal steroids that is required for epiphyseal closure.***

Laboratory Evaluation and Imaging

The laboratory evaluation of girls with delayed puberty is aimed first at differentiating primary (hypergonadotropic) from secondary (hypogonadotropic) hypogonadism, which typically can be accomplished by measuring the serum FSH, LH, and estradiol concentrations.

By mid-adolescence, gonadotropin levels, particularly FSH, are grossly elevated in girls with primary gonadal failure.[27] In patients with hypogonadism, low basal gonadotropin levels are consistent with the diagnosis of constitutional delay of puberty, but also with congenital GnRH deficiency or pituitary gonadotropin deficiency. Ultra-sensitive immunofluorometric assays for FSH and LH may help to distinguish the low but detectable concentrations typically observed in those with constitutional delay from the undetectable levels in patients with congenital GnRH deficiency, but these assays have not yet been widely validated for use in patients who are truly GnRH deficient.[23] ***GnRH agonist stimulation testing generally is not helpful or necessary.*** Whereas some have found that stimulation with a GnRH agonist (buserelin, nafarelin, triptorelin) can successfully discriminate constitutional delay of puberty from congenital GnRH deficiency in boys,[283–285] similar studies have not been conducted in girls with delayed puberty. Consequently, after excluding other causes, time and serial observations may be required to establish the correct diagnosis.

When the estradiol level is clearly low, a serum FSH level in the low normal range has the same interpretation and clinical implication as a frankly low FSH concentration. If the hypothalamic-pituitary-ovarian axis were intact and functioning normally, the FSH level should be high when estrogen levels are grossly low; therefore, a "normal" value is abnormally low in that clinical context and indicates hypothalamic-pituitary suppression or dysfunction. Moreover, although the level of immunoreactive FSH may be normal, the level of biologically active FSH may not be, because patients with hypogonadotropic hypogonadism may secrete gonadotropins having altered patterns of glycosylation and reduced biological activity.[286]

Further laboratory evaluation is directed toward determining the cause of hypogonadotropic or hypergonadotropic hypogonadism, once that is established.

Hypogonadotropic Hypogonadism

In girls with hypogonadotropic hypogonadism, measurement of the serum ***prolactin*** concentration is indicated to identify those with hyperprolactinemia, which can cause either delayed or arrested pubertal development, depending on when it arises. Hyperprolactinemia can result from excessive secretion by a pituitary lactotrope adenoma, from any other hypothalamic or pituitary tumor or disorder that interrupts the normal delivery of hypothalamic dopamine via the tuberoinfundibular tract, or from medications that interfere with the actions of dopamine. ***Therefore, hyperprolactinemia is an indication for imaging by MRI, except when it can be attributed confidently to medications.***

Measurement of the serum ***TSH and free thyroxine (T4)*** concentrations also is indicated to identify those who may have primary or secondary hypothyroidism, particularly if growth velocity has slowed and the bone age is grossly delayed.

The serum ***DHEA-S*** concentration may be helpful for distinguishing constitutional delay of puberty from congenital GnRH deficiency. Patients with congenital GnRH deficiency are more likely to have a normal adrenarche than those with constitutional delay, although values in the two groups frequently overlap.[287]

Other laboratory tests are aimed at identifying those who may have an occult chronic illness, such as chronic inflammatory bowel disease, liver disease, or anorexia nervosa, and should include a ***complete blood count, erythrocyte sedimentation rate, and liver function tests.***

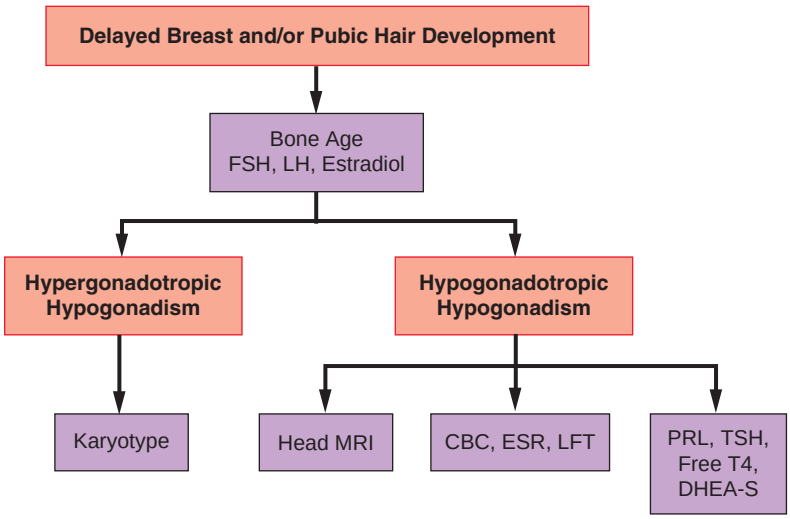
Although ***pelvic ultrasonography*** can be used to determine the presence or absence of a uterus in virginal girls, it must be interpreted cautiously because results can be misleading when the reproductive organs are immature and very small, and generally is unnecessary. Whereas müllerian anomalies are a common cause of primary amenorrhea, they are not associated specifically with delayed puberty.

A head MRI should be obtained in patients with hypogondotropic hypogonadism and those with with neurologic signs or symptoms. In addition to detecting mass lesions, imaging can reveal the presence or absence of the olfactory bulbs and tracts (absent in Kallmann syndromes).

Hypergonadotropic Hypogonadism

A karyotype should be obtained in all girls with hypergonadotropic hypogonadism to detect chromosomal abnormalities, except when a history of previous chemotherapy or gonadal radiation provides an obvious explanation. The most common disorder of this type is gonadal dysgenesis, with Turner syndrome (45,X) being the prototype. In addition to other structural X chromosome abnormalities (e.g., deletions, rings and isochromosomes), karyotype will identify those harboring a Y chromosome (e.g., 46,XY, Swyer syndrome), in whom gonadectomy will be indicated due to the significant risk for malignant transformation in occult testicular elements (20–30%).

In patients with hypergonadotropic hypogonadism and a normal (46,XX) karyotype, the diagnostic possibilities include 17α-hydroxylase deficiency, a rare steroidogenic enzyme defect associated with sexual infantilism and hypertension, and other uncommon causes of primary ovarian failure, all of which are discussed in detail in Chapter 11.



Treatment of Delayed Puberty

The first priority in the treatment of delayed puberty is to correct the specific cause, when that is possible, such as thyroid hormone therapy for hypothyroidism, dopamine agonist therapy for hyperprolactinemia, and excision of a craniopharyngioma or other operable central lesion. In those with no such identifiable cause, congenital GnRH deficiency must be distinguished from constitutional delay of puberty, but in most, a final diagnosis can be established only after serial observations; however, treatment options are the same in either case.

Patients with congenital GnRH deficiency or constitutional delay of puberty can be managed expectantly, providing reassurance and psychological support, or with hormone therapy, which may be appropriate for those with severe pubertal delay or serious psychosocial concerns that cannot be resolved with reassurance and education. ***In general, sex hormone therapy should be limited to girls over 12 years of age having few or no signs of sexual maturation causing significant distress or anxiety.***

The goals of short-term hormone therapy are to foster age-appropriate secondary sexual development and to induce a growth spurt and a normal adolescent increase in bone density without causing premature epiphyseal closure, which requires that bone age be monitored at 6-monthly intervals during treatment. In those who prove ultimately to have an isolated GnRH deficiency, the longer-term goals are to maintain sex hormone levels in the normal physiologic range and to induce ovulation with exogenous gonadotropin therapy when fertility becomes a priority. Oral or transdermal estrogen therapy can be used, beginning at doses well below those used for adults (e.g., 0.25–0.5 mg oral micronized estradiol or its equivalent), increasing gradually at intervals of 3–6 months according to response (Tanner stage, bone age), with the goal of completing sexual maturation over a period of 2–3 years. ***A progestin should not be added to the treatment regimen until there is substantial breast development and full contour breast growth has plateaued, because premature progestin treatment can adversely affect breast growth or contours.*** In general, progestin therapy can safely begin once menses have begun, or after 12–24 months of estrogen treatment. Once breast development has been accomplished and menses are established, hormone therapy can be discontinued for 1–3 months, at intervals, to observe whether spontaneous menses will begin, as can be expected in girls with constitutional delay of puberty. Persistent hypogonadism beyond 18 years of age clearly suggests congenital GnRH deficiency.

In general, GH therapy is best limited to those with documented GH deficiency. Serum GH and IGF-I levels typically are low in patients with constitutional delay of puberty but increase after treatment with estrogen, and usually are normal in those with congenital GnRH deficiency.

Growth Problems in Normal Adolescents

Perhaps the worst thing about an adolescent growth problem is that it makes the individual feel "different." It is probably true that, more than anyone else, the adolescent does not like to be different. Therefore, concerns over unusually short or tall stature deserve attention and should not be dismissed.

Growth in height is a continuous but not a linear process. There are three distinct phases of growth. The first is the infantile phase, which is characterized by rapid growth amounting to a total of 30–35 cm during the first 2 years of life. The next phase is the childhood phase, during which growth proceeds at a relatively constant pace of 5–7 cm/year, often slowing in late childhood. The last is the pubertal phase, which is characterized by a growth spurt at a rate of 8–14 cm/year, reflecting the effects of increasing levels of both GH and sex steroid hormones.[288, 289]

The contribution of heredity to final adult height is difficult to predict accurately, but a child's height potential can be estimated by calculating the ***midparental height;*** for girls, midparental height is calculated as follows:

$$\frac{(\text{father's height} - 13\text{ cm}) + (\text{mother's height})}{2}$$

Target height represents a range of heights encompassing the 3rd to 97th percentiles for expected adult height, equating with the midparental height ± 8.5 cm.[290] For children with delayed or accelerated growth, height should be adjusted to the appropriate percentile based on bone age, rather than chronologic age, to permit a more accurate judgment regarding whether growth is consistent with genetic potential.

The basic and essential laboratory test in the evaluation of perceived abnormal growth is a left hand/wrist x-ray for bone age. The Bayley-Pinneau tables (found at the end of this chapter) can be used to determine a ***predicted adult height***, based on current height and bone age, in reference to the Greulich-Pyle Atlas.[291] The predicted adult height is the number found where the column corresponding to the patient's current height intersects the row corresponding to her bone age. If bone age is within 1 year of chronologic age, the table for average girls should be used. If bone age is accelerated or delayed by 1 year or more, the tables for children with accelerated or delayed growth should be used, with one exception. Height predictions for girls with idiopathic GnRH-dependent precocious puberty are more accurate when using the table for average girls.[292]

Short Stature

Short stature is defined as height 2 or more standard deviations below the mean height for children of the same sex and chronologic age, as determined by plotting height on an appropriate growth chart. ***Accurate serial measurements of height and growth velocity***

are perhaps the most useful tool in the evaluation of children with growth failure.[293] ***The pattern of growth is more important than any single point measurement.*** Slowing growth that increasingly deviates from a previously defined pattern (percentile) is the key finding. ***Children should grow at a rate of at least 5 cm/year from age 4 year to the onset of puberty.***

The most common causes of short stature are familial (genetic) and constitutional delay of growth; both are characterized by a normal growth velocity. If growth velocity is abnormally low (<5th percentile for age), thorough evaluation for the many potential causes is warranted, as described below.

Short stature is a feature of a variety of chromosomal (Down syndrome, Turner syndrome) and other genetic disorders (Noonan syndrome, Russell-Silver syndrome), and commonly is associated with intrauterine growth restriction or infections and maternal exposures during pregnancy such as smoking and alcohol. Children with short stature relating to endocrine disorders such as Cushing syndrome, GH deficiency, and hypothyroidism usually are overweight for height. In contrast, those with malnutrition due to an eating disorder, excessive exercise, malabsorption or other systemic illness typically are underweight for height and merit evaluation for gastrointestinal,[294, 295] heart, pulmonary,[296] and renal disease.[297]

Idiopathic short stature describes children whose height is more than 2 standard deviations below the mean for age with no identifiable endocrine, metabolic, or other cause. Such children generally exhibit low normal growth velocity and have normal serum IGF-I levels. Those with genetic or familial short stature typically have normal bone age and a predicted adult height within the target range, whereas those with constitutional delay of growth have delayed bone age. Recent evidence has suggested that 2–15% of children with idiopathic short stature may have mutations in the *SHOX* (Short Stature Homeobox) gene, located at the distal tip of the short arm of the X chromosome (Xp22.33).[298–300] Affected children tend to have short forearms and lower legs, Madelung deformity of the forearm (a congenital subluxation or dislocation of the distal ulna), cubitus valgus (wide carrying angle of the arm), a high arched palate, and muscle hypertrophy, compared to those without such mutations.[301]

Evaluation

Although it is unlikely that a patient with congenital hypothyroidism will present undiagnosed and untreated as an adolescent, thyroid function always should be evaluated. Because both primary and secondary hypothyroidism can cause growth failure, both the serum TSH and free T4 concentration should be measured. Cushing syndrome (hypercortisolism) is rare in children, except when it results from excess glucocorticoid treatment. Children with congenital GH deficiency generally are not difficult to recognize, usually presenting as young children with severe growth failure, delayed bone age, and very low serum concentrations of IGF-I and its major binding protein, IGFBP-3. Provocative tests of GH secretion are required to establish the diagnosis.

The evaluation of children with short stature should include the following:

- Bone age
- A complete blood count and erythrocyte sedimentation rate
- Electrolytes, creatinine, bicarbonate, calcium, phosphate, alkaline phosphatase, albumin
- TSH, free T4, IGF-I (and IGFBP-3 in children under 3 years of age)
- Anti-endomysial antibodies (a serologic screen for celiac disease)
- Karyotype (to exclude Turner syndrome or other X chromosome abnormalities)

A head MRI is not required to establish the diagnosis of idiopathic short stature, but should be considered in children with known GH deficiency and those with signs or symptoms of hypothalamic-pituitary dysfunction.

Treatment

GH therapy for idiopathic short stature was approved by the United States Food and Drug Administration in 2003 and is considered indicated for girls whose height is more than 2.25 standard deviations below the mean for age whose epiphyses are not closed and whose predicted adult height is less than 59 inches. ***However, GH treatment of children with idiopathic short stature is controversial, because the response to therapy is unpredictable and typically quite modest,[302–305] and because evidence that short stature has significant psychosocial consequences is lacking.[306–308]***

Most children with idiopathic short stature, particularly those with constitutional delay of growth, exhibit catch-up growth during puberty without treatment.[302, 309] Moreover, the average increase in height is only approximately 4–6 cm after more than 5 years of treatment.[304] The optimal age for starting treatment is between age 5 years and early puberty.[310]

GH treatment has relatively few potential adverse effects.[302, 304, 311] Although high-dose treatment has been reported to advance the onset of puberty and epiphyseal closure,[312] lower doses do not.[313] Current evidence suggests that the impact of treatment correlates with serum IGF-I levels and that treatment can be optimized when the dose of GH is adjusted to maintain a normal IGF-I concentration.[310] ***However, GH therapy generally should be limited to children whose short stature is a significant disability and whose self-image and socialization are judged likely to improve significantly with an increase in height.[314, 315]*** The associated costs and potential benefits must be weighed carefully because the costs of GH therapy are extremely high. In the United States, the cost has been estimated to exceed $50,000 per inch gained in adult height![316]

An alternative approach to treatment involves the use of a long-acting GnRH agonist to delay pubertal development and epiphyseal fusion. However, the modest impact of such treatment (ranging up to only 4 cm in increased height) comes at the cost of a substantial decrease in bone mineral density accretion.[317] Although treatment with an aromatase inhibitor might also seem logical, such therapy actually slows growth in girls via profound inhibition of estrogen production.

Tall Stature

Tall stature is defined as height 2 or more standard deviations above the mean height for children of the same sex and chronological age. Although tall stature is nearly as common as short stature, it is perceived as more socially acceptable and less commonly perceived as a problem. Most children with tall stature, like those with short stature, represent the extremes of a normal distribution of heights and only a few have a specific growth abnormality.[318]

Abnormally rapid growth during childhood and adolescence can result from precocious pubertal development, GH excess,[319] hyperthyroidism,[320] sex hormone deficiency or insensitivity,[321, 322] or rare autosomal recessive disorders such as familial glucocorticoid deficiency[323] or resistance[324] and congenital total lipodystrophy.[325] Patients with Marfan syndrome,[326] homocystinuria,[327] and neurofibromatosis type I also can be unusually tall.[328]

The diagnosis of familial or constitutional tall stature generally is established by family history and the absence of dysmorphic features, distinguishing it from disorders of excessive growth. ***In most tall but otherwise normal children, a careful family history, physical examination, and bone age generally are all that is required to establish the diagnosis and to provide reassurance.*** Serial measurements of growth at 6–12-month intervals can help to confirm that growth is in the high normal range but not excessive. The Bayley-Pinneau tables (found at the end of this chapter) can be used to predict adult height in tall girls and become more accurate after the age of 12 years; in younger children they may tend to overestimate adult height.[329]

Treatment

As in otherwise normal short children and adolescents, the treatment of tall children and adolescents is controversial, generally discouraged, and should be limited to those whose tall stature is the cause of significant psychosocial problems.[330–332]

Sex steroids have been used to treat tall girls and boys for decades, the goal being to promote early epiphyseal fusion.[333, 334] ***The earlier treatment is started, the greater the likelihood that adult height will be diminished.*** The stage of secondary sexual development is relevant because the adolescent growth spurt precedes menarche and treatment must begin before menarche to be optimally effective.[318] Although that implies that treatment could begin as early as age 8 or 9 years, the usual age to begin treatment is between 10 and 12 years. However, treatment that begins after menarche can still achieve up to an inch of growth reduction.[335, 336] A typical starting dose is 15–30 μg of ethinyl estradiol, which can be administered in a low-dose oral contraceptive pill. Treatment should continue until the epiphyses are closed, as can be determined by serial measurements of bone age at 6–12-month intervals during treatment.[337] The mean adjusted reduction in height ranges up to 6 cm, but averages a more modest 1–2.5 cm,[329] and treatment is not without potential complications and future consequences. Common side effects include nausea, water retention and weight gain, and menorrhagia. Whereas an early study found no adverse effects of treatment on future fertility,[338] a more recent study involving 1,243 adult women with hereditary tall stature observed that the risk for future infertility was significantly increased and that cycle fecundability was decreased by approximately 40% in those treated with estrogen as adolescents, compared to those who received no treatment[339]; the mechanism responsible for the effect is unknown.

Bayley-Pinneau Table for Average Girls[291]

To predict height, find vertical column corresponding to skeletal age and horizontal row for the present height. The number at the intersection is the predicted height in inches. If figures do not fall at the whole inch or 6-month intervals, the predicted height must be extrapolated.

Skeletal Age		*6/0*	*6/6*	*7/0*	*7/6*	*8/0*	*8/6*	*9/0*	*9/6*	*10/0*	*10/6*	*11/0*	*11/6*	*12/0*
Height in inches	37	51.4												
	38	52.8	51.5											
	39	54.2	52.8	51.5										
	40	55.6	54.2	52.8	51.8									
	41	56.9	55.6	54.2	53.1	51.9								
	42	58.3	56.9	55.5	54.4	53.2	51.9							
	43	59.7	58.3	56.8	55.7	54.4	53.1	52.0						
	44	61.1	59.6	58.1	57.0	55.7	54.3	53.2	52.1	51.0				
	45	62.5	61.0	59.4	58.3	57.0	55.6	54.4	53.3	52.2				
	46	63.9	62.3	60.8	59.6	58.2	56.8	55.6	54.5	53.4	52.0			
	47	65.8	63.7	62.1	60.9	59.5	58.0	56.8	55.7	54.5	53.2	51.9	51.4	51.0
	48	66.7	65.0	63.4	62.2	60.8	59.3	58.0	56.9	55.7	54.3	53.0	52.5	52.1
	49	68.1	66.4	64.7	63.5	62.0	60.5	59.3	58.1	56.8	55.4	54.1	53.6	53.1
	50	69.4	67.8	66.1	64.8	63.3	61.7	60.5	59.2	58.0	56.6	55.2	54.7	54.2
	51	70.8	69.1	67.4	66.1	64.6	63.0	61.7	60.4	59.2	57.7	56.3	55.8	55.3
	52	72.2	70.5	68.7	67.4	65.8	64.2	62.9	61.6	60.3	58.8	57.4	56.9	56.4
	53	73.6	71.8	70.0	68.7	67.1	65.4	64.1	62.8	61.5	60.0	58.5	58.0	57.5
	54		73.2	71.3	69.9	68.4	66.7	65.3	64.0	62.6	61.1	59.6	59.1	58.6
	55		74.5	72.7	71.2	69.6	67.9	66.5	65.2	63.8	62.2	60.7	60.2	59.7
	56			74.0	72.5	70.9	69.1	67.7	66.4	65.0	63.3	61.8	61.3	60.7
	57				73.8	72.2	70.4	68.9	67.5	66.1	64.5	62.9	62.4	61.8
	58					73.4	71.6	70.1	68.7	67.3	65.6	64.0	63.5	62.9
	59					74.7	72.8	71.3	69.9	68.4	66.7	65.1	64.6	64.0
	60						74.1	72.6	71.1	69.6	67.9	66.2	65.6	65.1
	61							73.8	72.3	70.8	69.0	67.3	66.7	66.2
	62								73.5	71.9	70.1	68.4	67.8	67.2
	63								74.6	73.1	71.3	69.5	68.9	68.3
	64									74.2	72.4	70.6	70.0	69.4
	65										73.5	71.7	71.1	70.5
	66										74.7	72.9	72.2	71.6
	67											74.0	73.3	72.7
	68												74.4	73.8
	69													74.8
	70													
	71													
	72													
	73													
	74													

12/6	13/0	13/6	14/0	14/6	15/0	15/6	16/0	16/6	17/0	17/6	18/0	
												37
												38
												39
												40
												41
												42
												43
												44
												45
												46
												47
51.0												48
52.1	51.1											49
53.1	52.2	51.3	51.0									50
54.2	53.2	52.4	52.0	51.7	51.5	51.4	51.2	51.2	51.1	51.0	51.0	51
55.3	54.3	53.4	53.1	52.7	52.5	52.4	52.2	52.2	52.1	52.0	52.0	52
56.3	55.3	54.4	54.1	53.8	53.5	53.4	53.2	53.2	53.1	53.0	53.0	53
57.4	56.4	55.4	55.1	54.8	54.5	54.4	54.2	54.2	54.1	54.0	54.0	54
58.4	57.4	56.5	56.1	55.8	55.6	55.4	55.2	55.2	55.1	55.0	55.0	55
59.5	58.5	57.5	57.1	56.8	56.6	56.4	56.2	56.2	56.1	56.0	56.0	56
60.6	59.5	58.5	58.2	57.8	57.6	57.4	57.2	57.2	57.1	57.0	57.0	57
61.6	60.5	59.5	59.2	58.8	58.6	58.4	58.2	58.2	58.1	58.0	58.0	58
62.7	61.6	60.6	60.2	59.8	59.6	59.4	59.2	59.2	59.1	59.0	59.0	59
63.8	62.6	61.6	61.2	60.9	60.6	60.4	60.2	60.2	60.1	60.0	60.0	60
64.8	63.7	62.6	62.2	61.9	61.6	61.4	61.2	61.2	61.1	61.0	61.0	61
65.9	64.7	63.7	63.3	62.9	62.6	62.4	62.2	62.2	62.1	62.0	62.0	62
67.0	65.8	64.7	64.3	63.9	63.6	63.4	63.3	63.2	63.1	63.0	63.0	63
68.0	66.8	65.7	65.3	64.9	64.6	64.4	64.3	64.2	64.1	64.0	64.0	64
69.1	67.8	66.7	66.3	65.9	65.7	65.5	65.3	65.2	65.1	65.0	65.0	65
70.1	68.9	67.8	67.3	66.9	66.7	66.5	66.3	66.2	66.1	66.0	66.0	66
71.2	69.9	68.8	68.4	68.0	67.7	67.5	67.3	67.2	67.1	67.0	67.0	67
72.3	71.0	69.8	69.4	69.0	68.7	68.5	68.3	68.2	68.1	68.0	68.0	68
73.3	72.0	70.8	70.4	70.0	69.7	69.5	69.3	69.2	69.1	69.0	69.0	69
74.4	73.1	71.9	71.4	71.0	70.7	70.5	70.3	70.2	70.1	70.0	70.0	70
	74.1	72.9	72.4	72.0	71.7	71.5	71.3	71.2	71.1	71.0	71.0	71
		73.9	73.5	73.0	72.7	72.5	72.3	72.2	72.1	72.0	72.0	72
		74.9	74.5	74.0	73.7	73.5	73.3	73.2	73.1	73.0	73.0	73
					74.7	74.5	74.3	74.2	74.1	74.0	74.0	74

Bayley-Pinneau Table for Accelerated Girls[291]

To predict height, find vertical column corresponding to skeletal age and horizontal row for the present height. The number at the intersection is the predicted height in inches. If figures do not fall at the whole inch or 6-month intervals, the predicted height must be extrapolated.

Skeletal Age		6/0	6/6	7/0	7/6	8/0	8/6	9/0	9/6	10/0	10/6	11/0	11/6	12/0
Height in inches	37	51.4												
	38	52.8	51.5											
	39	54.2	52.8	51.5										
	40	55.6	54.2	52.8	51.8									
	41	56.9	55.6	54.2	53.1	51.9								
	42	58.3	56.9	55.5	54.4	53.2	51.9							
	43	59.7	58.3	56.8	55.7	54.4	53.1	52.0						
	44	61.1	59.6	58.1	57.0	55.7	54.3	53.2	52.1	51.0				
	45	62.5	61.0	59.4	58.3	57.0	55.6	54.4	53.3	52.2				
	46	63.9	62.3	60.8	59.6	58.2	56.8	55.6	54.5	53.4	52.0			
	47	65.8	63.7	62.1	60.9	59.5	58.0	56.8	55.7	54.5	53.2	51.9	51.4	51.0
	48	66.7	65.0	63.4	62.2	60.8	59.3	58.0	56.9	55.7	54.3	53.0	52.5	52.1
	49	68.1	66.4	64.7	63.5	62.0	60.5	59.3	58.1	56.8	55.4	54.1	53.6	53.1
	50	69.4	67.8	66.1	64.8	63.3	61.7	60.5	59.2	58.0	56.6	55.2	54.7	54.2
	51	70.8	69.1	67.4	66.1	64.6	63.0	61.7	60.4	59.2	57.7	56.3	55.8	55.3
	52	72.2	70.5	68.7	67.4	65.8	64.2	62.9	61.6	60.3	58.8	57.4	56.9	56.4
	53	73.6	71.8	70.0	68.7	67.1	65.4	64.1	62.8	61.5	60.0	58.5	58.0	57.5
	54		73.2	71.3	69.9	68.4	66.7	65.3	64.0	62.6	61.1	59.6	59.1	58.6
	55		74.5	72.7	71.2	69.6	67.9	66.5	65.2	63.8	62.2	60.7	60.2	59.7
	56			74.0	72.5	70.9	69.1	67.7	66.4	65.0	63.3	61.8	61.3	60.7
	57				73.8	72.2	70.4	68.9	67.5	66.1	64.5	62.9	62.4	61.8
	58					73.4	71.6	70.1	68.7	67.3	65.6	64.0	63.5	62.9
	59					74.7	72.8	71.3	69.9	68.4	66.7	65.1	64.6	64.0
	60						74.1	72.6	71.1	69.6	67.9	66.2	65.6	65.1
	61							73.8	72.3	70.8	69.0	67.3	66.7	66.2
	62								73.5	71.9	70.1	68.4	67.8	67.2
	63								74.6	73.1	71.3	69.5	68.9	68.3
	64									74.2	72.4	70.6	70.0	69.4
	65										73.5	71.7	71.1	70.5
	66										74.7	72.9	72.2	71.6
	67											74.0	73.3	72.7
	68												74.4	73.8
	69													74.8
	70													
	71													
	72													
	73													
	74													

12/6	13/0	13/6	14/0	14/6	15/0	15/6	16/0	16/6	17/0	17/6	
											37
											38
											39
											40
											41
											42
											43
											44
											45
											46
											47
51.9											48
53.0	51.9	50.9									49
54.1	52.9	51.9	51.4	51.0							50
55.2	54.0	53.0	52.5	52.0	51.7	51.5	51.4	51.3	51.1	51.0	51
56.3	55.0	54.0	53.5	53.1	52.7	52.5	52.4	52.3	52.1	52.0	52
57.4	56.1	55.0	54.5	54.1	53.8	53.5	53.4	53.3	53.1	53.0	53
58.4	57.1	56.1	55.6	55.1	54.8	54.5	54.4	54.3	54.1	54.0	54
59.5	58.2	57.1	56.6	56.1	55.8	55.5	55.4	55.3	55.1	55.0	55
60.6	59.3	58.2	57.6	57.1	56.8	56.5	56.4	56.3	56.1	56.0	56
61.7	60.3	59.2	58.6	58.2	57.8	57.6	57.4	57.3	57.1	57.0	57
62.8	61.4	60.2	59.7	59.2	58.8	58.6	58.4	58.3	58.1	58.0	58
63.9	62.4	61.3	60.7	60.2	59.8	59.6	59.4	59.3	59.1	59.0	59
64.9	63.5	62.3	61.7	61.2	60.9	60.6	60.4	60.3	60.1	60.0	60
66.0	64.6	63.3	62.8	62.2	61.9	61.6	61.4	61.3	61.1	61.0	61
67.1	65.6	64.4	63.8	63.3	62.9	62.6	62.4	62.3	62.1	62.0	62
68.2	66.7	65.4	64.8	64.3	63.9	63.6	63.4	63.3	63.1	63.0	63
69.3	67.7	66.5	65.8	65.3	64.9	64.6	64.4	64.3	64.1	64.0	64
70.3	68.8	67.5	66.9	66.3	65.9	65.7	65.5	65.3	65.1	65.0	65
71.4	69.8	68.5	67.9	67.3	66.9	66.7	66.5	66.3	66.1	66.0	66
72.5	70.9	69.6	68.9	68.4	68.0	67.7	67.5	67.3	67.1	67.0	67
73.6	72.0	70.6	70.0	69.4	69.0	68.7	68.5	68.3	68.1	68.0	68
74.7	73.0	71.7	71.0	70.4	70.0	69.7	69.5	69.3	69.1	69.0	69
	74.1	72.7	72.0	71.4	71.0	70.7	70.5	70.3	70.1	70.0	70
		73.7	73.0	72.4	72.0	71.7	71.5	71.4	71.1	71.0	71
		74.8	74.1	73.5	73.0	72.7	72.5	72.4	72.1	72.0	72
				74.5	74.0	73.7	73.5	73.4	73.1	73.0	73
						74.4	74.5	74.4	74.1	74.0	74

Bayley-Pinneau Table for Delayed Girls[291]

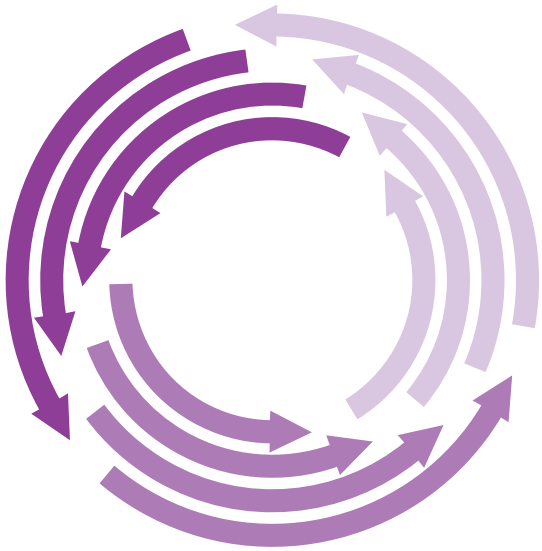
To predict height, find vertical column corresponding to skeletal age and horizontal row for the present height. The number at the intersection is the predicted height in inches. If figures do not fall at the whole inch or 6-month intervals, the predicted height must be extrapolated.

Skeletal Age		6/0	6/6	7/0	7/6	8/0	8/6	9/0	9/6	10/0	10/6	11/0	11/6
Height in inches	38	51.8											
	39	53.2	51.9										
	40	54.6	53.3	51.9									
	41	55.9	54.6	53.2	52.0								
	42	57.3	55.9	54.5	53.3	52.2	51.0						
	43	58.7	57.3	55.8	54.6	53.5	52.2	51.1					
	44	60.0	58.6	57.1	55.8	54.7	53.5	52.3	51.3				
	45	61.4	59.9	58.4	57.1	56.0	54.7	53.5	52.4	51.5			
	46	62.8	61.3	59.7	58.4	57.2	55.9	54.7	53.6	52.6	51.3		
	47	64.1	62.6	61.0	59.6	58.5	57.1	55.9	54.8	53.8	52.5	51.2	
	48	65.5	63.9	62.3	60.9	59.7	58.3	57.1	55.9	54.9	63.6	52.3	51.8
	49	66.9	65.2	63.6	62.2	60.9	59.5	58.3	57.1	56.1	54.7	53.4	52.9
	50	68.2	66.6	64.9	63.5	62.2	60.8	59.5	58.3	57.2	55.8	54.5	54.0
	51	69.6	67.9	66.2	64.7	63.4	62.0	60.6	59.4	58.4	56.9	55.6	55.1
	52	70.9	69.2	67.5	66.0	64.7	63.2	61.8	60.6	59.5	58.0	56.6	56.2
	53	72.3	70.6	68.8	67.3	65.9	64.4	63.0	61.8	60.6	59.2	57.7	57.2
	54	73.7	71.9	70.1	68.5	67.2	65.6	64.2	62.9	61.8	60.3	58.8	58.3
	55		73.2	71.4	69.8	68.4	66.8	65.4	64.1	62.9	61.4	59.9	59.4
	56		74.6	72.7	71.1	69.7	68.0	66.6	65.3	64.1	62.5	61.0	60.5
	57			74.0	72.3	70.9	69.3	67.8	66.4	65.2	63.6	62.1	61.6
	58				73.6	72.1	70.5	69.0	67.6	66.4	64.7	63.2	62.6
	59				74.9	73.4	71.7	70.2	68.8	67.5	65.8	64.3	63.7
	60					74.6	72.9	71.3	69.9	68.7	67.0	65.4	64.8
	61						74.1	72.5	71.1	69.8	68.1	66.4	65.9
	62							73.7	72.3	70.9	69.2	67.5	67.0
	63							74.7	73.4	72.1	70.3	68.6	68.0
	64								74.6	73.2	71.4	69.7	69.1
	65									74.4	72.5	70.8	70.2
	66										73.7	71.9	71.3
	67										74.8	73.0	72.4
	68											74.1	73.4
	69												74.5
	70												
	71												
	72												
	73												
	74												

12/0	12/6	13/0	13/6	14/0	14/6	15/0	15/6	16/0	16/6	17/0	
											38
											39
											40
											41
											42
											43
											44
											45
											46
											47
51.5											48
52.6	51.6										49
53.6	52.7	51.9	51.2								50
54.7	53.7	52.9	52.2	51.9	51.6	51.3	51.2	51.1	51.1	51.0	51
55.8	54.8	53.9	53.2	52.9	52.6	52.3	52.2	52.1	52.1	52.0	52
56.9	55.8	55.0	54.2	53.9	53.6	53.3	53.2	53.1	53.1	53.0	53
57.9	56.9	56.0	55.3	54.9	54.6	54.3	54.2	54.1	54.1	54.0	54
59.0	58.0	57.1	56.3	56.0	55.6	55.3	55.2	55.1	55.1	55.0	55
60.1	59.0	58.1	57.3	57.0	56.6	56.3	56.2	56.1	56.1	56.0	56
61.2	60.1	59.1	58.3	58.0	57.6	57.3	57.2	57.1	57.1	57.0	57
62.2	61.1	60.2	59.4	59.0	58.6	58.3	58.2	58.1	58.1	58.0	58
63.3	62.2	61.2	60.4	60.0	59.7	59.4	59.2	59.1	59.1	59.0	59
64.4	63.2	62.2	61.4	61.0	60.7	60.4	60.2	60.1	60.1	60.0	60
65.5	64.3	63.3	62.4	62.1	61.7	61.4	61.2	61.1	61.1	61.0	61
66.5	65.3	64.3	63.5	63.1	62.7	62.4	62.2	62.1	62.1	62.0	62
67.6	66.4	65.3	64.5	64.1	63.7	63.4	63.3	63.1	63.1	63.0	63
68.7	67.4	66.4	65.5	65.1	64.7	64.4	64.3	64.1	64.1	64.0	64
69.7	68.5	67.4	66.5	66.1	65.7	65.4	65.3	65.1	65.1	65.0	65
70.8	69.5	68.5	67.6	67.1	66.7	66.4	66.3	66.1	66.1	66.0	66
71.9	70.6	69.5	68.6	68.2	67.7	67.4	67.3	67.1	67.1	67.0	67
73.0	71.7	70.5	69.6	69.2	68.8	68.4	68.3	68.1	68.1	68.0	68
74.0	72.7	71.6	70.6	70.2	69.8	69.4	69.3	69.1	69.1	69.0	69
	73.8	72.6	71.6	71.2	70.8	70.4	70.3	70.1	70.1	70.0	70
	74.8	73.6	72.7	72.2	71.8	71.4	71.3	71.1	71.1	71.0	71
		74.7	73.7	73.3	72.8	72.4	72.3	72.1	72.1	72.0	72
			74.7	74.3	73.8	73.4	73.3	73.1	73.1	73.0	73
					74.8	74.4	74.3	74.1	74.1	74.0	74

All references are available online at: http://www.clinicalgynendoandinfertility.com

11 Amenorrhea



Few problems in gynecologic endocrinology can present a diagnostic challenge to clinicians like that of amenorrhea. The number, variety, and complexity of diseases and disorders that must be considered can seem daunting and, in many instances, includes unfamiliar organ systems. Moreover, some of the diagnostic possibilities can have serious consequences, if not recognized and treated effectively. Consequently, otherwise confident and experienced clinicians may view the problem as too complicated and time-consuming or may question their ability to perform or interpret the evaluation. However, when approached logically and systematically, the diagnostic evaluation of amenorrhea truly is straightforward, involving laboratory tests and procedures already familiar to almost all clinicians. With few exceptions, an evaluation can be completed quickly and without great expense.

The purpose of this chapter is to provide a systematic strategy for the evaluation of amenorrhea that will yield an accurate diagnosis, no matter how common or uncommon the cause. Once a diagnosis is established, additional corroborating evidence and the assistance of appropriate specialists (e.g., neurosurgeon, internist, endocrinologist, or psychiatrist) can be obtained, when necessary. However, the large majority of women with amenorrhea have relatively simple problems—polycystic ovarian syndrome (PCOS), hypothalamic amenorrhea, hyperprolactinemia, and ovarian failure—all of which can be managed easily by primary care clinicians.

The diagnostic evaluation described here is not new. With minor modifications, it has been applied successfully for several decades. Before describing the evaluation in detail, amenorrhea first must be defined, so as to identify the patients who warrant evaluation. A brief preliminary review of the physiologic mechanisms involved in menstruation provides the framework necessary to understand and follow the logical design of the diagnostic evaluation.

Definition of Amenorrhea

The age at which menarche should be expected varies with individual differences in the age at the onset of puberty. The normal pubertal progression is discussed in detail in Chaper 10 and is only briefly summarized here. In general, the first menses should occur within 2–3 years after the initiation of pubertal development. In most young girls (approximately 80%), the first sign of puberty is an acceleration of growth, followed by breast budding (thelarche), and the appearance of pubic hair (adrenarche). In the remainder (approximately 20%) adrenarche precedes thelarche by a brief interval, but the two events typically are closely linked. Consequently, menarche can occur as early as age 10 (when puberty begins at age 8), and rarely occurs later than age 16 (when puberty begins at age 13). On average, the mean ages for thelarche, adrenarche, and menarche in African-American girls are 6–12 months earlier than in Caucasian American girls. Once normal menstrual cycles have been established, they should occur at regular intervals ranging between 25 and 35 days. Therefore, patients fulfilling any of the following criteria should be evaluated for amenorrhea:

- No menses by age 14 in the absence of growth or development of secondary sexual characteristics.
- No menses by age 16 regardless of the presence of normal growth and development of secondary sexual characteristics.
- In women who have menstruated previously, no menses for an interval of time equivalent to a total of at least three previous cycles, or 6 months.

Having affirmed the traditional definition of amenorrhea, it is important to point out that strict adherence to these criteria can result in improper management of individual patients. For example, there is no reason to defer the evaluation of a young girl who presents with the classical phenotype of Turner syndrome. Similarly, a 14-year-old girl who has no vagina should not be advised to return in 2 years. All patients deserve a considerate evaluation whenever their anxieties, or those of their parents, are brought to the attention of a clinician. Finally, the possibility of pregnancy always should be considered.

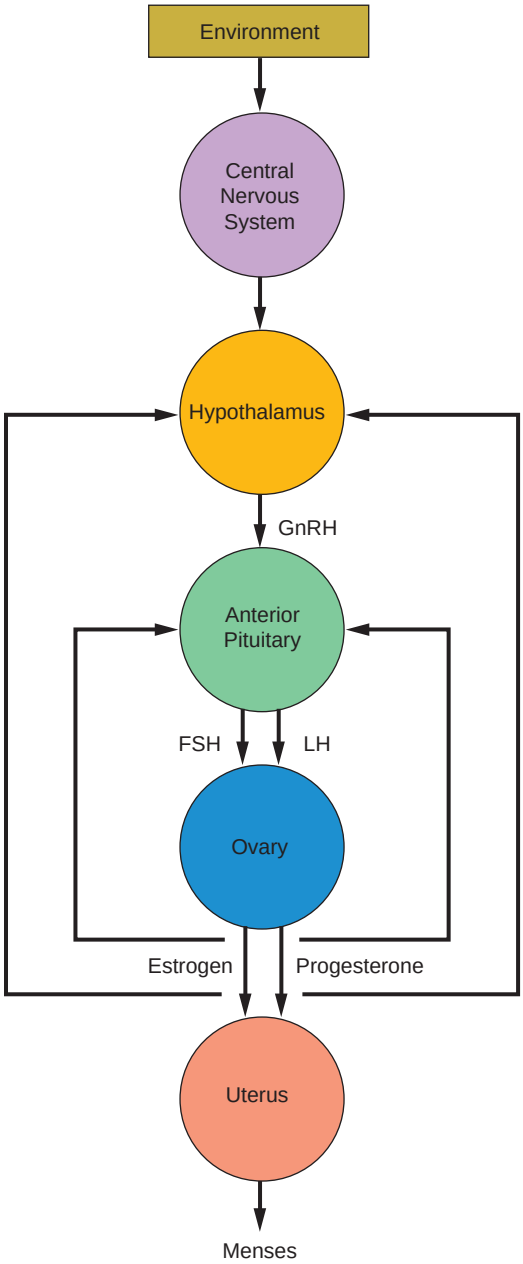
Traditionally, amenorrhea has been categorized as primary or secondary. Primary amenorrhea describes patients who never have menstruated and secondary amenorrhea describes those who have menstruated previously but now do not. The differential diagnoses of primary and secondary amenorrhea differ, but only to a limited extent. For example, the diagnosis of mullerian agenesis is possible only in patients with primary amenorrhea and premature menopause necessarily occurs only in women with secondary amenorrhea. However, besides helping to narrow the scope of diagnostic possibilities, the classical distinction between primary and secondary amenorrhea serves little practical purpose. Such preliminary categorization sometimes even can mislead the evaluation or its interpretation. In any case, the diagnostic approach recommended here can be applied effectively in all women with amenorrhea.

Basic Principles in Menstrual Function

The clinical demonstration of menstrual function requires visible external evidence of the menstrual discharge. For that to occur, the genital outflow tract must be anatomically intact with continuous connection between the vaginal orifice, the vaginal canal, the endocervix, and the uterine cavity. The uterus also must contain a functional endometrium that can respond to the actions of ovarian sex steroid hormones, estrogen and progesterone,

across the ovarian cycle of follicle development, ovulation, and corpus luteum function. The ovaries must contain viable follicles that can respond to stimulation by the gonadotropins, follicle-stimulating hormone (FSH) and luteinizing hormone (LH), released from the anterior pituitary. In turn, pituitary gonadotropin secretion depends on the action of gonadotropin-releasing hormone (GnRH), secreted from the medial basal hypothalamus into the portal vascular network that bathes the anterior pituitary. Finally, the pulsatile pattern of hypothalamic GnRH secretion is governed by input from higher centers that interpret and translate environmental stimuli, and modulated by the feedback effects of ovarian sex steroids. The entire system is highly regulated by a complex mechanism that integrates biophysical and biochemical information composed of interacting hormonal signals, autocrine/paracrine factors, and target cell reactions.

The basic requirements for normal menstrual function thus include four anatomically and functionally distinct structural components—the genital outflow tract including the



uterus, the ovary, the pituitary, and the hypothalamus—thus providing a natural and useful hierarchy for organizing the diagnostic evaluation of amenorrhea. Accordingly, the many causes of amenorrhea can be categorized according to the site or level of the disorder or disturbance:

- Disorders of the genital outflow tract and uterus
- Disorders of the ovary
- Disorders of the anterior pituitary
- Disorders of the hypothalamus or central nervous system

Amenorrhea can result from congenital or acquired disease or dysfunction at any level in the system and can involve more than one mechanism. For example, PCOS involves a number of interrelated pathophysiologic mechanisms operating at the ovarian, pituitary and hypothalamic levels.

Evaluation of Amenorrhea

The evaluation of amenorrhea, like any other complaint, begins with a careful medical history and physical examination, which always provide valuable diagnostic clues. Information gained from the medical history and physical examination clearly can exclude certain diagnostic possibilities, but first impressions also can be deceiving and lead to errors in judgment and to inappropriate, costly, and needless testing. A methodical, systematic approach to diagnosis therefore is best.

Logically, the recommended evaluation for amenorrhea is designed to separate the reproductive system into its distinct structural components—the genital outflow tract and uterus, the ovary, the pituitary, and the hypothalamus—and to test the functional integrity of each, beginning at the lowest level and progressing systematically to the higher levels of the system until the cause is determined.

Medical History

The menstrual history is, of course, key. Primary amenorrhea speaks for itself, but cyclic pelvic or lower abdominal pain or urinary complaints can be caused by developmental anomalies resulting in obstructed menstrual flow (cryptomenorrhea), as may be caused by an imperforate hymen, transverse vaginal septum, or cervical atresia. In women with secondary amenorrhea, the history surrounding the onset of amenorrhea can provide important diagnostic clues. Onset following curettage or other uterine surgery clearly suggests the possibility of damage to the reproductive tract. The menstrual history in women having the most common causes of amenorrhea is distinctly different and easily recognized. Women with PCOS classically present with infrequent and irregular menses dating from menarche or early adulthood, and gradually progressive hirsutism. In most with hypothalamic amenorrhea, the onset of amenorrhea temporally relates to events resulting in severe nutritional, physical, or emotional stress. Women with hyperprolactinemia or premature ovarian failure commonly notice a gradual decrease in their regular intermenstrual interval, followed by increasing oligomenorrhea, and finally, amenorrhea, one sometimes accompanied by galactorrhea, and the other by hot flushes.

Questions relating to past medical history, general health, and lifestyle can identify a severe or chronic illness such as diabetes, renal failure, or inflammatory bowel disease, previous head trauma, or evidence of physical or psychological stress. Specific history relating to weight loss or weight gain and to the frequency and intensity of exercise is highly relevant and often revealing. Headaches, seizures, vomiting, behavioral changes, or visual symptoms may suggest a CNS disorder. Vaginal dryness or hot flushes are evidence of estrogen deficiency and suggest ovarian failure. Progressive hirsutism or virilization is an indication of hyperandrogenism that may relate to PCOS, non-classical (late-onset) congenital adrenal hyperplasia (CAH), or an androgen-producing tumor of the ovary or adrenal. Symptoms of galactorrhea obviously suggest hyperprolactinemia. History relating to the time and duration of any treatment with oral contraceptive pills (OCP), progestins (e.g., depot-medroxyprogesterone acetate), GnRH agonists, or other medications or drugs that can affect central neurotransmitter secretion (phenothizines, reserpine derivatives, amphetamines, benzodiazepines, antidepressants, dopamine antagonists, opiates) also can provide important diagnostic clues.

Physical Examination

The overall body habitus often provides important information. Height, weight, and body mass index (BMI) should be determined and recorded. Short stature (less than 60 inches) and sexual infantilism are hallmarks of gonadal dysgenesis. Low body weight frequently is associated with hypothalamic amenorrhea resulting from poor nutrition (eating disorders, malabsorbsion syndromes) or physical, psychological, or emotional stress. Obesity or an increased waist-hip ratio (>0.85) are common features of women with PCOS.

Examination of the skin can reveal a soft, moist texture as seen in hyperthyroidism; a rapid pulse, exopthalmos or lid lag, a fine tremor, and hyperreflexia suggest the diagnosis of Graves' disease. Conversely, dry, thick skin, a slow pulse, diminished reflexes, and thinning of the hair suggest hypothyroidism. A goiter or thyroid nodule is further evidence of a thyroid disorder; both hypothyroidism and hyperthyroidism can be associated with amenorrhea. Orange discoloration of the skin, without scleral icterus, can result from hypercarotinemia associated with excessive ingestion of low calorie carotene-containing fruits and vegetables in dieting women. Acanthosis nigricans (velvety hyperpigmented skin observed most commonly at the nape of the neck, in the axillae, groin, and beneath the breasts) strongly suggests severe insulin resistance and the possibility of diabetes. Acne and hirsutism are indications of hyperandrogenism that may be associated with PCOS, non-classical CAH, or exposure to androgenic anabolic steroids. When accompanied by any sign of frank virilization (deepening of the voice, fronto-temporal balding, decrease in breast size, increased muscle mass, clitoromegaly), the possibility of ovarian hyperthecosis or an ovarian or adrenal neoplasm must be considered.

Examination of the breasts deserves careful attention. ***Breast development is a reliable indicator of estrogen production or exposure to exogenous estrogens.*** The Tanner stage of breast development should be noted (Chapter 10). A secondary arrest of breast development suggests a disruption of the hypothalamic-pituitary-ovarian (HPO) axis. When menarche has not followed breast development at the expected time, a developmental anomaly of the reproductive tract also should be considered. Breast examination should include gentle compression, beginning at the base and moving toward the nipple. ***Secretions that result from hormonal stimulation typically emerge from multiple duct openings in the nipple, whereas a discharge relating to breast pathology usually arises from a single duct.*** Microscopic examination of any expressed cloudy or white nipple secretions demonstrating lipid droplets confirms galactorrhea and suggests hyperprolactinemia.

Abdominal examination rarely may reveal a mass as may result from hematometra or an ovarian neoplasm. Growth of sexual hair in the infraumbilical region suggests hyperandrogenism. Abdominal striae raise the possibility of Cushing syndrome, but most often result from progressive obesity or previous pregnancy.

Careful examination of the external genitalia and lower genital tract is key. ***The presence of pubic hair growth reliably reflects androgen production or exposure.*** Because breast development and growth of pubic hair typically progress in a symmetrical manner, their Tanner stages should be consistent. Absent or scant growth of sexual hair can be expected in otherwise sexually infantile girls, but also is a classical sign of androgen insensitivity syndrome (AIS) when breast development is asymmetrically advanced. Attempts at office examination of the vagina in sexually infantile girls or those with a small hymeneal ring are generally unrewarding and often even counterproductive, but whenever feasible, speculum examination should be performed. A patent vagina and normal cervix excludes mullerian/vaginal agenesis, AIS, and obstructive causes of amenorrhea such as an imperforate hymen or transverse vaginal septum. In those with primary amenorrhea having an absent or infantile vaginal orifice, rectal examination should be performed to detect any distended hematocolpos that may form above the obstruction when the uterus is present and functional.

Evaluation of the Genital Outflow Tract and Uterus

Evaluation of the genital outflow tract and uterus can be organized easily based on the menstrual history and physical examination of the genital anatomy. Primary amenorrhea with a blind or absent vagina points directly to a developmental anomaly of the genital outflow tract. Primary or secondary amenorrhea with a patent vagina and visible cervix excludes abnormalities of the genital outflow tract, except in those with history of previous cervical or uterine surgery or infection in whom the possibilities of cervical stenosis and intrauterine adhesions or other endometrial damage must be considered.

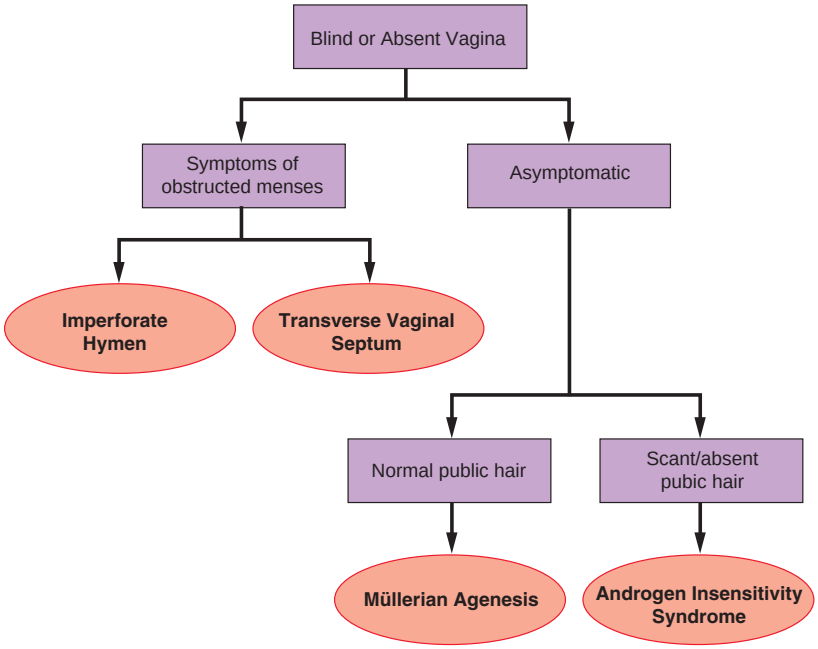
In premenarchial age girls with the incidental finding of an absent vaginal orifice, diagnosis can be more difficult, but also is seldom urgent. Although pelvic ultrasonography generally can determine whether a uterus is present, imaging must be interpreted cautiously because even abdominal/pelvic magnetic resonance imaging (MRI) can be misleading when the reproductive organs are immature and very small. Remaining alert to the diagnostic possibilities, careful observation over time is preferable to invasive investigations otherwise unnecessary in the asymptomatic prepubertal girl.

Abnormal Genital Anatomy

The embryology of the female genital tract is complex but generally well defined and is described in detail in Chapter 4. Briefly summarized, it involves both the medial migration and midline fusion of the müllerian (paramesonephric) ducts to form the uterus, cervix, and upper vagina, and the vertical fusion of that developing ductal system with the invaginating urogenital sinus that forms the lower vagina and the introitus. Outflow tract abnormalities that result from failure of müllerian duct development include vaginal/müllerian agenesis and AIS; the presence or absence of pubic hair distinguishes the two. Abnormalities caused by failure of vertical fusion include imperforate hymen and transverse vaginal septum or cervical atresia. Although all are uncommon and the clinician may have only limited or no previous experience with any of the four, each has unique and distinguishing features that, in most cases, point to the correct diagnosis at time of the initial visit. Only limited additional evaluation, described in a later section of this chapter, is required to firmly establish the diagnosis and to plan treatment.

Normal Genital Anatomy

In those with primary or secondary amenorrhea having a patent vagina and visible cervix, the likelihood of a genital outflow tract abnormality is very small. The only possibilities that need be considered are cervical stenosis and intrauterine adhesions (Asherman syndrome) or other endometrial damage that may result from surgical trauma or infection. Since these are acquired conditions, they result in secondary amenorrhea with an onset that typically correlates closely with the time of the previous insult. Both receive separate and thorough discussion in the later section of this chapter devoted to specific disorders of the genital outflow tract and uterus.



Evaluation of Ovarian Function

In women with normal genital tract anatomy and no relevant history to suggest the possibility of cervical stenosis or Asherman syndrome, disorders of the genital outflow tract and uterus can be excluded and further stepwise evaluation is required to determine the cause of amenorrhea. Attention now may be focused on the next level of the reproductive system, the ovary.

Abnormalities of ovarian function are the most common overall cause of amenorrhea and include a wide variety of disorders ranging from simple chronic anovulation, as in women with PCOS, obesity, thyroid disorders and hyperprolactinemia, to complete ovarian failure relating to chromosomal abnormalities or other genetic disorders such as Fragile X (*FMR1*) premutations and galactosemia, autoimmune disease, radiation or chemotherapy. These and other specific causes of ovarian failure and the mechanisms involved are discussed at length in a later section of this chapter. The focus here is on the evaluation of ovarian function and on the diagnosis and treatment of common chronic anovulatory disorders.

The most obvious measure of ovarian function is estrogen production. Unfortunately, one cannot rely on symptoms and signs of estrogen deficiency to identify hypogonadal women.

Genitourinary atrophy develops only gradually and is not observed commonly in young women, even when estrogen levels are clearly low, and vasomotor symptoms typically are absent in women with hypothalamic dysfunction. Other methods for assessing the level of ovarian estrogen production include measurement of the serum estradiol concentration and "bioassays" based on clinical observations of the amount and character of cervical mucus, the results of a "progestin challenge test," or measurement of endometrial thickness by transvaginal ultrasonography. ***Although each method is useful, each also has pitfalls, no one method is definitive, and more than one measure therefore is recommended.*** Overall, the duration of amenorrhea and other clinical history and features are as or more important and useful for assessing ovarian function.

Seum Estradiol Concentration

A serum estradiol measurement is easy to obtain, relatively inexpensive, and objective. Reasonably, one would expect to find normal estrogen levels in women with normal ovaries whose amenorrhea results simply from mild dysregulation and chronic anovulation, as in obese women and those with PCOS, and to find low estrogen levels in women with ovarian failure, pituitary disease, or more severe hypothalamic dysfunction. Unfortunately, serum estradiol concentrations can fluctuate erratically in all conditions, normal or low on any given day, and therefore can be misleading. A random estradiol concentration greater than approximately 40 pg/mL clearly suggests the presence of functional ovarian follicles but also is common during a premature or normal perimenopause and occurs sporadically in women with hypothalamic amenorrhea. A low random estradiol concentration may suggest ovarian failure, but also is typical of women with hypothalamic amenorrhea and may be observed in those with less severe forms of chronic anovulation, as in normal women during the early follicular phase.

Bioassays of Estrogen Production

The observation of ***"estrogenic" cervical mucus***—clear, watery, and relatively abundant—suggests a normal level of ovarian estrogen production, but its absence cannot be interpreted confidently because many normal women exhibit such mucus only during the late follicular phase of the cycle when estrogen levels are relatively high, or not at all.

The ***progestin challenge test*** is based on the premise that progestin treatment (e.g., medroxyprogesterone acetate 10 mg daily for 5–7 days, or progesterone in oil 200 mg i.m.) will induce menses only in those having normal circulating estrogen concentrations. ***A pure progestational agent must be used because endogenous estrogen status cannot be inferred from the response to an OCP that contains both estrogen and progestin.*** The more potent synthetic progestins such as medroxyprogesterone acetate are a better choice than oral micronized progesterone, which must be administered in relatively high doses (e.g., 300 mg daily) to achieve a response.[1] A positive test—bleeding within 2–7 days after completion of progestin treatment—implies normal estrogen production and ovarian function, and a negative test—no withdrawal menses—suggests hypogonadism. Scant withdrawal bleeding or spotting suggests marginal levels of endogenous estrogen production. However, the overall correlation between withdrawal bleeding and estrogen status is far from perfect; both false positive (withdrawal bleeding despite generally low levels of estrogen production) and false negative results (absent bleeding despite significant estrogen production) are relatively common. Up to 40–50% of women whose amenorrhea relates to stress, exercise, weight loss, hyperprolactinemia, or ovarian failure, in whom estrogen levels generally are low, exhibit withdrawal bleeding.[2, 3] Up to 20% of amenorrheic women

with significant estrogen production have no withdrawal bleeding,[4] in some because the endometrium is decidualized by high circulating androgen levels.

The ***endometrial thickness***, determined by transvaginal ultrasonography (the maximum 2-layer thickness in the mid-sagittal plane), is a measure of endometrial proliferation, which reflects the level of estrogen production. Endometrial thickness correlates with both the serum estradiol concentration and with the response to a progestin challenge in women with amenorrhea. In one study involving 44 women with secondary amenorrhea, endometrial thickness was significantly greater in 32 women who had withdrawal bleeding (10.3±4.1 mm) than in 12 who did not (5.0±1.3 mm); the serum estradiol level also was significantly greater (45.3±19.4 vs. 18.6±8.0 pg/mL), and an endometrial thickness measuring 6.0 mm or greater predicted withdrawal bleeding with 95% accuracy.[2] An added potential benefit of endometrial thickness as a measure of ovarian estrogen production is that it can help to identify individuals with chronic anovulation at low risk for having associated pathology such as hyperplasia or cancer.

Serum FSH Concentration

The serum FSH concentration is another obvious and useful, but indirect, measure of ovarian function. A normal or low serum FSH level indicates the presence of functional ovarian follicles and may be observed in a variety of conditions associated with amenorrhea, including chronic anovulation (e.g., PCOS), pituitary disease, and hypothalamic dysfunction. ***A high serum FSH concentration is a reliable indication of ovarian follicular depletion or failure.*** Exceptions are rare and include inactivating mutations involving the FSH or LH receptor, enzyme deficiencies (17α-hydroxylase, aromatase), and functional pituitary and ectopic FSH-secreting tumors. ***Because the clinical implications of an elevated FSH level are serious, one or more repeated measurements are warranted to confirm the finding.***

Clinical State	*Serum FSH*	*Serum LH*
Normal adult female	5–20 IU/L (midcycle peak ~ 2 times the basal level in ovulatory women	5–20 IU/L (midcycle peak ~ 3 times the basal level in ovulatory women
Hypogonadotropic state: Prepubertal, Hypothalamic or pituitary dysfunction	<5 IU/L	<5 IU/L
Hypergonadotropic state: Postmenopausal, Castrate, or Ovarian failure	>20 IU/L	>40 IU/L

Although certainly not inappropriate, generally it is not necessary or helpful to also measure the serum LH concentration because levels of the two gonadotropins typically move in parallel. The one notable and highly relevant exception in women with amenorrhea—the "monotropic" rise in FSH that signals a more advanced stage of follicular depletion—can be detected by measuring FSH alone. A moderately increased serum LH concentration frequently is observed in women with PCOS, but is not a diagnostic criterion and has no other clinical relevance. A "reversed" LH/FSH ratio (LH lower than FSH), like that seen in prepubertal girls, suggests but does not prove hypothalamic dysfunction. During the midcycle gonadotropin surge in ovulatory cycles, LH levels increase more than those of FSH, but that has little relevance in women with amenorrhea. Other conditions in which levels of the two gonadotropins diverge significantly are truly rare and include ectopic gonadotropin secretion by tumors outside the reproductive tract, single gonadotropin deficiencies resulting from mutations in genes encoding the β-subunit of LH or FSH, and the very rare functional gonadotroph adenoma that secretes clinically important amounts of one gonadotropin (FSH) but not the other.

Measurement of the serum FSH level traditionally has been recommended only for those having demonstrable evidence of hypogonadism (e.g., a negative progestin challenge), as the means to differentiate patients with gonadal failure from those with hypothalamic or pituitary causes of amenorrhea. However, routine measurement of serum FSH in the evaluation of amenorrhea is not difficult to justify because none of the available measures of ovarian estrogen production is completely reliable, for reasons already described. A serum FSH contributes to a more confident clinical assessment of ovarian function and helps differentiate patients with common chronic anovulatory conditions from those with more severe hypogonadism who otherwise may go unrecognized and who require further specific evaluation, counseling, or treatment. For example, when evaluation suggests marginal levels of estrogen production (e.g., serum estradiol 30–40 pgmL or scant bleeding after a progestin challenge), a low serum FSH can identify those who merit further evaluation to exclude pituitary and hypothalamic disease, as described below. Conversely, in women with normal levels of estrogen production, a moderately elevated FSH level (e.g., 10–15 IU/L) can reveal a diminished ovarian reserve in women who may be afforded the chance to actively pursue their reproductive goals before the opportunity is lost, if alerted to their advanced reproductive aging. Like the other measures of ovarian function, the serum FSH concentration must be interpreted carefully, in its clinical context. FSH levels can fluctuate unpredictably, particularly during the years immediately preceding the menopause, regardless whether it occurs prematurely or at the usual age.

Chronic Anovulation

When evaluation reveals clear evidence of normal ovarian estrogen production and the serum FSH level also is normal, the diagnosis of chronic anovulation is established. Hyperprolactinemia is one of the most common causes of anovulation and amenorrhea and, although less common, thyroid disorders are easily identified and treated. ***Measurement of the serum prolactin and thyroid-stimulating hormone (TSH) concentrations are therefore justified in all women with amenorrhea.*** For efficiency, both can be measured along with the serum FSH and estradiol levels, at the outset of the evaluation. When all are normal, no further evaluation is required.

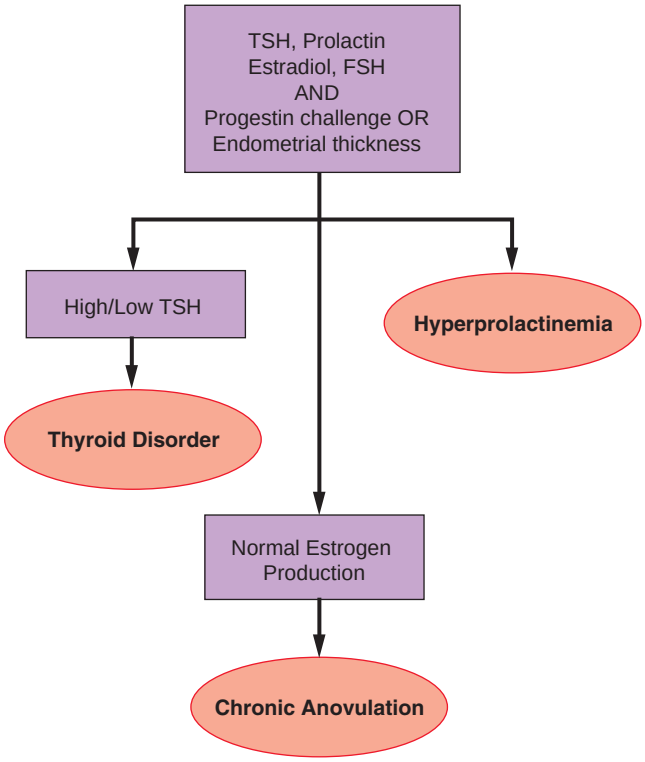
Besides thyroid and prolactin disorders, common and likely causes of chronic anovulation include PCOS, obesity, stress or exercise, and reproductive aging. In all but the last, anovulation can be attributed to a dysfunctional HPO axis in which gonadotropin secretion is sufficient to stimulate follicular development and estrogen production, but the system lacks the coordination required to achieve ovulation. Women with classical PCOS usually are easily recognized because they also exhibit signs of hyperandrogenism, unlike most women whose chronic anovulation relates solely to weight gain or obesity; the pathophysiology of the two disorders is complex and is discussed at length in Chapter 12 (PCOS) and Chapter 19 (obesity). Severe hirsutism or signs of virilization warrant additional specific evaluation to exclude enzyme deficiencies, androgen-secreting tumors, and Cushing's syndrome, as described in Chapter 13. The diagnosis of anovulation relating to emotional, nutritional, or physical stress is made by exclusion, but often is suggested by the medical history and physical examination. The management of chronic anovulation associated with thyroid disorders and hyperprolactinemia is summarized here.

Thryoid Disorders

The newest ultra-sensitive TSH assays now in common use detect both primary hypothyroidism (elevated TSH) and primary hyperthyroidism (low TSH); either may result in chronic anovulation and amenorrhea. Although only a few patients presenting with amenorrhea will have a thyroid disorder that is not clinically apparent, their exclusion and

treatment are so simple that routine measurement of TSH is justified; a return of ovulatory cycles typically follows the restoration of normal thyroid hormone levels. Any abnormal TSH value should be confirmed and accompanied by measurement of serum thyroxine (tetra-iodothyronine; T4, or free T4) to better define the nature and extent of the thyroid disorder. An elevated TSH with a normal free T4 concentration indicates a subclinical hypothyroidism, best viewed as a compensated state wherein normal levels of T4 are maintained, but only under increased levels of pituitary stimulation. ***Although observation and periodic re-evaluation are reasonable in patients with subclinical hypothyroidism, because not all will develop frank hypothyroidism, treatment is warranted in those with menstrual dysfunction or infertility***. In those with a low TSH and a normal free T4 level, serum tri-iodothyronine (T3) should be measured; an elevated T3 can identify hyperthyroidism that otherwise might escape detection. When the T3 also is normal, a subclinical hyperthyroidism is likely and should be followed carefully. On rare occasions, both TSH and free T4 levels are low, suggesting a secondary hypothyroidism of pituitary origin that requires additional evaluation to determine the cause and whether other pituitary functions also are affected.

A few women with hypothyroidism will develop a secondary hyperprolactinemia and even galactorrhea. The likelihood of hyperprolactinemia increases with the duration of hypothyroidism; galactorrhea is more common in young women with higher prolactin levels.[5] The mechanism probably involves both the gradual depletion of hypothalamic dopamine (the putative prolactin-inhibiting factor) and constant stimulation of pituitary lactotropes by thyrotropin-releasing hormone (TRH), which may cause pituitary hypertrophy or hyperplasia and sometimes even enlargement or erosion of the sella turcica.[6, 7] Although hormone levels rapidly normalize with appropriate treatment, the disappearance of breast secretions in those with galactorrhea is gradual and can take several months. Patients with primary hypothyroidism and hyperprolactinemia may present with either primary or secondary amenorrhea.[8]



Hyperprolactinemia

Hyperprolactinemia is among the most common causes of secondary amenorrhea and also may result in delayed puberty and primary amenorrhea when it arises before menarche. ***A serum prolactin concentration is therefore justified in all women with amenorrhea.*** A normal random measurement (<15–20 ng/mL in most clinical laboratories) excludes hyperprolactinemia. The prolactin level is fairly stable throughout the day but can increase transiently during sleep, exercise, breast stimulation, and meals. ***To avoid otherwise unnecessary and costly imaging, mildly elevated prolactin levels (20–40 ng/mL) are best repeated and confirmed before the diagnosis of hyperprolactinemia is made.***

The mechanism by which hyperprolactinemia results in anovulation and amenorrhea relates to a disruption or inhibition of the normal hypothalamic GnRH pulse rhythm, resulting in ineffective or frankly low levels of gonadotropin secretion. It may be that increased circulating prolactin levels stimulate a generalized increase in hypothalamic dopaminergic neuronal activity, intended to suppress prolactin secretion but also inhibiting GnRH neurons. In any case, the end result is anovulation or an even more profound hypogonadotropic hypogonadism, depending on the extent to which gonadotropin secretion is suppressed. Mild hyperprolactinemia (20–50 ng/mL) may cause only a short luteal phase, resulting from poor preovulatory follicular development.[9, 10] Moderate hyperprolactinemia (50–100 ng/mL) frequently causes oligomenorrhea or amenorrhea, and higher prolactin levels (>100 ng/mL) typically result in frank hypogonadism with low estrogen levels, and their clinical consequences (e.g., genitorurinary atrophy, osteopenia).[11, 12]

The symptom or finding of galactorrhea cannot reliably identify those whose amenorrhea results from hyperprolactinemia. Only about one-third of women with hyperprolactinemia exhibit galactorrhea, probably because breast milk production requires estrogen and hyperprolactinemia often results in anovulation or a more severe secondary hypogonadotropic hypogonadism and low circulating estrogen levels. The structural heterogeneity of prolactin offers another possible explanation. Prolactin circulates in various forms that have varying bioactivity (manifested by galactorrhea) and immunoactivity (recognition by immunoassay).[13–15] The predominant form (80–95%) is monomeric (molecular weight 23 kDa), which is more biologically active than larger glycosylated forms that may combine to form dimmers or trimers ("big prolactin," 50–60 kDa) and other even larger varieties (macroprolactin, >100 kDa), which result from the aggregation of smaller prolactin molecules bound together with immunoglobulins.[16] The larger molecular forms are cleared more slowly, predominate in women with hyperprolactinemia having normal menses, and result in minimal or no galactorrhea.[17] If suspected, the diagnosis of macroprolactinemia can be confirmed by asking the laboratory to pretreat the patient's serum with polyethylene glycol to precipitate the macroprolactin before performing the prolactin assay.[18] In women with mildly elevated prolactin levels, diagnosis of macroprolactinemia avoids unnecessary and costly imaging aimed at excluding pituitary and hypothalamic mass lesions.

Hyperprolactinemia has many causes that are discussed in detail in the chapter dedicated to the breast (Chapter 16) and briefly summarized here.[19] Hyperprolactinemia may result from hypothyroidism, prolactin-secreting pituitary adenomas, and other pituitary or hypothalamic tumors that may compress the pituitary stalk and disrupt the delivery of dopamine. A variety of drugs that lower dopamine levels or inhibit dopamine action may cause hyperprolactinemia, including amphetamines, benzodiazepines, butyrophenones, metaclopromide, methyldopa, opiates, phenothiazines, reserpine, and tricyclic antidepressants. Breast or chest wall surgery, cervical spine lesions, or herpes zoster (affecting the dermatome that includes the breast) may activate the afferent sensory neural pathway that stimulates prolactin secretion in a manner similar to suckling. Renal insufficiency and macroprolactinemia may cause hyperprolactinemia, due to decreased clearance. Rarely, hyperprolactinemia may result from ectopic prolactin secretion by pituitary tissue in the pharynx, by bronchogenic and renal-cell carcinomas, or by a gonadoblastoma or prolactinoma that may

arise in benign or malignant ovarian teratomas.[20–24] All possible causes must be considered and excluded; a careful history can eliminate most of the possibilities. When the cause reasonably may be attributed to a medication, a trial discontinuation or use of an alternative drug should be considered, in consultation with the prescribing physician. When that is not possible, further evaluation to exclude a pituitary or hypothalamic mass lesion is required.

Women with amenorrhea and hyperprolactinemia that cannot be attributed confidently to medication or another specific cause require further evaluation with imaging to exclude pituitary tumors and hypothalamic mass lesions. (see Evaluation of Pituitary Function, below). Pituitary adenomas and their management are discussed in detail in a later section of this chapter devoted specifically to pituitary causes of amenorrhea. Discussion here is limited to the treatment of hyperprolactinemia unassociated with any demonstrable sellar abnormality.

Treatment with a dopamine agonist restores ovulatory function and menses within several weeks in the large majority of women with hyperprolactinemia. Although the amount of breast secretions in those with galactorrhea decreases significantly over the same interval of time, complete cessation often takes considerably longer.[25] Both bromocriptine and cabergoline are highly effective. Bromocriptine has a relatively short half life, must be administered daily (at bedtime) or twice daily, and often is associated with gastrointestinal side effects such as nausea. Cabergoline is a selective dopamine receptor type 2 agonist having fewer side effects than bromocriptine, greater potency and a longer duration of action requiring less frequent administration (twice weekly), and can be effective in those who cannot tolerate or prove resistant to bromocriptine.[26, 27] However, cabergoline also has been associated with hypertrophic valvular heart disease when used in high doses (>3 mg daily) as in patients with Parkinson's disease; mitogenic stimulation of normally quiescent valve cells via activation of serotonin receptors is the suspected mechanism.[28, 29] Although the doses required for effective treatment of hyperprolactinemia are much lower, long-term use of even relatively low doses may increase the risk of valvular heart disease.[30, 31] Consequently, cabergoline should be used in the lowest dose required to normalize serum prolactin concentrations and a trial discontinuation of treatment should be attempted if prolactin levels have been normal for 2 or more years.[32] The dose of dopamine agonist treatment should be adjusted according to response, beginning with a low dose and increasing gradually as needed to normalize prolactin levels. In those who cannot tolerate oral treatment, vaginal administration is effective and associated with fewer size effects.[33, 34] Either drug may be used in women planning to conceive since both appear to be safe in early pregnancy.[35, 36]

Unfortunately, amenorrhea and galactorrhea often promptly recur within weeks after discontinuation of dopamine agonist treatment and most therefore require long-term therapy. Treatment with a dopamine agonist is the obvious choice when the objective is ovulation induction and pregnancy or the elimination of troublesome galactorrhea. However, for those with neither specific indication, alternative treatments deserve careful consideration. ***Although treatment with a dopamine agonist certainly is a logical choice, it is by no means the only choice or necessarily the best choice for all women with hyperprolactinemia and amenorrhea. It is important to remember that treatment should be focused on the patient, and not on the prolactin level.*** Hyperprolactinemia itself poses no particular health risks. In women not at risk for an unwanted pregnancy, cyclic progestin therapy will prevent the clinical consequences of chronic unopposed estrogen exposure in those who are not frankly hypogonadal, and in those who are, physiologic cyclic or combined estrogen/progestin treatment will prevent the consequences of chronic estrogen deficiency. In women who need contraception, treatment with a low-dose oral contraceptive achieves the same goals. In the past, treatment with exogenous estrogen was considered contraindicated for women with hyperprolactinemia due to fear it might aggravate the underlying pathophysiology or promote growth of a pituitary tumor, but experience has shown that

hormone therapy and oral contraceptives pose no such risks.[37, 38] The same treatments are useful in the management of women with medication-induced hyperprolactinemia and hypogonadism when the drug cannot be discontinued or another substituted. Dopamine agonists are best avoided in patients with medication-induced hyperprolactinemia, because they may interfere with or counteract the dopamine antagonist properties of their primary treatment.

General Management

All patients with chronic anovulation require management, and with the limited evaluation described here, treatment can be implemented immediately. Clinicians are keenly aware that normal endometrium can progress to hyperplasia, atypia, and cancer within a relatively short interval of time. However, too often they believe the problem is relevant only in older aged women. ***The critical factor is not age, but the duration of exposure to unopposed estrogen stimulation.*** Young women who remain anovulatory for long periods of time can, and do, develop endometrial cancer.[39–42] Although endometrial sampling is not indicated for all women with chronic anovulation, it should be considered seriously for those at greatest risk for endometrial pathology. Obese women and those with PCOS are the most likely candidates because obesity, hyperinsulinemia, and hyperandrogenism are known risk factors for endometrial neoplasia.[43, 44] Screening by endometrial thickness generally has poor positive predictive value for detecting endometrial pathology, but may be useful for identifying individuals at very low risk in whom biopsy safely may be omitted. No studies correlating endometrial thickness and histology in premenopausal women with amenorrhea have been performed. However, in premenopausal women with abnormal uterine bleeding, no serious pathology was found in those having an endometrial thickness less than 8 mm,[45] and in asymptomatic postmenopausal women, an endometrial thickness less than 5–6 mm has greater than 99% negative predictive value for endometrial disease.[46, 47] Whereas some studies have suggested that amenorrheic women also may be at increased risk for developing breast cancer,[48] the weight of available evidence suggests an inverse association between breast cancer risk and chronic anovulation (discussed in Chapter 16).[49]

At a minimum, women with chronic anovulation require periodic treatment with a progestin, to induce predictable menses and protect against the risk of developing endometrial pathology. For example, medroxyprogesteone acetate 5–10 mg daily can be administered for the first 12–14 days of each, or at least alternate, months; experience with varying hormone treatment regimens has demonstrated that treatment for an interval greater than 10 days is required to effectively counteract the growth-promoting effects of continuous estrogen exposure. ***It is important to note that cyclic treatment with a progestin, at physiologic doses, does not change the intrinsic rhythm of the HPO axis and will not prevent a sporadic ovulation.*** Therefore, if menses do not occur at the expected time, pregnancy must be considered and excluded. Absent bleeding after a course of progestin treatment also may indicate that estrogen production has fallen to grossly low levels and signal the need for further evaluation, as described in the section that follows. When reliable contraception is required, cyclic treatment with a low-dose oral contraceptive pill or a vaginal contraceptive ring is the obvious and better choice. There is no evidence that hormonal contraception has any impact, positive or negative, on menstrual cyclicity after treatment is discontinued.

For women with chronic anovulation having pregnancy as their goal, treatment should be aimed at inducing normal ovulatory cycles. Methods for ovulation induction are described in detail in Chapter 31. For women with thyroid disorders, specific treatment to restore normal thyroid function is indicated. For those with hyperprolactinemia, a dopamine agonist is the treatment of choice. Most having neither will respond to treatment with clomiphene citrate, reserving exogenous gonadotropin stimulation for those who do not.

Ovarian Failure

When evaluation reveals clear evidence of low ovarian estrogen production and the serum FSH level is consistently high, the diagnosis of ovarian failure is established. Although premature follicular depletion is the cause in almost all cases, additional specific evaluation is indicated to exclude chromosomal and other genetic abnormalities and autoimmune disease that may have important potential health implications for the patient and other members of her family. The elements and purpose of the expanded diagnostic evaluation are summarized here. The known causes of ovarian failure and the disorders deserving specific consideration and exclusion are discussed at greater length in a later section of this chapter.

Karyotype

In all patients under age 30 with a diagnosis of ovarian failure, a karyotype should be obtained to exclude chromosomal translocations, deletions, and mosaicism that might offer an obvious explanation. A karyotype also identifies those having a Y chromosome in whom gonadectomy is indicated due to the significant risk for malignant transformation in occult testicular elements (20–30%). Signs of virilization cannot reliably identify the subset of women at risk because many having a Y chromosome exhibit no signs of excess androgen production. In women over age 30, ovarian failure reasonably can be regarded as premature menopause. Karyotype after age 30 generally is unnecessary because most tumors in patients with a Y chromosome arise before age 20, and virtually all before the age of 30.[50, 51] After age 30, women with short stature or a family history of early menopause still merit a karyotype to exclude X chromosome deletions and translocations that may affect other family members.[52–55] Otherwise, pelvic ultrasonography can exclude the rare tumor not recognized previously.

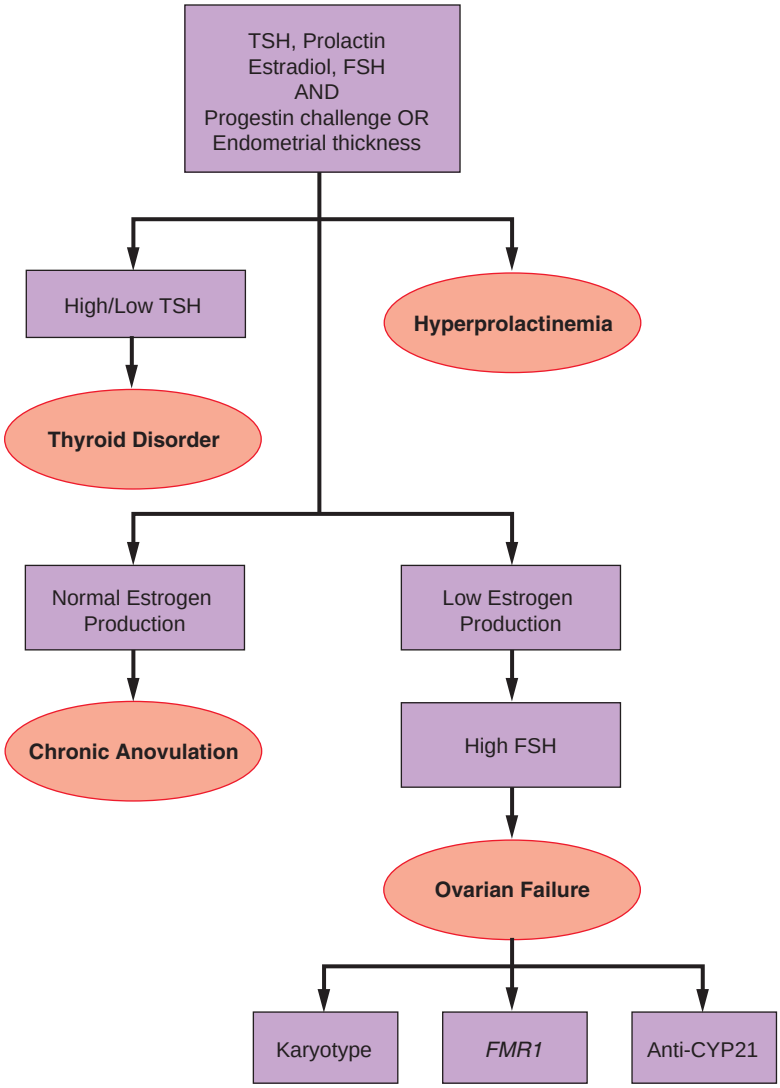
Fragile X (FMR1) Premutations

Fragile X syndrome is the most common inherited cause of mental retardation and autism and results from abnormal expansion of an unstable trinucleotide (CGG) repeat sequence in the *FMR1* (Fragile X Mental Retardation) gene, located on the long arm of the X chromosome (Xq27.3). The gene normally contains about 30 CGG repeats, but in those with Fragile X syndrome, the number exceeds 200. Convincing evidence has demonstrated an association between premature ovarian failure (POF) and fragile X "premutations," characterized by 55–200 CGG repeats. Whereas the full mutation silences the *FMR1* gene, resulting in little or no production of the corresponding mRNA or gene product (fragile X mental retardation protein, FMRP), the POF associated with premutations may reflect *FMR1* mRNA gain-of-function toxicity.[56] Women with premutations often exhibit endocrine evidence of early ovarian aging and up to one-third have an early menopause. The prevalence of premutations is approximately 14% in women with familial POF, and between 1% and 7% in sporadic cases of POF.[56, 57] ***Women with POF therefore should be offered testing for FMR1 premutations.***[58] Women who carry Fragile X premutations (and any also affected children or siblings) are at risk for having a child with Fragile X syndrome, because the length of the CGG repeat sequence is unstable and may expand to a full mutation when passed from mother to offspring. The inheritance and the implications of premutations are complex and affected women therefore should receive formal genetic counseling.

Autoimmune Screening

Ovarian failure sometimes may be the consequence of autoimmune disease.[59] Addison's disease (autoimmune adrenocortical insufficiency) has the strongest association with POF; the presence of autoantibodies to steroid-producing cells and observations of a lymphocytic infiltrate in the ovaries of affected individuals suggests the mechanism (autoimmune

oophoritis).[60] The prevalence of other autoimmune diseases (e.g., thyroid autoimmunity, Type I diabetes, and myasthenia gravis) is higher among women with POF than in the general population, but there is no direct or compelling evidence to indicate a cause and effect relationship. Autoimmune ovarian failure generally occurs as part of a specific autoimmune polyendocrine syndrome (APS) that includes adrenal insufficiency. However, because POF may precede onset of adrenal insufficiency by several years, the autoimmune cause may not be recognized when the diagnosis of POF is first made.[59] Women with POF should be tested for anti-adrenal antibodies (most easily demonstrated against the 21-hydroxylase enzyme, CYP21), and for anti-thyroid antibodies (anti-thyroid peroxidase and anti-thyroglobulain antibodies). ***The presence of anti-adrenal antibodies strongly implies autoimmune oophoritis as the cause of POF and idenifies women who should be carefully evaluated and followed to exclude adrenal insufficiency. The presence of thyroid autoantibodies does not prove autoimmune ovarian failure, but identifies women at risk for developing autoimmune thyroid disorders.*** Routine screening for other autoimmune endocrine disorders is unnecessary and can be reserved for those with clinical indications.[61]



Evaluation of Pituitary Function

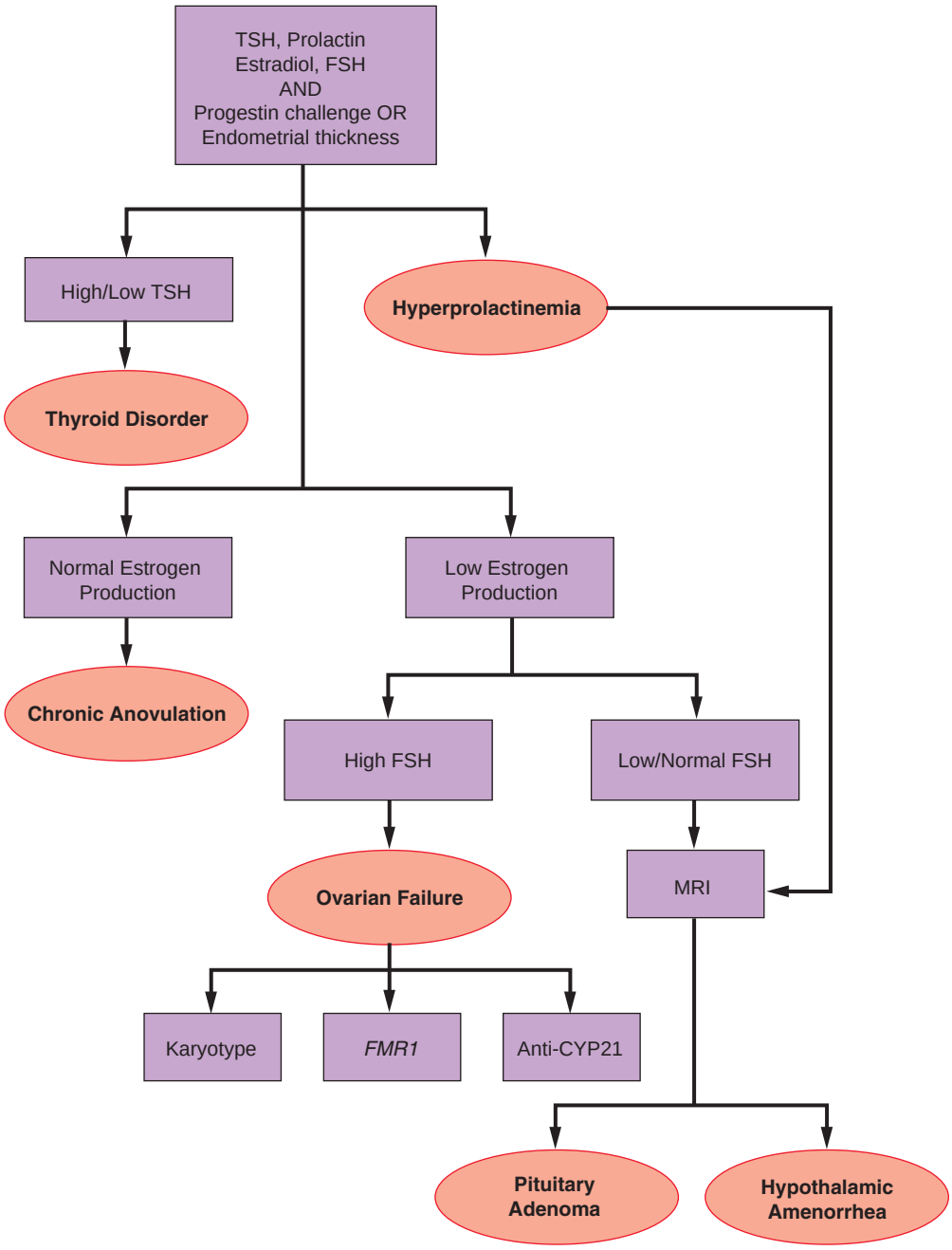
The normal feedback relationship between ovarian estrogen production and pituitary gonadotropin secretion dictates that low estrogen levels should cause a compensatory increase in FSH release to stimulate ovarian follicular development and estrogen secretion, just as they do during the early follicular phase of the normal cycle. When estrogen production is abnormally low, a low serum FSH concentration (<5 IU/L) indicates that inadequate or ineffective gonadotropin secretion is the cause and that even basic central feedback mechanisms in the HPO axis are not functioning. ***When estrogen levels are clearly low, a serum FSH level in the low normal range (5–10 IU/L) has the same interpretation and clinical implication***, for two reasons. First, because the FSH level should be high when estrogen production is grossly low, even a "normal" value is, in fact, abnormally low in that clinical context. Second, although the measured level of immunoreactive FSH may be normal, the level of biologically active FSH clearly is not; if it were, follicular growth and estrogen production would be maintained. The biological activity of glycoprotein hormones varies with their carbohydrate moieties (discussed in Chapter 2) and evidence indicates that women with hypogonadotropic hypogonadism may secrete gonadotropins having altered patterns of glycosylation and reduced biological activity.[62] Indeed, most women with hypogonadotropic hypogonadism have normal serum gonadotropin concentrations; extremely low or undetectable gonadotropin levels typically are observed only in those with large pituitary tumors or in patients with anorexia nervosa.

Imaging

When there is no clear explanation for hypogonadotropic hypogonadism (e.g., significant physical, nutritional, or emotional stress) or for hyperprolactinemia (e.g., medications), further evaluation with imaging is indicated to exclude tumors and to help distinguish between pituitary and hypothalamic causes. The method of choice is MRI (with gadolinium contrast) because it is more sensitive and accurate than other imaging techniques for detection of abnormalities within and near the sella turcica.[63] MRI can demonstrate the nearby optic chiasm and also can detect blood, allowing hemorrhage and vascular abnormalities to be distinguished from other sellar mass lesions. Most sellar masses are pituitary adenomas, which account for 10% of all intracranial neoplasms. Other less common mass lesions in or near the sella include benign tumors (craniopharyngioma, hamartoma, meningioma), pituitary hyperplasia (thryotroph or gonadotroph hyperplasia due to long-standing primary hypothyroidism or gonadal failure), malignant tumors (germ cell, sarcoma, chordoma, carcinoma, lymphoma), metastases (lung, breast), cysts (Rathke's cleft, arachnoid, dermoid), pituitary abscess, lymphocystic hypophpysitis, sarcoidosis, tuberculosis, and carotid arteriovenous fistula.

Although mass lesions are the most obvious abnormality to be excluded, other rare possibilities include Sheehan syndrome (pituitary infarct resulting from hypotension associated with postpartum hemorrhage), infiltrative hemosiderosis relating to frequent transfusions or hereditary hemochromatosis, traumatic brain injury,[64] and mutations in the GnRH receptor.[65]

In the absence of any demonstrable mass lesion in the sellar region or relevant history suggesting another specific cause for pituitary damage, there is no need to perform any additional specific pituitary function tests. The clinical signs and symptoms associated with different types of functional and nonfunctional pituitary tumors and other specific pituitary causes of gonadotropin deficiency are discussed in a later section of this chapter.



Evaluation of Hypothalamic Function

When imaging reveals no mass lesion and there is no reason to suspect other specific pituitary pathology, the diagnosis is ***functional hypothalamic amenorrhea***, by exclusion. The pathophysiology of the disorder relates to a suppressed or otherwise abnormal pattern of pulsatile hypothalamic GnRH secretion, resulting in decreased gonadotropin secretion, absent follicular development, anovulation, and low serum concentrations of estradiol. The serum FSH (and LH, if measured) concentration is low or in the normal range; often, but not always, the FSH level is higher than LH as in prepubertal girls.

Unfortunately, there is no simple way to test, manipulate, or measure hypothalamic function to prove a GnRH deficiency. Whereas one might anticipate that the LH response to a

bolus of exogenous GnRH (e.g., 100 μg, administered subcutaneously) would be revealing, experience has shown it may be normal (>10 IU/L) or low in women with pituitary or hypothalamic disease. The response to repeated bolus GnRH administration (e.g., 24 hours after the first) may be somewhat more informative, due to the self-priming effect that GnRH has on its own receptor.[66] The administration of exogenous pulsatile GnRH using a programmable infusion pump can restore normal gonadotropin secretion and menstrual function and induce ovulation in women with hypothalamic amenorrhea,[67, 68] but the considerable costs and logistical challenges of such treatment make it impractical and impossible to justify as a diagnostic test.

In most cases, the probable cause of hypothalamic amenorrhea can be identified, such as extreme emotional stress, acute weight loss or chronic malnutrition, or strenuous physical exercise. However, in others with hypothalamic amenorrhea, no obvious cause or precipitating event can be identified. Rare individuals with idiopathic hypogonadotropic hypogonadism may present with primary amenorrhea and infantile sexual development due to a congenital GnRH deficiency, resulting from the failure of GnRH neuronal development during embryogenesis or from mutations in the GnRH receptor, but specific evaluation to identify such abnormalities is not clinically necessary or indicated, except perhaps when other family members are affected. The causes of hypothalamic amenorrhea and its management are discussed at length in a later section of this chapter devoted to specific disorders of the hypothalamus.

Specific Causes of Amenorrhea

With only modest effort, time, and expense, the problem of amenorrhea has been dissected by systematic evaluation of the organ systems involved in menstrual function—the genital outflow tract and uterus, the ovary, the anterior pituitary, and the hypothalamus. Once the anatomic level of the disorder has been so defined, attention can turn to making a specific diagnosis. This section of the chapter considers each of the major causes of amenorrhea and their management, organized by organ system.

Disorders of the Genital Outflow Tract and Uterus

As the cause of amenorrhea, disorders of the genital outflow tract and uterus are relatively uncommon. Congenital developmental anomalies of the genital outflow tract and uterus result from the failure of vertical fusion (imperforate hymen, transverse vaginal septum or cervical atresia) or from failure of müllerian duct development (vaginal/müllerian agenesis, AIS) and generally present at or near the expected time of menarche with primary amenorrhea. The only disorders of the genital outflow tract or uterus associated with normal genital tract anatomy are cervical stenosis and intrauterine adhesions (Asherman syndrome) or other endometrial damage resulting from surgical trauma or infection. All are acquired conditions that present as secondary amenorrhea with an onset that typically correlates closely with the time of previous insult.

Imperforate Hymen

The hymen is formed by invagination of the posterior wall of the urogenital sinus and usually ruptures spontaneously during the perinatal period. Although most cases of

imperforate hymen occur sporadically, reports of families with several affected members suggest that some cases may have a genetic and heritable cause.[69]

Typically, patients with an imperforate hymen present at the expected time of menarche with complaints of cyclic perineal, pelvic or abdominal pressure or pain that results from the gradual accumulation of obstructed menstrual flow (cryptomenorrhea), and exhibit otherwise normal, symmetrical secondary sexual development for age. They also may present with acute urinary retention due to compression of the urethra and bladder by a grossly distended lower vagina.[70] ***The genital examination reveals no obvious vaginal orifice and a thin, often bulging, blue perineal membrane at the inferior limit of a palpable, fluctuant mass (hematocolpos).***

The treatment of women with an imperforate hymen centers on providing relief of symptoms related to accumulated menstrual fluid and debris. Definitive surgery should be accomplished as soon as possible because delay can lead to infertility due to inflammatory changes and to the development of severe endometriosis. Surgical correction of an imperforate hymen is straightforward. The classical procedure is to make a simple cruciate incision in the hymen to the base of the hymeneal ring and to excise its central portion to allow drainage of sequestered menstrual fluid and subsequent normal menstruation. Alternatively, to avoid any risk of damage to the hymeneal ring (a sign of virginity important to some individuals and in some cultures), a sterile puncture can be made in the center of the distended membrane and enlarged to approximately 0.5 cm in diameter to allow insertion of a 16F Foley catheter. After thorough drainage of the vagina via irrigation with sterile saline, the catheter is left in place for approximately 2 weeks to allow further drainage from the vagina and upper genital tract. A single dose of prophylactic antibiotics is prudent and estrogen cream applied locally to the hymeneal ring helps to encourage re-epithelialization.[71]

Transverse Vaginal Septum/Cervical Atresia

A transverse vaginal septum results when the vaginal plate, formed from the fused sinovaginal bulbs, fails to break down or canalize during embryogenesis. As could be expected, girls with a transverse vaginal septum or cervical atresia, like those with an imperforate hymen, generally present at or soon after the age of expected menarche with complaints of cyclic pelvic or abdominal pain due to obstructed menses and exhibit symmetrical, age-appropriate secondary sexual development. ***Physical examination reveals a normal vaginal orifice, a shortened vagina of varying length, no visible cervix, and a palpable hematocolpos in the proximal vaginal segment above the obstruction and/or a pelvic mass resulting from hematometra and hematosalpinges.*** A Valsalva maneuver will cause distention at the intoitus in those with an imperforate hymen, but not in those with a transverse vaginal septum or cervical atresia, and can help to distinguish the two. Imaging is necessary to define the anatomy of the disorder but laboratory investigation generally is not required. Pelvic ultrasonography can reveal the level and extent of the hematocolpos and any associated hematometra or hematosalpinges. However, abdominal/pelvic MRI provides greater anatomical detail and is recommended to more clearly define the length of the atretic segment between the lower and upper vaginas,[72, 73] information that is essential to planning surgical treatment. ***The temptation to insert a needle for diagnostic purposes must be resisted to avoid the risk of converting a hematocolpos into a pyocolpos.*** In rare instances, laparoscopy may be required to clarify the anatomy of the developmental anomaly. Transverse vaginal septum and cervical atresia may be accompanied by abnormalities of the upper reproductive tract, such as absent segments or atresia of the fallopian tubes or unilateral absence of the fallopian tube and ovary.[74] Unfortunately, chronic retrograde menstruation frequently results in pelvic endometriosis and adhesions, which can be severe.

In all instances, every effort should be made to incise and drain the sequestered menstrual fluid from below, at the level of the obstruction. Even in complicated circumstances, continuity of the lower genital tract usually can be achieved successfully. Operative excision of painful masses from above risks damage to the bladder, ureters, and rectum and unnecessarily removes distended but otherwise healthy reproductive organs. The surgical management of a transverse vaginal septum can be challenging, and because they also are infrequently encountered, often requires consultation with specialists having the necessary training and experience. Simply described, the procedure involves excision of the septum or dissection of the atretic segment and primary anastamosis of the margins of the lower and upper vaginal canals over the site of the defect. Atretic segments of greater length may require application of a graft to bridge the gap between the lower and uppper vaginas. ***Because septa that appear relatively thin by physical examination and MRI may be significantly larger after decompression of the proximal hematocolpos, preparations for surgery should consider the possibility that a graft may be required.***

The best surgical management for rare women with cervical atresia is controversial. Ideally, the goal would be to create a functional vagina and to preserve the uterus and fertility, but experience has proven that such heroic efforts may be associated with serious postoperative complications such as peritonitis and sepsis, recurrent obstruction, and persistent infertility, prompting many to view hysterectomy as the best management option. However, conservative surgical treatment is reasonable to consider in selected individuals. The best candidates are those recognized early, before they develop severe pelvic endometriosis and adhesions, having a well-developed lower vagina,[75] although successful reconstruction and pregnancy can be achieved even in those who also require vaginoplasty.[76, 77]

Müllerian Agenesis (Mayer-Rokitansky-Küster-Hauser Syndrome)

The failure of müllerian development is a relatively common cause of primary amenorrhea, much more frequently encountered than AIS and second only to gonadal dysgenesis in prevalence;[78] in Finland, the incidence is approximately 1 in 5,000 newborn girls.[79] The cause is unknown. Although usually sporadic, some cases of müllerian agensis are associated with chromosomal translocations or occur in familial aggregates, suggesting a genetic basis for the disorder. Logically, müllerian agenesis might be attributed to an activating mutation in the gene encoding antimüllerian hormone (AMH) or its receptor, causing excess AMH activity. Inactivating mutations in these genes causing persistence of müllerian structures in otherwise normally virilized males have been described.[80, 81] However, no activating mutations have been identified in patients with mullerian agenesis.[82] The prevalence of a mutation in the galactose-1-phosphate uridyl transferase (*GALT*) gene (different from that associated with classical galactosemia) is increased in daughters with müllerian agenesis and their mothers.[83] The observation suggests that errors in fetal or maternal galactose metabolism resulting in increased intrauterine galactose exposure may have adverse effects on müllerian development, consistent with studies in rodents wherein a high galactose diet during pregnancy delayed vaginal opening in female offspring.[84] Given the relationship between classical galactosemia and premature ovarian failure, patients with müllerian agenesis who carry such a variant *GALT* gene mutation may be at increased risk for the same.

Patients with müllerian agenesis typically present in late adolescence or as young adults, well after menarche was expected, with primary amenorrhea as their only complaint. ***They exhibit normal, symmetrical breast and pubic hair development, no visible vagina, and have no symptoms or signs of crytomenorrhea because the rudimentary uteri contain no functional endometrium.*** However, in approximately 10%, functional islands of endometrium may result in a hematometra and symptoms of cyclic pain.[85, 86] Two forms

of the disorder have been described. Type A is characterized by symmetrical, muscular, rudimentary uteri, and normal fallopian tubes, and Type B by asymmetrical rudimentary uteri and absent or hypoplastic fallopian tubes.[87] In the great majority of patients with müllerian agenesis, the ovaries are entirely normal, but one or both also may be undescended, hypoplastic, or associated with an inguinal hernia. Urologic anomalies are relatively common (15–40%), particularly in Type B müllerian agenesis, and include unilateral renal agenesis, ectopic or horseshoe kidney, and duplication of the collecting system(s).[85, 86] Skeletal malformations involving the vertebrae, the ribs, or the pelvis are observed in 10–15% of patients; some of the more common abnormalities include hemivetebrae leading to scoliosis and the Klippel-Feil syndrome, characterized by a short neck, low hairline, limited range of motion, and sometimes pain and neurologic symptoms, all relating to one or more fused cervical vertebrae.

Although müllerian agenesis usually can be diagnosed by medical history and physical examination alone, additional evaluation is warranted to establish the diagnosis and to identify any of the urologic (renal ultrasonography) and skeletal anomalies (spinal X-rays) associated with the disorder. After puberty, a serum testosterone concentration in the normal female range effectively excludes AIS (discussed below). ***However, because patients with müllerian agenesis can exhibit characteristics similar to those observed in some types of male pseudohermaphroditism, a karyotype is justified and definitive.*** When examination raises suspicion that a uterine structure may be present, imaging is indicated. Ultrasonography may help to define the size and symmetry of any pelvic reproductive organs, but MRI is more accurate and is indicated when doubt remains.[88, 89] Laparoscopy usually is not necessary for diagnosis of müllerian agenesis. Although imaging frequently does not agree completely with surgical observations, detailed knowledge of the pelvic anatomy is not often needed.[90] Surgery generally is indicated only in those with symptoms relating to hematometra, endometriosis, or a hernia into the inguinal canal.

The primary goal of treatment in women with müllerian agenesis—creation of a functional vagina—can be accomplished with a variety of methods, when the time is appropriate. In the large majority of cases, progressive vaginal dilation as originally described by Frank [91] and later by others,[92] is an appropriate and effective first choice. In motivated patients, the technique is highly successful and can create a functional vagina within 3 to 6 months.[93] The procedure involves applying pressure to the point of moderate discomfort for an interval of 20–30 minutes daily, using commercially available vaginal dilators. Initially, pressure is directed posteriorly, to create a shallow pouch. After approximately 2 weeks, pressure shifts to the usual axis of the vagina. After the desired depth is achieved, dilators of increasing diameter will expand the vagina to a functional size. A variation on the technique uses a tight-fitting garment to hold the dilator in place, maintaining pressure by leaning forward on a bicycle seat mounted on a stool, or even on a bicycle.[94]

Operative treatment of women with müllerian agenesis generally can be reserved for those who are unable or unwilling to dedicate themselves to a program of progressive vaginal dilation and for those in whom earnest efforts fail. The traditional McIndoe procedure for surgical creation of a neovagina involves dissection of the rectovaginal space and placement of a skin graft, held in place with a soft mold until the graft becomes established.[95] Subsequently, regular intercourse or vaginal dilation must be maintained to avoid risk of fibrosis and loss of function. The alternative Vecchietti operation involved internal stretching of the vaginal dimple after surgical abdominal and vaginal dissection of the vesicorectal space.[96] A modification of the Vecchietti operation, performed laparoscopically, has emerged as an attractive and effective option for surgical creation of a neovagina.[97, 98] The procedure employs a specially designed system including a spring-loaded traction device, positioned on the abdomen, connected to the tip of a dilator, positioned at the introitus, via paired threads introduced with a needle inserted through small incisions in the lower abdomen and guided beneath the peritoneum to penetrate

the introitus in the midline. Tension on the threads is adjusted to maintain traction on the dilator, which is drawn upward gradually, invaginating the tissue to produce a neovagina 7–8 cm in depth over an interval of 7–10 days. Thereafter, standard vaginal dilators are employed to further extend and expand the vagina to functional dimensions over a period of a few weeks. Evidence indicates that the procedure results in a quality of sexual experience comparable to that in a sample of healthy, age-matched women of equivalent cultural and social status.[98]

Reassurance and support are important elements of the management of vaginal/müllerian agenesis. Affected women should be counseled that although they are infertile, normal sexual function can be expected and that genetic offspring can be achieved by in vitro fertilization (IVF) using oocytes retrieved from their own normal ovaries and the sperm of their chosen partner, with subsequent transfer of embryos to a gestational surrogate. [99, 100] An analysis of 34 live births resulting from IVF in 58 women with müllerian agenesis revealed no evidence to suggest a dominant pattern of inheritance and demonstrated that IVF, combined with gestational surrogacy, is a realstic option for patients with the disorder.[100, 101]

Androgen Insensitivity Syndrome

Complete AIS (testicular feminization) is a form of male pseudohermaphroditism, the term referring to the gonadal sex (male) and the contrasting phenotype (female). The disorder, discussed in greater detail in Chapter 9 as a cause of abnormal sexual development, is the third most common cause of primary amenorrhea, after gonadal dysgenesis and müllerian agenesis. Patients with AIS have a normal male karyotype (46,XY) and testes that produce both testosterone and AMH. However, an inactivating mutation in the gene encoding the intracellular androgen receptor (located on the long arm of the X chromosome, Xq) results in an end organ insensitivity to androgen actions that prevents normal masculinization of the internal and external genitalia during embryonic development. Consequently, the external genitalia are those of a female (absent androgen action), the cervix and uterus are absent (due to normal AMH action), and the vagina is short and ends blindly (derived only from the urogenital sinus).

Patients with complete AIS appear normal at birth. Growth and development during childhood also are generally normal, although overall height usually is above average and the body habitus somewhat eunuchoid (long arms, large hands and feet). At puberty, the breasts develop, driven by estrogen derived from the peripheral conversion of high circulating testosterone levels, unopposed by the actions of androgen. The breasts may become relatively large and have subtle abnormalities; lacking the actions of progesterone, they have little glandular tissue, small nipples and pale areolae. The labia minora usually are underdeveloped and the vagina is short and ends blindly. Pubic and axillary hair does not develop, due to the absence of androgen stimulation. The testes may be intra-abdominal, but often are partially descended; more than half of patients with complete AIS have an inguinal hernia. The testes frequently are palpable in the inguinal canals, most commonly at the level of the external inguinal ring. They generally resemble any cryptorchid testes but may be nodular. After puberty, the testes contain immature seminiferous tubules lined by immature germ cells and Sertoli cells, with no evidence of spermatogenesis.

Patients with complete AIS most commonly present after the age of puberty in late adolescence or as young adults with primary amenorrhea. ***They exhibit asymmetrical secondary sexual development (breast development with absent or scant pubic hair), a short vagina with no visible cervix, and have no other symptoms or complaints.*** They also may be recognized at birth or in childhood when they may present with an inguinal mass or hernia,

particularly when the disorder is reasonably suspected because other family members such as a sister or maternal aunt are affected. Diagnosis usually is not difficult. Patients with complete AIS generally are easily distinguished from those with müllerian agenesis by the absence of pubic and axillary hair, and from those with an imperforate hymen or transverse vaginal septum by the absence of a uterus and symptoms relating to obstructed menstrual flow. In some cases, the presence of some pubic hair, due to incomplete penetrance, can be confusing or misleading. ***A serum testosterone concentration easily distinguishes patients with AIS because levels are normal or modestly elevated above the range observed in normal males and well above the normal range for females.*** Serum LH levels also are elevated, reflecting androgen insensitivity at the hypothalamic-pituitary level. A karyotype (46,XY) firmly establishes the diagnosis.

In rare cases of ***incomplete androgen insensitivity***, the sensitivity to androgens is greater. Consequently, pubic hair growth may accompany breast development and the clitoris may enlarge, or a phallus even may be present.[102] In other rare individuals having a deficiency of the enzyme ***17β-hydroxysteroid dehydrogenase*** (type 3), which catalyzes the conversion of androstenedione to testosterone in testicular Leydig cells, the clinical presentation may be similar, but due to impaired testosterone production rather than abnormalities in the androgen receptor. When necessary, the two disorders can be differentiated by molecular analysis of the genes encoding the androgen receptor (*AR*) and the enzyme (*17HSDB3*).[103, 104]

The treatment of patients with complete AIS has two major components, one focusing on creation of a functional vagina, and another relating to the risk for developing malignancy in the cryptorchid testes. In patients with AIS, the options for creation of a neovagina are the same as in those with müllerian agenesis—progressive vaginal dilation and vaginoplasty. The short but distinct vagina observed in most patients speeds the progress of efforts at vaginal dilation. Good results also can be expected with surgical treatment, when necessary.[105] Gonadectomy is indicated because the incidence of neoplasia in cryptorchid testes is relatively high. In one early series of 50 cases, 11 malignancies, 15 adenomas, and 10 benign cysts were observed: a 22% incidence of malignancy and a 52% overall incidence of neoplasia.[106] More recent series suggest a lower 5–10% overall incidence of gonadal tumors.[50, 102, 107, 108] Whereas gonadectomy is recommended at time of diagnosis in other intersex states such as XY gonadal dysgenesis (Swyer syndrome), it is better delayed in those with AIS, for two reasons. First, the smooth pubertal development that results from endogenous hormone production is difficult to achieve with exogenous hormone treatment, and second, gonadal tumors develop less often in patients with AIS and rarely before puberty. Therefore, gonadectomy and hormone therapy (physiologic estrogen treatment) generally are best postponed until after pubertal development is complete, by approximately age 16–18. ***Complete AIS is the only exception to the rule that gonads with a Y chromosome should be removed as soon as a diagnosis is made.*** Gonadectomy usually can be accomplished endoscopically with relative ease when the testes reside within the abdomen, and via inguinal incisions when they are partially descended.[109, 110] In patients with the incomplete form of AIS, surgery should not be postponed because prompt gonadectomy will prevent further unwanted virilization.

In years now past, conventional wisdom warned against unthinking and "needless" disclosure of the true gonadal and chromosomal sex to patients with complete AIS for fear of undermining gender identity, but that attitude has changed. Although infertile, patients with AIS have a completely female gender identity that should be reinforced, not questioned. We strongly advocate combining a truthful education with appropriate psychological counseling for both patient and parents. Patients want, deserve, and appreciate a fuller understanding of themselves and the disorder. Moreover, because the public now has ready access to sophisticated medical information, secrecy also is no longer a practical possibility. One excellent resource is the Androgen Insensitivity Syndrome Support Group, headquartered in the United Kingdom (http://www.aissg.org/).

Cervical Stenosis

Severe cervical stenosis with complete outflow obstruction is a rare complication of cervical conization procedures or other surgical treatments for cervical intraepithelial neoplasia. When cervical stenosis causes symptoms, worsening dysmenorrhea or prolonged light staining or spotting after menses are the most common complaints; amenorrhea is a rare occurrence. Compared to other methods for limiting blood loss during conization, elective suturing appears to increase the risk for subsequent cervical stenosis and amenorrhea.[111] In women with amenorrhea who have had a previous conization or other cervical surgery or ablative treatment, simple uterine sounding will establish the diagnosis of cervical stenosis and transvaginal ultrasonography will reveal any associated hematometra. The treatment for cervical stenosis is careful dilation, ideally performed under ultrasound guidance. Temporary placement of a urinary or specialized balloon catheter for an interval of approximately 2 weeks provides ongoing drainage of the uterine cavity and may help to prevent recurrence.[112]

Asherman Syndrome (Intrauterine Adhesions)

Asherman syndrome, first described by Joseph Asherman in 1948 and called "amenorrhoea traumatica,"[113] results from intrauterine adhesions that obstruct or obliterate the uterine cavity, as a consequence of trauma. Risk for developing intrauterine adhesions is increased by inflammation, as may result from endometritis or retained products of conception, and when the endometrium is relatively thin and inactive, as it is during the postpartum period. Consequently, most cases arise in close temporal proximity to a pregnancy and are associated with surgical trauma, primarily curettage.[114] In the original series of 29 cases described by Asherman, 11 had a previous postpartum hemorrhage, 15 a spontaneous abortion, 2 an elective termination, and 1 had a hydatidiform mole.[113] Asherman syndrome is an uncommon but recognized complication of cesarean section, abdominal or hysteroscopic myomectomy or metroplasy, and uterine artery embolization,[115] Elective endometrial ablation procedures for the management of menorrhagia frequently result in amenorrhea, by intent, but most do not have intrauterine adhesions. End organ endometrial damage and adhesions also may result from intrauterine infections such as tuberculosis and schistosomiasis, which are rare in the United States but not in other regions of the world.[116, 117]

Although Asherman syndrome from any cause may result in amenorrhea, most women with intrauterine adhesions present with dysmenorrhea, hypomenorrhea, infertility, or recurrent pregnancy loss, rather than amenorrhea. ***The diagnosis of Asherman syndrome is based primarily on a high index of suspicion, based on history. In women whose history suggests the possibility, scant or no withdrawal bleeding after sequential treatment with exogenous estrogen (e.g., conjugated equine estrogens 1.25 mg daily for 21 days) and progestin (e.g., medroxyprogesterone acetate 10 mg daily for the last 5–7 days) can demonstrate end organ endometrial failure and corroborate the clinical suspicion.*** However, some form of imaging ultimately is required to establish the diagnosis. Transvaginal or transabdominal ultrasonography may reveal a hematometra, but such findings are surprisingly rare. Sonohysterography or hysterosalpingography (HSG) provide more specific information regarding the location and extent of adhesions that partially or completely obliterate or obstruct the endometrial cavity or the cervical canal,[118] and hysteroscopy is definitive. A variety of classification schemes have been proposed to describe the extent of intrauterine adhesions and to predict treatment outcomes,[119–124] but none has been validated. Diagnosis of genital tuberculosis is made by endometrial biopsy (histopathology or culture) or with a nucleic acid-based test performed on an endometrial aspirate. Diagnosis of shistosomiasis is made by identifying the eggs of the parasite in urine, feces, rectal scrapings, menstrual discharge, or the endometrium.

Operative hysteroscopy is the primary method for treatment of intrauterine adhesions that may be lysed by scissors, electrodissection, or with a laser; most prefer sharp dissection, which may have less risk for causing further injury. Simultaneous laparoscopy or transabdominal ultrasonography provides useful guidance when dense scar tissue makes it difficult to enter the uterine cavity and can help to maintain orientation in a grossly distorted cavity, reducing the risk of uterine perforation. Most now advocate insertion of an intrauterine balloon catheter (left in place for approximately 7–10 days) after adhesiolysis to keep the walls of the uterine cavity separated during healing and decrease the risk of recurrence.[125] Treatment with a broad-spectrum antibiotic (e.g., doxycycline 100 mg twice daily) and a non-steroidal anti-inflammatory drug help to minimize risk of infection and uterine cramping while the catheter remains in place. High dose exogenous estrogen treatment (2.5 mg conjugated equine estrogens 2–3 times daily, or its equivalent) for approximately 4 weeks after surgery generally is recommended to encourage rapid endometrial re-epithelialization and proliferation; treatment with a progestin during the final week, if followed by menses, demonstrates a return of function. Despite these measures, recurrence rates are relatively high, ranging from 20% to over 60% in severe cases,[126] and repeated procedures often are required to restore a normal uterine cavity.[120] The surgical outcome can be assessed by postoperative HSG or by early "second-look" office hysteroscopy, which also provides the means to lyse any early recurring adhesions when still filmy.[114] Menstrual function can be restored in most cases (52–88%),[126] and among infertile women, live birth rates after hysteroscopic adhesiolysis generally have ranged between 25 and 35%.[127] As might be expected, outcomes tend to correlate with the severity of adhesions.[119, 120, 126, 128] In those who do achieve pregnancy, the risks of preterm labor, placenta accreta, placenta previa, and postpartum hemorrhage are increased.[114]

Disorders of the Ovary

Disorders of the ovary include the most common causes of amenorrhea and can present as primary or secondary amenorrhea. Those resulting only from chronic anovulation and relating to PCOS, obesity, and thyroid or mild prolactin disorders were discussed earlier, in the section of this chapter devoted to the evaluation of ovarian function. The focus here is on specific disorders that result in ovarian failure and their management.

Ovarian failure occurs when few or no follicles remain that are capable of producing estradiol in response to pituitary gonadotropin stimulation. Follicular depletion may occur during embryonic life with no follicles remaining by infancy or early childhood, after puberty has begun but before menarche, or at some later time before menopause normally would be expected. Consequently, depending on when the available supply of ovarian follicles is functionally depleted, puberty may not occur, it may begin normally but stop before the first menses, or it may progress normally to and beyond menarche with secondary amenorrhea having onset at some later point in time.

In a few women with rare genetic disorders, hypergonadotropic hypogonadism results from a functional ovarian failure due to abnormalities of follicular development, rather than from follicular depletion.

Gonadal Dysgenesis

Gonadal dysgenesis is defined as an incomplete or defective formation of the gonads, resulting from a disturbance in germ cell migration or organization, caused by structural or numerial sex chromosome abnormalities or mutations in the genes involved in formation of

the urogenital ridge and sexual differentiation of the bipotential gonad. Gonadal dysgenesis is among the most common causes of primary amenorrhea (approximately 30–40%). Due to the absence of ovarian follicles or their accelerated depletion during embryogenesis or the first few years of life, the gonads contain only stroma and appear as fibrous streaks. The large majority of patients with gonadal dysgenesis have an obvious abnormality involving an X chromosome. Aproximately 25% of affected individuals have a normal 46,XX karyotype and may harbor a more subtle abnormality invoving one or more specific genes on the X chromosome that are required for normal ovarian function; some with 46,XX gonadal dysgenesis also have neurosensory deafness, the combination known as Perrault syndrome. By far, the most common form of gonadal dysgenesis is Turner syndrome.

Turner Syndrome

Turner syndrome is a well known and thoroughly studied disorder, classically associated with a 45,X karyotype, but also with an assortment of other structural X chromosome abnormalities (deletions, ring and iso-chromosomes), any of which may be present in all or only in some of the cells of the body (mosaicism), depending on the stage of embryonic development at the time they arise. The disorder is discussed thoroughly in Chapter 9, as a cause of abnormal sexual development, and is more briefly summaried here.

The classical phenotype of Turner syndrome includes short stature, absent sexual development, a webbed neck, low set ears and posterior hairline, widely-spaced nipples ("shield chest") short fourth metacarpals, and an increased carrying angle at the elbow ("cubitus valgus"). Evidence indicates that the specific phenotype of patients with Turner syndrome relates, in part, to the parental origin of their X chromosome; most with a 45,X karyotype retain the maternal X.[129]

If not recognized by phenotype or poor growth during childhood, patients with Turner syndrome generally present at or near the time of expected puberty with primary amenorrhea and absent secondary sexual development. The diagnosis of Turner syndrome generally can be made easily, based on the phenotype and findings of hypergonadotropic hypogonadism. A karyotype is definitive, and specifically indicated, in part because it may reveal a cell line containing a Y chromosome otherwise not suspected or identified (e.g., 45,X/46,XY); approximately 5% of women with Turner syndrome have a karyotype containing all or part of a Y chromosome.[130] Futher analysis with fluorescence in situ hybridization (FISH) using one or more probes specific for segments of the Y chromosome will identify another 5% having occult Y chromosome material.[130, 131] Whereas it is important to identify a Y chromosome because affected individuals are at significant increased risk for developing gonadoblastoma (20–30%), that risk appears lower (5–10%) in women with Turner syndrome, and limited to those having detectable Y chromosome on their karyotype. ***FISH analysis is most clearly indicated for those exhibiting any evidence of virilization or having a chromosomal fragment of uncertain origin (discussed in chapter 9).***[132]

Mosaicism in women with Turner syndrome has important clinical implications besides those relating to a cell line containing a Y chromosome. In those with a mosaic 46,XX cell line (e.g., 45,X/46,XX), the gonad may contain functional ovarian cortical tissue, resulting in some degree of sexual development, or even menses and the possibility of pregnancy. Approximately 15% of patients with Turner syndrome begin but do not complete pubertal development and approximately 5% complete puberty and begin menstruation.[133] As might be expected, the phenotype varies, with some appearing normal and attaining normal stature before experiencing ovarian failure when the limited supply of follicles is exhausted. Natural pregnancies do occur in women with Turner syndrome, but they are rare

and associated with a relatively high risk for sex chromosome aneuploidy and spontaneous abortion.

Women with Turner syndrome may have a wide variety of medical problems having health implications as or more important than those relating to or resulting directly from hypogonadism.[132] Approximately one-third has cardiovascular anomalies, including a bicuspid aortic valve, coarctation of the aorta, mitral valve prolapse, and aortic aneurysm. Renal anomalies also are common and include horseshoe kidney, unilateral renal agenesis or pelvic kidney, rotational abnormalities, and partial or complete duplication of the collecting system(s). Autoimmune disorders are common in Turner syndrome and include thyroiditis, type 1 diabetes, autoimmune hepatitis and thrombocytopenia, and celiac disase. Hearing loss also is common.[134, 135] Consequently, additional and periodic medical evaluation is indicated and should include the following[132]:

- Echocardiography (at diagnosis, at least once between the ages of 12 and 15 years, and every 5 years if normal; more often if abnormal);
- Renal ultrasonography (once if normal, every 3–5 years if abnormal);
- TSH and free T4 (at diagnosis and every 1–2 years);
- Complete blood count, fasting glucose, lipid profile, renal function tests, and liver enzymes (every 2 years);
- Anti-endomysial antibodies, to detect celiac disease (at diagnosis);
- Audiometry (at diagnosis, at least once during the teen years or young adulthood, and every 10 years if normal).

Average intellectual performance is within the normal range,[136] although the prevalence of attention-deficit/hyperactivity disorder (ADHD) is increased in girls with Turner syndrome.[137] Overall mortality is increased approximately 3-fold and relates primarily to circulatory disease (e.g., hypertension), diabetes, liver and renal disease.[138] Overall cancer risks in women with Turner syndrome are similar to those in the general population, but the incidence of CNS tumors, bladder cancer, and endometrial cancer may be increased, and the risk of breast cancer is decreased.[139] With early diagnosis and treatment with growth hormone (GH), a final height greater than 150 cm (59 inches) can be achieved in most patients with Turner syndrome. The most important determinants of final height are the dose of GH and the duration of treatment before estrogen therapy begins. Treatment with GH generally should begin as soon as height falls below the fifth percentile of normal female growth and must be individualized, according to response.[132]

Treatment with estrogen must be timed carefully, with the goals of minimizing its adverse effects on growth and adult height and inducing puberty at an approximately normal age. Ideally, estrogen therapy should begin no later than age 15 and not before age 12 when growth is a priority, unless height already has been maximized. Estrogen therapy should begin at a low dose (e.g., 0.25–0.5 mg micronized estradiol or its equivalent), increasing gradually at intervals of 3–6 months according to response (Tanner stage, bone age), with the goal of completing sexual maturation over a period of 2–3 years. When vaginal bleeding first occurs, or after 12–24 months of estrogen therapy, a progestin (e.g., medroxyprogesterone acetate) should be added to the treatment regimen to complete development, prevent dysfunctional bleeding, and to protect the endometrium from the effects of unopposed estrogen.[132]

Oocyte donation offers the possibility of pregnancy to patients with Turner syndrome, but the cardiovascular demands of pregnancy pose unique and potentially serious risks that must be carefully considered. ***The risk of death during pregnancy is increased as much as 100-fold, primarily due to complications of aortic dissection or rupture.*** Risk is greatest for those with preexisting abnormalities such as a bicuspid aortic valve or a dilated aortic root, but even those without such findings remain at risk. Consequently, Turner syndrome generally should be regarded as a relative contraindication to pregnancy. Those expressing

serious interest in oocyte donation must receive thorough evaluation and counseling, and those having any significant cardiac abnormality should be strongly discouraged.[140]

Swyer Syndrome (46,XY Gonadal Dysgenesis)

Swyer syndrome is a distinctly different and less common form of gonadal dysgenesis, characterized by a 46,XY karyotype. Despite the presence of a Y chromosome, the phenotype is female because the dysgenetic (streak) gonads produce neither AMH nor androgens. Consequently, the vagina, cervix, uterus, and fallopian tubes develop normally and the internal and external genitalia do not masculinize.[141] In at least 10–15% of affected individuals, a mutation of the SRY gene (Sex-determining Region of the Y chromosome; located on the short arm, Yp11.3) is the cause.[142] In the remainder, no cause can be determined, although mutations in SRY regulatory elements or in other genes involved in the testis-determining pathway have been implicated (*SF1, SOX9, WT1, CMRT1*).[143–145]

Patients with Swyer syndrome generally present after the expected time of puberty with delayed sexual maturation and primary amenorrhea. The presence of pubic hair reflects a normal adrenarche. Evaluation reveals hypergonadotropic hypogonadism, prompting a karyotype that establishes the diagnosis. ***Gonadectomy is indicated soon after diagnosis due the significant risk for malignant transformation in occult testicular elements (20–30%).*** Gonadoblastoma is a premalignant germ cell tumor, unique to intersex states like Swyer syndrome, and may contain or give rise to other highly malignant tumors, including dysgerminoma, endodermal sinus tumor, embryonal and choriocarcinomas; evidence suggests they originate from clonal expansion of surviving germ cells in areas of undifferentiated gonadal tissue.[146]

Patients with Swyer syndrome exhibit normal growth and intellectual development, have no increased prevalence of any specific medical problems, and require no specific monitoring or treatment beyond that relating to hormone therapy aimed at inducing sexual maturation. The same sequential sex steroid treatment regimen described above for patients with Turner syndrome can be applied successfully in those with Swyer syndrome. Pregnancy, achieved with in vitro fertilization using donor oocytes, is a realistic expectation and has not been associated with any specific risks or complications.[147]

46,XX Gonadal Dysgenesis

Some individuals with primary amenorrhea and gonadal dysgenesis (streak gonads) have a normal 46,XX karyotype, providing indirect evidence that autosomal genes also play a critical role in ovarian differentiation. Affected women are normal in stature and, in most cases, have no apparent somatic anomalies. A wide variety of candidate genes has been identified, primarily via experiments involving murine knock-out models, including several that encode DNA and RNA binding proteins and transcription factors expressed during oogenesis.[148]

Premature Ovarian Failure

Premature ovarian failure (POF), traditionally defined as hypergonadotropic hypogonadism and amenorrhea arising before the age of 40, is a heterogeneous disorder that varies widely in cause and phenotype. Whereas the term POF is well-entrenched in the medical literature,

an alternative term—"premature ovarian insufficiency"—has been proposed to more accurately reflect the continuum of decreased ovarian function observed in affected women,[149] acknowledging that many exhibit intermittent ovarian function and ovulation and that 5–10% may conceive and deliver a pregnancy.[150, 151]

Premature ovarian failure (POF) generally results in secondary amenorrhea at some time after puberty is completed, but also may occur at any time before menarche and is distinguished from gonadal dysgenesis on the basis of ovarian morphology and histology; instead of streak gonads, the ovaries more closely resemble those of postmenopausal women. Approximately 1% of women will develop POF before the age of 40 years. In a cross-sectional survey of women aged 40–55 conducted at seven sites in the United States to determine eligibility for a community-based, multi-ethnic longitudinal sudy of the perimenopause (The Study of Women Across the Nation, SWAN), premature menopause was reported by 1% of Caucasians, 1.4% of African Americans and Hispanics, 0.5% of Chinese, and 0.1% of Japanese women.[152]

Important known causes of POF include numerical and structural chromosomal abnormalities, fragile X (*FMR*1) premutations, autoimmune disorders, radiation therapy, and chemotherapy. Whereas history alone can identify the latter two, the first three merit specific consideration and exclusion. In these and other rare disorders associated with ovarian failure such as galactosemia, the basic pathophysiology involves accelerated follicular atresia. In other rare genetic disorders, mutations in genes encoding intraovarian regulators, steroidogenic enzymes, gonadotropins, or their receptors result in impaired or abnormal follicular development, but the end result is much the same—functional ovarian failure. In most women with POF, a specific cause cannot be identified, but evidence implicating a number of genetic factors is growing rapidly.[148]

Numerical and Structural Chromosomal Abnormalities

A wide variety of numerical and structural chromosomal abnormalities may be identified in women presenting with ovarian failure. A review of karyotypes obtained in women with secondary amenorrhea revealed the spectum of possibilities and demonstrates the importance of a karyotype in women with POF.[153] Half of the observed abnormalities were numerical, involving X chromosome mosaicism (including 45,X, 46,XX, and 47,XXX cell lines) or Y chromosome mosaicism (including 46,XY, 47,XYY, and 47,XXY cell lines); the remainder included an assortment of X chromosome translocations, deletions and other structural abnormalities, and even some with a pure 46,XY karyotype.

Chromosomal deletions and translocations involving either the short (Xp) or the long arm (Xq) of the X chromosome may be identified in women with POF. Only about half of those with deletions involving the short arm of the X chromosome present with primary amenorrhea and gonadal dysgenesis; the remainder menstruates and often presents with POF. Presumably, the short arm of the X chromosome contains genes essential for ovarian germ cell function. Although the specific genes involved are unknown, likely candidates include *BMP15* (Bone Morphogenic Protein) and other members of the TGFβ (Transforming Growth Factor) superfamily.[148] The long arm of the X chromosome also contains genes crucial for normal ovarian function. Likely candidate genes, identified by the phenotypes associated with deletions and translocations involving Xq, include *XIST* (the X-Inactivation gene), *DACH2* (encoding a transcription factor), and *QM* (encoding a ribosomal protein).[148, 154]

Fragile X (*FMR1*) Premutations

A spectrum of clinically important disorders, including POF, involves a dynamic trinucleotide (CGG) repeat sequence mutation in the X-linked *FMR1* gene, located near

the terminal end of the long arm of the X chromosome (Xq27.3). The normal *FMR1* gene contains approximately 30 repeats. The fully expanded form of the mutation, characterized by more 200 CGG repeats, results in fragile X syndrome (FXS), the most common known genetic cause of mental retardation and autism. The premutation, characterized by 55–200 repeats, is associated with two disorders distinct from FXS. One is the fragile X-associated tremor/ataxia syndrome (FXTAS), a neurologic disorder that affects males primarily, as might be expected for an X-linked disorder. The other is POF, affecting approximately 15% of women who carry the premutation.[56]

In the full mutation, the expanded number of trinucleotide repeats in the 5′ untranslated region of the gene results in hypermethylation that extends into the promoter region and silences the gene. Consequently, there is little or no production of mRNA and its product, the fragile X mental retardation protein (FMRP), which is an RNA binding protein that functions as a translational suppressor. Ultimately, the result is an overexpression of mRNAs, leading to the clinical features of FXS.[155] The incidence of FXS is approximately 1 in 4,000 males and 1 in 4,000–8,000 females.[56] Although affected females are protected to some extent from the full impact of the mutation, due to X-inactivation, approximately 70% have a borderline or lower IQ or other functional deficits.[156–158]

In the premutation, the trinucleotide repeat sequence is not methylated, the gene functions, and FMRP is produced. However, the condition is quite different from the usual unaffected carrier state in two important ways: (1) premutation carriers are at risk for developing disorders different from FXS, and (2) the affected alleles are unstable and at risk for expansion from the premutation to the full mutation. FXTAS is a progressive neurodegenerative disorder affecting male premutation carriers, generally after age 50, causing intention tremor, ataxia, autonomic dysfunction, cognitive deficits, behavioral abnormalities, and peripheral neuopathy.[159] Female premutation carriers may develop the disorder, but do so infrequently; unfavorable X chromosome inactivation may increase the risk of FXTAS in women.[160] However, women with a *FMR1* premutation frequently develop POF. The prevalence of premutations is approximately 15% among women with familial POF and lower, but still significant (1–7%), in those having no family history of POF.[56] The variation in prevalence may be explained by the relationship between the probability of POF and the size of the CGG repeat sequence; risk for POF increases with the number of repeats, between 59 and 99, but rises no further and even decreases for those with more than 100.[161, 162] Women with a modestly expanded repeat sequence (41–58 repeats) also may be at increased risk for POF.[163, 164]

Women who carry the *FMR1* premutation often exhibit signs of early reproductive aging. The lengths of their cycle and follicular phase are shorter, FSH levels across all phases of the cycle are higher, and inhibin levels are lower, compared to those in normal women.[165] They also enter menopause approximately 5 years earlier than average.[161] Exactly why or how the *FMR1* premutation predisposes to POF is not entirely clear but may involve gain-of-function toxicity, due to overexpression of mRNAs, resulting in accelerated follicular atresia.

The inheritance of *FMR1* mutations and premutations follows the basic pattern of an X-linked disorder; females transmit the abnormality to 50% of their offspring and male carriers to all of their daughters and none of their sons. The pattern of inheritance is complicated by the meiotic instability of the trinucleotide repeat sequence, which has a tendency to expand when passed from the mother to her progeny, thereby increasing the risk for FXS with each generation, a phenomenon known as “anticipation.” The risk for expansion depends on the size of premutation; a repeat sequence numbering between 59 and 79 expands to the full mutation less than half the time, but one larger than 90 does so more than 90% of the time.[166] In contrast, the size of the premutation remains relatively stable when transmitted from fathers to their daughters and rarely expands to the full mutation,[167] possibly because large repeat sequences are highly unstable in developing sperm and only

smaller premutations can be transmitted. Intermediate or "gray zone" repeat sequences numbering between 45 and 54 may expand to premutation size across generations. ***Due to the complexity of the inheritance of FMR1 premutations, all carriers should receive formal genetic counseling.*** Issues that must be addressed include the risk for infertility and early menopause, the risk for and implications of transmitting the premutation, and the possibility that other family members may be affected, in a variety of ways, raising still other and obvious ethical concerns.

All women with POF should be offered screening for the FMR1 premutation. In sum, guidelines issued by the American College of Medical Genetics, the American College of Obstetricians and Gynecologists, and the American Society for Reproductive Medicine all recommend testing for women with unexplained POF.[58, 168] ***Whereas there is not yet a clear consensus on whether testing is indicated for all infertile women under age 40 with modestly elevated FSH levels suggesting a diminished ovarian reserve, all agree that testing should be offered to those having a family history of POF, FXS, or FXTAS or having relatives with unexplained mental retardation or autism.*** Arguments against more liberal or widespread population screening center on the limited resources available to provide the complicated counseling required and the uncertainties surrounding the risks associated with alleles in the intermediate range, which are common.

Up to 5–10% of women with POF who carry an *FMR1* premutation will conceive after diagnosis, without medical intervention, but there is no evidence that any treatment other than oocyte donation can increase the likelihood of pregnancy.[151]

Autoimmune Disorders

Autoimmune disease is one of the known causes of POF, accounting for approximately 4% of cases.[169] There is substantial evidence to indicate that autoimmunity is the cause of POF in women who also exhibit signs of adrenal autoimmunity.[60] Autoimmune oophoritis may occur as part of a Type I or II autoimmune polyglandular syndrome (APS) associated with autoantibodies to multiple endocrine and other organs. Type I APS (also known as autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy) presents in childhood, typically as hypoparathyroidism (89%) or mucocutaneous candidiasis (75%), and often is accompanied by adrenal insufficiency (60–80%) and POF (60%); the cause is a mutation in the *AIRE* (autoimmune regulator) gene, located on chromosome 21. Type II APS has an adult onset and is characterized by adrenal insufficiency (100%) and thyroid autoimmunity (70%) or type 1 diabetes (50%); 25% of women with the disorder have amenorrhea and 10% have POF.[59] POF also has been described in women with systemic lupus erythematosus and myasthenia gravis.

The mechanism that stimulates or causes ovarian autoimmunity is unknown, but might involve a virus or other cause of damage to ovarian tissue that renders it antigenic, or a basic failure in immune regulation resulting in a loss of tolerance to some component of ovarian tissue. The hallmark of autoimmune oophoritis is a lymphocytic infiltrate surrounding secondary and antral follicles but not primordial follicles,[169, 170] strongly suggesting that steroid hormone producting cells in the theca contain the inciting antigen. Observations of normal inhibin B production in otherwise hypogonadal women with autoimmune oophoritis further suggest that the theca is selectively targeted and that granulosa cells are spared.[171] Almost all women with documented autoimmune oophoritis have circulating antibodies directed against steroidogenic enzymes such as 21-hyroxylase, 17α-hydroxylase, and side chain cleavage.[60, 169] The diagnosis of autoimmune ovarian failure therefore hinges on the demonstration of autoantibodies against steroidogenic cells. ***A positive test for adrenal or 21-hydroxylase antibodies is sufficient to establish a diagnosis of autoimmune ovarian failure.*** A commercially available serum anti-ovarian antibody test had poor predictive value and an unacceptably high false-positive rate.[172] ***Ovarian biopsy solely for diagnosis***

of autoimmune ovarian failure is unnecessary and is not recommended. Whereas antibodies against membrane-bound receptors may cause diseases such as mayasthenia gravis, the failure to detect antibodies against the FSH receptor in women with POF suggests they rarely, if ever, are the cause.[173] Women with autoimmune ovarian failure may develop large luteinized follicular cysts, possibly due to increased gonadotropin stimulation in response to impaired follicle development and function.[171]

The strong association between autoimmune adrenal and ovarian failure justifies screening for anti-adrenal antibodies in all women with POF, at the time of diagnosis. The most sensitive and useful methods for their detection are indirect immunofluorescence assays using adrenal tissue as substrate and immunoprecipation assays for antibodies against the 21-hyroxylase enzyme (CYP21). ***Patients with positive anti-adrenal antibodies should be further evaluated to exclude asymptomatic adrenal insufficiency, by measuring the morning (6:00–9:00 A.M.) serum cortisol level.*** A value greater than 18 μg/dL effectively excludes clinical adrenal insufficiency; those with lower values require further evaluation with an ACTH stimulation test to determine whether ACTH reserve is sufficient to meet demand during times of stress. The test is performed by measuring the serum cortisol concentration before and 60 minutes after administering cosyntropin (synthetic ACTH 1–24; 0.25 mg) intramuscularly or intravenously; a stimulated cortisol concentration ≥18 μg/ dL is a normal response. ***Those with negative tests for adrenal antibodies should be followed and monitored with repeat testing at intervals, because POF may precede onset of adrenal insufficiency by up to several years and its autoimmune cause may not, at first, be recognized.***[59]

Other autoimmune disorders that may be identified in women with POF, such as autoimmune thyroiditis, rarely are associated with autoimmune oophoritis and therefore cannot be considered proof of autoimmune ovarian failure. ***However, because the prevalence of thyroiditis is relatively high among women with POF (14–27%) and the presence of thyroid autoantibodies identifies patients at risk for developing autoimmune thyroid disease, screening for thyroid peroxidase and thyroglobulin antibodies also is indicated.*** When positive, TSH levels should be monitored annually, less often (e.g., 5-year intervals) when negative. In the past, screening for other autoimmune endocrine disorders was recommended (serum calcium, phosphorus, fasting glucose, vitamin B12), but such screening has a very low yield in asymptomatic patients and can be safely reserved for those with clinical indications.[61]

Radiation Therapy

The adverse effects of radiation on the ovary depend on the age of the patient, the dose of radiation, and the radiation field.[174] In young women, radiation therapy may result only in transient amenorrhea that ends after a period of 6 to 18 months, probably reflecting the interval required to reestablish the mechanisms that govern the initiation of follicular growth, and the size of their follicular reserve. Transient suspension of normal ovarian cycling also has been observed in women who received radioactive iodine for the treatment of thyroid cancer.[175] However, some will suffer immediate and irreversible ovarian failure, and even those who recover may later exhibit early ovarian aging and an early menopause. The ovaries of older women are more sensitive to the effects of radiation. Whereas doses greater than 6 Gy (gray units, 1 Gy=100 rads) almost uniformly cause ovarian failure in women over age 40,[176] younger women have achieved successful pregnancies after having received far higher doses.

The radiosensitivity of human oocytes has been estimated at approximately 2 Gy, implying that approximately 50% of remaining oocytes will survive after such exposure.[177] Using the best available model for the rate of natural follicular depletion,[178] estimates for the mean and effective sterilizing doses of radiation (resulting in immediate and permanent ovarian

failure in 50% and 97.5% of individuals, respectively) can be calculated. The mean age of subsequent ovarian failure also can be estimated for any given age and dose of radiation. Representative values derived from such calculations are as follows:[179]

	Sterilizing Dose (Gy)	
Age at Treatment (yr)	Mean (50%)	Effective (97.5%)
0	18.8	20.3
10	17.0	18.4
20	15.0	16.5
30	12.0	14.3
40	8.0	11.3

	Mean Predicted Age of Ovarian Failure (yr)			
Age at Treatment (yr)	3 Gy	6 Gy	9 Gy	12 Gy
0	35.1	22.6	13.7	7.9
10	36.7	26.5	19.7	15.3
20	39.0	31.4	26.4	22.8
30	42.2	37.0	33.2	30.1

When the radiation field excludes the pelvis, there is no significant risk for permanent ovarian failure.[180, 181] Consequently, direct irradiation to the ovaries should be avoided whenever possible. Elective ovarian transposition (moving the ovaries out of the radiation field) can help to preserve gonadal function in patients receiving pelvic radiation without chemotherapy. It may be performed electively before radiation therapy begins, or at the time of surgical staging or the primary cytoreductive surgery.[174, 182, 183] However, the success of ovarian transposition for preserving fertility after radiation therapy has varied widely, from 16% to 90%, due to differences in treatment regimens, exposure to scatter radiation, shielding, and adjuvant chemotherapy. The procedure also has potential complications, including chronic ovarian pain and cysts, which may require additional surgery. Subsequently, pregnancy also may be difficult to achieve spontaneously or even with IVF, unless the ovaries are moved back into the pelvis.[183] Other fertility preservation strategies that might be considered before radiation therapy include urgent IVF and embryo cryopreservation, oocyte cryopreservation, and ovarian tissue cryopreservation, as discussed in detail in Chapter 32.

There is no evidence for an increased risk of birth defects in the offspring of women treated with radiation therapy, chemotherapy, or both.[184] However, the risk for pregnancy complications such as miscarriage, preterm labor, and low birthweight may be increased due to impaired uterine growth and blood flow.[185–187]

Chemotherapy

Most chemotherapeutic drugs target actively dividing cells and therefore might not be expected to have significant adverse effects on oocytes; nonetheless many do. In fact, the fixed supply of oocytes is extremely sensitive to cytotoxic drugs. ***Chemotherapy causes depletion of the primordial follicular pool in a drug- and dose-dependent manner and is a relatively common cause of POF.***

The ovarian toxicity of common chemotherapeutic agents varies significantly.[183] Alkylating agents such as cyclophosphamide, which alters base pairs and causes DNA cross-links and breaks, can affect both resting and dividing cells. The risk for ovarian failure after chemotherapy increases with the age of the patient, presumably because the size of the residual follicular pool declines progressively with advancing age and the damage caused by treatment becomes proportionately greater. For example, a dose of cyclophosphamide that uniformly causes amenorrhea in women over age 40 does so in only about half of younger women.[188] However, individual women vary significantly in their sensitivity or susceptibility to gonadal damage from chemotherapeutic agents. The common use of more than a single drug further complicates efforts to predict the likelihood of ovarian failure resulting from treatment.

High Gonadal Toxicity	*Moderate Gonadal Toxicity*	*Low Gonadal Toxicity*
Cyclophosphamide	Cisplatin	Bleomycin
Chlorambucil	Adriamycin	Actinomycin D
Melphalan		Vincristine
Busulphan		Methotrexate
Nitrogen mustard		5-flurouracil
Procarbazine		Taxanes

It is important to note that a large majority of studies examining the effects of chemotherapeutic agents on ovarian function have used the incidence of amenorrhea, from as little as 6 months to as long as 5 years after completion of treatment, as the measure of impact. Serum FSH and estradiol concentrations, or other accepted measures of "ovarian reserve" such as inhibin B, AMH, and ovarian volume or antral follicle counts generally have not been considered. However, it's clear that most women in the advanced stages of follicular depletion, due to aging or other causes, still menstruate even though the number and quality of their remaining oocytes and their fecundity are quite low.[189] Consequently, it seems certain that the amenorrhea rate associated with any given chemotherapeutic drug or combination grossly underestimates its true adverse impact on ovarian function and fertility. Indeed, observations derived from a few studies that have examined the effect of chemotherapy on markers of ovarian reserve (AMH in particular) suggest that chemotherapy induces accelerated ovarian aging and thereby increases the likelihood of POF well beyond the risk estimates based on observations of amenorrhea after the end of treatment.[190–192]

Long-acting GnRH agonists (e.g., leuprolide acetate) have been used widely to induce a hypogonadal state before chemotherapy in hopes to reduce or prevent the adverse effects of cytotoxic drugs on ovarian function and future fertility. Collected data from studies in rodents, one in nonhuman primates,[193] and numerous observational studies in women have suggested that GnRH agonist pretreatment may offer some protection against chemotherapy-induced ovarian follicular depletion.[194] Overall, less than 10% of patients treated with a GnRH agonist before chemotherapy have developed irreversible POF, compared to 40–70% of those receiving similar treatment but no GnRH agonist. Some have suggested that, at the least, GnRH agonist pretreatment may widen or extend the "fertility window" by 7 years or more for patients who receive chemotherapy.[194] Several mechanisms have been proposed to explain the potential protective actions of GnRH agonists: (1) a decreased number of primordial follicles entering development (due to suppression of FSH), when they become more vulnerable to damage; (2) decreased ovarian perfusion induced by hypoestrogenism, resulting in lower exposure to chemotherapeutic agents; and (3) a direct effect on the ovary, independent of gonadotropins, such as up-regulation of an intraovarian antiapoptotic factor or protection of germline stem cells.[194] Moreover, advocates of GnRH

agonist pretreatment have emphasized that it can decrease the risk of menorrhagia resulting from chemotherapy-induced thrombocytopenia and, at least in patients with hormone-sensitive breast cancers, also may improve survival rates.[195]

Questions concerning the value and even the safety of GnRH agonist chemoprophylaxis have stimulated an active and ongoing debate. Many have challenged the biological plausibility of the proposed protective mechanisms and raised concerns that the initial "flare" in gonadotropin secretion that follows GnRH agonist treatment may have effects opposite from those intended. Those who oppose GnRH agonist pretreatment argue that the return of menses after chemotherapy is an insensitive and unreliable measure of protection from drug-induced ovarian injury,[196] that the sustained decrease in serum AMH levels that follows chemotherapy indicates a loss of primordial and preantral follicles (the primary source of AMH),[192] and that GnRH agonist treatment cannot protect against damage from cytotoxic drugs because primordial and preantral follicles are not sensitive to gonadotropin stimulation.[197] They cite the methodologic weaknesses of previous observational studies and the results of the only published randomized trial examining the effects of GnRH agonist pretreatment on ovarian function after chemotherapy, which did not show a benefit.[198] The safety of GnRH agonists has been questioned because many tumors of the breast and reproductive tract express GnRH receptors mediating actions that might reduce the efficacy of chemotherapy or promote an even greater loss of follicles.[199–202] Finally, treatment with a GnRH agonist can be expected to induce acute symptoms of estrogen deficiency, in addition to those that invariably result from chemotherapy, thereby adding to the overall burden of treatment. Ultimately, more confident conclusions regarding the relative risks and benefits of GnRH agonist pretreatment must await the results of a number of ongoing clinical trials being conducted by the Southwest Oncology Group in the United States and by others in Germany, Italy, Spain, and the United Kingdom.

Galactosemia

Galactosemia is an autosomal recessive disorder of galactose metabolism caused by a deficiency of the enzyme galactose 1-phosphate uridyl transferase and is another, albeit very rare, cause of POF.[203] Affected women have fewer primordial follicles, presumably due to the cumulative toxicity of galactose metabolites on germ cell migration and survival.[204, 205] Diagnosis usually is made in the first few days of life after feeding with breast or cows' milk-based formulas begins, causing jaundice, vomiting, and failure to thrive.

Functional Ovarian Failure Resulting from Disorders of Follicular Development

Whereas accelerated follicular depletion is the underlying mechanism for the most common causes of POF, a variety of rare genetic disorders causing impaired or abnormal follicular development may result in a functional ovarian failure. Examples include disorders of intraovarian regulation, steroidogenic enzyme defects, and abnormalities in gonadotropins and their receptors. In 1969, Jones and de Moraes-Ruehsen described three patients with amenorrhea and hypergonadotropic hypogonadism who also were resistant to high doses of exogenous gonadotropins, although their ovaries contained numerous follicles. They called the disorder "resistant ovary syndrome" or "Savage syndrome," after the name of their first patient.[206] It now seems likely that the resistant ovary syndrome results from intrinsic defects in follicular development.

Abnormalities in any of the many identified paracrine regulators of ovarian function may interfere with or prevent a normal response to gonadotropin stimulation. For example, mutations in the gene encoding bone morphogenic protein-15 (*BMP15*), an oocyte-specific growth factor that stimulates folliculogenesis and granulosa cell proliferation (in the same family that includes activins and inhibins), have been identified in a small number of women with idiopathic POF. In functional assays, the mutant BMP15 was processed abnormally, associated with decreased granulosa cell growth, and antagonized the stimulation of

granulosa cell proliferation by wild-type BMP15.[207] In a study of women with POF who underwent genetic screening, 7/166 (4%) having heterozygous mutations in BMP15 were detected, compared to none in control groups containing women who had a natural menopause or women randomly selected from the general population.[208] The blelpharophimosis/ptosis/epicanthus inversus syndrome (BPES) is another example. BPES is a rare autosomal dominant disease characterized by eyelid malformations and POF, caused by a variety of mutations in the gene encoding a forkhead box transcription factor (*FOXL2*) required for normal granlulosa cell function.[209]

Rare steroidogenic enzyme defects that effectively block follicular development may result in hypergonadotropic hypogonadism due to functional ovarian failure. Mutations in the genes encoding the steroidogenic acute regulatory (StAR) enzyme (*STAR*), the 17α-hydroxylase enzyme (*CYP17A1*), and the aromatase enzyme (*CYP19A1*) are examples. The StAR enzyme transports cholesterol from the outer to the inner mitochondrial membrane where the side chain cleavage enzyme converts cholesterol to pregnenolone, the first and rate-limiting step in steroid biosynthesis; StAR mutations thus result in impaired synthesis of all adrenal and gonadal steroid hormones. Affected individuals present with congenital lipoid adrenal hyperplasia and exhibit severe adrenal insufficiency soon after birth or in early infancy; those who are recognized and treated early have absent pubertal development.[210] Females with 17α-hydroxylase deficiency, involving the enzyme complex that converts 21-carbon progestogens to 19-carbon androgens, usually present around the expected time of puberty with primary amenorrhea and sexual infantilism. Elevated levels of progesterone (the hormone immediately proximal to the enzyme block) are alternatively converted to mineralocorticoids (deoxycorticosterone, corticosterone), resulting in hypertension and hypokalemia. Those having a partial or less severe 17α-hydroxylase deficiency may exhibit varying degrees of sex steroid production and sexual development.[211] Rare females with aromatase deficiency, involving the enzyme that converts androgens to estrogens, classically present with ambiguous genitalia at birth, elevated androgen levels, and absent breast development at puberty. During pregnancy, fetal androgens cannot be aromatized to estrogen in the placenta, resulting in masculinization of both the fetus and mother. Aromatase mutations also can produce variable or nonclassic phenotypes characterized by varying degrees of breast development.[212]

Inactivating mutations in the β-subunit of LH or FSH can result in abnormal gonadotropin molecules having limited or no immunoreactivity or bioactivity. Affected individuals are hypogonadal and have a high level of one (the normal) gonadotropin, but a basal or undetectable level of the other.[213, 214] Inactivating mutations in the FSH or LH receptor may result in a failure of gonadotropin binding or signal transduction and thus again in hypergonadotropic hypogonadism due to a failure of follicular development.[215] One specific point mutation in the FSH receptor has a relatively high gene frequency in the Finnish population (0.96%).[216, 217] A search for the same mutation among women with POF in the U.S. Brazil, Switzerland, Denmark, Japan, and Singapore found only a single carrier,[218–221] but other inactivating mutations have been identified.[222, 223] A few women with similar mutations in the LH receptor have been described, presenting with amenorrhea, enlarged cystic ovaries, and a high serum LH but normal FSH concentration.[224, 225]

Management of Premature Ovarian Failure

It is important to emphasize that effective management of POF requires careful counseling and emotional support, as well as specific evaluation and medical treatment. Young women with POF understandably are not prepared for the diagnosis, and many are dissatisfied with the manner in which they were informed.[226] Affected women need, and deserve, sufficient time for thorough education and for planning their longer term management.

Attention should focus first on excluding those causes of POF having important potential health consequences for the patient or other members of her family. Those with chromosomal translocations or deletions or fragile X premutations should receive appropriate genetic counseling, and those with autoimmune disease will require careful monitoring over time to ensure that emerging and potentially serious health problems are promptly recognized and treated.

Hormone Therapy

The longer term management of women with POF centers on their hypogonadism and its sequellae. In the absence of exogenous estrogen treatment, they are at risk for developing osteopenia and osteoporosis,[227] and also early coronary heart disease.[228–231] Inevitably, they also will develop symptoms of estrogen deficiency such as vasomotor flushes and genito-urinary atrophy that can be debilitating. Therefore, unless there is a specific contraindication to its use, women with POF should receive exogenous estrogen therapy. Other strategies for protecting bone and heart health also deserve discussion, including exercise, diet, adequate calcium and vitamin D intake, and the avoidance of smoking.

Estrogen treatment in women with POF can take several forms. Physiologic levels of estrogen can be achieved using oral (e.g, micronized estradiol 1–2 mg daily or conjugated equine estrogens 0.625–1.25 mg daily) or transdermal treatment regimens (0.1 mg/24 hours).[232] Because most women with POF have an intact uterus, cyclic or continuous treatment with a progestogen is essential to prevent endometrial hyperplasia and neoplasia that can result from treatment with estrogen alone.[233] Cyclic treatment with a progestogen (e.g., micronized progesterone 200 mg daily or medroxyprogesterone acetate 10 mg daily for 12–14 days each month) is preferable for those still hoping to conceive. Oral contraceptives also may be used but contain substantially greater amounts of hormones than are required and may thus be reserved for those who want to prevent even the possibility of random ovulation and pregnancy. ***It is important to emphasize to young women with POF that they are distinctly different from older postmenopausal women and that the balance between the risks and benefits of hormone therapy for them also differs from that in postmenopausal women.*** Because they are significantly younger, their baseline risks for cardiovascular disease and breast cancer are much lower than those for older postmenopausal women. Moreover, without estrogen therapy, their risk for later coronary heart disease may be increased, rather than decreased.[228–231] Hormone therapy should continue up to at least age of 50, in much the same way as endogenous hormone production does in normal women.

Ovarian androgen levels (testosterone, androstenedione) may be somewhat lower in women with POF than in normal women of comparable age,[234] although in most they are maintained at least to the same extent as in normal postmenopausal women. Currently, there are no established criteria for diagnosis of androgen deficiency and there is no approved or validated method for providing physiologic androgen treatment in women.[235] Moreover, exogenous androgen treatment may cause acne and hirsutism, and when administered orally, dyslipidemia. The longer term clinical consequences of reduced androgen levels, if any, have not been studied and the safety of long-term androgen treatment has not been established. Consquently, androgen treatment cannot be recommended for women with POF.

Fertility

Although the likelihood of achieving pregnancy after diagnosis is only about 5–10%, some women with POF do conceive and approximately 80% of their pregnancies end in a healthy live birth.[151] ***However, there is no evidence that any form of treatment other than egg donation and IVF can increase the chance for pregnancy.***

Intermittent or episodic ovulation is by no means rare in women with POF, just as in perimenopausal women. Serial blood sampling and transvaginal ultrasonography can demonstrate developing follicles, but disordered patterns of folliculogenesis frequently are observed and premature luteinization, possibly a consequence of elevated LH levels, is common.[236] Evidence indicates that physiologic exogenous estrogen therapy allows, but does not improve, follicular development or ovulation.[150] Although ovulation induction with exogenous gonadotropins often has been attempted, women with established hypergonadotropic hypogonadism are, for obvious reasons, poor candidates. Attempts to improve ovulation rates achieved with gonadotropin therapy by pretreatment with estrogen or a GnRH agonist have met with some limited success,[237–240] but pregnancy and live birth rates remain extremely low. Women with POF from any cause generally are excellent candidates for IVF using donor oocytes.[241]

Psychological and Emotional Support

Profound grief and a sense of loss for the children they hoped for and expected are common in women after diagnosis of POF; validating those emotions can be both therapeutic and reassuring. It also is important to emphasize that the diagnosis of POF does not imply or predict premature aging in any other way, something that many women understandably may fear. Women with POF also are at risk for developing related depression and anxiety disorders.[226, 242] Consequently, referral to a support group (www.pofsupport.org) and to a therapist having expertise in counseling women and couples with reproductive failure can be very helpful.

Disorders of the Anterior Pituitary

A variety of disorders involving the anterior pituitary may be a cause of amenorrhea. Pituitary tumors are the most common by far and most are benign adenomas; in a large transsphenoidal surgical series, 91% of sellar and parasellar masses were pituitary adenomas.[243] Malignant pituitary tumors almost never are encountered. Other tumors that may arise in the sellar region include craniopharyngiomas, meningiomas, gliomas, metastatic tumors, and chordomas. Not all sellar masses are neoplastic;[244] cysts, tuberculosis, sarcoidosis, and fat deposits that compress the pituitary have been reported. Nearby lesions, such as internal carotid artery aneurysms also can cause amenorrhea.

Other pituitary causes of hypogonadotropic hypogonadism include damage from surgery or radiation, ischemia and infarction (e.g., Shehan's syndrome), and infiltrative diseases such as lymphocytic hypophysitis and hemochromatosis.

Pituitary Adenomas

Pituitary adenomas are true neoplasms but almost always are benign. The large majority are monoclonal, suggesting that somatic cell mutations precede clonal expansion and play an important role in tumorigenesis.[245, 246] Specific genetic mutations are known to be involved in the development of some pituitary tumors. Autosomal dominant inactivating mutations in the *MEN1* gene (encoding menin, a putative tumor suppressor) predispose to the development of parathyroid, pituitary, and entero-pancreatic adenomas (insulinomas, gastrinomas, carcinoid tumors) in patients with one of the multiple endocrine neoplasia (MEN) syndromes (Type 1).[247] Activating mutations in the *GNAS1* gene (encoding the stimulatory G-protein alpha subunit involved in the signal transduction pathway that links

receptor-ligand interactions with activation of adenyl cyclase) are identified in approximately 40% of GH-secreting somatotroph adenomas.[248] The pituitary tumor transforming gene *PTTG1* is overexpressed in most all pituitary adenomas, and to an even greater degree in those that extend outside of the sella.[249]

Pituitary adenomas are classified by cell type and size, and may be functional (hormone-secreting) or nonfunctional. The large majority of pituitary adenomas are functional prolactin-secreting lactotroph adenomas or nonfunctional adenomas, most of which derive from gonadotrophs. Functional thyrotroph adenomas (secreting TSH and causing hyperthyroidism), somatotroph adenomas (secreting GH and causing acromegaly) and corticotroph adenomas (secreting ACTH and causing Cushing's disease) are rare, particularly in women presenting with amenorrhea. If large enough, even nonfunctioning adenomas can have functional consequences, by compressing the pituitary stalk and interfering with the delivery of hypothalamic releasing or inhibiting factors, or by compressing surrounding cells. Tumors less than 10 mm in size are called microadenomas and those 10 mm or larger are called macroadenomas. Pituitary adenomas may be discovered during evaluation for neurologic symptoms or symptoms of hormone deficiency (such as amenorrhea) or excess, or when the head is imaged for other reasons. MRI (with gadolinium contrast) is the best method for imaging the pituitary gland and surrounding region and may reveal tumor extending outside of the sella, into the cavernous sinuses or sphenoid sinus, or causing elevation of the optic chiasm.

The most common neurologic symptom associated with pituitary tumors is visual impairment. The classical complaint is bitemporal hemianopsia (tunnel vision), caused by upward pressure on the optic chiasm at its center (affecting those portions of the optic nerves correlating with the nasal retinas and the temporal visual fields), but one or both eyes may be affected and to varying degrees. Decreased visual acuity develops with more severe compression of the chiasm and diplopia (blurred vision) results from lateral extension and compression of the oculomotor nerve. The onset of visual symptoms is so gradual they often go unrecognized for months or years. Other neurologic symptoms include nonspecific headaches (from expansion of the sella), cerebrospinal fluid rhinorrhea (from inferior extension of the tumor), and pituitary apoplexy (caused by sudden hemorrhage into the adenoma).

The endocrine consequences of pituitary adenomas depend on whether they are functional or nonfunctional, and on their size. Functional microadenomas and macroadenomas secrete excessive amounts of hormone, according to their specific cell type, and cause symptoms that result from overstimulation of the target organ or tissues, as discussed below. Functional and nonfunctional macroadenomas also may cause pituitary hormone deficiencies, due to their mass effects on the pituitary stalk and surrounding cells. Whereas gonadotropin deficiency results in symptoms of hypogonadism such as amenorrhea and vaginal atrophy, the symptoms of TSH deficiency are those of hypothyroidism and include fatigue, lethargy, cold intolerance, decreased appetite, constipation, dry skin, bradycardia, and anemia. GH deficiency causes short stature in children. In adults, GH deficiency has been associated with decreased muscle mass and increased fat mass,[250, 251] decreased bone density,[252] and an increased risk for cardiovascular disease,[253] but causes no symptoms other than a decreased sense of well-being.[254] Prolactin deficiency has no known symptoms other than failed lactation after delivery. The symptoms of ACTH deficiency are those of cortisol deficiency and include postural hypotension and tachycardia, fatigue, anorexia, weight loss, hypoglycemia, and eosinophila. Whereas primary adrenal insufficiency also results in salt wasting, volume contraction, and hyperkalemia due to aldosterone deficiency, and to hyperpigmentation due to the compensatory increase in ACTH secretion, secondary adrenal insufficiency caused by ACTH deficiency does not. Both primary and secondary adrenal insufficiency can cause hyponatremia due to inappropriate secretion of antidiuretic hormone (vasopressin), which results from cortisol deficiency. ***Moderate deficiencies of TSH, GH, and ACTH often cause few or no recognizable symptoms and easily can go unrecognized if not suspected and specifically excluded.***

Pituitary Function Tests

In women with macroadenomas, pituitary function tests are indicated to exclude other pituitary hormone deficiencies that may have important health implications. Although the required additional evaluation generally is not difficult to perform or interpret and only a few with abnormal results may require more complicated dynamic tests, the generalist reasonably may want to consult with or refer such patients to a reproductive or medical endocrinologist more familiar with the indicated tests. ***The routine endocrine evaluation of women with amenorrhea includes the measurement of serum TSH, prolactin, and FSH. Women with pituitary macroadenomas require additional evaluation, including a serum free T4, IGF-1, and morning cortisol level (6:00–9:00 A.M.).***

When the serum TSH level is low or normal, a low serum free T4 demonstrates secondary hypothyroidism. ***It is important to understand that measurement of serum TSH alone is not sufficient in patients who may have hypothalamic or pituitary disease.*** A normal TSH level excludes hypothyroidism only when there is every reason to believe that the hypothalamic-pituitary-thyroid axis is intact and functioning normally. In women with a sellar mass and hypogonadotropic hypogonadism, indicating a dysfunctional HPO axis, normal function of the thyroid axis cannot be assumed. High levels of both TSH and free T4 suggest a rare functional thyrotroph adenoma (causing hyperthyroidism).[255]

Screening for GH deficiency or excess is best accomplished by measuring the serum IGF-1 level, because the basal serum GH concentration is not reliable in adults. IGF-1 is produced in the liver, in response to GH stimulation, and a concentration below the age-specific normal lower limit has greater than 95% specificity for diagnosis of GH deficiency.[256] Provocative tests of GH secretion such as the insulin-induced hypoglycemia test or the combined administration of arginine and GH-releasing hormone are more sensitive,[256] but seldom are necessary. An elevated serum IGF-1 concentration suggests a GH-secreting somatotroph adenoma (causing acromegaly).[257]

A very low morning cortisol level (<3–5 μg/dL) indicates adrenal insufficiency and implies ACTH deficiency; a value ≥ 15–18 μg/dL demonstrates normal cortisol secretion. Intermediate values require additional evaluation to determine whether ACTH reserve is sufficient to meet demand during times of stress. ACTH reserve can be evaluated most easily by performing an ACTH stimulation test, based on the premise that chronic ACTH deficiency results in adrenal atrophy and the inability to increase cortisol secretion normally in response to an acute ACTH stimulus. The test is performed by measuring the serum cortisol concentration before and 60 minutes after administering cosyntropin (synthetic ACTH 1–24; 0.25 mg) intramuscularly or intravenously; a stimulated cortisol concentration ≥ 15–18 μg/dL is a normal response. A grossly elevated basal serum cortisol concentration may suggest the possibility of hypercortisolism (Cushing's syndrome) resulting from an ACTH-secreting corticotroph adenoma, but additional more specific evaluation is required for diagnosis of Cushing's syndrome and to determine its cause,[258] as described below (see Corticotroph Adenomas).

Testing for prolactin deficiency is unnecessary because the only clinical consequence relates to nursing and there is no effective treatment for failed lactation due to prolactin deficiency. Whereas hyperprolactinemia obviously suggests a functional lactotroph adenoma, prolactin levels also are commonly elevated in women with nonfunctional pituitary adenomas and other sellar masses.

Gonadotroph Adenomas

The large majority of gonadotroph adenomas are nonfunctional, do not secrete significant amounts of FSH or LH, and do not cause clinical symptoms; 80–90% of all

nonfunctional pituitary adenomas derive from gonadotrophs. Moreover, even moderately elevated gonadotropin levels do not cause specific symptoms. Consequently, even large gonadotroph adenomas most often present with headaches and visual disturbances rather than with amenorrhea.[259–261] Rare FSH-secreting adenomas may cause anovulation and spontaneous ovarian hyperstimulation, resulting in amenorrhea, multiple large ovarian cysts, and high serum FSH and estradiol levels;[262–264] in prepubertal girls, they may cause breast development and vaginal bleeding.[265] However, most patients with gonadotroph adenomas have normal or low serum gonadotropin concentrations because the tumors are nonfunctional and disrupt menstrual function only indirectly, via compression of the pituitary stalk or surrounding cells. They may inhibit gonadotropin secretion by interrupting the delivery of hypothalamic GnRH or by compressing normal gonadotrophs. Alternatively, they may cause hyperprolactinemia by interfering with the inhibitory actions of dopamine on lactotrophs, resulting in a secondary suppression of hypothalamic GnRH secretion and amenorrhea. Although such "null cell" tumors rarely produce clinically significant amounts of gonadotropins, they do exhibit gonadotropin production *in vitro* or gene expression at the mRNA level.[266, 267] Gonadotroph adenomas also may secrete large amounts of the α-subunit common to all of the pituitary glycoprotein hormones (having no intrinsic biological activity and therefore causing no symptoms).

A nonfunctioning gonadotroph adenoma should be suspected in patients having a serum prolactin level under 100 ng/mL and no signs of hyperthyroidism, acromegaly, or Cushing's syndrome. In the presence of low or normal gonadotropin levels, an elevated level of free α-subunit also suggests a nonfunctioning gonadotroph adenoma.[266] In rare women with functioning gonadotroph adenomas, the serum FSH may be elevated when the LH level is low. In premenopausal women, elevated levels of both FSH and estradiol, associated with multicystic ovaries and endometrial hyperplasia, strongly suggest a functional FSH-secreting gonadotroph adenoma causing ovarian hyperstimulation.[262, 268–270] Treatment with a GnRH agonist fails to down-regulate gondadotropin secretion in women with such tumors and may even cause or exacerbate ovarian hyperstimulation.[263]

In patients with neurological symptoms or clinically important excess gonadotropin secretion, transsphenoidal surgical resection of a gonadotroph or other nonfunctioning adenoma can provide rapid relief. If surgery succeeds in removing the adenoma but not the normal pituitary, gonadotropin secretion and ovarian function should return to normal. However, surgery also may result in additional pituitary hormone deficiencies or a more global panhypopituitarism.[271] Serious complications of transphenoidal surgery are uncommon, occurring in less than 5% of patients, but include worsening vision, hemorrhage, and cerebrospinal fluid rhinorrhea leading to meningitis. Variations in the secretion of antidiuretic hormone (ADH), causing diabetes insipidus or the opposite problem, the syndrome of inappropriate andidiuretic hormone (SIADH), are more common but also usually only transient. Conventional or sterotactic radiation therapy is useful when post-operative imaging reveals significant residual tumor or progressive regrowth. In patients having no neurologic symptoms or clinically important gondotropin secretion, nonfunctioning microadenomas and macroadenomas can be monitored carefully by serial imaging at annual intervals (beginning 6 months after diagnosis for macroadenomas) to detect progressive growth, and with decreasing frequency in the absence of change. Any associated pituitary hormone deficiencies should be replaced.

Thyrotroph Adenomas

Functional thryrotroph adenomas are a rare cause of hyperthyroidism, accounting for less than 1% of all functional pituitary tumors. Most patients present with the typical signs and symptoms of hyperthyroidism; other clinical manifestations include a diffuse goiter, visual impairment, menstrual disturbances, and galactorrhea. Because the TSH secreted by thyrotroph adenomas can vary greatly in biological activity and in immunoactivity, serum TSH concentrations also can vary widely, ranging from normal (but still inappropriately

high in the presence of hyperthyroidism) to markedly elevated.[272] The majority of patients with functional thyrotroph adenomas also have elevated serum concentrations of free α-subunit.

Transphenoidal surgery is the standard treatment for patients with functional thyrotroph adenomas, but has yielded mixed results, primarily because most such tumors are macroadenomas and surgery often is not curative. For those with persistent tumor, treatment with the somatostatin analog octreotide is effective in most,[272] so much so that 6–12 months of preoperative treatment with octreotide may be considered for patients with large functioning thyrotroph adenomas. Antithyroid therapy is not indicated, because the decrease in thyroid hormone may be expected to stimulate TSH secretion and tumor growth.

Somatotroph Adenomas

Functional GH-secreting somatotroph adenomas are the cause of more than 95% of cases of acromegaly.[257] GH stimulates excess hepatic IGF-1 secretion, which, in turn, causes most of the clinical features of the disorder. The characteristic signs of acromegaly are an enlarged jaw, and enlarged and swollen hands and feet, resulting in increasing shoe and ring size, and joint symptoms relating to hypertrophic arthropathy. However, the onset and progression of acromegaly is extremely slow and typically evolves over a period of several years. At the time of diagnosis, most patients have macroadenomas, many with extension outside of the sella.

Serum IGF-1 levels do not vary with meals, the time of day, or exercise, but do vary with age, being highest during puberty and decreasing gradually thereafter. The IGF-1 level therefore must be interpreted according to established age-specific norms. Most patients with acromegaly also have elevated serum GH levels, but concentrations fluctuate widely in response to a variety of stimuli, including fasting, exercise, stress, and sleep; levels also may be elevated in those with poorly controlled diabetes, liver disease, and malnutrition. Because random serum GH measurements can be difficult to interpret confidently, the most specific test for diagnosis of acromegaly is an oral glucose tolerance test. In normal individuals, GH levels fall to very low levels within 2 hours after ingesting 75 g of glucose (<0.3 ng/mL when measured with a modern highly sensitive immunoradiometric or immunochemiluminescent assay); a GH level greater than 0.3 ng/mL is abnormally high.[273]

Transsphenoidal surgery is the treatment of choice for patients with somatotroph adenomas. The outcomes achieved with surgery generally are very good; in 80–90% of patients with microadenomas, GH secretion declines to normal and other pituitary functions are preserved; results are more variable in those with macroadenomas.[274, 275] Clinical symptoms rapidly improve after successful surgery. Effective medical treatments are available for treatment of persistent or recurrent tumor and symptoms, including somatostatin analogs (e.g., octreotide, lanreotide), which inhibit GH secretion, and GH receptor antagonists (e.g., pegvisomant), which lower IGF-1 concentrations.

Corticotroph Adenomas

Functional ACTH-secreting corticotroph adenomas are the specific cause of Cushing's disease and one cause of the more general disorder, Cushing's syndrome, which results from an excess of circulating glucocorticoids. The most common cause of Cushing's syndrome is the ingestion of prescribed glucorticoids (e.g., prednisone), although oral, injected, topical, and inhaled glucocorticoids also may cause the disorder.[276–278] Other causes include cortisol-secreting adrenal adenomas and carcinomas and ectopic production of ACTH or corticotrophin-releasing hormone (CRH) by brochial carcinoids and other rare tumors.

The classical clinical manifestations of Cushing's disease result primarily from hypercortisolism, caused by increased ACTH stimulation of the adrenals, and vary with the duration

and extent of excess cortisol secretion. The most common features are progressive central obsesity, those resulting from excess fat accumulation in the cheeks ("moon face") and nuchal fat pad ("buffalo hump"), those caused by atrophy of the skin and subcutaneous tissue (easy bruising and purple striae on the abdomen and flanks), and hyperpigmentation (caused by excess ACTH), which is most noticeable in areas exposed to light (the face, neck, and back of the hands) or chronic mild trauma, friction, or pressure (the elbows, knees, knuckles, and shoulders). Menstrual abnormalities are common, affecting 80% of women with one third developing amenorrhea.[279] Other common signs and symptoms result from mild androgen excess (hirsutism, acne) and the effects of excess cortisol on skeletal muscle (proximal muscle wasting and weakness), bone (osteoporosis), and glucose metabolism (insulin resistance, glucose intolerance, diabetes). Functional corticotroph adenomas are usually quite small and may be difficult to image.

The diagnostic evaluation for suspected Cushing's syndrome begins by measuring 24-hour urinary free cortisol excretion (twice), the late-night (11:00 P.M.) salivary cortisol level (twice), or by performing an overnight or low-dose dexamethasone suppression test; the three screening tests have similar diagnostic accuracy.[280] Creatinine excretion should be measured in the same specimen to judge compliance with instructions because a reliable 24-hour urine collection can be difficult to obtain. The urinary cortisol excretion and the late-night salilvary cortisol level are interpreted by comparison to established normal laboratory ranges. The overnight dexamethasone suppression test is performed by administering 1.0 mg of dexamethasone between 11:00 P.M. and midnight and measuring the serum cortisol at 8:00 A.M. the following morning; a value less than 1.8 μg/dL is a normal result.[258] The low-dose dexamethasone suppression test is performed by administering 0.5 mg of dexamethasone every 6 hours over 2 days for a total of 8 doses (e.g., 8:00 A.M., 2 P.M., 8:00 P.M., and 2:00 A.M.), measuring serum cortisol 2 or 6 hours after the last dose; as with the overnight test, a value less than 1.8 μg/dL is a normal result.[258] If the first screening test is abnormal, a second, different, test should be performed. Those with concordant abnormal results require additional evaluation to determine the cause of Cushing's syndrome, as discussed in detail in Chapter 13. Those with discordant results also merit further testing. Those with normal results require no further evaluation, except when the clinical suspicion is high, based on the clinical presentation.

The treatment of choice for Cushing's disease is transsphenoidal surgery; among experienced surgeons, the permanent cure rate is approximately 70%. Radiation therapy is an option for those not cured by surgery and is effective in approximately 45% of adult patients. Bilateral total adrenalectomy, requiring lifelong daily glucocorticoid and mineralocorticoid treatment, is the final and definitive cure.[281]

Lactotroph Adenomas (Prolactinomas)

Functional lactotroph adenomas are common, accounting for approximately 40% of all clinically recognized pituitary adenomas. Most arise from clonal expansion of a single cell, presumably due to a somatic mutation. They also may occur as part of the MEN1 syndrome, a possibility that can be effectively excluded by measuring the serum calcium level since primary hyperparathyroidism (hypercalcemia associated with an inappropriately high parathyroid hormone level) is the most common clinical manifestation of the disorder. Approximately 10% of adenomas that secrete prolactin also secrete GH, leading some to recommend measuring the serum IGF-1 concentration, even in women with microadenomas.[282] In women with lactotroph adenomas, serum prolactin concentrations generally correlate with the size of the adenoma. Microadenomas usually are associated with serum prolactin concentrations less than 200 ng/mL and macroadenomas with higher levels, but exceptions in either case are not uncommon. In some women with large lactotroph macroadenomas, prolactin levels are only modestly elevated because the tumor is largely cystic, or due to an assay artifact that may occur when a test sample contains a

massive excess of antigen that saturates both the capture and signal antibodies, preventing them from forming a "sandwich" in immunoradiometric and chemiluminescent assays (known as the "hook effect").[283]

Hyperprolactinemia commonly results in menstrual disturbances and is the cause of secondary amenorrhea in up to 30% of women.[284] The mechanism relates to inhibition of hypothalamic GnRH secretion, which, in turn, results in decreased pituitary gonadotropin secretion and in anovulation or a more severe hypogonadotropic hypogonadism, depending on the level of hyperprolactinemia and the extent to which gonadotropin secretion is suppressed. Chronic hypogonadism may result in progressive osteopenia that improves after normal prolactin levels are restored, but bone mineral density does not always return to normal.[285] Hyperprolactinemia also may result in galactorrhea, but most hyperprolactinemic women do not have galactorrhea, primarily because their estrogen levels are abnormally low. Although prolactinomas are much more common in adults, they can cause growth failure and primary amenorrhea in children.[286] Postmenopausal women with prolactinomas do not exhibit the classical symptoms and often are recognized only when a large tumor causes neurological symptoms.

Medical Treatment

Dopamine agonists are the first treatment of choice for women with functional prolactin-secreting lactotroph adenomas of all sizes because they effectively lower prolactin levels and decrease the size of more than 90% of such tumors.[287] Bromocriptine and cabergoline both are highly effective. Cabergoline, a selective dopamine receptor type 2 agonist, has fewer side effects, greater potency, and also is more effective than bromocriptine in restoring normal prolactin levels in women with lactotroph adenomas.[26, 27] Consequently, most consider it the better choice. However, even at relatively low doses, long-term use of cabergoline may increase the risk of hypertrophic valvular heart disease.[30, 31] Bromocriptine poses no such risk and some therefore consider it a somewhat safer choice, reserving cabergoline for those patients who prove intolerant or resistant to bromocriptine. To minimize size effects, treatment with either drug should begin with a low dose (e.g., bromocriptine 1.25 mg at bedtime; cabergoline 0.25 mg twice weekly) and increase gradually, guided by serial prolactin levels obtained at approximately monthly intervals. In general, the higher the prolactin level, the greater the dose of dopamine agonist required to restore normal concentrations. In women with large macroadenomas and very high prolactin levels, doses of bromocriptine up to 5 mg twice daily and of cabergoline up to 1.5 mg twice weekly may be required. Either drug also may be administered vaginally in women who cannot tolerate oral treatment.[33, 34]

Prolactin levels typically decrease within 2–3 weeks after treatment begins and can be normalized in virtually all patients with microadenomas and in nearly all of those with macroadenomas.[288] In women with macroadenomas causing visual impairment, significant improvement can be observed within days after treatment begins and increases gradually over a period of months.[289, 290] Significant shrinkage of adenomas may begin in as little as 6 weeks but can be assessed more confidently by repeating the MRI after 3–6 months of treatment. Longer durations of treatment often achieve further shrinkage in the size of the adenoma.[288, 291, 292] In women with macroadenomas, size generally decreases in parallel with the prolactin level. However, the overall response does not always correlate with the basal prolactin level, the absolute or relative decrease in concentrations, or even the normalization prolactin levels. ***The failure of a tumor to shrink significantly in size despite a normalization of prolactin levels strongly suggests it is a nonfunctioning adenoma, rather than a functional lactotroph adenoma.*** Menses, ovulation, and fertility typically return when normal prolactin levels are restored.[26] For women seeking pregnancy, the safety of both drugs is now established.[35, 36] In those who cannot tolerate dopamine agonist treatment and therefore remain hyperprolactinemic and anovulatory, ovulation induction can be induced with exogenous gonadotropins.

In women with microadenomas, the dose of dopamine agonist often can be reduced after approximately one year of treatment, and a trial discontinuation can be attempted if prolactin levels have been normal for 2 years or more and MRI reveals no evidence of the adenoma. In those with macroadenomas, an MRI should be repeated after 6 and 12 months of treatment to determine the extent to which the adenoma has decreased in size. If the prolactin level has been normal for a year or more and the adenoma has decreased significantly in size, the dose of treatment can be reduced gradually as long as the prolactin level remains normal.[293] As with microadenomas, a trial discontinuation of treatment can be attempted after prolactin levels have been normal for 2 years if MRI reveals no persistent tumor. Frequently, but not always, prolactin levels rise again after discontinuation of treatment. In a large retrospective study, recurrent hyperprolactinemia was observed in 75% of patients with microadenomas and in 84% of those with macroadenomas treated with bromocriptine.[294] However, in a prospective study involving 105 patients with microadenomas and 70 with macroadenomas, treatment was discontinued when prolactin levels were normal and MRI demonstrated no residual tumor, or more than a 50% decrease in size with no cavernous sinus invasion and more than 5 mm separation from the optic chiasm; hyperprolactinemia returned in only 31% of patients with microadenomas and in 36% of those with macroadenomas and no tumor regrowth was observed in any patient after 2–5 years of observation.[295] The best candidates for discontinuation of treatment are those with normal serum prolactin levels and very little or no visible residual tumor on MRI.

A dopamine agonist is the best initial treatment for women with prolactin-secreting macroadenomas and is certainly a suitable choice, but not the only treatment option, for women with microadenomas. As in hyperprolactinemic women having no adenoma, treatment can be tailored to the patient's needs and goals. Treatment with a dopamine agonist is the clear choice for women ready to attempt pregnancy (for ovulation induction) and those with troublesome galactorrhea. Others may be treated with physiologic cyclic estrogen/progestin therapy or with combined hormonal contraception (orally or vaginally administered), according to their contraceptive needs; such treatments pose little or no risk for stimulating tumor growth.[37, 38] However, it is prudent to monitor serum prolactin levels approximately every 6 months and to repeat the MRI 1 and 2 years after diagnosis to reassess the size of the adenoma.

Surgical Treatment

In patients with lactotroph adenomas who cannot tolerate or prove resistant to dopamine agonist treatment, transsphenoidal surgery is an appropriate alternative. Surgery might also be considered for women with very large macroadenomas (e.g., >3 cm) who want to attempt pregnancy, even when their tumor responds to medical treatment. Although surgery offers the possibility of a permanent cure, it also has important limitations and potential consequences. Incomplete resection of the tumor and persistent hyperprolactinemia are unfortunately common, more so in patients with macroadenomas (up to 90%) than in those with microadenomas (approximately 30%), as might be expected, and depending upon the skill and experience of the neurosurgeon. Recurrent hyperprolactinemia and tumor also may be observed within 5 years after surgery.[296, 297] In general, the higher the initial prolactin level, the lower the permanent cure rate. Not surprisingly, better outcomes can be expected when the adenoma lies entirely within the sella, emphasizing the importance of preoperative medical treatment for patients with extrasellar lesions.[298] The best predictor of a long-term cure is the prolactin level on the day following surgery. In the largest series with long-term follow-up on more than 400 women, the recurrence rate was 26% when postoperative prolactin levels were 20 ng/mL or less.[299] The risks and potential consequences of transsphenoidal surgery otherwise are the same as for other types of pituitary adenomas (see Gonadotroph Adenomas, above). For those with persistent hyperprolactinemia and tumor, management options include dopamine agonist treatment and radiation therapy.

Radiation Therapy

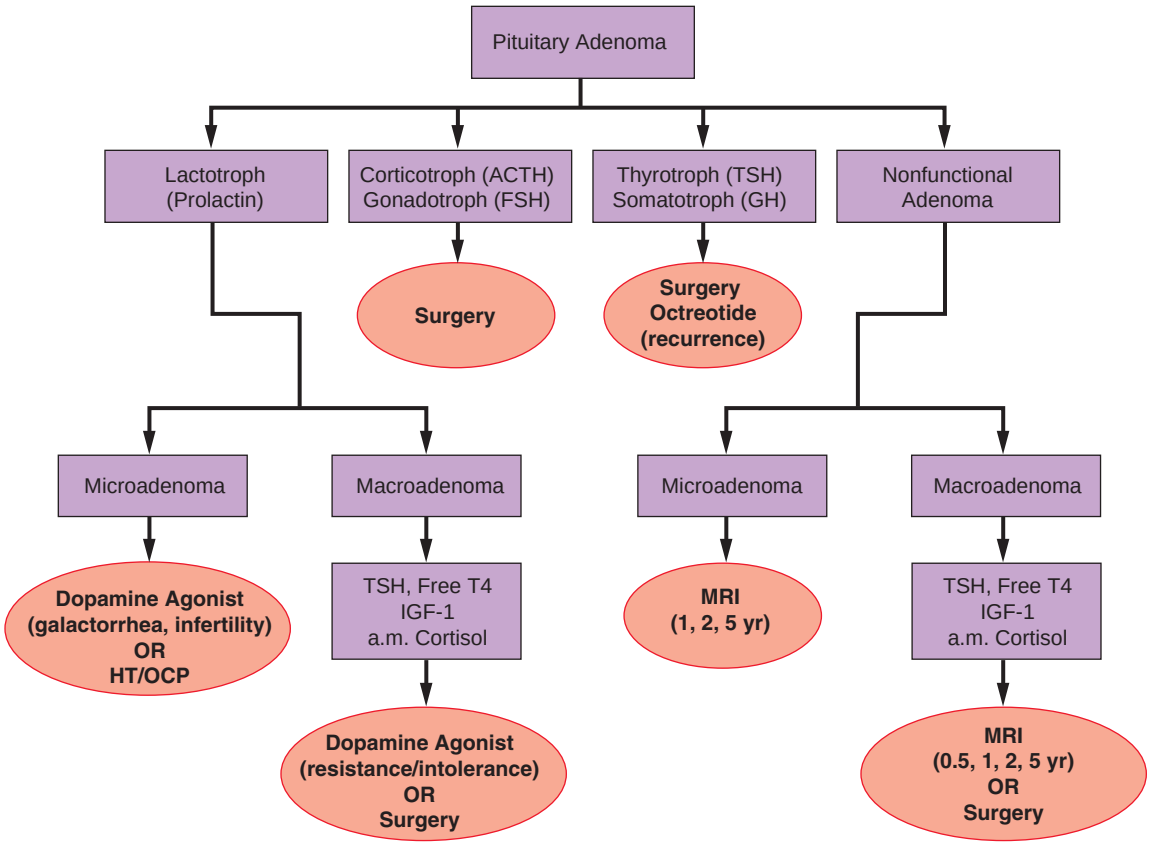
Radiation therapy can decrease the size of lactotroph adenomas, but prolactin levels and tumor size decrease only very slowly, over a period of several years.[300, 301] Therefore, radiation therapy is used primarily in the management of women with large macroadenomas having significant residual tumor after surgery. Approximately half of patients treated with radiation slowly develop panhypopituitarism over the 10 years following treatment.[302] Consequently, patients treated with radiation must be followed carefully over time, alert to the signs and symptoms of pituitary hormone deficiencies.

Management During Pregnancy

Not surprisingly, given the effectiveness of dopamine agonist treatment in restoring ovulation and fertility in women with hyperprolactinemia, many women with functional lactotroph adenomas become pregnant. Overall, approximately 80% of hyperprolactinemic women, with or without adenomas, achieve pregnancy with dopamine agonist treatment.[303–305] Understandably, because the normal pituitary gland approximately doubles in size by the third trimester of pregnancy[306] and estrogen levels are quite elevated throughout, there is some increase in the risk for tumor growth during pregnancy. The risk for clinically significant growth in women with microadenomas is extremely low—only approximately 1–2%.[307] About 5% will develop asymptomatic tumor enlargement (as determined by imaging), and essentially none will ever require surgical intervention. The risk is significantly higher (approximately 15–20%) in those with macroadenomas.[304, 307] Nevertheless, serial prolactin measurements during pregnancy are unnecessary.

Regardless of the size of the adenoma, there is no indication for treatment with dopamine agonists or for imaging during pregnancy in the absence of symptoms; treatment may be safely discontinued when pregnancy is established. In women with microadenomas, the serum prolactin should be measured approximately 2 months after delivery or the cessation of nursing and, if still elevated, treatment with a dopamine agonist, hormone therapy, or hormonal contraception can resume, according to the patient's needs. In women with macroadenomas, an interval of treatment with a dopamine agonist before pregnancy is advisable, to shrink the tumor and reduce the associated risk. In those with macroadenomas that fail to shrink with treatment, pregnancy should be avoided until after surgical debulking since medical treatment is not likely to be effective if symptoms develop.

In the few women with macroadenomas who experience significant tumor growth during pregnancy, headaches usually precede visual disturbances, and both may occur in any trimester. The headaches have no specific characteristics and vary in intensity, location, and character. Bitemporal hemianopsia (tunnel vision) is the classic visual impairment, but other defects can occur. The symptoms typically regress promptly with resumption of dopamine agonist treatment. Although seldom necessary, such treatment poses no risk to the fetus.[35, 303, 308–310] Amniotic fluid prolactin, with its presumed actions on the regulation of amniotic fluid water and electrolyte balance, derives from the decidua and its secretion is controlled by estrogen and progesterone, not by dopamine, which has no effect on amniotic fluid prolactin levels. ***Breastfeeding poses no significant risk for tumor growth in women with microadenomas or macroadenomas that remain asymptomatic during pregnancy, but is contraindicated for those with neurologic symptoms at the time of delivery.***[311] Dopamine agonist treatment should not resume until after cessation of nursing.



Pituitary Incidentaloma

MRI has high sensitivity for detecting small lesions and may reveal a pituitary tumor when the head is imaged for reasons other than suspected pituitary disease. In sum, 13 autopsy studies involving the examination of more than 10,000 pituitary glands identified more than 1,000 unsuspected microadenomas and 3 macroadenomas, yielding an overall prevalence of pituitary "incidentaloma" of approximately 10%.[312]

Relatively little is known about the fate of such common asymptomatic sellar masses. However, in one series involving 506 patients, all having serum prolactin levels less than 100 ng/mL and no evidence of other pituitary hormone hypersecretion or deficiency, 258 had immediate surgery and 248 received expectant management for an average of 27 months. In those receiving surgery, 209 (81%) of the lesions were nonfunctional adenomas and 41 were cysts. Among those followed with serial imaging, the lesion grew in 30 (12%), shrunk in 29 (12%) and remained unchanged in all others. Among the lesions that grew in size, 20 were initially ≥10 mm, 10 were less than 10 mm, and only 3 grew to more than 10 mm over 65–84 months of observation.[313] In 2 smaller studies involving a total of 98 patients with pituitary incidentalomas, 10/41 (25%) tumors ≥10 mm and only 1 of 57 smaller tumors grew over 2–11 years of observation.[314, 315]

Based on the available data, it generally is recommended that patients with pituitary incidentalomas ≥10 mm in size be evaluated in the same way as those with symptomatic macroadenomas, with pituitary function tests and visual field examination, because such tumors are more likely to be associated with abnormalities in pituitary hormone secretion and to enlarge over time. Those with functional macroadenomas should be treated

accordingly and those with nonfunctional tumors can be followed with serial testing and imaging at 6 months, 1 and 2 years, and less frequently thereafter; any associated hormone deficiencies should be treated appropriately. In patients with smaller incidentalomas (<10 mm), a serum prolactin should be obtained, but other endocrine evaluation is unnecessary when there are no clinical signs or symptoms to suggest a functional tumor.[316] In the absence of any clinical or endocrine abnormalities, MRI can be repeated once after 1–2 years to identify the very few that may exhibit significant growth.

Empty Sella Syndrome

The "empty sella syndrome" is a misnomer because the sella turcica is not, in fact, empty. The sella is enlarged and appears empty on imaging because it contains cerebrospinal fluid, still within the subarachnoid space but extending downward into the pituitary fossa. The disorder is included here, in our discussion of pituitary adenomas, because it most commonly results from the previous removal or destruction of a pituitary adenoma by surgery, radiation, or infarction. Alternatively, it may result from a congenital defect in the sellar diaphragm ("primary empty sella"). In either case, the remaining or otherwise normal pituitary tissue is flattened against the sellar floor, which may become demineralized due to the increased pressure within the pituitary fossa.

In autopsy studies, the prevalence of an empty sella is approximately 5%, and approximately 85% are in women.[317] The prevalence in women with amenorrhea and galactorrhea is between 4% and 15%.[284, 318] Not surprisingly, the syndrome may coexist with an adenoma and, less commonly, with deficiencies in pituitary hormone secretion that can be severe.[319] However, the condition usually is quite benign and does not progress to pituitary failure. There is no convincing evidence to indicate that a primary empty sella causes pituitary insufficiency.

Because of the possibility of a coexisting adenoma, patients with hyperprolactinemia and an empty sella should undergo annual surveillance (prolactin assay and imaging) for a few years to detect any evidence of tumor growth. Treatment for the condition is dictated by the associated disturbances in pituitary hormone secretion.

Sheehan's Syndrome

Acute infarction and ischemic necrosis of the pituitary gland resulting from postpartum hemorrhage and hypovolemic hypotension is known as Sheehan's syndrome and is one of the most common causes of hypopituitarism in underdeveloped or developing countries.[320, 321] Failed lactation after delivery is the classical presenting symptom. The rest of the clinical picture varies with the severity of the pituitary insult, ranging from severe hypopituitarism soon after delivery, manifesting as lethargy, anorexia, and weight loss, to secondary amenorrhea, loss of sexual hair, and less severe symptoms of fatigue that emerge weeks and months later.[322–324] Deficiencies in GH, prolactin, and gonadotropins are most common, although the majority also exhibit ACTH and TSH deficiencies. Approximately one-third of patients may have hyponatremia, but diabetes insipidus is almost never observed.[323, 324] A partially or completely empty sella is a common later finding.

The evaluation and treatment of Sheehan's syndrome is no different from that for other causes of hypopituitarism, with one caveat. Any ACTH stimulation test that may be performed to detect a secondary adrenal insufficiency should be postponed until approximately 6 weeks

after delivery. Adrenal atrophy due to chronic ACTH deficiency, resulting in the inability to increase cortisol secretion normally in response to an acute ACTH stimulus, requires time to develop. Earlier testing therefore may yield inaccurate (false negative) results.

Infiltrative Pituitary Lesions

Infiltrative pituitary lesions that may cause hypogonadotropic hypogonadism include hemochromatosis and lymphocytic hypophysitis.

Hemochromatosis

Hereditary hemochromatosis is an inherited autosomal recessive disorder caused by mutations in the *HFE* gene (on chromosome 6) that alter the size or shape of the HFE (hemochomatosis) protein, preventing its transport to the cell surface where it normally interacts with the transferrin receptor, which plays an important role in regulating the amount of iron that enters the cell.[325] Excessive absorption of dietary iron is the consequence, leading to parenchymal iron overload and subsequent tissue damage. In the pituitary, gonadotrophs are the most common cell type affected, resulting in hypogonadotropic hypogonadism;[326] TSH and ACTH deficiencies are less common. An acquired form of hemochromatosis may result from frequent transfusions in individuals with severe anemias (sickle cell disease, beta thalassemia major, aplastic anemia).

The best screening test for hereditary hemochromatosis is a fasting transferrin saturation (a ratio of serum iron to total iron binding capacity, expressed as a percentage;[327] values greater than 45% are are an indication for *HFE* genotyping.[328] Early diagnosis and treatment (phlebotomy, chelation therapy) help to prevent serious disease relating to iron deposition in the liver, pancreas, anterior pituitary and heart. Although hereditary hemochromatosis is an uncommon cause of hypogonadotropic hypogonadism, some have suggested that iron studies should be performed in all patients with hypopituitarism and normal imaging.[329]

Lymphocytic Hypophysitis

Lymphocytic hypophysitis is a rare autoimmune disorder causing enlargement of the pituitary that mimics a pituitary tumor, most often occurring during pregnancy or in the first 6 months postpartum.[330, 331] The chronic inflammatory process results in focal or diffuse adenohypophysial destruction of varying severity and subsequent fibrosis. In the initial phase of hypophysitis, hyperprolactinemia is common, followed by progressive hypopituitarism. The disorder should be considered in women with sellar enlargement soon after pregnancy and in those with hypogonadism and a coexisting autoimmune disorder. Patients with symptoms and signs of pituitary enlargement and surpasellar extension can be treated by transsphenoidal surgery, dopamine agonists, anti-inflammatory or immunosuppressive drugs, or by pituitary radiotherapy.[332]

Disorders of Hypothalamic Function

Hypothalamic dysfunction is one of the most common causes of secondary amenorrhea. In its most severe form, commonly known as ***hypothalamic amenorrhea***, the HPO axis is profoundly suppressed—abnormally low levels of hypothalamic GnRH secretion stimulate only basal amounts of pituitary gonadotropin secretion, which, in turn, fail to stimulate ovarian follicular development, resulting in very low levels of estrogen production.[333] The clinical manifestations of dysfunctional hypothalamic GnRH secretion depend on the extent to which gonadotropin secretion is suppressed, much like in women with hyperprolactinemia

and probably involving similar mechanisms. A minor disturbance in pulsatile hypothalamic GnRH secretion may result only in poor luteal function due to decreased LH stimulation of corpus luteum progesterone secretion, and more significant dysfunction in disordered follicular development and chronic anovulation presenting as oligomenorrhea or amenorrhea, as discussed in an earlier section of this chapter devoted to the evaluation of ovarian function (see Chronic Anovulation). The primary focus here is on the more severe disorder, hypothalamic amenorrhea, characterized by overt hypogonadotropic hypogonadism. Other rare hypothalamic causes of amenorrhea include genetic mutations resulting in a congenital GnRH deficiency and infiltrative diseases involving the hypothalamus such as lymphoma, Langerhans cell histiocytosis, and sarcoidosis.

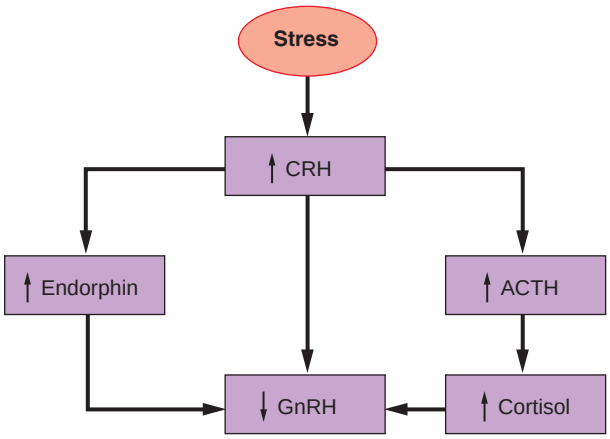
Hypothalamic Amenorrhea

Hypothalamic amenorrhea is a diagnosis of exclusion, based on findings of a low or normal serum FSH concentration, despite low levels of estrogen production, in the absence of any sellar mass lesion or reason to suspect other rare pituitary causes of hypogonadotropic hypogonadism.

Hypothalamic amenorrhea frequently is associated with extreme physical, nutritional, or emotional stress, suggesting it represents a functional suppression of reproduction as a psychobiologic response to life events.[334] Affected women frequently are underweight (e.g., ≥10% below ideal body weight), report recent weight loss, or engage in regular strenuous exercise, although the critical threshold weight and the amount of weight loss or exercise required to induce amenorrhea varies significantly among individual women. Women with hypothalamic amenorrhea often exhibit abnormal eating patterns and generally expend more calories in aerobic activity and have a higher fiber intake and lower percent body fat than normally cycling women.[335] Many have endocrine, metabolic, and psychological characteristics that suggest a subclinical eating disorder.[336, 337] However, in others with hypothalamic amenorrhea, no obvious cause or precipitating event can be identified.

Regardless of the cause, the large majority of women with hypothalamic amenorrhea exhibit an abnormal pattern of hypothalamic GnRH secretion, as inferred from the pattern of pulsatile gonadotropin secretion. Some have no detectable gonadotropin secretory pulses (8%) and others exhibit low frequency/amplitude (27%), low amplitude/normal frequency (8%), low frequency/normal amplitude (43%), or normal frequency/amplitude pulses (14%); different patterns may be observed over time.[338] Differences in the glycosylation of secreted gonadotropins, resulting in reduced bioactivity, also help to explain why those with normal immunoactive serum gonadotropin levels are nonetheless hypogonadal.[62, 339] Almost half exhibit augmented LH secretion during sleep and higher serum levels of FSH than LH, as observed in pubertal girls.

Observations of increased cortisol secretion in women with idiopathic hypothalamic amenorrhea suggest that stress may interrupt reproductive function indirectly, by activating the hypothalamic-pituitary-adrenal axis.[340–342] Evidence suggesting the mechanism derives from experiments in nonhuman primates demonstrating that corticotropin-releasing hormone (CRH) inhibits gonadotropin secretion, probably by augmenting endogenous central opioid secretion (endorphins).[343] Interestingly, cortisol levels return to normal months before the return of menses in women with hypothalamic amenorrhea, further suggesting the importance of stress, mediated via the adrenal axis.[344] Other studies indicate that increased hypothalamic dopaminergic inhibition of pulsatile GnRH secretion may be involved in at least some women with hypothalamic amenorrhea.[345]



The natural history of hypothalamic amenorrhea has been examined in a few studies that have followed women with the disorder over time. In those with hypothalamic amenorrhea associated with psychological stress or weight loss, spontaneous recovery with resumption of menses was observed in over 70% after 6–8 years of observation.[346, 347] Women who recovered had higher BMIs and lower cortisol levels than those who did not. Others have observed that women with an established cause or inciting event for their hypothalamic amenorrhea (stress, weight loss, eating disorder) were more likely to recover than those in whom no cause could be identified.[348] In patients with eating disorders, a return of menstrual function is associated with weight gain, and recurrent amenorrhea with weight loss.[349]

Eating Disorders

St. Wilgefortis was the seventh daughter of the King of Portugal, living around the year 1000. When confronted with an arranged marriage (despite her vow of virginity to join a convent), she turned to intense prayer, became ascetic and anorexic, and developed a diffuse growth of body hair, including a beard. Confronted with this new appearance, the King of Sicily withdrew his offer of marriage, and Wilgefortis's father had her crucified. Around 1200, the legend of Wilgefortis spread throughout Europe.[350]

St. Wilgefortis became a symbol, a woman who liberated herself from the burdens of womanhood, and became a protectress of women with sexual problems, including those associated with childbirth. Women who wished to free themselves from their husbands prayed to her, because she had successfully resisted both a father and an unwelcome suitor. In England, she was known as St. Uncumber because women believed she could uncumber them from their husbands. The medieval explanation (with ascendancy to sainthood) for a young girl's response (anorexia nervosa) to her fears of marriage and sexuality is still, to some extent, viable today, as anorexia can be the expression of the need to exert control over some facet of a rapidly changing and increasingly demanding and stressful life.

Our contemporary culture is obsessed with weight loss, clearly exhibited by the constant introduction and wide promotion of new fad diets, exercise regimens and apparatus, and the public adulation of models, actors, and athletes. Taken together, they send an unequivocal and unfortunate message to maturing adolescents, as reflected by the results of a national Youth Risk Behavior Survey taken in 2005: 38% of adolescent girls (in grades 9–12) considered themselves overweight, 62% were trying to lose weight, 17% had recently fasted

for 24 hours or more, 8% had taken diet pills or similar products, and 6% had recently tried to induce vomiting or taken laxatives to lose weight or keep from gaining weight.[351] Although the estimated lifetime prevalence of eating disorders in women is relatively low, ranging between 0.3% and 2%,[352–355] these disturbing statistics reveal the size of the population at risk.

In addition to cultural influences, other psychological, biological, genetic, and social factors likely contribute to the development of eating disorders. Several have been associated with their development, including a history of dieting,[356] preoccupation with weight,[357] athletic and artistic pursuits that favor leanness or involve subjective judging, and possibly, sexual abuse.[358] Young women having a first degree relative with an eating or affective disorder or alcoholism are at increased risk for developing an eating disorder. Linkage analyses have identified possible susceptibility loci for anorexia nervosa on chromosome 1 and for bulimia nervosa on chromosome 10.[359, 360] Affective, anxiety, and obsessive-compulsive disorders, personality disorders, and substance abuse are common in women with eating disorders.[361] Family stresses relating to high perceived parental expectations (for success, achievement, and appearance), poor communication, and marital tension also may play a role.

The clinical spectrum of eating disorders ranges from a limited period of amenorrhea associated with a crash diet in otherwise normal women, to the grossly underweight anorexic having a distorted body image, and the bulimic who cycles regularly between binge eating and purging behaviors. The specific diagnostic criteria for anorexia nervosa and bulimia nervosa are defined in the Diagnostic and Statistical Mannual of Mental Disorders (DSM-IV) and are briefly summarized here. The DSM-IV categorizes those with clearly abnormal eating patterns and weight control habits that do not meet the specific criteria for anorexia nervosa or bulimina nervosa as having an “eating disorder, not otherwise specificed”.[362]

Anorexia Nervosa

1. Refusal to maintain body weight within a normal range for height and age (<85% of ideal body weight);

2. Fear of gaining weight or becoming fat, even though underweight;

3. Distorted body image, with undue importance on weight or shape;

4. In postmenarcheal females, amenorrhea.

Two subtypes of anorexia nervosa have been defined—restricting, and binge/purging. In the first, a restriction of food intake is the primary method for achieving weight control. In the second, binging and purging, via self-induced vomiting or the use of laxatives or diuretics, is the primary method for controlling weight. In both types, compulsive exercise may be an additional behavioral strategy for maintaining or losing weight.

The diagnostic criteria for bulimia nervosa are distinct from those for anorexia nervosa, primarily in that they do not include low body weight or amenorrhea.

Bulimia Nervosa

1. Episodic binge eating, consuming abnormally large amounts of food, with a sense of lack of control;

2. Recurrent compensatory behavior, including self-induced vomiting or misuse of laxatives, diuretics, or other medications, fasting, or excessive exercise;

3. Binging and purging behaviors occurring at least twice per week, on average for an interval of 3 months or more;

4. Dissatisfaction with body weight or shape;

5. The behavior does not occur exclusively during episodes of anorexia nervosa.

Two subtypes are again described—purging, and non-purging. In the first, the purging behavior includes regular self-induced vomiting or the misuse of laxatives or diuretics. In the second, other compensatory behaviors predominate, such as exercise.

The clinical symptoms of restrictive anorexia nervosa include weight loss that frequently dates from a specific event like an illness, insensitive comment, rebuke, or loss. Amenorrhea typically precedes weight loss, which begins with dieting and specific restriction of fat intake. Affected women often admit fatigue, nausea, early satiety or bloating after meals. They exhibit a distorted body image, denial, and disordered thinking, and frequently use exercise as an additional weight control strategy. In women with anorexia nervosa, the physicial examination may reveal hypotension, bradycardia, low body temperature, dry skin, and lanugo (fine, soft hair on the back, buttocks and extremities). Those with bulimia nervosa exhibit impulsive and addictive qualities to their behaviors, an inability to control binge eating and purging, and frequently use cigarettes, alcohol, and other drugs. Many have irregular menses, but not amenorrhea, and in most, weight fluctuates but is not abnormally low. Those with bulimia nervosa may have parotid gland hypertrophy and erosion of their teeth enamel (from frequent vomiting).

The metabolic abnormalities associated with anorexia nervosa reflect dysfunctional hypothalamic regulation of appetite, thirst, temperature, sleep, autonomic balance, and endocrine secretion.[363] The clinical consequences can be severe and even life-threatening. The associated endocrine abnormalities include low serum FSH, LH, estradiol, IGF-1, and leptin concentrations, and increased cortisol levels; prolactin, TSH and T4 levels are normal, but the T3 level is low, and reverse T3 (rT3) is high (rT3 is an isomer of T3, derived from T4, that binds but does not activate thyroid hormone receptors). With weight gain, all of the metabolic and endocrine abnormalities resolve. Even though normal gonadotropin secretion may be restored with weight gain, approximately one-third remains amenorrheic, likely reflecting persistent hypothalamic dysfunction.[364]

The treatment of anorexia nervosa and bulimia nervosa is complex, but generally requires nutrition, medical monitoring, and cognitive behavioral therapy in a multidisciplinary approach involving a physician, a dietitian, and a mental health professional. Antidepressant medications (primarily selective serotonin reuptake inhibitors) can be an important part of the management of bulimia nervosa, but have less value in anorexics. In those with anorexia nervosa, weight gain is the key to a successful outcome. Patients with severe medical or psychological symptoms may require hospitalization. Approximately half of patients with anorexia nervosa have good outcomes, as defined by weight gain and a return of menses, and about 25% improve but also have relapses; outcomes are poor in the remaining 25%.[365] Poor outcomes have been associated with older age at onset, longer duration of illness, lower nadir in weight, and lower body fat after weight gain.[366]

Osteopenia and osteoporosis are among the most serious complications of anorexia nervosa, the inevitable consequence of the combined adverse effects of severe malnutrition and estrogen deficiency. Although both contribute to bone loss, malnutrition clearly has the greater effect and importance.[367, 368] In those who do not gain weight or resume menses, bone mineral density decreases approximately 2.5% annually in both the spine and the hip. Conversely, in those who gain weight and resume menses, bone density increases by approximately the same amount in both locations. In women who resume menses, spine

bone density increases, independent of weight gain, and in women who gain weight, hip bone density increases, regardless whether they resume menses.[369] By itself, hormone therapy has little or no benefit. Treatment with oral contraceptives cannot prevent progressive bone loss in women with anorexia nervosa and generally has little value,[369–371] except perhaps in those with extremely severe illness (those below 70% of ideal body weight).[372] Undernourished women with eating disorders who do not achieve the rapid gain in bone density that normally occurs during adolescence may always have reduced bone density, even if they recover and resume normal menses.[373]

Eating Disorders and Pregnancy

The caloric requirements of pregnancy, the changes in body shape, and the need to gain weight present particular challenges for women with eating disorders. For some, pregnancy may offer an opportunity or reason for recovery, but for others, it increases stress and risk. In general, women with restrictive behaviors tend to gain relatively little weight during pregnancy and those with bulimic behaviors often gain excessive weight.

Caloric restriction and compensatory behaviors such as purging pose risks for both mother and fetus. Anorexia nervosa increases the risk for maternal malnutrition, fetal growth restriction, and low birth weight. A low pre-pregnancy weight and inadequate weight gain during pregnancy each independently increase the risk of preterm birth.[374] The incidence of hyperemesis gravidarum, miscarriage, preterm birth, cesarean delivery and postpartum depression is increased in pregnant women with eating disorders.[375–378] Women with anorexia nervosa in remission at time of conception gain more weight and have higher birth weights than those with active illness.[379] In most women with bulimia nervosa, symptoms improve during pregnancy, but often recur or worsen postpartum.[380]

Overall, the best outcomes are achieved when illness is well-controlled before pregnancy. Consequently, treatment for anovulatory infertile women with eating disorders should focus first on management of their underlying illness before turning to ovulation induction.[381] The goal of pregnancy can be a strong motivational force and provides a unique opportunity for effective intervention.

During pregnancy, diet, weight gain, and fetal growth should be carefully monitored and nutritional counseling provided. Calcium supplementation is particularly important since most women with established eating disorders have osteopenia, which can worsen during pregnancy and lactation. As in nonpregnant women with eating disorders, a team management approach works best. Active management must continue into the postpartum period, when symptoms may resurface, and should involve the pediatrician because women with eating disorders are more likely to underfeed their babies.[382]

Exercise and Amenorrhea

Soranus of Ephesus in the first century AD observed in his famous treatise, “On the Diseases of Women,” that amenorrhea is frequently observed in the youthful, the aged, the pregnant, in singers, and in those who take much exercise. In contemporary culture, a substantial proportion of reproductive age women engage in some kind of regular exercise. Whereas exercise clearly offers significant health benefits, it also may result in infertility or amenorrhea, and in adolescents may cause delayed puberty.[383]

Women who are involved in strenuous recreational exercise or other forms of demanding physical activity, such as dance, have a high prevalence of menstrual irregularity and

amenorrhea. As much as two-thirds of runners who menstruate exhibit a short luteal phase or anovulatory cycles.[384, 385] Typically, previously normal cycles become irregular after exercise begins and progress to amenorrhea as the intensity of exercise increases, particularly when accompanied by weight loss. Physical training that begins before menarche may delay its onset by as much as 3 years. ***Exercise alone does not cause amenorrhea; the specific type of exercise is important.*** Physical activities associated with a low body weight and high lean body mass (running, dance, gymnastics, figure skating) are associated with a higher incidence of amenorrhea than others like swimming. Low body weight, by itself, also does not cause amenorrhea, because the effect of exercise on menstrual pattern varies significantly among women having a similar BMI. ***The potential adverse effects of exercise and body weight on menstrual function are synergistic.***

The critical weight hypothesis holds that the onset and regularity of menstrual function require that weight remains above a critical threshold level, with a corresponding critical level of body fat, which is estimated at 17% for menarche, and at 22% for regular menstruation.[386] According to the hypothesis, excessive exercise or malnutrition may decrease the amount of body fat to below threshold values, resulting in delayed menarche in adolescents and in amenorrhea in adults. Logically, those at or near their critical weight and body fat content would be at greatest risk for loss of menstrual function.[387] Critics of the hypothesis acknowledge the correlation between body fat content and menstrual function, but deny any cause-effect relationship.[388] Indeed, normal and abnormal menstrual function may be observed at widely varying levels of body weight and fat content. However, the discovery of leptin (secreted by adipocytes in proportion to body fat stores) and of leptin receptors in the hypothalamus, revealing a feedback mechanism for the central regulation of body fat content, has stimulated renewed interest in the critical weight hypothesis (Chapter 19).

In addition to the influence of weight and body fat on menstrual function, stress and energy expenditure exert important and independent effects, illustrated by the observation that menses often return in dancers during intervals of rest associated with injury, in the absence of any change in body weight or fat content.[389] Therefore, it is not surprising that women with low body weight and fat content who engage in strenuous physical activity are highly susceptible to anovulation and amenorrhea. ***It appears that a negative energy balance, resulting when the level of energy expenditure exceeds the available supply (derived both from intake and available stores), predisposes to a disruption in pulsatile gonadotropin secretion and a loss of menstrual function.***[390]

Other mechanisms that may help to explain the effects of exercise on menstrual function involve the actions of endogenous opioids, activation of the adrenal axis, and leptin. Substantial evidence suggests that secretion of endogenous hypothalamic opioids, which inhibit GnRH secretion, increases after exercise.[391–394] Cortisol levels are increased in hypothalamic amenorrhea (including that relating to exercise),[341, 395, 396] suggesting increased activity in the adrenal axis, mediated by CRH, which inhibits GnRH secretion.[343] Leptin levels are low in exercising amenorrheic women, lower than can be attributed to body fat content alone.[397] Exercising amenorrheic women do not exhibit a normal diurnal leptin rhythm.[398] Moreover, treatment with exogenous recombinant human leptin can restore gonadotropin pulsatility, follicular development, and ovulatory function in exercising amenorrheic women.[399] A decrease in leptin levels, resulting from low body fat stores and a negative energy balance, may thus suppress the HPO and thyroid axes and stimulate the adrenal axis, thereby producing many of the endocrine features observed in amenorrheic female athletes.

A unifying hypothesis focuses on the importance of energy balance.[400, 401] When energy demands are high, as in exercise, or supplies are insufficient, as in eating disorders, reproduction is suspended in favor of more essential metabolic functions. Teleologically, the concept makes sense; adaptations to stress also inhibit menstrual function because conditions do not favor successful reproduction.

Regardless whether the critical factor is body weight, percent fat content, or energy balance, and whether endogenous opioids, CRH, or leptin mediate the effect, the mechanism by which exercise disturbs normal menstrual function relates directly to changes in the pattern of hypothalamic GnRH secretion, which may vary with body composition. In amenorrheic athletes with low body weight, both LH pulse frequency and amplitude are decreased; pituitary sensitivity to exogenous GnRH also is increased, further indicating that decreased gonadotropin levels result from decreased endogenous GnRH stimulation.[336] In contrast, when exercise is not weight-bearing and leanness is therefore less important, as in competitive swimmers, LH levels often are modestly elevated rather than suppressed, and estradiol concentrations are normal, suggesting a less severe disturbance in GnRH pulse rhythm that results in chronic anovulation and amenorrhea, but not in hypogonadotropic hypogonadism (hypothalamic amenorrhea) as more often occurs in runners and dancers.[402]

Because most women with exercise-induced amenorrhea are estrogen deficient, they are at risk for its natural clinical consequences, including genitourinary and breast atrophy and osteopenia. Most do not have associated vasomotor symptoms, reflecting their underlying hypothalamic dysfunction. ***It is important to emphasize that the beneficial effects of weight-bearing exercise on bone are not sufficient to prevent the adverse effects of estrogen deficiency, particularly in adolescents.*** In general, hypogonadism has greater detrimental effects on trabecular bone than on cortical bone. However, because skeletal loading patterns differ with the type of exercise, exercising estrogen-deficient women exhibit site-specific differences in bone density. Gymnasts have a higher bone density in the spine than runners, despite similar menstrual patterns and body fat.[403] In ballet dancers, cortical bone density may be normal or increased at weight-bearing sites, such as the proximal femur, but trabecular bone density is decreased in the lumbar spine.[404] Rowers may accrue greater bone density in the lumbar spine due to the mechanical loading produced in their exercise.[405] However, the normal stress-induced compensatory changes in bone density also may be impaired or prevented by estrogen deficiency.[406] Fractures, particularly stress fractures, are common in athletes, but more common in those with abnormal eating patterns,[407, 408] possibly because a negative energy balance results in low rates of bone turnover and favors resorption.

For some women, an explanation of the need to maintain caloric intake at a level in balance with energy expenditure encourages behavioral modifications (increased caloric intake and/or a decrease in exercise) that may restore menses.[407] However, most women with exercise-induced hypothalamic amenorrhea understandably are resistant to the suggestion that they gain weight or reduce or stop exercising. For many, exercise also is an important stress management strategy. It is helpful to explain the effects of diet, exercise, and hormones, and to emphasize the importance of building bone early in life; almost all of hip and vertebral bone mass accumulates by late adolescence (age 18) and the early postmenarcheal years (age 11–14) are especially important.[409, 410]

Bone loss is the most obvious, immediate, and demonstrable consequence of exercise-induced hypothalamic amenorrhea. ***A baseline bone density measurement revealing significant osteopenia, and the associated increased risk for debilitating stress fracture, can help to illustrate and to emphasize the need for a change in habits, or for hormone therapy.*** Subsequent periodic measurements of bone density are useful for assessing their effectiveness. The increase in bone mineral density that typically accompanies a return of normal menses is significantly greater than can be achieved by treatment with estrogen or oral contraceptives,[411–414] probably because normal bone metabolism requires both adequate nutrition (supporting bone formation) and estrogen (decreasing bone resorption). Nonetheless, in those unable or unwilling to make the lifestyle changes that might restore gonadal function, cyclic or combined continuous estrogen/progestin treatment or hormonal contraception is indicated, according to the patient's needs and preferences (regarding the importance of maintaining amenorrhea). Those choosing physiologic hormone therapy must be cautioned that treatment will not prevent ovulation and pregnancy if and when

normal function returns; menstrual bleeding at other than expected times suggests a return of function and that treatment may be discontinued for an interval of observation. For those who require reliable contraception, a low-dose oral contraceptive is the better choice. Women should be informed that hormone therapy may cause modest increases in weight and fat mass, but also reassured that such minor changes in body composition likely will have little or no impact on physical performance.[411]

Although supplemental calcium (1,000–5,000 mg daily) and vitamin D (1,000–2,000 IU daily) should be encouraged, bisphosphonates are not a good choice for the prevention and treatment of osteopenia in women with exercise-induced hypothalamic amenorrhea, for two reasons. First, they have low levels of bone formation and bone turnover and therefore respond poorly to anti-resorptive therapy.[413] Second, bisphosphonates remain in bone for 10 years or more and constantly leech into the circulation, most women with exercise-induced hypothalamic amenorrhea have not completed childbearing, and the effects of bisphosphonates on fetal skeletal development are unknown.

Women with exercise-induced hypothalamic amenorrhea should be reassured that treatment aimed at inducing ovulation and restoring normal fertility will be available, at the appropriate time. For those ready to pursue pregnancy, treatment with exogenous gonadotropins likely will be required to induce ovulation because clomiphene citrate typically is ineffective in women with hypogonadotropic hypogonadism. Logically, if low endogenous estrogen levels have failed to stimulate the appropriate compensatory increase in gonadotropin secretion, there is little reason to believe that treatment with an estrogen antagonist will do so. It may be useful to advise such women that whereas weight gain and a decrease in exercise may not restore spontaneous ovulatory function and fertility, they may nonetheless result in sufficient improvement in hypothalamic function to allow them to respond to clomiphene and to avoid the rigors and risks of gonadotropin treatment.

Congenital GnRH Deficiency

In rare individuals, hypothalamic amenorrhea results from a congential GnRH deficiency relating to specific genetic mutations that prevent normal GnRH neuronal migration during embryogenesis or to mutations in the pituitary GnRH receptor. Although seldom necessary, the diagnosis can be inferred by demonstrating a complete lack of pulsatile LH secretion and little or no LH secretion in response to exogenous pulsatile GnRH treatment.[415, 416] Congential GnRH deficiency is more common in males than in females (5:1). In affected women, the gonads respond normally to exogenous gonadotropin stimulation, which can be used to induce ovulation and restore fertility, as in women with more common causes of hypothalamic amenorrhea.

Kallmann's Syndrome

When congenital GnRH deficiency is associated with anosmia or hyposmia (an absent or grossly impaired sense of smell), the disorder is known as Kallmann's syndrome.[417] The classical X-linked form of the disorder is caused by a variety of genetic mutations in the *KAL* gene (located on the short arm of the X chromosome, Xp22.3) encoding anosmin-1, a neural adhesion molecule that promotes migration of GnRH neurons, and olfactory neurons, from the olfactory placode into the hypothalamus during embryonic development.[418] Obligate female carriers in families with the X-linked form of the disorder have no recognizable specific phenotype. Kallmann's syndrome also can be inherited in an autosomal dominant or recessive fashion. The autosomal dominant form has been linked to an inactivating mutation in the gene encoding the fibroblast growth factor-1 receptor (*FGFR1*). Mutations similar to but distinct from that in Kallmann's syndrome, affecting only GnRH neuronal migration, offer a potential explanation for an isolated gonadotropin deficiency, unaccompanied by anosmia.

At puberty, both males and females with Kallmann's syndrome usually present with delayed growth and sexual development. The presence of pubic hair, reflecting a normal adrenarche, helps to distinguish them from those with a constitutional delay of puberty in whom adrenarche typically also is delayed. However, the most distinguishing feature of Kallmann's syndrome is the inability to perceive odors, such as coffee or perfume. Patients with the disorder also may have a family history of delayed puberty and other abnormalities, including cleft lip/palate, urogenital tract anomalies, or syndactaly.

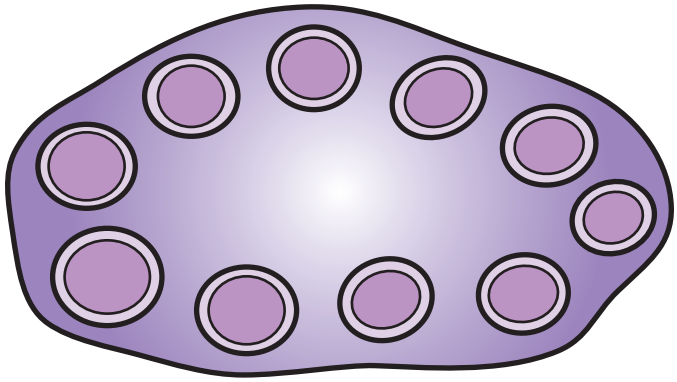
GnRH Receptor Mutations

More than 20 inactivating mutations in the GnRH receptor gene (*GNRHR*) have been described.[419, 420] Some effectively prevent GnRH binding, some interfere with normal signal transduction, both resulting in resistance to GnRH stimulation, and others predispose to abnormal folding at the site of synthesis within the endoplasmic reticulum, leading to degradation before transport to the cell surface membrane.[65]

The phenotypic spectrum observed in individuals with GnRH receptor mutations ranges from a complete absence of sexual development to delayed puberty, and generally correlates with the LH-secretory response to exogenous GnRH treatment. Stardard treatments for hypogonadotropic hypogonadism, including hormone therapy and ovulation induction with exogenous gonadotropins, are effective. Specific evaluation aimed at identifying GnRH mutations might be considered when other family members are affected, for research purposes, but is not clinically necessary or indicated, at least not yet. Interestingly, a nonpeptide cell permeant GnRH receptor antagonist (IN3) has shown the ability to restore ligand binding and function for several naturally occurring mutant GnRH receptors *in vitro*, apparently by acting as a molecular chaperone, helping to properly arrange the mutant receptor and to usher it to the cell surface membrane. The discovery suggests that such "pharmacochaperones" may someday have therapeutic applications in patients with GnRH receptor mutations.[65]

All references are available online at: http://www.clinicalgynendoandinfertility.com

12 Chronic Anovulation and the Polycystic Ovary Syndrome



Anovulation is very common and has a number of different clinical manfestations, including amenorrhea, dysfunctional uterine bleeding, and hirsutism. The condition also has serious potential consequences, such as infertility and an increased risk for developing endometrial hyperplasia and neoplasia. In many anovulatory women, the pathophysiology involves insulin resistance, which increases the risks for developing diabetes mellitus and cardiovascular disease. In others, overt hypogonadism increases the risk for developing early osteoporosis. All clinicians who care for women must therefore be thoroughly familiar with the evaluation and management of anovulatory women.

Normal ovulatory function requires coordination at all levels of the hypothalamic-pituitary-gonadal axis, and anovulation can result from disruption at any level. This chapter considers the variety of mechanisms that can cause anovulation and the clinical consequences of chronic anovulation, focusing on the most common anovulatory disorder, the polycystic ovary syndrome, and its management.

Causes of Anovulation

The complex interaction of neuroendocrine, intra-ovarian, and endometrial mechanisms that regulate the normal ovulatory menstrual cycle are discussed in detail in other chapters in this text (Chapters 5 and 6). They are summarized briefly here, to provide the foundation for subsequent discussion of the pathophysiology of anovulation.

As the corpus luteum regresses and the menstrual cycle draws to a close in the late luteal phase, serum concentrations of estradiol, progesterone, and inhibin A decline to basal levels, releasing the hypothalamic-pituitary axis from their collective negative

feedback effects. Accordingly, the frequency of hypothalamic gonadotropin-releasing hormone (GnRH) secretion increases, stimulating an increase in pituitary follicle-stimulating hormone (FSH) secretion, which serves to "recruit" a new cohort of small antral follicles or, more accurately, to rescue a group of follicles from otherwise programmed demise via apoptosis. During the early follicular phase, the serum concentration of inhibin B, secreted by the recruited pool of small antral follicles, rises progressively.

During the mid-follicular phase, ovarian autocrine and paracrine mechanisms involving activin and insulin-like growth factors enhance FSH-stimulated aromatase activity in granulosa cells to help create and sustain the estrogenic microenvironment required for continued follicular growth and development. Whereas an estrogenic follicular milieu fosters further growth, an androgenic milieu promotes atresia. As the serum inhibin B concentration reaches its peak, estradiol and inhibin A levels, derived from the granulosa cells of growing follicles in the cohort, begin increasing steadily. In response to their combined inhibitory effects, luteinizing hormone (LH) pulse amplitude decreases and pulse frequency increases (presumably reflecting the pattern of hypothalamic GnRH secretion), and serum concentrations of both FSH and LH fall gradually. Declining FSH levels remain sufficient to support continued growth of the selected dominant follicle, which has more granulosa cells and FSH receptors and a more advanced microvasculature, but become inadequate to support further development in smaller follicles in the cohort.

During the late follicular phase, inhibin A and insulin-like growth factors combine to promote LH-stimulated androgen production in theca cells, which provides substrate for aromatization to estrogen in the proliferating mass of granulosa cells within the preovulatory follicle. FSH and estradiol then combine to induce expression of LH receptors on the granulosa cells that will mediate luteinization and ovulation when the follicle reaches full maturity. UItimately, serum levels of estradiol derived from the preovulatory follicle exceed the threshold concentration required to exert positive feedback effects centrally, acting primarily on the pituitary to induce the midcycle LH surge. The LH surge completes follicular maturation and triggers a cascade of events resulting in extrusion of the oocyte and formation of the corpus luteum. The oocyte completes the first meiotic division and local secretion of plasminogen activator and other cytokines mediates erosion of the follicular wall, allowing the oocyte to emerge with its surrounding investment of cumulus cells. The mural granulosa cells begin to luteinize and produce progesterone.

After ovulation, serum estradiol concentrations fall precipitously, but only transiently, before rising again in parallel with progesterone and inhibin A produced by the corpus luteum. Progesterone transforms the endometrium from a proliferative to a secretory morphology and stimulates a still uncharacterized cascade of biochemical events that renders the endometrium receptive to embryo implantation. As the progesterone level rises to its peak during the mid-luteal phase, LH pulse frequency decreases again and gonadotropin levels fall progressively to their nadir in the late luteal phase. Unless pregnancy intervenes and rapidly rising levels of human chorionic gonadotropin (hCG) rescue the corpus luteum and stimulate continued high level progesterone secretion, the corpus luteum regresses, estradiol and progesterone levels fall, support for the endometrium is withdrawn, and menses ensue.

Central Defects

Although difficult to demonstrate, hypothalamic dysfunction offers both a logical and likely explanation for ovulatory failure. A normal pituitary response to feedback signals from the follicle requires pulsatile GnRH secretion within a critical range. The onset of

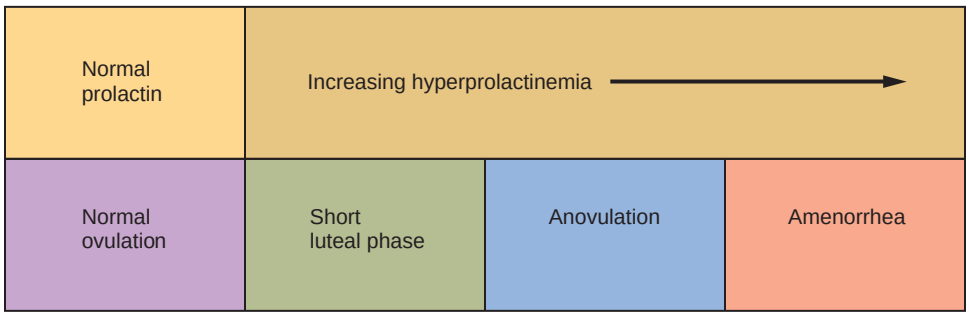
puberty in girls results from decreasing central inhibition of GnRH neuronal activity and increasing pulsatile GnRH secretion, which stimulates a progressive increase in pituitary gonadotropin release and, in turn, ovarian follicular growth and estrogen production (Chapter 10). After menarche, cycle length and menstrual characteristics in adolescent girls typically vary until the hypothalamic-pituitary-ovarian axis matures and the positive feedback relationship between estradiol and gonadotropin secretion and ovulation becomes established. ***Factors that reactivate central inhibitory mechanisms, such as emotional, nutritional (weight loss, eating disorders), or physical stress (excessive exercise), can suppress GnRH neuronal activity, leading to dysfunctional patterns of gonadotropin secretion that fail to promote progressive follicular development, resulting in anovulation.*** Although such patients more commonly present with amenorrhea (Chapter 11), lesser degrees of GnRH neuronal suppression can result in homeostatic levels of pituitary-ovarian function and a euestrogenic chronic anovulatory state.

Pituitary Tumors

Pituitary tumors can cause anovulation by inhibiting gonadotropin secretion. They may compress pituitary gonadotrophs directly, or interrupt delivery of hypothalamic GnRH by compression of the pituitary stalk. Alternatively, they may cause hyperprolactinemia by interfering with the inhibitory actions of hypothalamic dopamine (the putative prolactin inhibitory hormone) on pituitary lactotrophs, resulting in a secondary suppression of pulsatile GnRH secretion.

Hyperprolactinemia

Hyperprolactinemia is another specific example of anovulation resulting from a central defect.[1] The mechanism involves disruption or inhibition of the normal GnRH pulse rhythm, resulting in ineffective or frankly low levels of gonadotropin secretion. It's possible that elevated prolactin levels stimulate a generalized increase in hypothalamic dopaminergic activity, intended to suppress prolactin secretion but also inhibiting GnRH neurons. In any event, increasing prolactin levels can result in a spectrum of ovulatory dysfunction, ranging from a short luteal phase to anovulatory cycles to amenorrhea and hypogonadotropic hypogonadism, depending on the extent to which gonadotropin secretion is disturbed or suppressed. Mild hyperprolactinemia may cause only a short luteal phase, resulting from inadequate preovulatory follicular development.[2, 3] Moderate hyperprolactinemia frequently causes oligomenorrhea or amenorrhea, and higher prolactin levels typically result in frank hypogonadism with low estrogen levels.[4, 5] ***A breast examination with gentle compression looking for evidence of galactorrhea and measurement of the serum prolactin concentration are important parts of the evaluation of all anovulatory women.***



Abnormal Gonadotropin Secretory Dynamics

Many, but not all anovulatory women with polycystic ovaries exhibit abnormal gonadotropin secretory dynamics. The most common abnormality is an increase in mean serum LH levels, due to an increase in both LH pulse frequency and amplitude.[6, 7] Serum concentrations of FSH typically are normal or low. The pattern could result from a decrease in hypothalamic dopamine or opioid inhibition of pulsatile GnRH secretion,[8] or from abnormalities in steroid hormone feedback, including the lack of progesterone (due to anovulation)[9] or increased circulating androgen levels.[10] Other evidence from studies in nonhuman primates and women suggests strongly that prenatal exposure to increased androgen concentrations induced by genetic and/or environmental factors may program the GnRH pulse generator in the female fetus in such a way as to result in increased pituitary LH secretion, causing disordered follicular development and ovarian hyperandrogenism.[11–13] The increased prevalence of chronic anovulation and polycystic ovaries in women with epilepsy offers another example of how central nervous system dysfunction can disrupt the hypothalamic-pituitary-ovarian axis and result in anovulation.[14, 15]

Abnormal Feedback Signals

Anovulation can result from abnormal estrogen feedback signals from the periphery, in two ways. Chronically elevated estrogen levels may not permit the increase in FSH secretion required to stimulate or sustain progressive follicular development. Conversely, poor follicular development may not generate or sustain the estradiol level required to induce the ovulatory LH surge.

Chronically Elevated Estrogen Concentrations

The fall in estradiol levels that normally occurs during the late luteal phase (as the corpus luteum regresses) is a prerequisite for the inter-cycle rise in FSH that drives the wave of new follicular development. Sustained high levels of estrogen negative feedback caused by increased production or decreased clearance and metabolism can prevent any significant increase in FSH levels, resulting in a chronic anovulatory state.

Pregnancy is the most common and obvious example of anovulation resulting from sustained high levels of estrogen production. Rare estrogen producing ovarian tumors (e.g., granulosa cell tumors) can have the same effect. Although the adrenals do not normally secrete appreciable amounts of estrogen directly into the circulation, they contribute via their secretion of androgens (androstenedione, dehydroepiandrosterone and its sulfate), which can be converted to estrogen in the periphery. Adipose tissue has significant aromatase activity, which converts androgens to estrogens,[16] thereby providing at least one mechanism for the well-known association between obesity and chronic anovulation (see below).

The clearance and metabolism of estrogen can be impaired in a variety of conditions, such as thyroid or hepatic disease. Both hyperthyroidism and hypothyroidism can cause chronic anovulation by altering the metabolic clearance and peripheral inter-conversion of steroid hormones.[17–19] ***Hypothyroidism can be associated with elevated prolactin levels, providing the rationale for measuring serum thyroid-stimulating hormone (TSH), as well as prolactin, in the evaluation of anovulatory and amenorrheic women.*** Hepatic disease also disturbs the normal clearance and metabolism of sex steroids.[20]

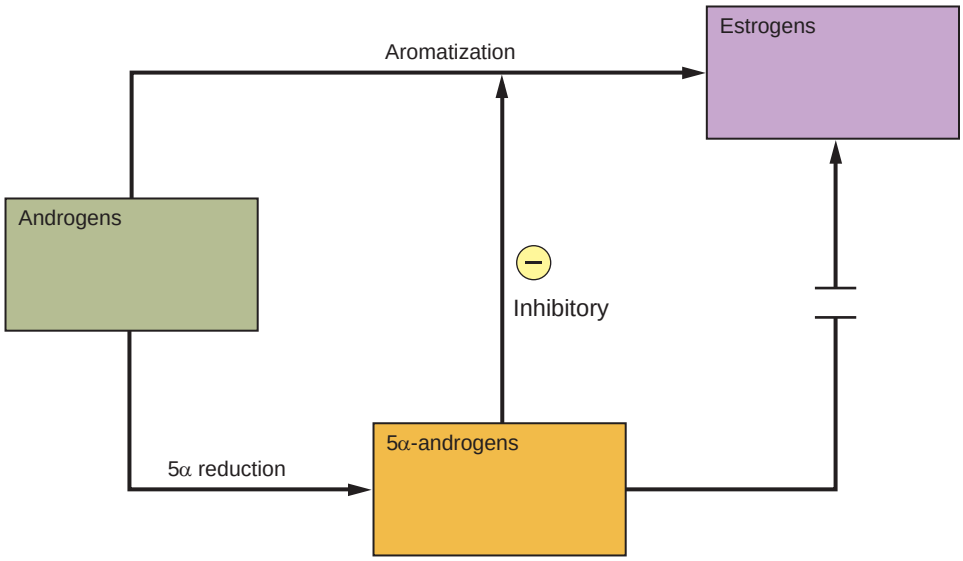
Failure of the LH Surge

The rising tide of estradiol arising from the preovulatory follicle in the late follicular phase induces the midcycle LH surge that stimulates ovulation. Quite obviously, women with gonadal dysgenesis or ovarian failure are anovulatory because they have no remaining functional ovarian follicles and no significant estrogen production. More commonly, clinicians encounter patients with normal serum levels of gonadotropins and estradiol who do not ovulate, whose anovulation results from the failure to achieve complete follicular development and to generate and sustain the level of estradiol required to induce the LH surge. Normal women typically also become anovulatory during the years immediately preceding the menopause, probably reflecting intrinsic deficiencies in aging follicles that impair normal follicular maturation.

Local Ovarian Conditions

A disturbance in one or more of the delicately balanced intra-ovarian regulatory mechanisms that serve to select the dominant follicle and allow it alone to grow and develop in the face of declining levels of FSH may lead to anovulation. Activins, inhibins, and insulin-like growth factors act via local autocrine and paracrine mechanisms to first enhance the action of FSH by increasing the concentration of FSH receptors within the dominant follicle, then combine to induce the appearance of LH receptors required to mediate the actions of LH during the midcycle surge that drives the final stages of follicular maturation and stimulates ovulation. A follicle can thus fail to grow and ovulate due to a failure or interference with any of these local mechanisms (Chapter 6).

The "two-cell, two gonadotropin" concept of ovarian follicular development (Chapters 2 and 6) emphasizes the critical importance of local androgen concentrations. At low levels, androgens serve as substrate for FSH-induced aromatization and estrogen production. At higher concentrations, androgens are converted alternatively to more potent 5α-reduced androgens, which cannot be converted to estrogen and also inhibit aromatase activity and FSH induction of LH receptors on granulosa cells. ***Consequently, abnormally high local androgen concentrations, from any cause, impede follicular maturation, promote atresia, and predispose to a chronic anovulatory state.***



Obesity

The prevalence of obesity in women with chronic anovulation and polycystic ovaries is high, ranging between 35% and 60%.[21–24] Obesity predisposes to chronic anovulation in at least three distinct ways:

1. Increased peripheral aromatization of androgens, resulting in chronically elevated estrogen concentrations.

2. Decreased levels of hepatic SHBG production, resulting in increased circulating concentrations of free estradiol and testosterone.

3. Insulin resistance, leading to a compensatory increase in insulin levels that stimulates androgen production in the ovarian stroma, resulting in high local androgen concentrations that impair follicular development.

Combined, these effects can be difficult to overcome, but even modest weight loss, which results in decreased circulating insulin and androgen concentrations, frequently restores ovulatory function and normal menstrual cyclicity.[25–28]

Defining the Cause of Anovulation

Whereas the cause of anovulation may be relatively clear in women with ovarian failure, pituitary tumors, eating disorders, hyperprolactinemia, or obesity, frequently it is not possible to isolate the specific mechanism responsible. However, it also is often not necessary. Regardless of its cause, the clinical manifestations and consequences are predictable, easily documented, and generally not difficult to manage. Women with absent or abnormal menstrual function who are otherwise healthy can be categorized as follows:

1. Ovarian failure. Hypergonadotropic hypogonadism, reflecting the inability of the ovary to respond to gonadotropin stimulation, due to follicular depletion (Chapter 11).

2. Central defects. Hypogonadotropic hypogonadism, reflecting hypothalamic or pituitary failure or suppression (Chapter 11).

3. Hypothalamic-pituitary-ovarian dysfunction, resulting in asynchronous gonadotropin and estrogen production, having a wide variety of causes and clinical manifestations that depend on the level of ovarian function, including amenorrhea (Chapter 11), hirsutism (Chapter 13), dysfunctional uterine bleeding (Chapter 15), endometrial hyperplasia and cancer (Chapter 18), and infertility (Chapters 27 and 31).

The polycystic ovary syndrome (PCOS) is the most obvious and common condition associated with chronic anovulation, affecting 4–6% of reproductive age women.[29, 30] Several mechanisms contribute to the pathophysiology of anovulation in PCOS, operating at every level of the reproductive system. ***It is inaccurate to state that PCOS is the most common "cause" of anovulation, because PCOS does not cause anovulation; rather, PCOS is the consequence of chronic anovulation, which can result from a wide variety of causes. In that context, the disorder is described more accurately as chronic anovulation with***

polycystic ovaries. Although the term PCOS is now firmly established in our scientific and clinical lexicon, it is important to emphasize that PCOS is not a discrete or specific endocrine disorder having a unique cause or pathophysiology. Instead, the condition is best viewed as a final common pathway in the chronic anovulatory state.

The Polycystic Ovary Syndrome

Multicystic or "sclerocystic" ovaries were recognized as early as the mid-18th century, but associated primarily with pelvic pain or menorrhagia. In the early 20th century, prevailing hypotheses viewed them as resulting from inflammation due to infection, congestion due to pressure or partial torsion that disrupted normal blood flow to the ovary, or from dystrophy due to abnormalities in ovarian nutrition.[31]

In 1935, Irving F. Stein and Michael L. Leventhal first described a symptom complex associated with anovulation.[32] Both gynecologists were born in Chicago, both were graduates of Rush Medical College, and both spent their entire professional careers at the Michael Reese Hospital.[33] Stein and Leventhal described seven patients (four being obese) with amenorrhea, hirsutism, and enlarged, polycystic ovaries. They reported that all seven resumed regular menses and that two became pregnant after bilateral ovarian wedge resection, involving the removal of one-half to three-fourths of each ovary. Stein and Leventhal developed the wedge resection procedure after observing a resumption of menses following ovarian biopsy in several patients with amenorrhea. They speculated that the thickened ovarian capsule prevented follicles from reaching and escaping from the surface of the ovary.

Careful histologic studies of the "Stein-Leventhal ovary" revealed that they had twice the cross-sectional area of normal ovaries, the same number of primordial follicles, double the number of developing and atretic follicles, a 50% thicker and more collagenized tunica, a 5-fold thicker subcortical stroma, and a 4-fold greater number of hilar cell "nests" than normal ovaries. These studies further suggested that "hyperthecosis," characterized by an abundance of such nests and a markedly increased stroma, was likely just a later or more advanced stage of a progressive process.[34]

The pathophysiology responsible for development of polycystic ovaries has puzzled gynecologists and endocrinologists for many years and proven very difficult to define. However, there is an answer that is very simple, logical, and clinically useful. ***The characteristic polycystic ovary develops when a chronic anovulatory state persists for a sufficient length of time.*** A cross section of anovulatory women at any one point in time will demonstrate that approximately 75% have multicystic or polycystic ovaries.[24, 35] ***Because there are many causes of anovulation, there are many causes of polycystic ovaries.*** Any of the causes of anovulation outlined earlier can yield the same or a similar clinical presentation. ***The polycystic ovary results from a functional derangement, not from a specific central or local defect.***

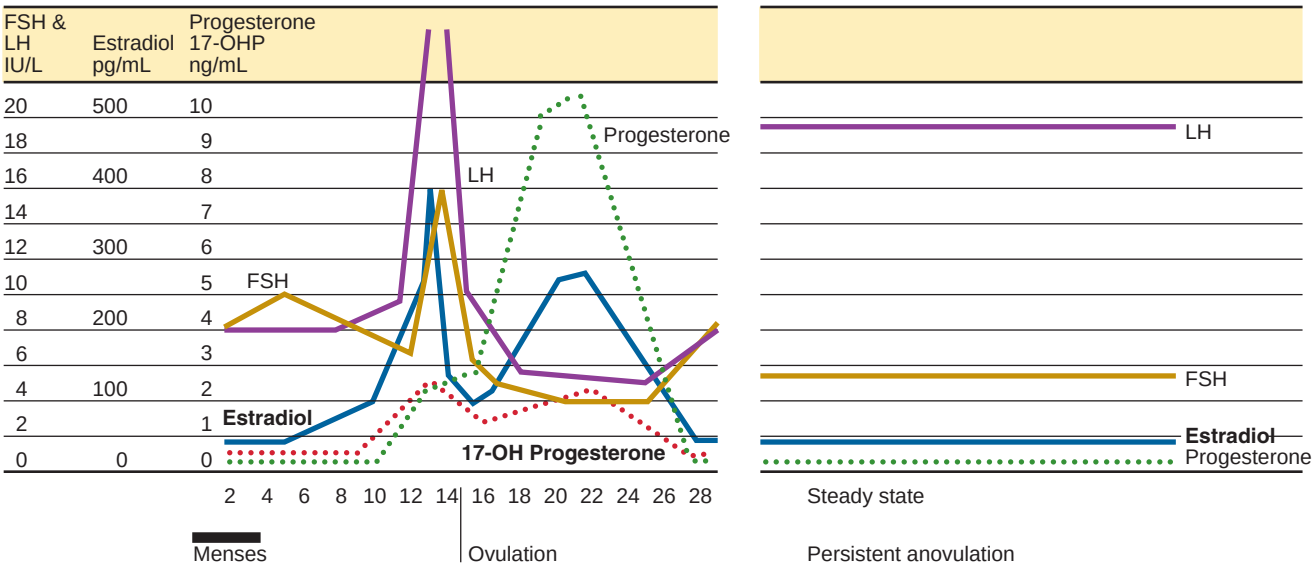
Pathophysiology

Whereas the morphological characteristics of polycystic ovaries were attributed at first to pathological changes in the ovaries themselves that prevented ovulation,[34] they now are recognized as reflecting the disordered endocrine milieu that results from chronic

anovulation. In contrast to the cyclic pattern of hormone concentrations that occurs during the normal cycle, the endocrine milieu in women with chronic anovulation is characterized by a "steady state" in which gonadotropin and sex steroid concentrations vary relatively little, by comparison.

The average daily production of both androgens and estrogens is increased in women with PCOS, as reflected by elevated serum concentrations of testosterone, androstenedione, dehydroepiandrosterone (DHEA), dehydroepiandrosterone sulfate (DHEA-S), 17α-hydroxyprogesterone (17-OHP), and estrone. The results of treatment with a long-acting GnRH agonist (aimed at suppressing gonadotropin-dependent ovarian steroid production) indicate that the increases in serum testosterone, androstenedione, and 17-OHP derive from the ovary and are LH-dependent, whereas those in DHEA and DHEA-S derive from the adrenal.[36–39] Serum estrone concentrations are modestly elevated, due to peripheral conversion of increased amounts of androstenedione. In contrast, serum estradiol levels in women with PCOS fluctuate but generally remain within the range typically observed in the early follicular phase,[40] reflecting continued low-level production from limited follicular development.[41, 42]

The endocrine milieu in women with PCOS reflects the chronic anovulatory state, which may result from a wide variety of causes. ***Current perspectives view PCOS as a complex disorder, similar to cardiovascular disease and type 2 diabetes mellitus, wherein numerous genetic variants and environmental factors interact, combine, and contribute to the pathophysiology.***[43] Not surprisingly, attention has focused on identifying genetic variants involving the regulation of gonadotropin secretion and action, insulin secretion and action, weight and energy regulation, and androgen synthesis and action.



Gonadotropin Secretion and Action

Stein and Leventhal suggested that polycystic ovaries were likely to result from abnormal anterior pituitary hormonal stimulation,[32] based on earlier observations that treatment with a urinary extract of anterior pituitary hormones could induce changes similar to those in polycystic ovaries.[44] Subsequent studies employing an LH bioassay (based on the ovarian

response of immature female rats or the prostatic response of hypophesectomized male rats to urinary extracts) demonstrated excessive LH activity in women with PCOS,[45–48] which later was confirmed by studies using a radioimmunoassay.[49]

Compared to normally cycling women, those with PCOS generally exhibit increased serum LH concentrations, low-normal FSH levels, and increased LH:FSH ratios.[7, 50, 51] The increase in serum LH levels results from abnormal LH secretory dynamics, characterized by an increase in LH pulse frequency, and to a lesser extent, also in pulse amplitude.[6, 52–54] The decrease in FSH levels results from the increase in GnRH pulse frequency, the negative feedback effects of chronically elevated estrone concentrations (derived from peripheral aromatization of increased androstenedione), and normal or modestly increased levels of inhibin B (derived from small follicles).[55, 56]

LH pulse frequency in women with PCOS does not exhibit the normal cyclic variation seen in ovulatory women and is relatively constant, at approximately one pulse per hour. The pattern presumably reflects a similar increase in hypothalamic GnRH pulse frequency, which favors secretion of LH more than FSH.[57–59] The LH response to an acute exogenous GnRH stimulus also is exaggerated in women with PCOS, but to a lesser extent in obese than in lean women; accordingly, LH pulse amplitude and serum LH levels generally are somewhat lower in obese than in lean women with PCOS.[7, 60] Elevated serum LH concentrations in women with PCOS also exhibit increased bioactivity in bioassay systems *in vitro*, reflecting a difference in glycosylation with a predominance of more basic (alkaline) LH isofoms, which have greater bioactivity.[53, 61–63]

The approximate hourly LH pulse frequency in women with PCOS is within the range of frequencies usually observed across the normal ovulatory cycle, suggesting it results from a failure of the mechanisms that normally slow the GnRH pulse generator rather than from an abnormal acceleration in pulse frequency. The increased pulse frequency might reflect intrinsic hypothalamic dysfunction, the effects of abnormal feedback signals from the periphery, or both.[64]

Since dopamine and opioids normally inhibit hypothalamic GnRH neuronal activity, the higher GnRH pulse frequency observed in women with PCOS could be caused by a decrease in dopaminergic or opioidergic neuronal stimulation. However, experimental evidence from studies involving treatment with medications that stimulate or inhibit these pathways does not support either mechanism. Treatment with a dopamine agonist has no discernible effect on the pattern of gonadotropin secretion in women with PCOS.[65, 66] Treatment with a progestin slows LH pulse frequency,[9] just as progesterone does during the normal luteal phase, indicating that the opioid-dependent process that normally mediates the effects of progesterone is operating,[67–69] and suggesting that any decrease in opioid tone results primarily from the lack of progesterone feedback, due to anovulation.

Infusion of exogenous insulin[70–72] and treatments that decrease insulin levels (metformin, thiazoladinediones) have no significant effect on the pattern of LH secretion in women with PCOS.[72, 73] LH levels also are lower in obese than in lean women with PCOS, even though insulin levels are higher in the obese.[7, 60] These observations suggest that hyperinsulinemia has no significant direct effect on LH secretion.

Treatment with exogenous estrone does not increase basal or GnRH-stimulated LH concentrations in women with PCOS,[74] and treatment with an aromatase inhibitor does not decrease LH pulse frequency,[75] indicating that increased circulating levels of estrone may exert negative feedback effects on FSH, but probably do not have any important direct influence on LH secretion in women with PCOS. Whereas the lack of progesterone feedback resulting from anovulation undoubtedly contributes to the higher LH pulse frequency,[9] evidence suggests that the GnRH pulse generator also is less sensitive to the feedback inhibition of sex steroids.

Treatment with an estrogen-progestin contraceptive or with physiologic doses of exogenous estrogen and progesterone slows LH pulse frequency in women with PCOS, but to a lesser extent than in normal women.[76–78] However, after pretreatment with flutamide (an androgen receptor antagonist), the effects of estrogen and progesterone on LH pulse frequency in women with PCOS are the same as in normal women,[10] suggesting that increased circulating androgen levels help to sustain the higher LH pulse frequency observed in women with PCOS by decreasing sensitivity to estrogen and progestin feedback.

Androgens also may contribute more directly to the abnormal pattern of gonadotropin secretion in women with PCOS. Evidence from studies in rats, sheep, monkeys, and women indicates that prenatal exposure to increased androgen concentrations may affect GnRH pulse generator programming, predisposing to an increased pulse frequency and LH secretion.[12, 13, 79–82] At least in rodents, prenatal androgen treatment also decreases basal and estrogen-induced hypothalamic progesterone receptor concentrations,[81] offering a mechanism to explain how androgens might decrease hypothalamic sensitivity to progesterone feedback. It could be that hyperandrogenemia from any cause, arising during fetal life (maternal hyperandrogenism, classical congenital adrenal hyperplasia), adolescence (premature adrenarche, nonclassical congenital adrenal hyperplasia), or in adulthood (obesity, hyperinsulinemia) induces abnormalities in the feedback control of pulsatile GnRH secretion, resulting in increased LH secretion, which stimulates increased ovarian androgen production, in a self-perpetuating cycle.

The primary evidence indicating that excessive LH stimulation plays an important role in the pathophysiology of PCOS comes from studies examining the effects of treatment with GnRH antagonists and long-acting GnRH agonists. In women with PCOS, treatment with a GnRH antagonist induces an acute dose-dependent decrease in both LH and testosterone concentrations,[54] and long-term treatment with an agonist can suppress ovarian androgen production to postmenopausal levels.[83, 84] However, normally cycling women with polycystic ovaries exhibit higher androgen and insulin levels and lower SHBG concentrations than women with normal ovarian morphology, even though LH levels and secretory dynamics are not different.[85] ***These observations suggest that excessive LH secretion or stimulation may be an important cause of disordered follicular development and anovulation, but is not the proximate cause of polycystic ovaries or of increased ovarian androgen production in women with PCOS.***

Insulin Secretion and Action

An association between glucose intolerance and hyperandrogenism was first recognized by Archard and Thiers in 1921, in a famous report describing a bearded diabetic woman.[86] Insulin resistance was first described in diabetic patients who required progressively higher doses of insulin to maintain effective glucose control, most often because they developed antibodies to preparations of insulin derived from animal sources.[87] Today, we recognize insulin resistance as a feature of a wide variety of disorders and conditions, ranging from extreme insulin-resistance syndromes (auto-antibodies to the insulin receptor, insulin receptor mutations, lipodystrophic states)[88–90] to common problems such as type 2 diabetes, obesity, stress, infection, pregnancy, and PCOS. The importance of insulin resistance, hyperinsulinemia, and insulin action in the pathogenesis of PCOS was first suggested by a study conducted in 1980, demonstrating significant correlations between basal levels of plasma insulin, androstenedione, and testosterone, and between insulin and testosterone levels after an oral glucose load.[91]

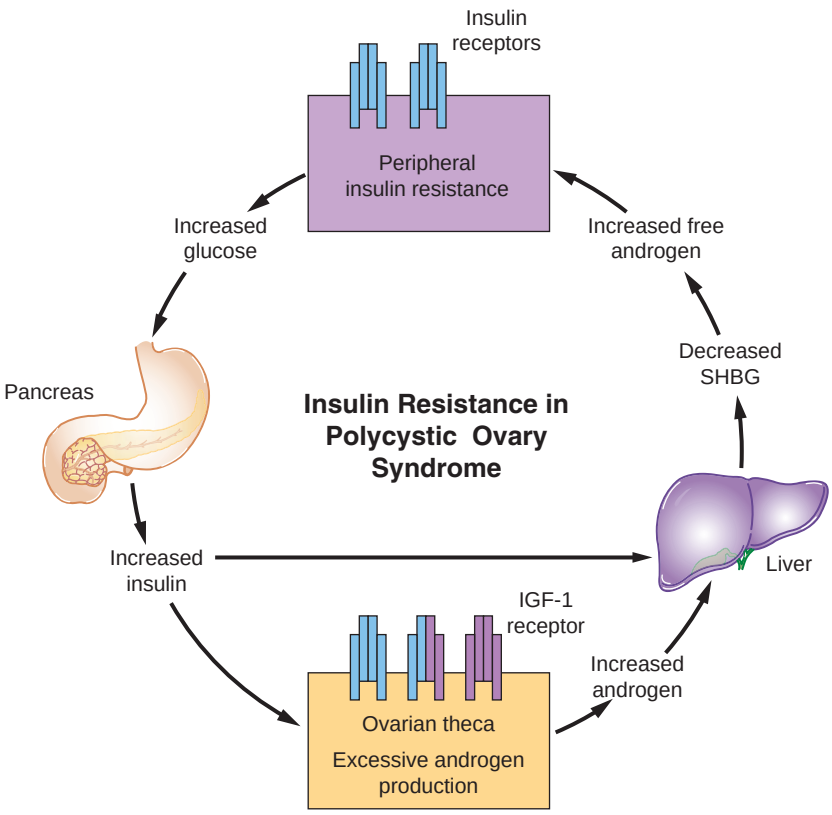
Insulin resistance is a common feature in obese and, to a lesser extent, lean women with PCOS; the overall prevalence ranges between 50% and 75%.[92–95] Insulin sensitivity is decreased by an average of 35–40% in women with PCOS, compared to normal

women, similar to what is observed among women with non-insulin dependent diabetes mellitus.[96–98] ***Up to 35% of women with PCOS exhbit impaired glucose tolerance and 7–10% meet criteria for type 2 diabetes mellitus.***[99, 100] Conversely, women with type 2 diabetes are 6-fold more likely than non-diabetic women of similar age and weight to have PCOS.[101]

Insulin resistance is a condition in which endogenous or exogenously administered insulin has less than normal effects on fat, muscle, and the liver.[102] In adipose, insulin resistance results in increased hydrolysis of stored triglycerides and elevated circulating free fatty acid levels. Decreased glucose utilization (primarily in muscle) and increased hepatic gluconeogenesis (which insulin normally inhibits) result in increased blood glucose concentrations and a compensatory hyperinsulinemia (in those with adequate pancreatic reserve). ***Increased circulating insulin levels cause or contribute to hyperandrogenism in women with PCOS in at least two important ways, by stimulating increased ovarian androgen production, and by inhibiting hepatic SHBG production.***

Numerous studies have demonstrated that insulin stimulates androgen production in ovarian theca cells *in vitro*.[103] Theca cells from women with PCOS also exhibit increased sensitivity to insulin, compared to those from normal women. Physiologic levels of insulin can stimulate androgen synthesis in theca cells of women with PCOS, whereas higher insulin concentrations are required in normal theca cells.[104, 105] ***Because insulin also potentiates the action of LH,***[106] ***insulin and LH act synergistically to stimulate androgen production.***[104, 107]

Clinical investigations in women with PCOS have demonstrated that insulin also stimulates ovarian androgen production *in vivo*. Notably, the cumulative sum insulin response during an oral glucose tolerance test correlates positively with the rise in serum androstenedione and testosterone above baseline concentrations.[108] Moreover, suppression of serum insulin



levels by treatment with diazoxide or an insulin-sensitizing agent (troglitazone) decreases serum androstenedione and testosterone levels in women with PCOS.[109, 110]

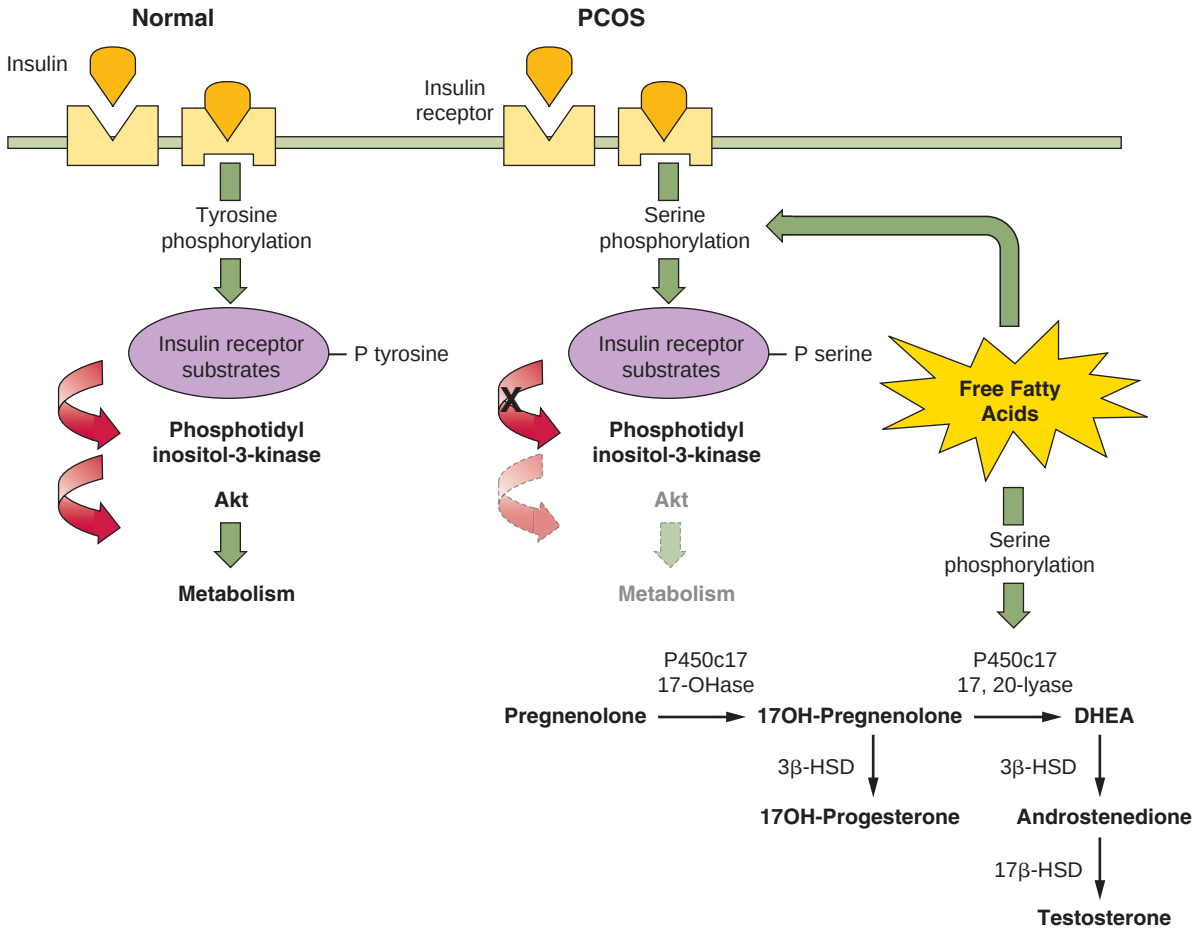
High insulin concentrations also inhibit hepatic SHBG production,[111, 112] as do high androgen concentrations. ***The combined actions of insulin and androgens lower SHGB concentrations, yielding increased free androgen levels, which aggravate the underlying insulin resistance.[113] Ultimately, these conditions foster a self-propagating positive feedback loop that can increase in severity over time.***

Insulin stimulates ovarian androgen production acting via insulin receptors on theca/interstitial cells in the ovarian stroma.[105, 114] At high concentrations, insulin also binds to IGF-1 receptors (and possibly hybrid receptors) which are structurally similar and use a similar signaling mechanism.[96] However, evidence indicates that insulin acts primarily via its own receptor, by activating a signaling system separate from that involved in glucose transport. Whereas an anti-insulin receptor antibody effectively inhibits insulin-stimulated steroidogenesis in cultured human granulosa cells, an anti-IGF receptor antibody has no effect.[105, 115, 116] A d-chiro-inositol containing glycan increases theca cell testosterone production *in vitro*, and preincubation with an anti-inositolglycan antibody blocks insulin stimulation, but not that of hCG.[105] These observations suggest inositolphosphoglycan mediators act as second messengers in signal transduction for insulin stimulation of theca cell androgen synthesis and that the mechanism differs from that mediating the actions of LH.[117]

What causes insulin resistance in women with PCOS is not entirely clear. Not surprisingly, given the complexity and polygeneic nature of the disorder, evidence suggests that more than one mechanism may be involved.

The classical actions of insulin are mediated via its receptor and two distinct intracellular pathways. The phosphatidyl-inositol 3-kinase (PI-3K) pathway mediates the metabolic effects of insulin, and the mitogen-activated protein kinase (MAPK) pathway mediates the proliferative actions of insulin. Normally, insulin binding to its receptor induces a conformational change, resulting in tyrosine phosphorylation of the receptor and protein substrates, which bind and serially activate PI-3K and Akt, an effector molecule that plays the major role in signal transduction for glucose regulation and metabolism.[118, 119] Akt activation potentiates the translocation of glucose transporter 4 (GLUT4) from intracellular compartments to the plasma membrane, thereby increasing glucose uptake. Other effector molecules mediate insulin inhibition of gluconeogenesis and glycogenolysis,[120, 121] stimulation of lipid synthesis, and inhibition of lipid catabolism.[122, 123] Luteinized granulosa cells obtained from women with PCOS display both a selective increase in insulin activation of the mitogenic pathway, via MAPK, and resistance in the PI-3K-mediated metabolic pathway of insulin action.[124] ***These and similar observations illustrate how insulin actions can be selectively inhibited and enhanced at the same time, via different signaling pathways,[96, 124] explaining how insulin can stimulate hyperandrogenism in women who are "insulin-resistant."***

Studies in cultured skin fibroblasts, muscle, and adipocytes from women with PCOS indicate that insulin resistance results from defects early in the post-receptor signaling pathway.[98, 125–127] The number and affinity of insulin receptors in both obese and lean women with PCOS are not decreased,[128, 129] but insulin receptors exhibit a constitutive increase in phosphorylation of serine residues and a decrease in insulin-stimulated phosphorylation of tyrosine residues. Serine phosphorylation of insulin receptor substrates prevents their binding with PI-3K and thereby inhibits insulin signaling. Increased serine phosphorylation can be induced by intracellular metabolites of free fatty acids,[126, 130] which are increased in most women with PCOS and have been demonstrated to cause insulin resistance *in vivo*.[131, 132] High circulating free fatty acid levels also can increase androgen production in women,[133] by inducing serine phosphorylation of P450c17, which results in increased 17,20 lyase



Adapted from Baptiste.[118]

activity.[134, 135] These observations offer a mechanism for insulin resistance that also further helps to explain the link between insulin and hyperandrogenism in women with PCOS.

Although hyperandrogenism can decrease insulin sensitivity, the effect is relatively modest.[113] ***Insulin resistance and hyperinsulinemia are the primary factors; they are the cause, not the result, of hyperandrogenism.*** Treatment with a GnRH agonist can normalize elevated serum androstenedione and testosterone levels in women with PCOS, but has limited or no effect on insulin resistance.[37, 136–138] Similarly, although bilateral ovarian cautery can decrease serum androgen concentrations by nearly 50% in women with PCOS, glucose utilization (insulin sensitivity) remains unchanged.[139]

Accumulating evidence suggests that deficiency or dysfunction in downstream signaling mediated by inositolphosphoglycans also may contribute to insulin resistance in women with PCOS.[140–143] Finally, obesity is a common feature of women with PCOS, representing yet another important mechanism contributing to the development of insulin resistance, as discussed below.

Insulin resistance and hyperinsulinemia are undoubtedly an important part of the pathophysiology of PCOS. However, it is important to emphasize that 25–50% of women with PCOS have no demonstrable insulin resistance. Moreover, among all women with insulin resistance, the prevalence of PCOS is relatively low (approximately 15%).[144] Therefore, insulin resistance and hyperinsulinemia are not the primary cause or pathogenic factor in all women with PCOS.

Weight and Energy Regulation

The risk for developing PCOS rises with increasing obesity,[144–146] as does the severity of insulin resistance, hyperinsulinemia, and ovulatory dysfunction, and the prevalence of metabolic syndrome, glucose intolerance, risk factors for cardiovascular disease, and sleep apnea.[100, 147–150]

Obesity, by itself, is associated with insulin resistance and compensatory hyperinsulinemia. Insulin resistance is most highly correlated with intra-abdominal obesity, because visceral fat is more active metabolically than subcutaneous fat, more sensitive to lipolysis, releases more free fatty acids, and produces a number of cytokines involved in insulin resistance, such as tumor necrosis factor-α (TNF-α), interleukin-6, leptin, and resistin.[151] The accumulation of free fatty acids in tissues causes lipotoxicity and insulin resistance, in part via TNF-α, which increases serine phosphorylation and thereby inhibits insulin signaling.[152] Insulin resistance due to obesity also induces leptin resistance and decreases adiponectin levels, thereby decreasing fatty acid oxidation and promoting lipotoxicity.[151, 153] Obesity in women with PCOS typically is distributed centrally, with a greater increase in visceral than in subcutaneous fat.[154–157] However, even lean women with PCOS have an increased percentage of body fat, a higher waist-hip ratio, and greater intra-abdominal, peritoneal and visceral fat, compared to normal women matched for body mass index (BMI).

The overall prevalence of obesity, and in women with PCOS, varies among different patient populations;[158] in the United States, approximately 35% of all adult women and 60% of women with PCOS are obese.[159, 160] However, the overall prevalence of PCOS among different populations is quite similar (approximately 7%).[30, 161–163] Moreover, the prevalence of PCOS among unselected women varies relatively little with increasing BMI: 8.2% in underweight women (BMI < 18.5), 9.8% in normal-weight women, 9.9% in overweight women (BMI 25.0–30.0), 9.0% in obese women (BMI ≥ 30.0), 12.4% in those with a BMI between 35.0 and 40.0, and 11.5% in morbidly obese women (BMI > 40.0).[146] ***Combined, these observations indicate that obesity relates primarily to genetic and environmental factors and is a common, but not essential, feature of PCOS. Obesity contributes modestly to the risk for developing PCOS and adds to the pathophysiology in already affected women by aggravating the degree of insulin resistance and hyperinsulinemia.***[164, 165] It also is possible that PCOS itself may, to some degree, predispose to weight gain and obesity.

The prevalence of menstrual irregularity, dysfunctional bleeding, hirsutism, and infertility is higher in obese than in lean women with PCOS,[165–167] as is the risk for developing glucose intolerance and diabetes.[100, 168] Moreover, obese women have a higher prevalence of miscarriage, gestational diabetes, and pre-eclampsia, regardless whether they also have PCOS.[169]

Androgen Synthesis and Action

Hyperandrogenism is the key feature of PCOS, resulting primarily from excess androgen production in the ovaries and, to a lesser extent, in the adrenals.[170] In women with PCOS, approximately 60% of circulating androstenedione derives directly from the ovaries and the remainder from the adrenals; similarly, 60% of circulating testosterone is secreted directly by the ovaries, with most of the remainder deriving from peripheral conversion of androstenedione.[171]

The primary mechanisms driving increased ovarian androgen production in PCOS include increased LH stimulation resulting from abnormal LH secretory dynamics and

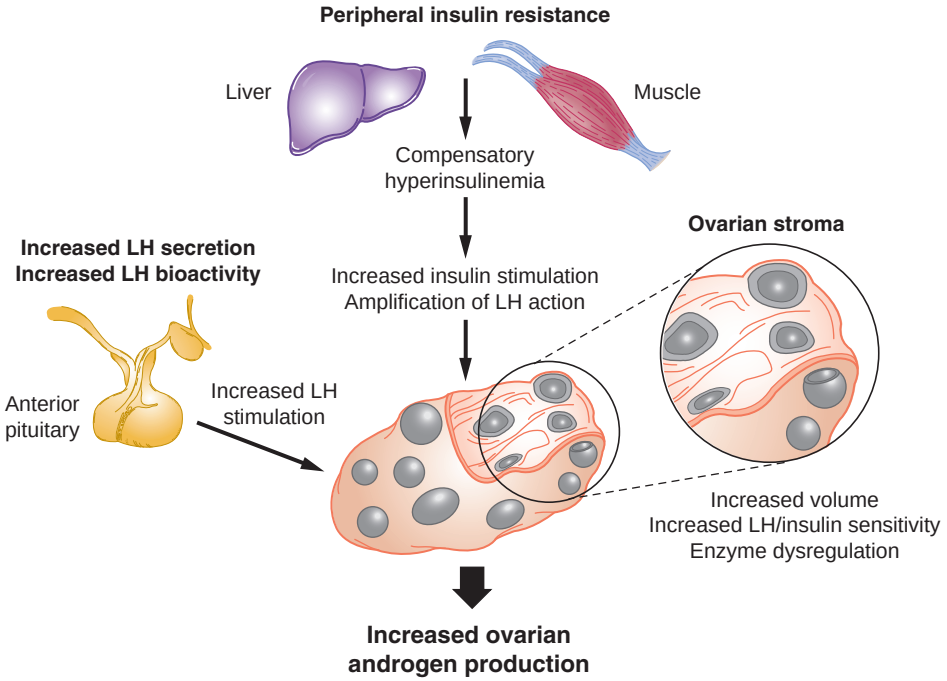
increased LH bioactivity, and hyperinsulinemia due to insulin resistance, which potentiates the action of LH and is worsened by obesity. Other evidence indicates that increased ovarian androgen synthesis in women with PCOS probably also relates to the increased volume of theca cells in an expanded ovarian stroma, and to increased sensitivity to LH stimulation,[172, 173] possibly due to over-expression of LH receptor in theca and interstitial (stromal) cells.[174] Androgen production remains elevated in long-term cultures of theca cells from women with PCOS,[175] even after suppression of LH levels by treatment with a GnRH agonist,[176] suggesting that increased androgen production in women with PCOS, to some extent, also reflects an intrinsic dysregulation of key steroidogenic enzmes, such as 3β-hydroxysteroid dehydrogenase (3β-HSD) and 17,20-lyase,[104, 172, 177–179] which may have a genetic foundation.[180]

Adrenal androgen production (androstenedione, DHEA, DHEA-S) also is increased in women with PCOS; over half exhibit moderately increased circulating DHEA-S levels.[181] When ovarian synthesis is suppressed by treatment with a long-acting GnRH agonist, adrenal androgen levels remain higher in women with PCOS than in normal women.[92, 172, 176, 182] Adrenal androgens have little or no intrinsic androgenic activity, but contribute to the pathophysiology of PCOS via conversion to testosterone in the periphery.

A number of potential mechanisms for the increase in adrenal androgen production have been investigated, but the explanation remains uncertain. Chronic estrogen stimulation due to anovulation could decrease adrenal 3β-HSD activity, as it does in the fetal adrenal cortex, but evidence for the mechanism is conflicting.[84, 183–186] Increased pituitary ACTH secretion or increased sensitivity to ACTH could provide an explanation, but neither can be demonstrated.[181, 187, 188] In some, but clearly not all women with PCOS, adrenal androgen excess might result from intrinsic upregulation of P450c17 17,20 lyase activity,[39, 135, 189–191] or from hyperinsulinemia.[192–196] ***In sum, no one mechanism explains the moderate adrenal androgen excess commonly observed in women with PCOS.***

High local androgen concentrations contribute to the polycystic morphogenesis of the ovaries, via conversion to more potent 5α-reduced androgens, which cannot be aromatized to estrogen and inhibit both aromatase activity and FSH induction of LH receptors on granulosa cells, thereby impeding or preventing progressive follicular development. Granulosa cells obtained from polycystic ovaries are not functionally impaired. They are sensitive to FSH and insulin-like growth factors and produce estrogen,[197–201] but cannot generate and maintain the estrogenic follicular milieu required to achieve more advanced stages of development. Consequently, new follicular growth continues but arrests long before full maturation is achieved, resulting in multiple small follicular cysts (typically measuring 2–10 mm in diameter), surrounded by hyperplastic theca cells, which often become luteinized due to increased LH stimulation. Atretic follicles ultimately contribute to an expanding ovarian stroma that increases in volume over time, further increasing the cellular mass producing androgens, in yet another self-propagating cycle that predisposes to chronic anovulation.

The importance of high local ovarian androgen concentrations in the pathophysiology of PCOS is demonstrated by the results of ovarian wedge resection and by observations in women with other conditions associated with hyperandrogenemia. Wedge resection results in a sustained decrease in androgen levels that precedes the return of ovulatory cycles, indicating that high intraovarian androgen concentrations effectively inhibit follicular development and prevent ovulation.[202–205] The success of ovarian wedge resection correlates with the amount of androgen-producing stromal tissue that is removed; even a unilateral oophorectomy can restore menstrual cyclicity and ovulation in anovulatory women with polycystic ovaries.[206] Although laparoscopic procedures such as ovarian “drilling” with an electrosurgical needle or a laser have replaced the classical wedge resection, the results achieved are similar. Polycystic ovaries also have been observed in women with androgen-producing ovarian and adrenal tumors,[207–209] and in female-to-male transsexuals



treated with exogenous androgens.[210, 211] ***These observations again illustrate the important point that polycystic ovaries are not a characteristic feature of a specific endocrine disorder. They result from a functional derangement in follicular development induced or sustained by increased intraovarian androgen levels as a consequence of chronic anovulation, whatever the cause.***

Genetic Considerations

Familial clustering of hyperandrogenism, anvoulation, and polycystic ovaries suggests an underlying genetic basis or cause. At least one group of patients with a heritable X-linked form of PCOS has been described, albeit with a widely varying phenotype.[212] Studies in large families have suggested autosomal-dominant inheritance, with premature balding as the male phenotype.[213, 214] Other studies of siblings and parents of women with PCOS have observed a high prevalence of hyperinsulinemia and hypertriglyceridemia, PCOS in females, and premature balding in males.[215, 216] Nearly 50% of sisters of women with PCOS have elevated total or bioavailable testosterone concentrations,[217] and approximately 35% of mothers also are affected.[218, 219] The first degree relatives of women with PCOS also exhibit other metabolic abnormalities such as dyslipidemia, which may predispose to an increased risk for cardiovascular disease.[220–223] These observations further suggest a genetic predisposition or susceptibility.

Understandably, efforts to identify genes associated with a susceptibility to anovulation and polycystic ovaries have focused on genes relating to the insulin receptor and substrates[224–226] and the genes encoding the P450 side-chain cleavage (*CYP11*) and P450c17 (*CYP17*) enzymes.[227–231] However, it seems likely that PCOS is a polygenic disorder involving the interaction of numerous genomic variants and the influence of environmental factors.[232] Candidate genes include the long list of molecules that participate in any of the metabolic and reproductive pathways affected in the syndrome, emphasizing yet again that PCOS is not a specific endocrine disorder, but a result of chronic anovulation due to a wide variety of causes.

Summary of Key Points

- Polycystic ovary syndrome is not a specific endocrine disorder having a unique cause. Rather, it is a complex disorder wherein numerous genetic variants and environmental factors interact, combine, and contribute to the pathophysiology.
- Polycystic ovaries and the clinical features of polycystic ovary syndrome reflect a functional derangement in follicular development, resulting in chronic anovulation. Because there are many causes of anovulation, there are many causes of polycystic ovaries and the polycystic ovary syndrome.
- Women with polycystic ovary syndrome generally exhibit increased serum LH concentrations, low-normal FSH levels, and increased LH:FSH ratios. The increase in serum LH levels results from abnormal LH secretory dynamics, characterized by increases in LH pulse frequency and amplitude, reflecting the pattern of pulsatile GnRH secretion. The decrease in FSH levels results from the increase in GnRH pulse frequency and from the negative feedback of chronically elevated estrone concentrations (derived from peripheral aromatization of increased androstenedione) and normal or increased levels of inhibin B (derived from small follicles).
- Insulin resistance and compensatory hyperinsulinemia are common features in women with polycystic ovary syndrome and play an important role in the pathophysiology. Up to 35% of women with polycystic ovary syndrome exhibit impaired glucose tolerance and up to 10% meet criteria for type 2 diabetes mellitus.
- Increased LH and insulin stimulation drives ovarian androgen production, and androgens and insulin combine to inhibit hepatic SHBG production, yielding increased free androgen, which aggravates underlying insulin resistance, in a self-propagating positive feedback loop that can increase in severity over time.
- Obesity contributes to the risk for developing polycystic ovary syndrome and adds to the pathophysiology in already affected women by aggravating the degree of insulin resistance and hyperinsulinemia.
- Hyperandrogenism is a major feature of polycystic ovary syndrome, resulting primarily from excess androgen production in the ovaries and, to a lesser extent, in the adrenals. Increased LH and insulin stimulation are the primary mechanisms driving increased ovarian androgen production. Others include an expanded ovarian stoma having increased sensitivity to insulin and LH, and intrinsic dysregulation of key steroidogenic enzymes.
- Polycystic ovary syndrome is a polygenic disorder likely involving the interaction of numerous genomic variants and the influence of environmental factors. Candidate genes include all of the molecules that participate in the affected metabolic and reproductive pathways.

Diagnosis of Polycystic Ovary Syndrome

It is generally accepted that PCOS is not a specific endocrine disease but a syndrome represented by a collection of signs and symptoms, and that no one sign, symptom, or test is diagnostic. Not surprisingly, establishing criteria for diagnosis of PCOS has proven both challenging and controversial.

It has been argued that having a clear and specific definition for PCOS is important because affected women are at increased risk for a variety of problems (infertility, dysfunctional bleeding, endometrial cancer, obesity, type 2 diabetes mellitus, dyslipidemia, hypertension, and cardiovascular disease), because the diagnosis can have health implications for other family

members, and because the need for life-long treatment may affect access to health care insurance in systems like that in the United States.[43, 233] ***In our view, the primary advantage to having specific diagnostic criteria for PCOS relates to research, because varying criteria cloud the conclusions and question the generalizability of results from studies involving women with "PCOS." In clinical medicine, simply knowing and understanding the health implications and consequences of chronic anovulation and methods for their effective management are far more important than assigning a specific diagnosis of PCOS.***

The basis for diagnosis of PCOS has changed with time and advances in medicine and related technology. The earliest descriptions of the disorder were based on findings of enlarged ovaries, hirsutism, and menstrual dysfunction.[32] The advent of hormone assays moved the focus to serum gonadotropin and androgen concentrations.[234] More recent advances in ultrasonography and recognition of the importance of insulin resistance in the pathophysiology have turned attention to ovarian morphology[235] and to the metabolic consequences of the disorder.

There have been three separate and distinct efforts to establish or refine the diagnostic criteria for PCOS. The first was a conference sponsored by the National Institute of Child Health and Human Development (NICHD) in 1990, concluding that the major criteria for diagnosis of PCOS (in order of importance) were (1) hyperandrogenism and/or hyperandrogenemia, (2) menstrual dysfunction, and (3) exclusion of other known disorders having a similar clinical presentation.[233] The second was a conference co-sponsored by the European Society for Human Reproduction and Embryology (ESHRE) and the American Society for Reproductive Medicine (ASRM), convened in Rotterdam, The Netherlands, in 2003, concluding that diagnosis of PCOS should be based on at least two of three major criteria, including (1) oligo/anovulation, (2) clinical or biochemical signs of hyperandrogenism, and (3) polycystic ovaries (as identified by ultrasonography), also excluding other androgen excess disorders.[236, 237] The third was a task force appointed by the Androgen Excess and PCOS Society (AE-PCOS) in 2006, concluding that diagnosis of PCOS requires (1) hyperandrogenism (hirsutism and/or hyperandrogenemia), (2) ovarian dysfunction (oligo/anovulation and/or polycystic ovaries), and (3) exclusion of other androgen excess or related disorders.[43]

The original 1990 NICHD diagnostic criteria were based on traditional concepts of PCOS, requiring evidence of *both* hyperandrogenism (hyperandrogenemia and/or hirsutism) and menstrual dysfunction (oligo/amenorrhea). The 2003 ESHRE/ASRM ("Rotterdam") criteria sought to recognize and accommodate a broader spectrum of the disorder, regarding polycystic ovaries as evidence of ovarian dysfunction and including women having *neither* hyperandrogenemia *nor* hirsutism. The 2006 AE-PCOS Society criteria allowed that polycystic ovaries could be considered a sign of ovarian dysfunction, but again emphasized that PCOS is characterized, first and foremost, by hyperandrogenism, including women with *either* oligo/amenorrhea or polycystic ovaries, but excluding those having *neither* hyperandrogenemia *nor* hirsutism.

Ironically, although the purpose of the consensus conferences and task force was to rigorously define PCOS for purposes of research, it can be argued that the differing sets of criteria succeeded only in creating controversy and confusion where clarity was needed most. ***Published clinical trials involving women with PCOS must be carefully reviewed to determine which diagnostic criteria were applied in selecting the study population.***

Hyperandrogenemia

Biochemical evidence of hyperandrogenism is based on the finding of elevated circulating androgen concentrations. Testosterone is the most important androgen produced by the

ovary and the usual basis for diagnosis of hyperandrogenemia. Other androgens that may be elevated in women with PCOS include androstenedione, DHEA, and DHEA-S.

Testosterone levels are elevated in most, but not all, women with PCOS. The free testosterone level is more sensitive for diagnosis of hyperandrogenic disorders,[43] but measurements of free testosterone have several limitations. Direct radioimmunoassays (RIA) for free testosterone are highly inaccurate,[238–240] particularly in the lower range and in women with decreased SHBG levels.[241] More sophisticated and accurate methods (equilibrium dialysis, gas or liquid chromatography-mass spectrometry) are technically complex, costly, and not widely available.[242] Moreover, testosterone is converted in androgen-sensitive tissues to dihydrotestosterone (DHT), which has a longer duration of action than testosterone, so serum total testosterone concentrations do not necessarily reflect androgen bioactivity.

For purposes of clinical research, the free testosterone concentration can be calculated, using equations derived from the laws of mass action, the serum concentrations of total testosterone, SHBG, and albumin, and the association constants for the interactions of testosterone with SHBG and albumin.[239, 243] Calculated values generally correlate well with those determined by equilibrium dialysis,[238, 239] although accuracy varies with the specific assays used to measure total testosterone and SHBG. ***For clinical purposes, measurement or calculation of the free testosterone level, or even measurement of the serum total testosterone concentration, usually is unnecessary.[244] In most cases, hirsutism provides ample evidence of hyperandrogenism, and if not severe, sudden in onset, rapidly progressive, or associated with symptoms or signs of virilization, there is little reason for concern about an androgen-producing tumor (Chapter 13).***

Measurement of the serum androstenedione concentration could yield evidence of hyperandrogenemia, but limited data suggest that levels are elevated in less than 20% of women with PCOS.[29] Measurement of serum DHEA also has little or no diagnostic value because levels are relatively low, exhibit a diurnal pattern and high between-subject variability, and are sensitive to stress.[43, 245]

The serum DHEA-S concentration is the traditional marker for adrenal androgen excess,[246–248] because it derives almost exclusively from the adrenal,[249–251] and concentrations are relatively high and remain stable across the day and cycle.[252, 253] Overall, the serum DHEA-S concentration is moderately elevated in over half of women with PCOS.[181] Some have an isolated increase in serum DHEA-S, suggesting a deficiency in 3β-HSD, but no genetic mutation in the enzyme has been found.[254, 255] ***Although the AE-PCOS Society regards an elevated serum DHEA-S level as sufficient evidence of hyperandrogenism to support the diagnosis of PCOS,[43] the test has very limited or no clinical value in our view.*** First, the test lacks both sensitivity and specificity for identifying women with adrenal causes of hyperandrogenism.[256] Second, DHEA-S, like DHEA and androstenedione, has little or no intrinsic androgenic activity and requires conversion to testosterone to exert androgenic effects. Third, the DHEA-S concentration can be grossly elevated (≥700 μg/dL) in women with rare androgen-secreting tumors, but in almost all such patients, the serum testosterone level also is greatly elevated,[257] due to peripheral conversion of high circulating DHEA-S levels, or because the tumor also secretes testosterone.

Clinical Hyperandrogenism

Clinical evidence of hyperandrogenism includes hirsutism, acne, and androgenic alopecia, all of which relate to the effect of androgens on the pilosebaceous unit. Because the sensitivity of the pilosebaceous unit varies significantly among individuals, the correlation between these clinical features and biochemical measures of hyperandrogenism is relatively poor.[258, 259]

Hirsutism is the growth of terminal hairs on the face or body in a male pattern. ***Hirsutism is the most obvious clinical indicator of androgen excess and is an important feature of PCOS.*** Whereas hirsutism affects 65–75% of White, Black and Southeast Asian women,[43, 260] its prevalence is lower in racial or ethnic groups having relatively little body hair.[260–262] The modified Ferriman-Gallwey score is the most common method for grading the extent of hirsutism, assigning a score from 0–4 in each of 9 androgen-sensitive areas, as illustrated and described in Chapter 13.[263, 264] The threshold value that defines hirsutism is not firmly established, but generally has ranged between 6 and 8.[29, 260, 264] ***The modified Ferriman-Gallwey score is the accepted standard for assessing the severity of hirsutism in clinical investigations. However, in clinical practice, the easiest and most practical way is to determine the method and frequency of hair removal (e.g., shaving, plucking, waxing), which also provides a clinically relevant measure for assessing the response to treatment.***

Acne can be another manifestation of hyperandrogenism. Like hirsutism, its prevalence among women with PCOS varies with ethnicity. The prevalence of acne is 12–14% among White women with PCOS,[159, 161, 262] higher in Asian Indians and women of Mediterranean descent (approximately 25%),[262, 265] and lower among Pacific Islanders.[261] However, it is unclear whether acne is any more prevalent among women with PCOS than in the general population. Approximately 20% of women under age 20, 15% of those ages 20 to 30, and 10% of women ages 30 to 40 complain of acne.[266–270] The extent to which PCOS may increase risk for developing acne, if at all, is therefore uncertain.

Androgenic alopecia, describing scalp hair loss in women, also can result from hyperandrogenism and is a recognized, but uncommon, feature of PCOS;[23, 159, 271–273] less than 5% of women with PCOS complain of hair loss. Typically, the hair loss is limited to the crown and does not involve the frontal hair line.[272, 274] Androgenic alopecia may be more common than is recognized, because 25% or more of scalp hair must be lost before thinning becomes apparent.[159, 274]

Ovulatory and Menstrual Dysfunction

Normal cyclic menses result from normal ovulatory function. The normal inter-menstrual interval ranges between 24 and 35 days and menses that occur less or more often are an indication of ovulatory dysfunction. Cyclic menses occurring at normal intervals strongly suggest, but cannot be regarded as proof of ovulation.

The majority of women with PCOS, approximately 60–85%, exhibit gross menstrual dysfunction.[43, 158, 161] The most common abnormalities are oligomenorrhea and amenorrhea. Polymenorrhea (regular cycles occurring at intervals less than 25 days) is very uncommon, observed in less than 2% of untreated women with PCOS.[159] Classically, menstrual dysfunction in women with PCOS has a premenarcheal onset, but many report regular cycles for varying intervals preceding the onset of oligo/amenorrhea.

In general, anovulatory women seldom have regular menses.[275] However, regular cycles are somewhat more common in anovulatory hyperandrogenic women.[43, 276] In studies of menstrual function in women with hyperandrogenism, approximately 15–40% are eumenorrheic, despite evidence of oligo-anovulation.[277–280] The prevalence of eumenorrhea among women with PCOS is significantly increased if the Rotterdam diagnostic criteria are applied, because hirsute eumenorrheic women with polycystic ovaries are included. The absence of any recognizable pattern of premenstrual molimina suggests anovulation in eumenorrheic women.

Polycystic Ovaries

PCOS takes its name from the enlarged polycystic ovaries so commonly observed in women with hyperandrogenic chronic anovulation.[32] Observations of mild hyperandrogenemia and insulin resistance in some asymptomatic women with polycystic ovaries provided the rationale for including polycystic ovaries among the Rotterdam diagnostic criteria for PCOS, as a sign of ovarian dysfunction.[85, 281–284]

Polycystic ovaries typically exhibit increased size and stromal volume and an increased number of small follicles. The Rotterdam criteria consider only the total number of follicles, requiring 12 or more measuring 2–9 mm in diameter (mean of both ovaries).[236, 237, 285–287] Others have defined polycystic ovaries on the basis of volume (>7.0–7.5 mL) and architecture.[288, 289] The prevalence of polycystic ovaries is quite high among women with androgen excess (>80%).[235, 277, 290–295] ***However, from 8% to 25% of normal women, and even 14% of women using oral contraceptives, also meet the ultrasonographic criteria for polycystic ovaries.***[281, 296–299 300] Moreover, polycystic ovaries are commonly observed during normal pubertal development, and even in women with hypothalamic amenorrhea and hyperprolactinemia.[301, 302]

The 2003 Rotterdam diagnostic criteria expanded the population of women that might be assigned a diagnosis of PCOS by approximately 50%, compared to the criteria earlier recommended by the NICHD, due entirely to the inclusion of polycystic ovaries.[303] The change in criteria ignited considerable controversy, primarily because polycystic ovaries are so commonly observed in normal women and in other conditions. Moreover, the finding, by itself, has little clinical significance. ***Otherwise normal women with polycystic ovaries generally have regular menstrual cycles, exhibit normal serum gonadotropin and ovarian steroid hormone levels, and are not infertile.***[85, 297, 304–306]

Again, the important point is that PCOS is a functional disorder in which polycystic ovaries result from chronic anovulation. Although present in most women with chronic hyperandrogenic anovulation, polycystic ovaries do not establish and are not required for diagnosis of PCOS.[163, 182, 296]

Other Features of the Polycystic Ovary Syndrome

PCOS has other common features besides hyperandrogenism and ovulatory dysfunction that are not included in any diagnostic criteria, including abnormal patterns of gonadotropin secretion, insulin resistance, and related metabolic abnormalities, such as dyslipidemia.

Abnormal Gonadotropin Secretion

Abnormal patterns of gonadotropin secretion have long been recognized as a common characteristic of women with PCOS. As discussed earlier in the section of this chapter devoted to the pathophysiology of the disorder, increased serum LH concentrations, low-normal FSH levels, and increased LH:FSH ratios are typical, but more so in lean than in obese women with PCOS. In the past, an increased LH:FSH ratio (e.g., >2:1) has been regarded as a marker of PCOS, but the ratio varies with the assays used to measure gonadotropin concentrations, and the prevalence of obesity is high among women with PCOS.[7, 60, 307–309] ***Consequently, gonadotropin levels or ratios are not a reliable diagnostic criterion; they neither make, nor exclude, the diagnosis.***

Insulin Resistance

Insulin resistance and hyperinsulinemia are common but not universal features of women with PCOS, no matter what method is used to assess insulin sensitivity.[92, 94, 310] ***The overall prevalence of insulin resistance among women with PCOS is between 50% and 75%, and greater in obese than in lean women with PCOS.***

Most women with PCOS and insulin resistance are young and have ample pancreatic β-cell reserve. Consequently, they are able to generate a compensatory hyperinsulinemia, allowing them to maintain normal glucose homeostasis, at least in the fasting state.[100] Although most therefore display an exaggerated insulin response to a glucose challenge, some also exhibit evidence of β-cell dysfunction,[311–313] particularly those having a family history of type 2 diabetes mellitus.[314]

Whereas there is no debate that insulin resistance and hyperinsulinemia play an important role in the pathophysiology of PCOS, or that the prevalence of unrecognized diabetes is sufficiently high to warrant testing to exclude the diagnosis in women with PCOS, the practical importance of detecting insulin resistance and what tests, if any, should be performed for that purpose remain highly controversial.

The gold standard method for measuring insulin sensitivity, to which all other methods are compared, is the ***hyperinsulinemic euglycemic clamp***.[315] The technique involves a fixed-rate intravenous infusion of insulin and a simultaneous intravenous glucose infusion, varying the rate as needed to establish a steady state plasma glucose level within the normal fasting range. The glucose infusion rate at steady state estimates the rate of glucose uptake in tissues at the defined plasma insulin concentration and is inversely proportional to the degree of insulin resistance; the lower the glucose infusion rate at steady state, the greater the degree of insulin resistance. Insulin sensitivity is defined as the ratio of the glucose disposal rate to the steady state insulin concentration (glucose disposal rate [mmol/kG] × min per mU/L × 100). The clamp technique and other methods involving intravenous infusions of glucose and/or insulin (e.g., the frequently sampled intravenous glucose tolerance test, the insulin sensitivity test, the insulin tolerance test, and the continuous infusion of glucose with model assessment) have been used extensively in clinical investigations of glucose and insulin dynamics. However, they have no real practical clinical application because they are time-consuming, invasive, costly, and require experienced personnel.

The complexities of clamp techniques and other methods requiring intravenous infusions and multiple blood samplings spurred efforts to find an uncomplicated and inexpensive quantitative method for evaluating insulin sensitivity. A number of fasting state (homeostatic) measures have been described, all based on the fasting glucose and insulin concentrations and using straightforward calculations.[316] One common weakness that all such methods have is that they assume a linear relationship between glucose and insulin that is, in fact, parabolic.

The ***fasting serum insulin concentration*** is easy to obtain and requires no calculations;[317] in euglycemic White women with PCOS, values greater than 20–30 μU/mL suggest insulin resistance. The ***fasting glucose/insulin ratio*** has been used widely as an index of insulin sensitivity in women with PCOS; a ratio less than 4.5 has reasonable sensitivity and specificity for insulin resistance.[94] The ***homeostatic model assessment of insulin resistance (HOMA-IR)*** is another measure of insulin sensitivity commonly used in larger epidemiologic studies. The HOMA-IR is calculated by dividing the product of the fasting glucose (mg/dL) and insulin (μU/mL) concentrations by a constant: [glucose (mg/dL)][insulin (μU/mL)]/405, or [glucose (mmol/L)][insulin (μU/mL)]/22.5.[318, 319] The HOMA-IR value correlates relatively well with results from clamp studies,[320, 321] and unlike the fasting insulin concentration and the glucose/insulin ratio, compensates for fasting hyperglycemia;

values greater than 3.2–3.9 generally indicate insulin resistance.[43, 95, 322] The ***quantitative insulin sensitivity check index (QUICKI)*** is yet another method for assessing insulin sensitivity in clinical investigations. Like the HOMA-IR, QUICKI can be applied in both euglycemic and hyperglycemic patients.[323] The QUICKI value is the inverse of the sum of the fasting glucose and insulin concentrations, expressed logarithmically: (1/[log(Glucose)+log(Insulin)]); values greater than 0.33 indicate insulin resistance.[43, 324] Still other methods use a weighted combination of the fasting insulin and triglyceride concentrations and BMI to estimate insulin sensitivity.[315] All of the calculated indices have limitations, primarily the lack of a standardized insulin assay. ***As the sheer number of different measures of insulin resistance demonstrates, there is no uniformly accepted test for measuring insulin resistance in a clinical setting. Consequently, routine screening for insulin resistance is not recommended.***

The standard ***oral glucose tolerance test*** (OGTT) is the mainstay of methods for diagnosis of impaired glucose tolerance and diabetes mellitus and also can be used to assess insulin sensitivity, when indicated (discussed below). Although techniques vary, all involve measures of plasma glucose and insulin at intervals over 2 to 4 hours after a 75-g or 100-g oral glucose load. ***A baseline 2-hour OGTT is recommended for all women with PCOS, as up to 35% exhibit impaired glucose tolerance and up to 10% have diabetes mellitus.***[99, 237, 325, 326]

Screening for glucose intolerance also is recommended for girls with premature adrenarche or menstrual irregularity that persists for more than 2 years after menarche because hyperinsulinemia often is the cause and they are at high risk for developing diabetes and severe hyperandrogenism.[327–330] In this population, specific screening for insulin resistance also is warranted because evidence indicates that early intervention in those affected can prevent progressive debilitating disease.[329, 331, 332] Specific screening for insulin resistance also is recommended for women with markely elevated serum androgen levels (≥150 ng/dL), to differentiate the severe insulin resistance syndromes (discussed below) from androgen-secreting tumors.

Interpretation	*2-hour Glucose*	*2-hour Insulin*[43]
Normal	<140 mg/dL	
Impaired glucose tolerance	140–199 mg/dL	
Diabetes mellitus	≥200 mg/dL	
Normal		<80–100 µU/mL
Insulin resistance		>80–100 µU/mL
Severe insulin resistance		>300 µU/mL

Dyslipidemia

Dyslipidemia is perhaps the most common metabolic abnormality observed in women with PCOS. Applying the National Cholesterol Education Program guidelines, nearly 70% have at least one borderline or elevated lipid level,[333] although many women with PCOS have entirely normal lipid profiles.[334–337] Insulin resistance and hyperinsulinemia are associated with decreased high-density lipoprotein (HDL) cholesterol and elevated triglyceride levels, and numerous studies have observed such abnormalities in women with PCOS.[334, 338, 339] Some also have observed elevated low-density lipoprotein (LDL) concentrations,[333, 335–337] which are not usually associated with insulin-resistant states and may result from hyperandrogenism or reflect a genetic or dietary influence.[223, 340, 341]

Obesity

Obesity is a common feature of PCOS. The prevalence of obesity is approximately 50% overall,[159] but varies significantly with country of origin. The prevalence is highest in the United States, probably reflecting the higher overall prevalence of obesity;[161] in other countries, women with PCOS generally are leaner.[294, 342–344]

The risk for PCOS increases with obesity.[145, 345] Although the effect appears relatively modest,[146] it is clear that obesity adds to the pathophysiology of PCOS in affected or predisposed women by aggravating the degree of insulin resistance and hyperinsulinemia.[164, 165] As discussed earlier in the section of this chapter devoted to the pathophysiology of PCOS, high insulin levels stimulate ovarian androgen production and suppresses hepatic SHBG production, thereby increasing bioavailable androgen levels. In turn, high androgen concentrations and chronically elevated estrogen levels (derived from aromatization of androgens in adipose), help to induce or perpetuate an abnormal pattern of gonadotropin secretion (increased LH, low FSH) by increasing LH pulse frequency and amplitude and inhibiting FSH secretion.

Exclusion of Other Androgen Excess Disorders

PCOS is a diagnosis of exclusion, after considering and eliminating other causes of chronic anovulation (primarily thyroid disorders and hyperprolactinemia) and androgen excess. Together, congenital adrenal hyperplasia, androgen-secreting tumors, severe insulin resistance syndromes, Cushing syndrome, and idiopathic hirsutism account for about 10–30% of hyperandrogenism in women.[159, 274, 279, 346] ***Whereas all should be considered and excluded, few actually warrant specific testing.***

Thyroid Disorders

Thyroid disorders are associated with menstrual dysfunction and also can have serious adverse impact on pregnancy outcomes and child development.[347–352] ***The overall high prevalence of thyroid dysfunction in women warrants specific testing to exclude the diagnosis (serum thyroid-stimulating hormone, TSH) in all anovulatory women, including those with hyperandrogenism, but not for diagnosis of PCOS.***

Hyperprolactinemia

Hyperprolactinemia is highly associated with menstrual dysfunction and is one of the most common causes of secondary amenorrhea. The many causes of hyperprolactinemia are considered at length elsewhere in this text (Chapters 11 and 16). Hyperprolactinemia is associated with increased adrenal androgen production *in vivo* and *in vitro*,[353, 354] but its prevalence among women who present with hyperandrogenism is quite low, and generally less than 3%.[159, 274, 342, 346, 355–358] ***The high prevalence of hyperprolactinemia among women with menstrual dysfunction justifies specific testing to exclude the diagnosis in all anovulatory women, but not for diagnosis of PCOS.***

Nonclassical Congenital Adrenal Hyperplasia

Congenital adrenal hyperplasia (CAH) is caused by adrenal steroidogenic enzyme defects that result in excessive adrenal androgen production. The most common cause is 21-hydroxylase deficiency; other enzyme defects (11β-hydroxylase, 3β-hydroxysteroid dehydrogenase) are relatively rare. In all, the pathophysiology stems from decreased cortisol production, which stimulates a compensatory increase in pituitary ACTH secretion, causing adrenal hyperplasia; increased levels of steroid hormones proximal to the enzyme block seek an alternative metabolic pathway, resulting in increased production of androgens. The disorder is inherited in an autosomal recessive fashion and is discussed in detail in Chapters 9, 10, and 13.

Females with classical CAH (both salt-wasting and simple virilizing forms) typically present at birth with ambiguous genitalia,[359–361] and thus would rarely be confused with PCOS, but those with the nonclassical or "late-onset" form of CAH present later, during childhood or early adolescence with precocious puberty, or as young adults with signs of hyperandrogenism, very much like those with PCOS.[362] ***Whereas it is logical to recommend that nonclassical CAH be excluded specifically in all women with hyperandrogenism,[43] we believe that specific testing can be safely reserved for those having an early onset of hirsutism (pre- or peri-menarcheal, including girls with premature adrenarche), women with a family history of the disorder, and those in high-risk ethnic groups (Hispanic, Mediterranean, Slavic, Ashkenazi Jewish, or Yupic Eskimo heritage).*** The yield from routine screening is very low, because the disorder is uncommon.[158, 159, 363] The prevalence of nonclassical CAH among American White and Hispanic women with hyperandrogenism is between 1 and 4%.[364] In other countries, the reported prevalence has ranged from as low as 0.3% among Northern Italians to as high as 6–10% among women from Israel, India, and Jordan.[365, 366] Moreover, a diagnosis of nonclassical CAH generally will not change the best choice of treatment, because glucocorticoids are less effective than estrogen-progestin contraceptives and/or antiandrogens for the treatment of chronic anovulation and hirsutism in women with nonclassical CAH.[367, 368] Whereas it is important to identify women at risk for conceiving a child with the more severe classical form of the disorder, the risk among women with hyperandrogenism is limited to those who carry one classic mutation and a variant allele associated with mild enzyme deficiency (compound heterozygotes), also having a male partner who carries an occult classic mutation. Data from neonatal screening programs for detection of classical CAH indicate that the overall prevalence of classical CAH is approximately 1 in 15,000 live births and varies with ethnicity, ranging from 1 in 28,000 Chinese[369] and between 1 in 5,000 and 1 in 23,000 Caucasians,[370, 371] to as high as 1 in 280 Yupic Eskimos.[372] In the United States, the prevalence of classical CAH is 1 in 15,500 White and 1 in 42,000 African Americans.[373]

Regardless whether universal or selective screening for nonclassical CAH is performed, a follicular phase morning serum 17-OHP concentration less than 200 ng/dL excludes, and a level greater than 800 ng/dL all but establishes the diagnosis.[374–376] Concentrations between the two threshold values suggest the possibility, which can be confirmed by performing an ACTH stimulation test, obtaining blood samples before and 60 minutes after administering cosyntropin (synthetic ACTH 1–24; 0.25 mg intramuscularly, or intravenously); in most women with nonclassical CAH, the 17-OHP concentration will rise above 1,500 ng/dL.[363, 365, 377]

Androgen-Secreting Ovarian and Adrenal Tumors

Androgen-secreting ovarian and adrenal tumors are rare. The prevalence of ovarian androgen-producing tumors is between 1 in 300 and 1 in 1,000 among women with hyperandorgenism.[159,346,357,378] Androgen-secreting adrenal tumors are even less common.[356]

In addition, androgen-secreting tumors almost always are accompanied by severe or rapidly progressive hirsutism or symptoms or signs of virilization (deepening of the voice, temporal or male pattern balding, breast atrophy, increased muscle mass, and clitoromegaly). ***The possibility of a tumor is excluded primarily by the clinical history and physical examination. Very few women will require specific evaluation to exclude the diagnosis.***

The recommended evaluation for women suspected of having an androgen-secreting tumor is discussed at length in Chapter 13 and briefly summarized here. A ***serum total testosterone concentration*** greater than 150 ng/dL identifies almost all women with a potential androgen-producing tumor. However, a tumor still should be suspected and excluded in women with rapidly progressive hirsutism or signs or symptoms of virilization, even when the serum testosterone concentration is below the threshold value. ***Transvaginal ultrasonography*** will identify almost all solid ovarian mass lesions, although very small tumors located in the hilar region can escape detection. ***Adrenal computed tomography (CT)*** is extremely sensitive for detecting rare androgen-secreting adrenal tumors, most of which are malignant. ***Selective ovarian venous catheterization*** can be considered for the rare patient having no demonstrable ovarian or adrenal mass lesion, but should be reserved only for those in whom a tumor is strongly suspected.

Severe Insulin Resistance Syndromes

Severe insulin resistance is a specific characteristic of a variety of uncommon clinical disorders. The type A insulin resistance syndrome results from defects in the insulin receptor and affects primarily lean women. The type B syndrome is an autoimmune disorder affecting the insulin receptor. The type C syndrome is a variant of type A and is characterized by marked acanthosis nigricans, hyperandrogenism, obesity, and the absence of insulin receptor defects, and also is known as the hyperandrogenic-insulin resistant-acanthosis nigricans (HAIR-AN) syndrome. Other rare disorders involving severe insulin resistance include leprechaunism, the Rabson-Mendenhall Syndrome, and a variety of lipodystropic syndromes.[43, 379]

Although the type C syndrome might reasonably be viewed as a severe form or phenotype of PCOS, the more profound insulin resistance and related metabolic abnormalities in the syndrome distinguishes the two.[380, 381] Ovarian hyperthecosis, characterized by distinct clusters of luteinzed theca cells scattered throughout the ovarian stoma and associated with severe hyperandrogenism,[382, 383] frequently is observed in women with severe insulin resistance syndromes. Skin tags and acanthosis nigricans (a gray-brown, velvety, sometimes verrucous, discoloration of the skin, usually involving the neck, groin, axillae, and the area beneath the breasts) are other common features of the severe insulin resistance syndromes. The mechanism responsible for their development is uncertain.

Although specific diagnostic criteria for the severe insulin resistance syndromes have not been established, the diagnosis can be substantiated by findings of markedly elevated levels of insulin, typically greater than 80 μU/mL fasting, or greater than 300 μU/mL 2 hours after an oral glucose load.[380, 381] As might be expected, most patients will have normal glucose levels in the early stages of the disorder, but are at high risk to develop β-cell failure, diabetes, and dsylipidemia. Accordingly, they require careful long-term follow-up and treatment.

Cushing Syndrome

Cushing syndrome results from excess adrenal cortisol secretion and can be ACTH-dependent (pituitary and ectopic ACTH-secreting tumors) or ACTH-independent (adrenal adenomas,

exogenous glucocorticoid treatment). ***The disorder has features commonly observed in women with PCOS, including menstrual dysfunction, hyperandrogenism, and central obesity. However, the prevalence of Cushing syndrome in women presenting with hyperandrogenism is extremely low, well below 1%.***[158, 159, 274, 346, 384] ***Consequently, routine screening is not justified and should be limited to the very few patients who also have distinct signs and symptoms of hypercortisolism.*** These include hypertension, severe fatigue and muscle weakness, atrophy of the skin and subcutaneous tissue (easy bruising and purple striae on the abdomen and flanks), hyperpigmentation (caused by excess secretion of α-melanocyte-stimulating hormone, as a byproduct of ACTH synthesis from pro-opiomelanocortin, the common precursor molecule) in areas most exposed to light (the face, neck, and back of the hands) or chronic mild trauma, friction, or pressure (the elbows, knees, knuckles, and shoulders), diabetes, and cognitive impairment.

Methods of screening for Cushing syndrome, specific tests for those who screen positive, and evaluation to differentiate among the causes of Cushing syndrome are discussed in detail in Chapter 13. ***The overnight dexamethasone suppression test is the best single screening test because of its simplicity and ability to discriminate.*** The test is performed by administering 1.0 mg of dexamethasone between 11:00 P.M. and midnight and measuring the serum cortisol at 8:00 A.M. the next morning; values less than 1.8 μg/dL are normal.[385]

Idiopathic Hirsutism

Idiopathic hirsutism is defined classically as hirsutism accompanied by normal ovulatory and menstrual function, in the absence of hyperandrogenemia. Using that definition, the prevalence of idiopathic hirsutism among hirsute women is approximately 5–7%.[158, 159, 279, 280, 386] If the 2003 Rotterdam criteria for diagnosis of PCOS are used, the definition would also include the absence of polycystic ovaries, further decreasing the prevalence of idiopathic hirsutism.

By definition, diagnosis of idiopathic hirsutism requires measurement of serum androgen levels, which otherwise is not necessary for those with mild hirsutism (Chapter 13). It is generally assumed that idiopathic hirsutism results from increased peripheral 5α-reductase activity, which amplifies the action of normal circulating testosterone concentrations via increased intracellular conversion to the more potent androgen, dihydrotestosterone (DHT). Given that many hirsute eumenorrheic women will be found oligo-ovulatory on closer scrutiny, a test of ovulation (e.g., serum progesterone during the putative luteal phase) further helps to differentiate women with PCOS from those with idiopathic hirsutism.

Exclusion of Androgen Excess Disorders other than PCOS

Diagnosis	*Method of Exclusion*
Nonclassical CAH	Follicular phase morning serum 170HP <2ng/mL (Early onset hirsutism, family history of CAH, high- risk ethnicity)
Androgen-secreting tumor	Primarily by clinical history and physical examination; serum testosterone
Servere insulin resistance syndrome	Primarily by clinical history and physical examination; 2-hour OGTT (glucose, insulin levels)
Cushing syndrome	Primarily by clinical history and physical examination; overnight dexamethasone suppression test
Idiopathic hirsutism	Menstrual history, serum progesterone (putative luteal phase), serum testosterone

Summary of Key Points

- Polycystic ovary syndrome is not a specific endocrine disease but a syndrome represented by a collection of signs and symptoms, and no one sign, symptom, or test is diagnostic.
- Diagnosis of polycystic ovary syndrome is based primarily on the clinical history and physical examination. The major clinical features of polycystic ovary syndrome are hyperandrogenism and menstrual dysfunction.
- Although present in most women with chronic hyperandrogenic anovulation, polycystic ovaries do not establish and are not required for diagnosis of polycystic ovary syndrome.
- Gonadotropin levels or ratios are not a reliable criterion for diagnosis of polycystic ovary syndrome.
- Knowing and understanding the health implications and consequences of chronic anovulation and methods for their effective management are far more important than assigning a specific diagnosis of PCOS.
- Evaluation of women with suspected polycystic ovary syndrome should include:

1. Serum thyroid-stimulating hormone (TSH)
2. Serum prolactin
3. 2-hour oral glucose tolerance test
4. Fasting lipid profile
5. Endometrial sampling (in women whose history indicates potential long-term exposure to unopposed estrogen stimulation)
6. Serum testosterone (in women with moderate or severe hirsutism)
7. Morning follicular phase serum 17-hydroxyprogesterone (in women with a pre- or perimenarcheal onset of hirsutism, a family history of congenital adrenal hyperplasia, or high-risk ethnicity)
8. Overnight dexamethasone suppression test (in women with signs or symptoms of hypercortisolism)

Clinical Management

The management of women with PCOS should seek to correct or prevent both its immediate and longer-term clinical consequences, which may include all of the following:

- Menstrual abnormalities.
- Increased risk for developing endometrial hyperplasia and neoplasia.
- Hyperandrogenism (hirsutism, acne, alopecia).
- Infertility.
- Increased risk for developing type 2 diabetes.
- Increased risk for developing cardiovascular disease.

In many cases, lifestyle changes will be an important part of the clinical management, requiring careful education, counseling, encouragement, and follow-up. For patients having no immediate desire to attempt pregnancy, estrogen-progestin contraceptives provide effective management for menstrual dysfunction and protect against the risk for development of endometrial hyperplasia and cancer. Estrogen-progestin contraceptives and antiandrogens help to prevent or decrease hyperandrogenism. Those seeking to conceive are candidates for ovulation induction. Women with impaired glucose intolerance at risk for developing type 2 diabetes or having features of the metabolic syndrome, indicating a high risk for developing cardiovascular disease, may warrant treatment with insulin sensitizing agents or other medications aimed specifically at reducing those risks. The important point to emphasize is that women with chronic anovulation require comprehensive clinical management that addresses their immediate needs, but also considers their longer-term health and incorporates appropriate risk reduction strategies.

Lifestyle Changes

The strong association between obesity, hyperandrogenism, impaired glucose tolerance, menstrual abnormalities, and infertility emphasizes the importance of addressing lifestyle issues in women with PCOS, focusing on nutrition and exercise. At least 50% of women with PCOS are obese. ***It is important to stress that even a small reduction in weight (2–5%) can result in significant improvements in metabolic and reproductive function.***[387–392] The loss of abdominal fat may be the best predictor of the effects of weight loss.

Weight reduction is the first best treatment for obese women.[393] Weight loss increases SHBG concentrations, thereby reducing free androgen levels and decreasing androgen stimulation of the hair and skin. Weight loss also improves ovulatory function, thereby increasing conception rates and also possibly decreasing the risk for miscarriage. ***A significant overall decrease in caloric intake is more important than the specific composition of the diet; there is no compelling evidence to indicate that a low carbohydrate diet is better than a low fat diet.***[394–396] ***Although treatment with metformin can facilitate weight loss,***[390, 397–399] ***primarily by suppressing appetite,***[400] ***the overall effect is modest and inconsistent.***[401–410] ***Consequently, metformin should not be used primarily for the purpose of weight reduction.***

The benefits of exercise for improving diabetes and cardiovascular health have been demonstrated in the general population. Incorporation of moderate activity into daily activities appears as effective for reducing the risk of developing diabetes and cardiovascular disease as that achieved with vigorous physical activity, is more likely to be sustained, and is essential for maintaining weight loss over time.[411]

Menstrual Abnormalities and Risk for Developing Endometrial Cancer

Oligomenorrhea is the most common presentation of women with chronic anovulation, although many present with amenorrhea or dysfunctional uterine bleeding, and some even have regular menses. The typical patient presents with irregular or infrequent menses or with amenorrhea, making any formal assessment of ovulatory function (e.g., basal body temperature, serum progesterone measurement) unnecessary. The overall number of menstrual cycles is less important than preventing abnormal bleeding and the other potential consequences of chronic anovulation. The evaluation and treatment of amenorrhea are discussed in depth in Chapter 11. Dysfunctional uterine bleeding is the focus of Chapter 15.

Chronic anovulation, obesity, and hyperinsulinemia all are associated with risk for developing endometrial cancer.[412–416] Presumably, the mechanism relates to constant, unrelenting estrogen stimulation of the endometrium, predisposing to abnormal patterns of growth. Endometrial hyperplasia, and even endometrial cancer can be encountered in young anovulatory women.[417–419] Overall, the risk for developing endometrial cancer may be increased by as much as 3-fold. Consequently, for those with long-standing anovulation, endometrial sampling to exclude endometrial hyperplasia is a prudent precaution. ***The decision on whether to perform an endometrial biopsy should not be based on the patient's age, but on the duration of potential exposure to unopposed estrogen stimulation.*** Whereas a grossly increased endometrial thickness (greater than 12 mm) clearly suggests the possibility of endometrial hyperplasia,[420] a normal thickness does not exclude the diagnosis.[421, 422]

Estrogen-progestin contraceptives are the most common treatment for the menstrual abnormalities associated with chronic anovulation because they induce regular cyclic menses and attenuate endometrial growth, thereby preventing dysfunctional uterine bleeding and also eliminating the risk for developing endometrial hyperplasia and neoplasia. In those who refuse or have a contraindication to the use of estrogen-progestin contraceptives, the same can be achieved with cyclic or continuous treatment with progestins alone. However, progestin treatment forfeits some of the other important actions of estrogen-progestin contraceptives that help in the treatment of hyperandrogenism, as discussed below. Metformin is another alternative that can restore ovulatory menses in many women with PCOS. However, results vary widely and may require up to 6 months of treatment before they are known.[423–430]

Hirsutism

True virilization is rare, but nearly 70% of anovulatory women complain of cosmetically disturbing hirsutism, the severity relating primarily to the level of hyperandrogenemia, but also to the genetic sensitivity of the individual's hair follicles to androgens. Hirsutism is more common in obese anovulatory women, because free androgen levels increase with BMI, due to insulin resistance, hyperinsulinemia, and the combined inhibitory effects of insulin and androgens on hepatic SHBG production. Skin and hair disorders can be both physically and psychologically very damaging. The spectrum of treatments for hirsutism is discussed in Chapter 13 and summarized here.

Mild focal hirsutism can be managed effectively with cosmetic measures (shaving, plucking, waxing, depilatories), but most who present with a complaint of hirsutism are already using one or more such methods and will require treatment. Medical management options include primarily estrogen-progestin contraceptives and antiandrogens (e.g., spironolactone).

Estrogen-progestin contraceptives are an effective treatment for hirsutism primarily because they suppress LH-dependent ovarian androgen production and stimulate hepatic SHBG production.[431–436] Some have questioned the wisdom and safety of estrogen-progestin contraceptives in women with PCOS, primarily because they have been associated with modest decreases in insulin sensitivity in some studies.[437–441] However, the overall weight of available evidence supports their safety in women with PCOS, with and without insulin resistance.[407, 442–452]

Antiandrogens are effective for the treatment of hirsutism, but generally should be used in combination with an estrogen-progestin contraceptive or another highly reliable method (e.g, an intrauterine device) because of their potential to adversely affect sexual development in a male fetus if the patient were to conceive unexpectedly. Options include spironolactone (50–100 mg twice daily),[453, 454] cyproterone acetate (12.5–100 mg daily, or in combination oral contraceptives containing the progestin),[455] and flutamide (62.5 mg daily).[456]

Although insulin sensitizing agents (metformin, thiazolidinediones) decrease circulating insulin and androgen levels in women with PCOS,[110, 457–463] a systematic review including 9 placebo-controlled trials concluded that they have no important benefits for the treatment of hirsutism,[464] and guidelines issued by the Endocrine Society suggest against their use for the treatment of hirsutism.[244]

Infertility

Chronic anovulation is one of the most common causes of infertility. In women with PCOS, other factors relating to oocyte quality or endometrial and implantation abnormalities also might contribute.[465] Infertile anovulatory women who want to conceive are candidates for ovulation induction. Methods for ovulation induction are the subject of Chapter 31 and are outlined briefly here.

The first drug of choice is clomiphene citrate, which is typically administered in an empiric incremental fashion to identify the lowest effective dosage (50–150 mg daily × 5 days, beginning on cycle day 3–5). The cumulative pregnancy rate with clomiphene treatment is approximately 50% after 3 induced ovulatory cycles, and approaches 75% within 6–9 cycles of treatment.[466] The risk for multiple gestation is approximately 5–8%. Approximately 20% of patients prove refractory to clomiphene treatment, most of those having severe hyperandrogenism or obesity.[467]

Treatment with insulin sensitizing agents (metformin, thizolidinediones, D-chiro-inositol) can increase ovulation rates in some women with PCOS.[463, 468, 469] Metformin has been used widely for that purpose, but there is no practical way to predict reliably those who will respond. Preliminary evidence suggests that a response to metformin may be less likely in women having a polymorphism of a gene encoding a hepatic serine-threonine kinase (*STK11*).[470] Fasting insulin concentrations and glucose:insulin ratios do not predict response to metformin,[471] and overall, metformin appears most effective in patients who also respond to clomiphene.[469, 472]

A 2003 meta-analysis of studies involving treatment with metformin in women with PCOS concluded that its efficacy for improving ovulatory function compared favorably with that of clomiphene.[471] However, subsequent randomized multicenter trials comparing the two drugs, alone and in combination, have found clomiphene clearly superior to metformin and observed that combined treatment offers no significant additional benefit.[473–475] In the largest trial, clomiphene yielded a significantly higher live birth rate than metformin (22.5% vs. 7.2%), and the results of combined treatment were not significantly better (26.8%).[474] In a few small studies involving clomiphene-resistant anovulatory women with PCOS, combined treatment has increased ovulation and pregnancy rates over those achieved with clomiphene alone.[476–479] A 2008 meta-analysis including 17 randomized trials concluded that combined treatment with metformin and clomiphene achieves higher ovulation and pregnancy rates than treatment with clomiphene alone.[469] Although there is no convincing evidence that combined treatment with metformin and clomiphene can increase live birth rates over those achieved with clomiphene alone,[480] the attempt seems justified for women having few alternatives besides ovarian drilling or treatment with exogenous gonadotropins. Limited evidence indicates that combined treatment with metformin and roziglitazone,[481] or with clomiphene and rosiglitazone,[482] is no more effective than metformin alone. ***In summary, clomiphene should be the first choice of therapy for ovulation induction in women with PCOS, and in those who prove resistant, combined treatment with metformin and clomiphene deserves consideration before proceeding to ovarian drilling or treatment with gonadotropins.***

Although there is no evidence that metformin treatment during pregnancy is associated with any increased risk for major fetal malformations,[483] the safety of its use during pregnancy

is not yet established. Some have advocated metformin treatment to reduce the increased risk for miscarriage in women with PCOS, which might relate to an underlying metabolic disorder.[150, 484–486] However, no difference in the miscarriage rates of women who did or did not receive metformin treatment have been observed in large randomized trials.[473–475] Metformin treatment during pregnancy also has been advocated to reduce the risk for developing gestational diabetes and other pregnancy complications in women with PCOS.[487] In diabetic women, treatment with metformin during pregnancy has been associated with an increased prevalence of pre-eclampsia and increased perinatal mortality in some studies,[488] but not in others.[489] Currently, routine metformin treatment during pregnancy is not recommended for women with PCOS.[472]

Induction of ovulation with exogenous gonadotropins is highly effective, but requires careful monitoring to avoid the intrinsic risks of multiple pregnancy and ovarian hyperstimulation syndrome (OHSS). Many women are highly sensitive to low doses of medication and exhibit a relatively narrow therapeutic range.[490–495] Although whether metformin treatment can improve outcomes for women with PCOS in gonadotropin-stimulated[485, 496, 497] or in vitro fertilization (IVF) cycles[403, 498, 499] remains unclear, evidence indicates the risk for OHSS may be decreased.[500]

Laparoscopic ovarian drilling with laser or diathermy also can be effective for restoring ovulatory function in women with PCOS, but has risk for causing postoperative adnexal adhesions and decreased ovarian reserve.[501] There is no evidence that metformin treatment improves outcomes achieved with ovarian drilling.[500]

Metabolic Abnormalities and Associated Health Risks

Women with chronic anovulation commonly exhibit insulin resistance and other risk factors for the development of type 2 diabetes and cardiovascular disease. These observations have focused a great deal of attention on the importance of incorporating risk reduction strategies into the clinical management of women with PCOS.

Insulin resistance results in compensatory hyperinsulinemia, which predisposes to a progressive decline in pancreatic β-cell reserve, leading to glucose intolerance, and ultimately, type 2 diabetes mellitus. In women with PCOS, pancreatic β-cell dysfunction can be demonstrated even before glucose intolerance becomes apparent, and the rate of progression from glucose intolerance to diabetes is increased;[99, 502, 503] up to 10% of women with PCOS develop diabetes by the age of 40.[99, 100] Obesity adds to the risk, by aggravating the underlying insulin resistance. Overall, the risk for developing impaired glucose tolerance or type 2 diabetes is increased 3- to 7-fold in women with PCOS, compared to women of comparable age without PCOS.[99, 100, 503]

Although direct evidence for an increased incidence of cardiovascular disease in women with PCOS is lacking, the prevalence of known risk factors is substantially increased.[504] Insulin resistance and hyperinsulinemia are associated with chronic low-grade inflammation, as reflected by elevations in C-reactive protein, interleukin-6, leukocyte count, and other inflammatory markers.[505–513] Hyperinsulinemia also is associated with hypertension and increased production of plasminogen activator inhibitor type-1 (PAI-1), the principal inhibitor of tissue plasminogen activator (tPA) and urokinase, thereby inhibiting fibrinolysis.[514, 515] Elevated androgen levels predispose to increased LDL-cholesterol and aggravate underlying insulin resistance. Consequently, many women with PCOS have some degree of dyslipidemia, such as decreased HDL-cholesterol and increased total and LDL-cholesterol and triglycerides.[338, 516] Many also have central obesity, and some even meet criteria for the diagnosis of the metabolic syndrome, predicting a high risk for developing cardiovascular disease.[517–520]

The ***metabolic syndrome***, originally known as syndrome X,[521] represents a constellation of closely related cardiovascular risk factors, and several studies have observed an increased prevalence of metabolic syndrome in women with PCOS.[517, 522] A number of different definitions for the metabolic syndrome have been proposed, varying in emphasis on abnormalities in glucose metabolism (insulin resistance, hyperinsulinemia, glucose intolerance, diabetes mellitus), central obesity, and cardiovascular risk factors (hypertension, increased triglycerides, decreased HDL cholesterol).[523–526] Although all of the definitions yield comparable estimates of the overall prevalence of the metabolic syndrome, they identify different populations in different ethnic groups.[527] For example, the risk for type 2 diabetes increases at much lower levels of body fat in Asians than in Europids (White people of European origin).[528] The definition proposed by the International Diabetes Federation (IDF) in 2005 attempted to reconcile the differences in definitions and to produce a consensus definition that would be useful for identifying those at risk for developing cardiovascular disease in all populations, and also allow comparative long-term studies.[526] The IDF definition views central obesity (as defined by waist circumference) as an essential component of the metabolic syndrome, because of the strength of the evidence linking waist circumference with cardiovascular disease and the other components of the syndrome, and the strong likelihood that central obesity is an early step in the pathophysiologic cascade leading to full expression of the metabolic syndrome.[526] ***In general, the diagnosis of metabolic syndrome requires three of the following five clinical characteristics:***[529]

- **Increased waist circumference (population specific, >88 cm in the United States)**
- **Increased blood pressure (≥130 mm Hg systolic; ≥85 mm Hg diastolic)**
- **Increased triglycerides (≥150 mg/dL)**
- **Decreased HDL-cholesterol (<50 mg/dL)**
- **Increased fasting glucose (≥100 mg/dL) or previously established diabetes mellitus**

Our recognition of the central role of insulin resistance in the pathophysiology of PCOS and our knowledge of its potential longer-term health consequences have focused a great deal of attention on the benefits of insulin sensitizing mediations and other drugs aimed at reducing the risks for developing diabetes and cardiovascular disease.

Metformin is a biguanide oral insulin-sensitizing agent and currently is the most widely used drug in the world for the treatment of type 2 diabetes mellitus. Metformin decreases hepatic glucose production, decreases intestinal glucose uptake, increases peripheral insulin sensitivity, and also inhibits lipolysis, resulting in decreased circulating concentrations of free fatty acids, which further helps to reduce hepatic gluconeogenesis.[472, 530, 531] Metformin's mechanism of action is not entirely clear, but involves activation of the adenosine monophosphate-activated protein kinase pathway in the liver and skeletal muscle.[532–536]

Metformin is available in both a regular and a sustained release form that may be associated with fewer gastrointestinal side effects (nausea, vomiting, diarrhea, constipation, bloating, flatulence, heartburn, indigestion, unpleasant metallic taste). To improve tolerance and decrease side effects, it is generally recommended that metformin treatment begin with a low dose (250–500 mg daily), increasing gradually over an interval of 4–6 weeks until the desired dose is attained. The drug also can interfere with intestinal absorption of vitamin B12, so patients should be alerted to symptoms of vitamin B12 deficiency, which include numbness, paresthesia, macroglossia, memory loss, behavioral changes, and pernicious anemia. Lactic acidosis is a rare complication of metformin treatment, but for that reason, the drug should not be administered to those with renal insufficiency, liver disease, or alcohol abuse.[537]

A large number of trials have observed beneficial effects of metformin in women with PCOS; in most, the dose has ranged between 1,500 and 2,000 mg daily. ***In general, metformin treatment increases insulin sensitivity,***[402, 410, 481, 538–541] ***decreases weight and BMI,***[402, 406, 410, 542] ***and decreases blood pressure and LDL-cholesterol.***[401] A meta-analysis

of 31 trials concluded that metformin increases insulin sensitivity up to 20%, decreases weight and BMI by 3–5%, decreases fasting glucose by about 5%, and increases HDL-cholesterol and decreases triglycerides by approximately 10% in patients at increased risk for developing diabetes.[543] Insulin resistance improves during metformin treatment, no matter how severe, and in lean and overweight women with PCOS as well as in those who are obese.[410, 538, 539, 541] Weight loss enhances the effects of metformin.[410] Metformin appears to decrease levels of C-reactive protein and soluble vascular cellular adhesion molecules (sVCAM), which reflect the low level of chronic inflammation associated with insulin resistance.[506, 544, 545] Indirect evidence suggests metformin also may improve vascular endothelial function and coronary flow rate in women with PCOS.[546–549]

Thiazolidinediones are another type of insulin-sensitizing agent that has been used to improve insulin resistance in women with PCOS. They include rosiglitazone, pioglitazone, and, formerly, troglitazone (withdrawn from the market due to concerns about liver toxicity). Thiazolidinediones are synthetic agonists for the peroxisome proliferator-activated receptor gamma (PPARγ), which serves as a nuclear transcription factor in the regulation of genes involved in carbohydrate, lipid, and protein metabolism (free fatty acids and eicosanoids are the natural receptor ligands). In trials involving women with PCOS, treatment with troglitazone improved insulin sensitivity and glucose tolerance in a dose-dependent manner.[110, 458, 468] Similar observations have emerged from studies examining the effects of rosiglitazone and pioglitazone in women with PCOS.[481, 482, 550–553] However, overall experience with thiazolidinediones is quite limited and they have been associated with cardiac complications. Metformin improves insulin sensitivity as much or more than thiazolidinediones and currently remains the preferred insulin-sensitizing agent for women with PCOS.[472, 481, 539, 540]

Although the benefits of estrogen-progestin contraceptives in the treatment of women with PCOS are undisputed, they generally do not correct any of the metabolic abnormalities commonly observed in women with PCOS.[407, 554–558] Although preparations containing drospirenone may have some limited impact,[456] other evidence suggests that even estrogen-progestin contraceptives containing antiandrogenic progestins may aggravate an underlying chronic inflammatory state.[506, 508, 557] Not surprisingly, combination therapies aimed at more comprehensive treatment are now emerging, including estrogen-progestin contraceptives and metformin, and low doses of metformin (850 mg daily) and an antiandrogen (flutamide, 62.5 mg daily), with or without an estrogen-progestin contraceptive.[456] In women receiving an estrogen-progestin contraceptive, the addition of metformin improves insulin resistance and further reduces hyperandrogenism.[409, 444, 559] The combination of low doses of metformin (850 mg daily) and an antiandrogen (flutamide, 62.5 mg daily) improves body composition (loss of fat and gain of lean mass) and lipid levels and increases levels of adiponectin, a anti-inflammatory protein secreted from adipose that modulates glucose regulation and fatty acid metabolism.[332] Combined treatment with an estrogen-progestin contraceptive and an antiandrogen has similar effects that are further enhanced if metformin also is added to the treatment regimen.[508] ***Overall, these observations demonstrate that the spectrum of metabolic abnormalities that accompanies PCOS can be improved significantly by treatment with low doses of metformin and antiandrogen in adolescents, and by their addition to estrogen-progestin contraceptives in young women.***[456] At least in theory, alternative or adjunctive treatment with metformin and antiandrogens is attractive because it may improve or reverse "upstream" abnormalities and help to prevent their "downstream" consequences. ***Experience with these combination treatment regimens is still limited, but is growing steadily, suggesting they may soon find their way into clinical practice.***

Dyslipidemia is common in women with PCOS, with many having decreased HDL-cholesterol or increased total and LDL-cholesterol or triglycerides.[516] Because metformin treatment does not have any important or consistent impact on lipid levels,[408, 541] interest has turned to the potential benefits of treatment with statins. In the first clinical trial involving

women with PCOS, lipid profiles improved more in women randomized to treatment with an oral estrogen-progestin contraceptive and simvastatin (20 mg daily) than in those receiving only the contraceptive.[560] In those receiving simvastatin, markers of systemic inflammation and endothelial function also improved, and serum testosterone levels decreased to a significantly greater extent.[560, 561] In a placebo-controlled trial, treatment with atorvastatin resulted in a significant decrease in serum testosterone, C-reactive protein, and insulin resistance, and improved lipid profiles.[562] In a trial comparing the effects of simvastatin and metformin, the two drugs decreased testosterone and improved markers of systemic inflammation and endothelial function to a similar extent, but lipid profiles and insulin sensitivity improved only in those receiving simvastatin; results of combined treatment were not different from those of treatment with simvastatin alone.[563]

Statins exert their effects primarily by inhibiting 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase, the rate-limiting step in the mevalonate pathway leading to cholesterol synthesis.[564] The effects of statins on testosterone concentrations may relate to decreased availability of products of the mevalonate pathway (including cholesterol), to inhibition of the mitogen-activated protein kinase (MAPK) pathway that mediates the proliferative actions of insulin, or to other mechanisms regulating ovarian steroidogenesis.[563, 565] ***Statins may offer a promising new approach to the treatment of women with PCOS at risk for developing diabetes and cardiovascular disease. However, it is important to emphasize that statins may be teratogenic and are contraindicated in pregnancy.***[566]

Indications for Treatment with Metformin

The best overall approach to the treatment of chronic anovulation and PCOS currently is somewhat controversial. For decades, estrogen-progestin contraceptives have been standard therapy for women who are not immediately interested in seeking pregnancy, for good reasons. However, the central role of insulin resistance in PCOS and limited evidence suggesting that estrogen-progestin contraceptives may aggravate insulin resistance have raised concerns that such treatment might increase long-term risks for diabetes and heart disease in women already predisposed.[555]

There is no question that the majority of both lean and obese women with PCOS are insulin resistant,[22, 567–569] and that the prevalence of impaired glucose tolerance and diabetes is increased in women with PCOS.[99, 100, 502] In the lean patient, insulin resistance is intrinsic, but poorly understood,[97, 126, 570] and the obese patient carries an additional metabolic burden.[571] There also is no debate that insulin resistance and PCOS are associated with increased risk for developing hypertension,[157, 572, 573] dyslipidemia,[335, 520, 574, 575] and with a number of other surrogate markers and risk factors for heart disease.[333, 458, 505, 521, 525, 576–578] Understandably, some view PCOS as an early sign, or even as a component, of the metabolic syndrome in women.[555]

Some studies have observed that estrogen-progestin contraceptives decrease insulin sensitivity.[404, 440, 442, 579] Overall, the evidence suggests that the effects of estrogen-progestin contraceptives on insulin resistance and glucose tolerance vary with the dose of ethinyl estradiol, with the dose and type of progestin, and with phenotype, and in general, are not clinically important.[407, 409, 580] No large studies have examined the risk for developing type 2 diabetes in women with PCOS specifically. Studies in healthy women have observed a modest increase in the relative risk for past and current users of estrogen-progestin contraceptives, compared with never users, although the differences were not significant.[581, 582] ***Estrogen-progestin contraceptives may decrease insulin sensitivity or glucose tolerance, to some extent, in some women, but concerns that the risk may be substantially higher in women with PCOS and underlying insulin resistance are unsubstantiated.***

Similarly, no large studies have examined the effect of estrogen-progestin contraceptives on the risk for developing cardiovascular disease in women with PCOS specifically. Alterations in vascular and endothelial function have been described and increased death rates from cardiovascular disease have been observed in women with history of menstrual irregularity.[583] Case-control studies have observed that estrogen-progestin contraceptives are associated with an increased risk for myocardial infarction, but events are rare and risk is almost entirely limited to women with hypertension and those who smoke.[584–588] ***Concerns that use of estrogen-progestin contraceptives might pose greater risk for women with PCOS are understandable, but there is no convincing evidence that they do.***

Treatment with metformin might indeed decrease the risk for developing diabetes and heart disease in women with PCOS, but evidence for those benefits is indirect, inferred primarily from studies in patients with impaired glucose tolerance,[393] and from studies examining surrogate markers and risk factors for heart disease; conclusive evidence is lacking. The two most common and classical features of PCOS are anovulation and hyperandrogenism, and metformin has little impact on either. Metformin improves menstrual cyclicity and ovulatory function in some women with PCOS, but not in most, and standard treatments for anovulatory infertility are clearly more effective. Consequently, for the large majority of women with PCOS, metformin treatment alone will not suffice; treatment with estrogen-progestin contraceptives and antiandrogens, or with clomiphene citrate, will be required. ***The most clinically relevant question is who can benefit most from metformin treatment.***

The most logical candidates for treatment with metformin (aimed at preventing or slowing progression to type 2 diabetes and at reducing longer-term risks for cardiovascular disease) are women with impaired glucose tolerance or diabetes, those with obvious evidence of severe insulin resistance (acanthosis nigricans),[589] and women having other features of the metabolic syndrome, such as central obesity, hypertension, and dyslipidemia. ***All women with PCOS should therefore be screened with an oral glucose tolerance test at the time of presentation, and every 2 years thereafter, and those with impaired glucose tolerance warrant annual screening.[326] Evaluation also should include blood pressure, waist circumference, and a lipid profile, to help identify those with features of the metabolic syndrome.*** There is good evidence from the Diabetes Prevention Trial that metformin treatment can decrease the risk for progression to diabetes in those with impaired glucose intolerance, by approximately 30% (although better results were observed in those receiving intensive lifestyle interventions).[393] In a retrospective study of 50 women with PCOS treated with metformin (including 11 with impaired glucose tolerance at baseline), impaired glucose tolerance persisted in 5/11 (45%), and reverted to normal in the remainder (6/11, 55%), over an average of 43 months of follow-up.[590]

Anovulatory adolescent girls are another group that warrants periodic screening for glucose intolerance, and specific screening for insulin resistance, particularly if they are obese or had low birthweight.[591, 592] Both characteristics are associated with premature adrenarche and the development of PCOS during adolescence, and evidence indicates that hyperinsulinemia is a key pathogenic factor.[327–330] Although menstrual irregularity is common for a time after menarche, those in whom it persists for more than 2 years merit greater scrutiny. There is substantial evidence that early treatment with metformin can decrease hyperinsulinemia and hyperandrogenism, and restore ovulatory menstrual function in girls with demonstrable insulin resistance, at least in those who are not obese.[329, 331] Addition of a low dose of antiandrogen has additional beneficial effects on body composition and lipid levels.[332] ***These observations are compelling and indicate that metformin treatment, alone or in combination with antiandrogens, can have enormous impact and benefits for this important population.***

Although most women with PCOS have insulin resistance, at least 25% does not, no matter what method is used to assess insulin sensitivity.[92, 94, 310] ***Routine screening for insulin resistance is not recommended, primarily because there currently is no validated test***

for measuring insulin resistance in a clinical setting. The most accurate methods have no clinical application because of their complexity, and calculated indices are limited by the lack of a standardized insulin assay and any data demonstrating that markers of insulin resistance predict response to treatment. Routine treatment with metformin is difficult to justify for women with PCOS who do not have abnormal glucose tolerance, acanthosis nigricans, or the features of metabolic syndrome. However, continued surveillance and periodic screening is warranted, and recommended.

Conclusion

We clearly are in a new era in our understanding and management of women with PCOS. In the past, we treated the specific problems of infertility, dysfunctional uterine bleeding, and hirsutism effectively. We now have the opportunity, indeed the obligation, to offer interventions that can help prevent or reverse some of the metabolic consequences of the disorder that have an important impact on overall health and on the quality and quantity of life.

All references are available online at: http://www.clinicalgynendoandinfertility.com

13 Hirsutism

Hirsutism, defined as excessive male-pattern facial and body hair, affects between 5 and 10% of reproductive age women.[1] Hirsutism can be the initial or only sign of androgen excess and usually is a consequence of chronic anovulation. Virilization describes the signs and symptoms of more severe androgen excess, which include deepening of the voice, temporal balding (androgenic alopecia), breast atrophy, changes in body habitus, and clitoromegaly. Virilization is rare and most commonly results from congenital adrenal hyperplasia or androgen-producing tumors of the ovary or adrenal.

Hirsutism is both an endocrine and a cosmetic problem and deserves a concerned and sympathetic response. Excessive hair growth on the face, chest, or abdomen is understandably disturbing and raises a number of concerns and questions about the possibility of underlying disease, effects on sexuality and fertility, and available treatments.

This chapter reviews the biology of hair growth and the causes and pathophysiology of hirsutism, and presents a straightforward, effective approach to the diagnostic evaluation and clinical management.

The Biology of Hair Growth

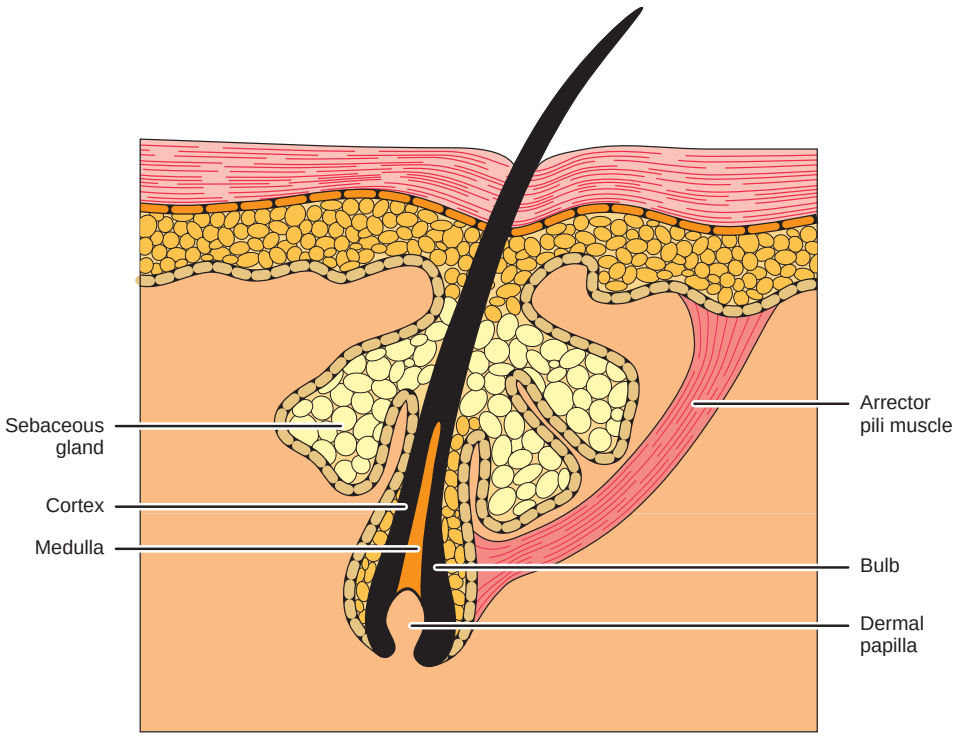
Hair is a distinguishing characteristic of mammals and serves a wide range of functions, including thermoregulation, physical protection, sensory activity, and social interactions.

Androgens are required for development of sexual hair and sebaceous glands, but numerous other factors are involved, including growth hormone, insulin, insulin-like growth factors, glucocorticoids, estrogen, and thyroid hormone.[2, 3]

Embryology

Hair follicles develop at approximately 8–10 weeks of gestation from a small group of epidermal cells overlying undifferentiated mesenchyme. Members of the transforming growth factor beta superfamily, activins and bone morphogenetic proteins in particular, play an important role in the communication between the epithelial and mesenchymal compartments during normal hair follicle development.[4, 5] Initially, the follicle is composed of a solid column of cells that proliferates from the basal layers of the epidermis and extends downward into the dermis. As the column elongates, it encounters a cluster of mesenchymal cells (the dermal papilla) that envelop its bulbous tip (bulb). The solid epithelial column then hollows to create a hair canal, and the ***pilosebaceous unit*** (a hair follicle, sebaceous gland, and arrector pili muscle) is formed. Hair color is determined by pigments produced by melanocytes located in the bulb.

One's total endowment of hair follicles is determined by 22 weeks of gestation and no new hair follicles develop de novo thereafter. The concentration of hair follicles in facial skin does not differ significantly between the sexes, but does differ between races and ethnic groups. Whereas Asian and Native American women generally have little body hair, women of Mediterranean descent typically have increased amounts of body hair, although serum androgen concentrations are similar in the three groups.[6] Differences in hair growth among races and ethnic groups probably also reflect differences in the local levels of 5α-reductase activity, the enzyme that converts testosterone to the more potent and active androgen, dihydrotestosterone (DHT).[7]



The Hair Growth Cycle

Hair growth is cyclic, rather than continuous, and exhibits three distinct phases, known as ***telogen*** (quiescent phase), ***anagen*** (growth phase), and ***catagen*** (involution phase).[2] In the resting phase (telogen), the hair is relatively short and loosely attached to the base (the bulb) of the epithelial canal. As growth (anagen) begins, cells in the epithelial matrix at the base of the hair follicle begin to proliferate, extending downward into the dermis in a column that elongates to approximately 4–6 times its length during telogen. With continued rapid growth, the epithelial column also pushes upward to the skin surface, breaking its tenuous contact with the previous hair, which is shed. The most superficial epithelial cells differentiate to form a keratinized column and growth continues as long as active mitosis persists in the basal epithelial cells. As the growth phase comes to a close, the column rapidly shrinks and the bulb shrivels (catagen) before the hair follicle again enters a quiescent phase (telogen).

The length of hair is determined primarily by the duration of the growth phase. Scalp hair remains in anagen for 2–5 years and spends only a relatively short time in telogen. Elsewhere, such as on the forearm, the hair cycle has a short anagen and a long telogen, yielding a short hair of relatively stable length. The outward appearance of continuous growth or periodic shedding reflects the extent to which hair follicles act in synchrony with others in the area. Typically, scalp hair is asynchronous and, therefore, always appears to be growing; the resting phase of some hairs (approximately 10–15%) is not apparent. If a larger proportion of hairs becomes synchronous and enters telogen simultaneously, noticeable shedding may occur, a process known as ***telogen effluvium***. Although women occasionally may notice and complain of scalp hair loss, the interval of shedding usually lasts no longer than 6–8 months. Growth resumes when asynchrony again becomes established. Telogen effluvium can be precipitated by pregnancy, certain drugs, and by febrile illness.

Hair is categorized as ***vellus*** (fine, soft, short, and unpigmented) or ***terminal*** (long, coarse, and pigmented).[3] The vellus hair that covers the body of infants is called ***lanugo.*** ***Hypertrichosis*** describes an uncommon condition characterized by a generalized increase in vellus body hair, usually associated with certain drugs (e.g., phenytoin, penicillamine, diazoxide, minoxidil, cyclosporin), systemic illness (e.g., hypothyroidism, anorexia nervosa, malnutrition, porphyria, dematomyositis) or malignancy (as a paraneoplastic syndrome). Hirsutism implies a transformation from vellus to terminal hair.

The Control of Hair Growth

The fate of a hair follicle depends on the health and function of the dermal papilla. Despite major injury to its epithelial component (e.g., freezing, x-rays, or a skin graft), the hair follicle will regenerate and re-grow hair if the dermal papilla survives intact. Serious injury or degeneration of the dermal papilla (e.g., electrolysis or laser hair removal) results in permanent hair loss.

Sexual hair is that which responds to sex steroids and grows primarily on the face, chest, lower abdomen, the pubis, and in the axillae. In androgen-sensitive areas, androgen stimulates hair follicles, inducing the growth of thicker, longer, and darker hairs. ***Thereafter, the hair exhibits typical cycles of growth, involution, and rest, but does not change in character, even if high androgen levels are not sustained.*** Because androgen stimulation of hair follicles requires the conversion of testosterone to DHT, the sensitivity of hair follicles to androgens is determined, in part, by the local level of 5α–reductase activity, helping to explain the varying extent of hirsutism observed in women with similar levels of androgen excess.[8]

Based on data from animal studies and on patterns of human disease, the following summarizes the effects of steroid hormones on hair growth:

1. Androgens, particularly testosterone, stimulate growth and increase the diameter and pigmentation of hair. Androgens also increase the proportion of time terminal hairs spend in anagen,[9] except on the scalp, where androgen decreases the duration of anagen.

2. Estrogens have actions opposite those of androgens, generally resulting in slower growth of finer and lighter hair.

3. Progestins have little or no direct effect on hair growth.

4. Pregnancy, characterized by high levels of both estrogen and progesterone, can induce greater synchrony among hair follicles, leading to periods of growth or shedding.

Observations in studies of the effects of male castration demonstrate an important clinical characteristic of hair growth. Males castrated before puberty do not grow a beard or other sexual hair, but when castrated after puberty is completed, the beard and sexual hair continue to grow, albeit more slowly and with finer caliber hair.

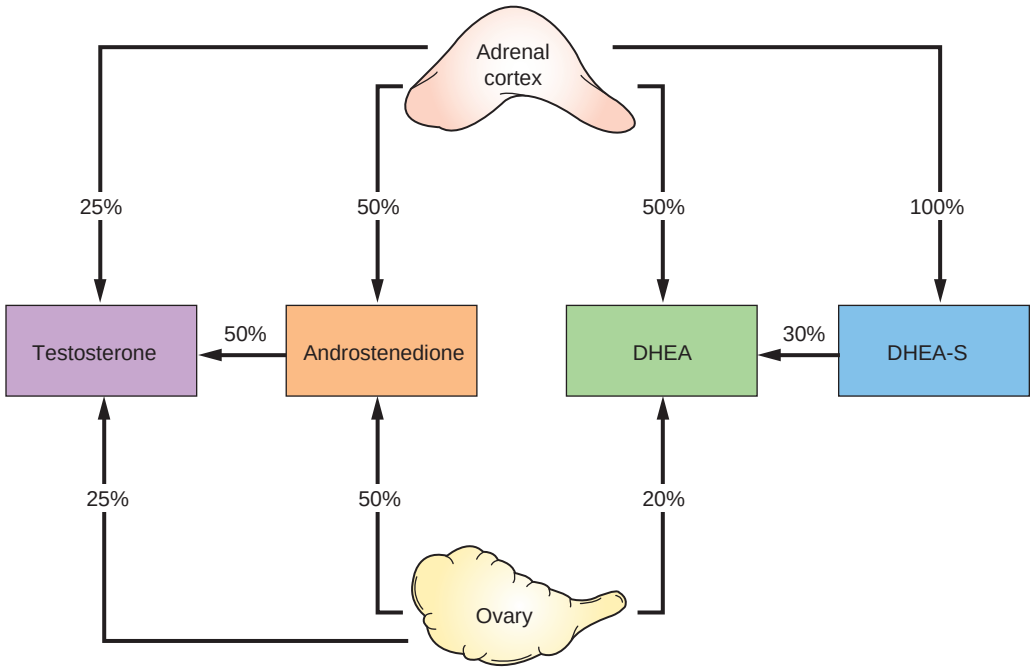
Endocrine disorders can affect the growth of both sexual and nonsexual hair. Hair growth is markedly reduced in individuals with hypopituitarism. Approximately 10–15% of patients with acromegaly also are hirsute. Hypothyroidism sometimes is associated with hair loss on the scalp, the pubis, in the axillae, and, curiously, the lateral third of the eyebrows. Hyperthyroidism generally results in finer hair that is lost easily. Insulin-like growth factor-1 (IGF-1), which stimulates 5α-reductase activity,[10] often is increased in women with chronic anovulation, insulin resistance, and hyperinsulinemia.

Hair growth also can be influenced by other factors, such as local skin temperature, blood flow, and edema. Hair grows faster in the summer than in the winter.[11] Hair growth also can be observed in association with central nervous system pathology (e.g., encephalitis, cranial trauma, multiple sclerosis), and with certain drugs.

Androgen Production

Hirsutism reflects the interaction between circulating androgen levels and the sensitivity of hair follicles to androgen stimulation. In women, the major circulating androgens (in descending order of serum concentration) are dehydroepiandrosterone sulfate (DHEA-S), dehydoepiandrosterone (DHEA), androstenedione, testosterone, and DHT.[12] ***DHEA-S, DHEA, and androstenedione can be considered pre-hormones because they have little or no intrinsic androgenic activity and require conversion to testosterone to exert androgenic effects.***

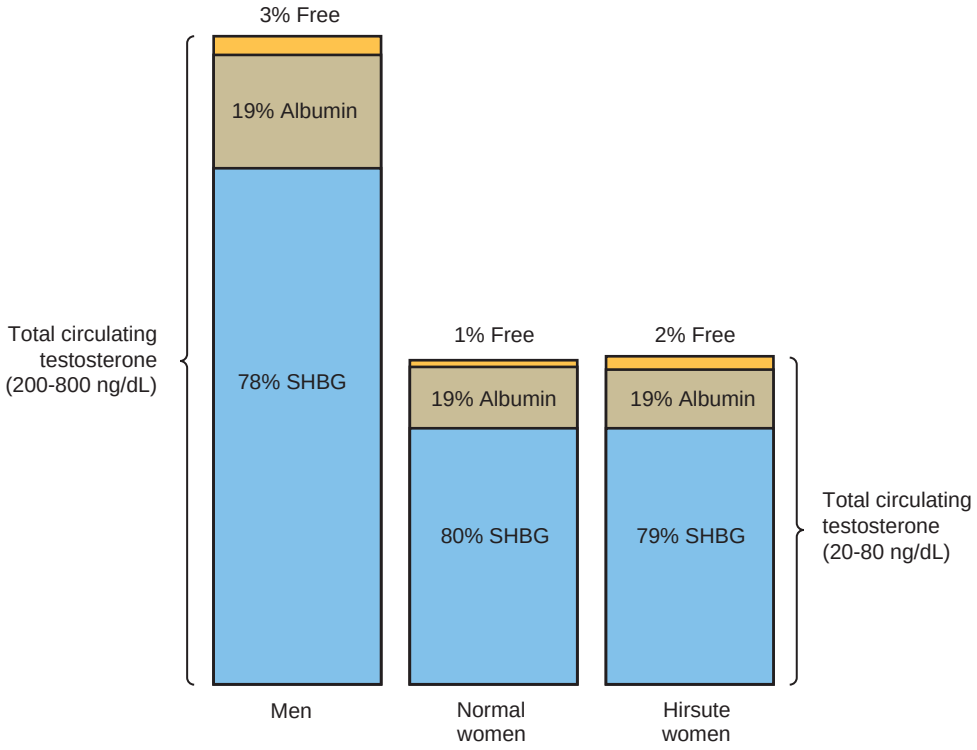
DHEA-S is produced almost exclusively by the adrenal glands, at a rate ranging between 3.5 and 20 mg/day;[13] the normal serum concentration is 100–350 μg/dL in most laboratories. DHEA is produced by both the adrenals (50%) and the ovaries (20%), and from the peripheral conversion of DHEA-S (30%). The production rate of DHEA is between 6 and 8 mg/day[14] and normal serum concentrations range between 1 and 10 ng/mL. Androstenedione production is divided equally between the ovaries and the adrenals; the production rate is



between 1.4 and 6.2 mg/day and the normal serum concentration is 0.5–2.0 ng/mL.[15, 16] Serum immunoassays for DHEA-S, DHEA, and androstenedione generally reflect the amount of biologically available hormone because none of the three is protein-bound to any significant extent.

Testosterone production derives from the adrenals (25%), the ovaries (25%), and from peripheral conversion of androstenedione (50%). The production rate ranges between 0.1 and 0.4 mg/day and the normal serum concentration is 20–80 ng/dL; levels do not fluctuate widely, but are lowest during the early follicular phase, and approximately 20% higher at midcycle.[14] In normal women, about 80% of circulating testosterone is bound to a beta globulin known as ***sex hormone-binding globulin (SHBG)***, another 19% is loosely bound to albumin, leaving only about 1% unbound or free. Routine serum immunoassays for testosterone measure the total testosterone concentration, including both bound and unbound hormone. However, the androgenic actions of testosterone relate primarily to the amount of free hormone and, to a limited extent, to the fraction associated with albumin. Anything that affects the SHBG concentration also affects the concentration of free/active testosterone.

Androgens themselves decrease SHBG production in the liver. Consequently, testosterone binding capacity in men is lower than in normal women; approximately 3% of total testosterone circulates in the free, active form in men. Whereas insulin and glucocorticoids also decrease SHBG levels, estrogens and thyroid hormone increase SHBG production. Therefore, binding capacity is increased in women with hyperthyroidism, in pregnancy, and during treatment with estrogens. In hirsute women, excess androgen production (and hyperinsulinemia, when present) depresses SHBG levels, increasing the amount of free/active testosterone to approximately 2%, even though the total testosterone level may remain within the normal range. ***Although specific assays to measure the level of free testosterone are available, they are costly and rarely necessary. The very presence of hirsutism or virilization indicates androgen excess. In hirsute women with "normal" serum total testosterone levels, decreased binding capacity and increased free testosterone can be assumed.***



In women with hirsutism, only about 25% of circulating testosterone arises from peripheral conversion, most coming from direct glandular secretion, with the ovary being the primary source of both increased testosterone and androstenedione.[17] By far, the most common cause of hirsutism is chronic anovulation and excess androgen production by the ovaries. Adrenal causes of hirsutism are very uncommon.

Although testosterone is the major circulating androgen, DHT is the major nuclear androgen in many androgen-sensitive tissues, including hair follicles and sebaceous glands. DHT is produced only in the periphery, by intracellular conversion of testosterone (via 5α-reductase). Circulating levels of DHT are, therefore, very low and do not reflect the level of 5α-reductase activity.[18] 3α-androstanediol is the peripheral tissue metabolite of DHT, and its glucuronide conjugate, 3α-androstanediol glucuronide (3α-AG), can be used as a marker of peripheral androgen metabolism.[19, 20] ***Serum 3α-AG levels correlate highly with levels of 5α-reductase activity in genital skin and are elevated almost uniformly in hirsute women,[21] including those with normal serum androgen levels, indicating that "idiopathic" hirsutism likely results from increased peripheral 5α-reductase activity. However, assays for serum 3α-AG have little clinical utility, primarily because results have no significant impact on the diagnosis and treatment of hirsutism.***

After the menopause, the production rate and serum concentration of androstenedione fall by about half, with approximately 80% derived from the adrenals.[22] Testosterone production and serum levels also decline, primarily due to the decrease in peripheral production, via the conversion of androstenedione.[23, 24] Ovarian testosterone production is largely maintained after the menopause, as demonstrated by the 40–50% decrease in serum testosterone levels after oophorectomy in postmenopausal women.[25, 26] Because the decrease in estrogen production far exceeds that in androgen production after menopause, the postmenopausal ovary is primarily an androgen-producing organ.[27] Elevated gonadotropin levels stimulate androgen synthesis in ovarian hilar and stromal cells.[28, 29] Adrenal androgen production also declines progressively with age; serum DHEA concentrations in women between ages 40 and 50 years are approximately half those in younger women.[30]

Causes of Hirsutism

The causes of hirsutism include specific endocrine disorders, such as androgen-secreting tumors, classical and nonclassical congenital adrenal hyperplasia (CAH), Cushing syndrome, and the hyperandrogenic insulin-resistant acanthosis nigricans (HAIR-AN) syndrome, as well as disorders of exclusion, including polycystic ovary syndrome (PCOS) and idiopathic hirsutism. In a case series of 873 women presenting with symptoms of androgen excess, the prevalence of these disorders was as follows:[31]

Diagnosis	*Number*	*Prevalence (%)*
Specific Disorders		
Androgen-secreting neoplasm	2	0.23
Classical congenital adrenal hyperplasia	6	0.69
Nonclassical congenital adrenal hyperplasia	18	2.06
HAIR-AN syndrome	33	3.78
Disorders of Exclusion		
Polycystic ovary syndrome	716	82.02
Idiopathic hirsutism	39	4.47
Hyperandrogenemia, hirsutism, and normal ovulation	59	6.75
Total	873	100.00

PCOS is by far the most common cause of androgen excess in women. The diagnostic criteria, clinical features, and treatment of PCOS are considered in depth in Chapter 12. The prevalence of PCOS among populations of hirsute women varies with differences in the diagnostic criteria proposed by the National Institutes of Health Conference on PCOS (NIH, 1990),[32] the European Society of Human Reproduction and Embryology/American Society for Reproductive Medicine Consensus Workshop Group (ESHRE/ASRM, 2003),[33] and the Androgen Excess Society (AES, 2009).[34] All include oligo/anovulation and clinical or biochemical evidence of hyperandrogenism as diagnostic criteria, two of the three (ESHRE/ASRM, AES) regard polycystic ovarian morphology (as defined by ultrasonographic imaging) as a criterion, and all three require exclusion of other specific diagnoses (e.g., congenital adrenal hyperplasia, Cushing syndrome, androgen-secreting tumors, and hyperprolactinemia). In three large case series of women presenting with symptoms of androgen excess, the prevalence of PCOS ranged between 57% and 82%.[31, 35, 36]

Although the "consensus" criteria all were developed in an effort to unify opinion and to standardize the diagnosis of PCOS, primarily for purposes of clinical investigation, ironically, they have created more clinical confusion and controversy. Many have objected to the ESHRE/ASRM criteria (also known as the Rotterdam criteria, where the conference was held) because they permit the diagnosis of PCOS in women with polycystic ovarian morphology, in the absence of hyperandrogenism. In contrast, both the NIH and the AES criteria *require* hyperandrogenism for diagnosis of PCOS. In this chapter, focused on the evaluation and treatment of hirsutism, the term PCOS describes only women with oligo/anovulation associated with hyperandrogenism and having no other specific diagnosis. ***In truth, there is no evidence that PCOS is a specific endocrine disorder having one unique cause. Rather, it is a common condition with features that develop as a direct consequence of chronic anovulation, which can result from a wide variety of causes.*** Whereas, in that context, the disorder might be more accurately described as "chronic anovulation with polycystic ovaries," the term PCOS is firmly entrenched in our scientific and clinical lexicon.

Hyperandrogenic insulin-resistant acanthosis nigricans (HAIR-AN) syndrome has the same clinical features as PCOS, but in the extreme. The primary underlying pathology is severe insulin resistance, with acanthosis nigricans being an epiphenomenon.[37, 38] A compensatory chronic and severe hyperinsulinemia stimulates a marked increase in ovarian androgen production, via theca cell receptors for insulin and insulin-like growth factor-1 (IGF-1), and induces a marked decrease in serum SHBG concentrations, yielding a large increase in free testosterone levels. In turn, high circulating androgen levels exacerbate the underlying insulin resistance, resulting in a self-propagating positive feedback loop that increases in severity over time, ultimately causing severe hirsutism and, in many, virilization (temporal balding, deepening of the voice, changes in body habitus, clitoromegaly). ***Ovarian stromal hyperthecosis*** is a histologic diagnosis, based on the observation of distinct clusters of luteinized thecal cells scattered throughout the ovarian stoma.[39] Patients with hyperthecosis typically are obese, severely hirsute, and often virilized, most having serum testosterone concentrations greater than 150 ng/dL and exhibiting severe insulin resistance and hyperinsulinemia.[40] It is likely that most, if not all, patients with the HAIR-AN syndrome have ovarian hyperthecosis, but hyperthecosis also can arise in postmenopausal women.[41–43]

Idiopathic hirsutism describes hirsute women with regular menstrual cycles and normal serum androgen levels.[8, 44, 45] Although some may have subtle forms of ovarian or adrenal enzymatic dysfunction,[46] an increased sensitivity to androgens, mediated by increased peripheral 5α-reductase activity,[21] is the most logical explanation. In affected women, normal circulating androgen levels stimulate hair growth. Many women previously assigned a diagnosis of idiopathic hirsutism would now be considered to have PCOS, according to some criteria.[33]

Congenital adrenal hyperplasia (CAH) is a specific but uncommon cause of hirsutism. The clinical features, diagnostic criteria, and treatment of CAH are discussed in detail in Chapter 10. Whereas females with classical CAH usually are recognized at birth or during early infancy, the nonclassical form of the disorder (also known as late-onset CAH) presents later, at or after puberty, with hirsutism and menstrual irregularity or amenorrhea. In different studies, the prevalence of nonclassical CAH has ranged between 1 and 15%.[47–50] The most common cause of both classical and nonclassical CAH is an adrenal 21-hydroxylase (P450c21) deficiency, resulting in excess production of 17α-hydroxyprogesterone (17OHP), which is the substrate for 21-hydroxylase in adrenal cortisol synthesis and a precursor for androgen synthesis (Chapter 9).

Androgen-secreting neoplasms of the ovary or adrenal are a rare cause of androgen excess and hirsutism. Androgen-secreting tumors account for only 5% of all ovarian tumors. Most are Sertoli-Leydig cell tumors, lipid- and theca-cell (stromal) tumors, or hilus-cell tumors, most are associated with frankly elevated serum testosterone concentrations greater than 150–200 ng/dL,[51–53] and most can be imaged by transvaginal ultrasonography. Although some adrenal adenomas secrete testosterone, most androgen-secreting adrenal tumors are carcinomas that secrete DHEA, DHEA-S and cortisol, in addition to testosterone.[54]

Some women with hirsutism also have mild ***hyperprolactinemia***. Elevated serum prolactin concentrations can be associated with increased serum DHEA-S levels,[55, 56] prolactin receptors have been identified in the human adrenal, and prolactin can increase adrenal DHEA production *in vitro*.[57] Although DHEA-S is a weak androgen, it can be converted in the periphery to testosterone and, in turn, to DHT. Hirsutism in women with hyperprolactinemia may result directly from prolactin stimulation of adrenal androgen production, but also could result from excessive ovarian androgen production due to chronic anovulation, caused by hyperprolactinemia.

Virilization during pregnancy should raise suspicion for a ***pregnancy luteoma***, which is a hyperplastic mass of luteinized ovarian cells rather than a true tumor. Although most luteomas produce little androgen or have little or no androgenic effect, serum concentrations

of androstenedione, testosterone, and dihydrotestosterone can be increased, sometimes dramatically;[58–60] only approximately one-third of reported pregnancy luteomas have been associated with maternal hirsutism or virilization,[58, 61] probably because any increase in serum free testosterone is limited by the large increase in sex hormone-binding globulin (SHBG) levels that occurs during pregnancy. Typically, luteomas are solid masses ranging between 6 and 10 cm in size; in approximately half of cases, they are bilateral.[59, 62] Pregnancy luteomas typically regress promptly after delivery, suggesting that human chorionic gonadotropin (hCG) plays a role in stimulating or perpetuating their androgen production,[63] although most are identified late in gestation, long after the peak in maternal serum hCG concentrations. In contrast, functional androgen-producing ***theca-lutein cysts (hyperreactio luteinalis)*** can develop in women with multiple pregnancies, isoimmunized or diabetic mothers, and those with molar pregnancies or gestational trophoblastic disease, all of which are associated with increased maternal serum hCG concentrations. Rarely, mothers with pre-existing hirsutism related to PCOS or ovarian stromal hyperthecosis also can develop theca-lutein cysts and become hirsute or virilize.[64–66] In those who do, serum concentrations of testosterone and androstenedione are elevated.

Evaluation of Women with Hirsutism

Accepting that the large majority of women with hirsutism have PCOS or idiopathic hirsutism, the evaluation of hirsute women is aimed at identifying the few having other causes that require additional specific evaluation and/or treatment. As always, evaluation should begin with a careful history and physical examination, which always provide important diagnostic clues. Laboratory investigation and imaging are used primarily to exclude other rare or potentially serious possibilities.

History and Physical Examination

The key elements of the medical history in women with hirsutism include the menstrual history and the age at onset and the rate of progression of hirsutism; the family and medication history also provide important information.

The ***menstrual history*** should include age at menarche, the regularity of menses, a characterization of premenstrual molimina, and information regarding any previous pregnancies and methods of contraception. Whereas women with PCOS typically report menstrual irregularity beginning at or soon after menarche, an abrupt departure from a previously established pattern of regular menses suggests another diagnosis. Although the clinical presentation of nonclassical CAH can closely resemble that of PCOS, hirsutism tends to be more severe in women with CAH.[67, 68] A later age at onset of hirsutism (after age 25) or rapid progression over a period of months suggests an androgen producing neoplasm. ***It is important to correlate changes in menstrual pattern with changes in weight and to remember that previous hormonal contraception could have obscured or delayed the onset of symptoms of menstrual dysfunction or androgen excess.*** Hirsutism in childhood usually is caused by classical CAH or an androgen-secreting tumor. Rare genetic causes of hirsutism, such as Y-chromosome mosaicism or incomplete androgen insensitivity, usually present with signs of androgen excess at puberty.

A ***family history*** of hirsutism, oligo/amenorrhea, obesity, and infertility are consistent with a familial predisposition to PCOS or, occasionally, nonclassical CAH, which is more common in women with Hispanic, Mediterranean, Slavic, or eastern European Jewish (Ashkenazi) heritage.[69] Drugs that can stimulate hair growth include methyltestosterone,

anabolic steroids (e.g., norethandrolone), phenytoin, diozoxide, danazol, cyclosporin, and minoxidil. DHEA or androstenedione, which are available as food supplements, can increase testosterone levels in women and cause hirsutism and acne, even at relatively low doses. The hair growth caused by medications, other than androgens, typically is diffuse and vellus in nature (hypertrichosis).

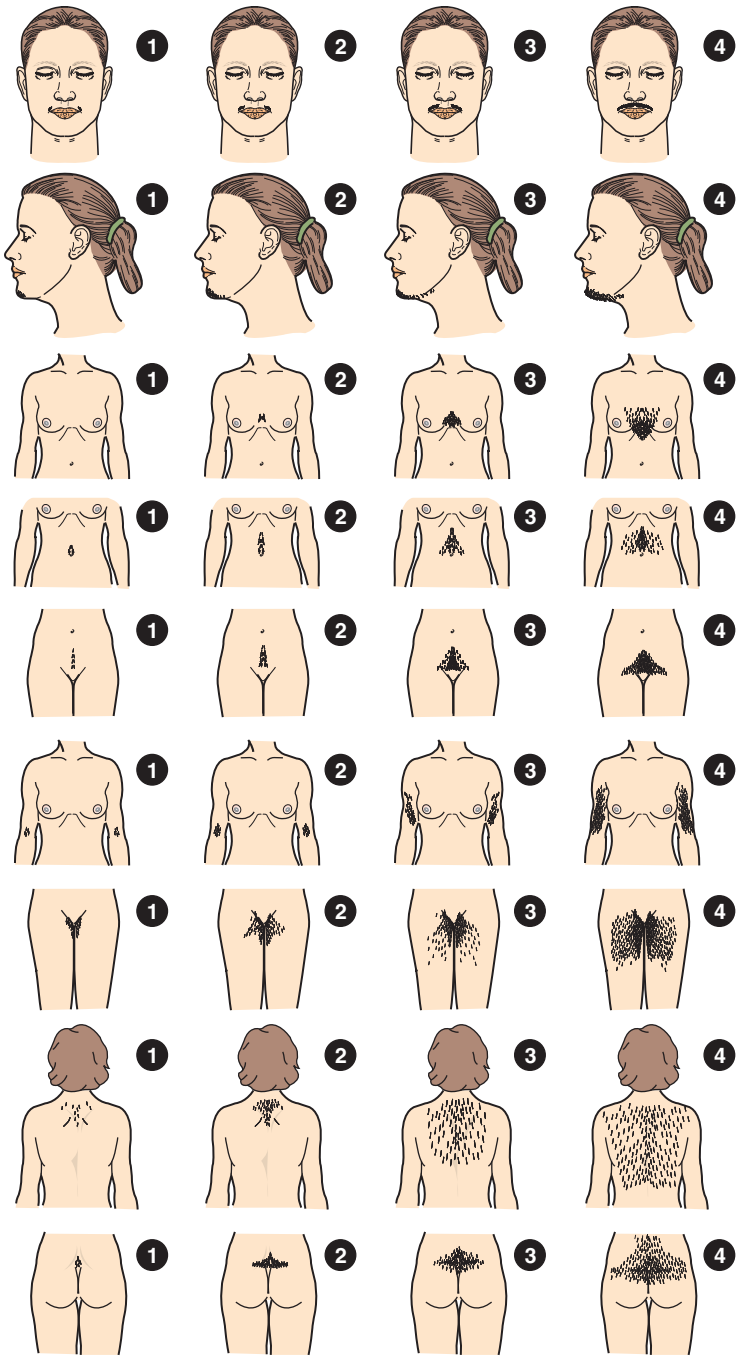
The ***physical examination*** should include a calculation of the body mass index (BMI) and document the distribution and extent of hirsutism. The modified ***Ferriman-Gallwey score*** is the most common method for grading the extent of hirsutism in clinical investigations.[70, 71] The method derives from studies in white women and scores hair growth from 0–4 in each of 9 androgen-sensitive areas, including the upper lip, chin, chest, upper and lower abdomen, upper arm, thighs, and the upper and lower back. Scores less than 8, 8–15, and greater than 15 generally indicate mild, moderate, and severe hirsutism, respectively. Approximately 95% of women have a modified Ferriman-Gallwey score less than 8. However, because the distribution of scores is not normally distributed and is skewed far to the left (half having a score of 0), scores of 3 or higher fall outside of the norm. Approximately 22% of women have scores of 3 or higher, 70% of which complain of hirsutism.[72] Notably, approximately 15% of women with scores less than 3 also consider themselves hirsute.

Overall, approximately 25% women use some sort of cosmetic treatment for excess hair, such as bleaching, plucking, shaving, waxing, or electrolysis; the frequency of self-treatment correlates positively with the Ferriman-Gallwey score. ***There are no significant differences between white and black women with regard to the distribution of scores or the proportions that complain of hirsutism or use some method of hair removal.***[72] Taken together, these data indicate that it is quite normal for most women to have at least some hair growth in androgen-sensitive areas and that a score of 8 or higher reflects significant androgen excess that warrants evaluation. ***Although the modified Ferriman-Gallwey score is the accepted standard for clinical investigations involving hirsute women, it is difficult to use clinically, primarily because most women who seek medical attention for the complaint already are using one or more methods of hair removal. Moreover, the score is unreliable for women in racial or ethnic groups having relatively little body hair; although less likely to develop hirsutism, they can exhibit other signs of androgen excess, such as acne and thinning or loss of hair. The easiest and most practical way to assess the severity of hirsutism is to determine the methods used to remove hair (e.g., shaving, plucking, waxing) and the frequency of their use, which also provides a clinically relevant measure for assessing the response to treatment.***

The physical examination also should note other relevant skin manifestations and any signs of virilization. Acne, seborrhea, and temporal balding are signs of androgen excess. Acanthosis nigricans (a gray or brown velvety discoloration of the skin, most commonly observed at the nape of the neck, the groin and axillae) indicates insulin resistance, and thin skin, striae, or bruising are signs of hypercortisolism. In addition to frontal or crown balding, signs of virilization include deepening of the voice, increased muscle mass, breast atrophy, and clitoromegaly. Clitoral size varies significantly among women[73]; in one study, the mean length of the glans clitoris was 5.1 ± 1.4 mm and the mean width was 3.4 ± 1.0 mm.[74] Clitoromegaly generally is defined by a clitoral length greater than 10 mm or by a clitoral index (length times width) greater than 35 mm[2,75] Other relevant physical findings include spontaneous or expressible galactorrhea, suggesting hyperprolactinemia, and abdominal or pelvic masses that may represent an androgen-secreting tumor. The large majority of functional ovarian tumors are palpable.

The physical manifestations of androgen excess generally reflect the extent to which androgen levels are elevated. Hirsutism is the most common complaint associated with androgen excess and essentially all women with hirsutism have an increased production rate of testosterone and androstendione.[76] Acne, increased libido, clitoromegaly, and virilization reflect progressively higher serum androgen levels.

Modified Ferriman-Gallwey Scoring System for Hirsutism



Alopecia can be a vexing problem for both patient and clinician. In many cases, alopecia is only temporary, resulting from telogen effluvium induced by some transient change that synchronizes a larger than normal proportion of scalp hair follicles, such as pregnancy or a febrile illness, and resolves after a period of 6–8 months. In a series of 109 consecutive women presenting with a complaint of diffuse alopecia, two-thirds had no clinical evidence of hirsutism or menstrual dysfunction, two had nonclassical CAH, and two had hyperprolactinemia associated with a pituitary adenoma.[77] Of the 42 (38.5%) that had elevated serum androgens, 11 were ovulatory and not hirsute, 13 were ovulatory and hirsute, and 18 had both

oligomenorrhea or amenorrhea and hirsutism.[77] Women with complaints of alopecia deserve evaluation for hyperandrogenism that can be treated successfully. Laboratory evaluation also should exclude thyroid disorders (serum TSH) and chronic illness. However, because alopecia can reflect increased scalp 5α-reductase activity, normal circulating androgen levels do not necessarily preclude effective treatment.[78, 79] Up to 60% of women with acne and normal serum androgen concentrations exhibit evidence of increased peripheral 5α-reductase activity and may benefit from treatment for hyperandrogenism.[80] Hair loss also is a normal consequence of aging, beginning in both sexes around the age of 50 years.[81]

Laboratory Evaluation

The relationship between hirsutism and circulating androgen concentrations is not entirely clear. Whereas some studies have found a correlation between hirsutism and androgen levels,[82] others have observed that only 50% of women with mild hirsutism have elevated free testosterone levels, and that 33% of women with moderately elevated free testosterone concentrations have no hirsutism, 40% have mild, and 27% have moderate hirsutism.[83] These observations suggest strongly that other factors, such as insulin and individual variations in androgen sensitivity, have substantial influence on the development and severity of hirsutism.

Laboratory evaluation is indicated for many but not all women with hirsutism.[84] The primary aim is to identify those having potentially serious endocrine disorders requiring specific treatment (nonclassical CAH, androgen-secreting tumors, Cushing syndrome). Thyroid disorders and hyperprolactinemia should be excluded in women with menstrual dysfunction. ***Laboratory evaluation is recommended for women with moderate or severe hirsutism, or hirsutism that is sudden in onset, rapidly progressive, or associated with symptoms or signs of virilization.***[85] ***Routine laboratory evaluation of women with mild hirsutism is neither necessary nor cost-effective.*** In women with oligo/amenorrhea, mild hirsutism can be attributed confidently to increased ovarian androgen production resulting from chronic anovulation. In women with regular menses, hirsutism most likely reflects an increased sensitivity to androgens relating to increased peripheral 5α–reductase activity.

The serum total testosterone concentration provides the best overall measure of androgen production and is the only hormone that need be measured in most women with hirsutism who merit evaluation. Testing for nonclassical CAH can be safely reserved for patients with an early onset of hirsutism (pre- or peri-menarcheal onset, including those with premature adrenarche), women with a family history of the disorder, and those in high-risk ethnic groups (Hispanic, Mediterranean, Slavic, or Ashkenazi Jewish heritage). Additional evaluation also is indicated for those with hirsutism having onset before puberty or after age 25, rapidly progressive hirsutism, or hirsutism that is accompanied by signs of virilization or hypercortisolism (Cushing syndrome).

The Serum Testosterone Concentration

Serum testosterone levels (normal 20–80 ng/dL) are elevated in most (70%), but not all, women with chronic anovulation and hirsutism. The total testosterone concentration can be normal in hirsute women because SHBG levels are depressed by androgen and insulin, thereby increasing the amount of unbound or free testosterone. Indeed, free testosterone levels are approximately twice normal (an increase from 1% to 2%) in women with PCOS.[86]

Laboratory testing for elevated androgen levels should begin with a serum total testosterone concentration. Total testosterone assays measure free testosterone, albumin-bound testosterone, and testosterone bound to SHBG. Although the free testosterone level is a more sensitive indicator of androgen excess, direct immunoassays of free testosterone are inaccurate,[87] yielding values only 20–60% of those measured by other, more precise methods.[88, 89] The best method for measuring the free testosterone level is equilibrium dialysis (a laborious, time-consuming, and costly method).[87] The free testosterone concentration also can be calculated using equations derived from the laws of mass action, knowing the serum total testosterone, SHBG, and albumin concentrations, and the association constants for the interactions of testosterone with SHBG and albumin.[90, 91] Calculated values generally correlate well with those determined by equilibrium dialysis, although accuracy varies with the specific assays used to measure total testosterone and SHBG. However, measurement or calculation of the free testosterone level generally is unnecessary, because the total testosterone level readily identifies women who may have an androgen-producing tumor.

A serum total testosterone concentration greater than 150 ng/dL identifies almost all women with a potential androgen-producing tumor.[35, 51, 52, 54, 92, 93] However, because serum testosterone concentrations can vary significantly in women with and without tumors,[52] a tumor still should be suspected, and excluded, in women with rapidly progressive hirsutism or signs or symptoms of virilization, even when the serum testosterone concentration is below the threshold value. Nearly all women with PCOS have a testosterone level less than 150 ng/dL, as do all women with idiopathic hirsutism, by definition. The suggested threshold value has very high sensitivity and negative predictive value, indicating that it captures virtually all women with tumors and can effectively exclude the diagnosis.[93] The positive predictive value of a serum total testosterone greater than 150 ng/dL is quite low, indicating that few women who meet the criterion will, in fact, have a tumor, primarily because such tumors are very rare; the large majority will have PCOS or hyperthecosis. Taken together, the clinical history (age at onset and rate of progression of hirsutism), physical examination (pelvic masses), and the serum total testosterone concentration will identify women with androgen-producing tumors.

It also is important to remember that testosterone levels are elevated significantly during normal pregnancy. Concentrations are greater than 100 ng/dL during the first trimester and can reach 500–800 ng/dL by term,[94] primarily due to the estrogen-induced increase in SHBG. Mother and fetus normally are protected from virilization because free testosterone levels rise only modestly and are rapidly converted to estrogen via placental aromatization. Because testosterone levels normally are lower in postmenopausal women, concentrations greater than 100 ng/dL should raise suspicion for a tumor.

The Serum DHEA-S Concentration

DHEA-S circulates in higher concentration than any other steroid and derives almost exclusively from the adrenal gland. It is, therefore, a direct measure of adrenal androgen activity. The upper limit of normal in most laboratories is approximately 350 μg/dL, but ranges vary among laboratories. DHEA-S serves primarily as a pre-hormone, providing substrate for conversion to testosterone and dihydrotestosterone in the periphery.[95]

Although the serum DHEA-S concentration would seem useful for identifying women with adrenal causes of hyperandrogenism, the test lacks both sensitivity and specificity for that purpose. DHEA-S levels frequently are not grossly elevated in women with nonclassical CAH or Cushing syndrome, and often are elevated in women with PCOS. Moreover, diagnosis of nonclassical CAH and Cushing syndrome requires other, more specific tests, as discussed below.

The serum DHEA-S concentration is moderately elevated in over half of women with PCOS.[96] The reasons remain unclear, despite extensive investigation. Some have argued that the increase in DHEA-S levels results from a 3β-hydroxysteroid dehydrogenase deficiency, induced by chronic anovulation and estrogen stimulation, similar to the mechanism that operates in the fetal adrenal cortex.[97, 98] Whereas there are data to support such a mechanism,[99–103] evidence is conflicting.[104–108] Notably, ACTH levels are not elevated in women with PCOS,[109, 110] exaggerated adrenal androgen secretion cannot be attributed to any increase in sensitivity to ACTH,[96] and ovarian (and estrogen) suppression by treatment with a long-acting gonadotropin-releasing hormone (GnRH) agonist has no consistent effect on DHEA-S levels in women with PCOS.[111–114] Although increased adrenal P450c17 17,20 lyase activity could cause adrenal androgen excess,[114–116] the patterns of steroidogenic response to ACTH stimulation in women with PCOS do not support the hypothesis.[117] The prevalence of adrenal androgen excess is comparable to that of insulin resistance among women with PCOS, suggesting that hyperinsulinemia might be the cause of increased adrenal androgen production.[6] However, insulin infusion studies indicate that insulin does not stimulate, and actually impairs, 17,20 lyase activity in both normal and hyperandrogenic women.[118–120] In sum, no one mechanism explains the moderate adrenal androgen excess commonly observed in women with PCOS.

The serum DHEA-S concentration can be grossly elevated (≥700 μg/dL) in women with rare androgen-secreting adrenal tumors. However, in almost all such patients, serum testosterone levels also are greatly elevated,[121] via peripheral conversion of high circulating DHEA-S levels, or because the tumor also secretes testosterone. A serum DHEA-S concentration may be useful in women whose clinical presentation suggests strongly the possibility of a tumor, but the test otherwise has little or no clinical utility in the evaluation of hirsutism.

Evaluation for a Suspected Androgen-Producing Tumor

When the serum total testosterone concentration (≥ 150 ng/dL) or the clinical presentation suggests the possibility of a rare androgen-producing ovarian or adrenal tumor (rapidly progressive hirsutism or symptoms or signs of virilization), evaluation is indicated to exclude the diagnosis or localize the lesion.

Androgen-secreting tumors of the ovary (in decreasing order of prevalence) include Sertoli-Leydig tumors, lipid-cell tumors, hilar-cell tumors, and rare androgen-producing theca-cell and Brenner tumors; virtually all are associated with grossly elevated serum total testosterone levels (≥ 150 ng/dL). Occasionally, virilization results from a nonfunctioning tumor due to stimulation of the surrounding stroma.[122] Most functioning ovarian tumors are palpable on pelvic examination, but small tumors easily can go unrecognized. ***Transvaginal ultrasonography*** can identify ovarian follicles and cysts as small as 3–5 mm in diameter and almost all solid ovarian mass lesions, although very small tumors located in the hilar region still can escape detection.

Adrenal CT imaging is extremely sensitive for detecting the rare androgen-producing adrenal adenoma or carcinoma, when pelvic examination and transvaginal ultrasonography fail to reveal an ovarian tumor.[123] Most androgen-secreting adrenal tumors are malignant.[124–126] Adrenal adenomas typically are smaller (<4 cm in diameter) than carcinomas and have smooth borders and characteristically low unenhanced CT attenuation values; irregular margins, necrosis, hemorrhage, or calcification suggest a carcinoma.[127] When needed, additional information to help define the nature of an adrenal mass lesion can be obtained

by magnetic resonance imaging (MRI), by functional nuclear imaging with scintigraphy using a labeled cholesterol analog (^{131}I-6-iodomethyl norcholesterol),[128] or positron emission tomography (PET) scanning.

Findings of bilateral disease require further evaluation to distinguish among the causes, which include metastatic cancer (most commonly from breast, kidney, or lung), adrenal hyperplasia (caused by long-term stimulation by pituitary or ectopic sources of ACTH and by rare forms of ACTH-independent macro- and micronodular disease),[129, 130] infection (tuberculosis and fungal), hemorrhage, pheochromocytoma, and amyloidosis.

Routine adrenal imaging is not recommended and can be misleading, because nonfunctioning adrenal masses (incidentalomas) are common and their incidental detection demands additional, otherwise unnecessary, evaluation.[131] In autopsy studies, the prevalence of incidental adrenal adenomas approaches 10%.[132, 133] In 2 case series of patients having abdominal CT scanning for a variety of indications, the prevalence of adrenal incidentaloma was 3–4%.[134, 135]

Incidentally detected adrenal masses require evaluation to determine whether they are functional; up to 15% secrete excess hormones, such as cortisol, catecholamines, and aldosterone.[136] The requisite tests include a 24-hour urine collection for fractionated metanephrines and catecholamines (pheochromocytoma), blood samples for fractionated metanephrines, testosterone and DHEA-S (adrenal carcinoma), plasma aldosterone and rennin activity (primary aldosteronism), and an overnight dexamethasone suppression test (Cushing syndrome).[131] Fine-needle aspiration (FNA) biopsy may be indicated when there is reason to suspect a malignancy outside of the adrenal gland or in patients undergoing staging evaluation for a known cancer.[133, 137] Although FNA is a relatively safe procedure, potential complications include adrenal hematoma and abscess, abdominal pain, hematuria, pancreatitis, and pneumothorax.[138, 139] Because inadvertent FNA of a pheochromocytoma can precipitate an acute hypertensive crisis, the diagnosis always should be excluded by biochemical tests before FNA is performed.[140] When testing detects no evidence of hormone function and there is no reason to suspect a cancer, expectant management is appropriate; current recommendations include repeated imaging after 6, 12, and 24 months (to detect evidence of progressive growth), and repeated endocrine evaluation annually (to detect autonomous function not identified at baseline) for at least 4 years.[131, 141, 142]

Selective ovarian venous catheterization can be considered for the rare patient having no demonstrable ovarian or adrenal mass lesion.[53, 92, 143–145] ***However, because the overall clinical utility of the technique is still uncertain,[146] the procedure should be reserved only for those in whom a tumor is strongly suspected.*** An analysis of results obtained in 136 reported patients with hirsutism who had selective ovarian venous sampling yielded a number of important observations.[146] A right:left ovarian venous effluent testosterone ratio greater than 1.44 correctly identified 90% of right-sided tumors, and lower values correctly identified 86% of women with left-sided or bilateral lesions. In three women with a left-sided tumor, the left:right testosterone ratio was greater than 15. The differing anatomy of the right (draining into the vena cava) and left ovarian vein (draining into the left renal vein) and the related technical difficulty of catheterization might explain why venous sampling was more effective for identifying right-sided tumors.

When suspicion for a tumor is insufficient to warrant ovarian venous catheterization, or the procedure reveals no significant gradient in testosterone concentrations, the likelihood of an occult ovarian tumor is very small, leaving the HAIR-AN syndrome or stromal hyperthecosis as the most likely cause of severe hyperandrogenism. Both are associated with severe insulin resistance, which can be documented by performing an ***oral glucose tolerance test including insulin levels***, as discussed below. Rarely, open surgical exploration of

the ovaries may be necessary to establish a diagnosis; laparoscopic inspection and biopsy are not sufficient.

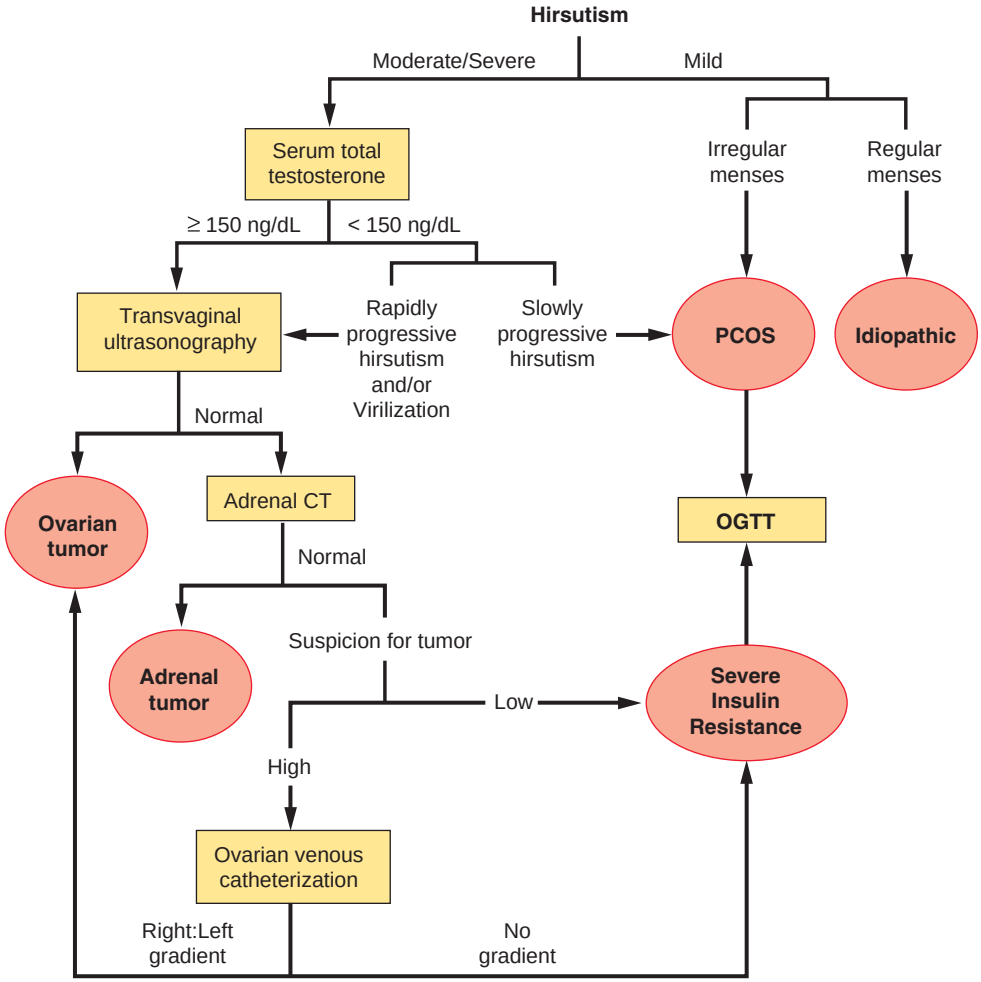
Dynamic endocrine evaluation, using dexamethasone, contraceptive steroids, or a GnRH agonist in attempts to isolate adrenal or ovarian androgen production, is not recommended, because results are unreliable and can be misleading.[122, 147–149] ***Ovarian androgen-secreting tumors are sensitive to LH stimulation and thus respond to ovarian suppression and stimulation.***[150–152]

Insulin Resistance

Insulin resistance is a common feature of women with PCOS and a key component of the HAIR-AN syndrome and ovarian stromal hyperthecosis. ***Although high circulating androgen concentrations decrease insulin sensitivity, the primary pathology in women with HAIR-AN and hyperthecosis is severe insulin resistance, resulting in grossly elevated insulin levels that stimulate ovarian androgen production in theca cells (via insulin, IGF-1, and hybrid receptors) and markedly decrease SHBG production, thereby greatly increasing the amount of free androgen.*** Insulin resistance and hyperinsulinemia also explain the occasional elderly woman who presents with severe progressive hirsutism. The problem does not reflect an ovarian response to elevated gonadotropin levels, but the development of hyperinsulinemia and hyperthecosis. Insulin appears to have a direct effect on the severity of hirsutism and a synergistic interaction with testosterone.[153]

The numerous methods that can be used to assess insulin sensitivity are discussed in the chapter devoted to chronic anovulation and PCOS (Chapter 12) and are summarized only briefly here. The hyperinsulinemic euglycemic clamp technique is the gold standard method for measuring insulin sensitivity but has no practical clinical application because it is time-consuming, labor-intensive, invasive, costly, and requires experienced personnel.[154] A number of less complicated, inexpensive, "homeostatic" measures have been described, all based on the fasting glucose and insulin concentrations and using straightforward calculations.[155] These include the fasting insulin concentration,[156] the fasting glucose/insulin ratio,[157] the homeostatic model assessment (HOMA),[158, 159] the quantitative insulin sensitivity check index (QUICKI),[160] and others. ***As the sheer number of different measures of insulin resistance illustrates, there is currently no uniformly accepted, validated, simple test for measuring insulin resistance in clinical practice.*** All of the measures have limitations, primarily the lack of a standardized insulin assay,[161] and the absence of data indicating that such measures can predict the response to treatment. Moreover, treatment with insulin-sensitizing agents has no important benefits for the treatment of hirsutism.[162] ***Consequently, routine assessment of insulin sensitivity in the evaluation of hirsutism is not recommended.***

A baseline 2-hour oral glucose tolerance test (75-g glucose load) is recommended for all women with PCOS,[33, 163, 164] because up to 35% exhibit impaired glucose tolerance (glucose 140–199 mg/dL) and up to 10% have non-insulin-dependent diabetes mellitus (glucose ≥ 200 mg/dL).[165] In patients with severe hyperandrogenism having no evidence of an androgen-secreting tumor, the corresponding fasting and 2-hour insulin concentrations can be used to document the degree of insulin resistance, in support of the diagnosis of HAIR-AN syndrome or hyperthecosis; most have grossly elevated insulin levels.[40] The 2-hour glucose/insulin ratio (mg/dL/ μU/mL) provides an estimate of insulin sensitivity, with values less than 1.0 indicating insulin resistance. Plasma insulin concentrations that exceed an upper limit of normal or a defined threshold value (e.g., a 2-hour plasma insulin > 100 μU/mL) also have been used as a qualitative test for insulin resistance.



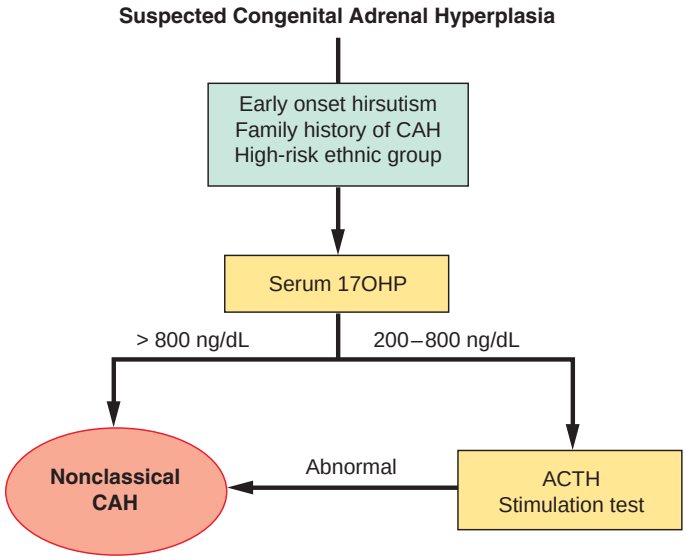
Nonclassical Congenital Adrenal Hyperplasia

Congenital adrenal hyperplasia (CAH) is caused by adrenal steroidogenic enzyme defects that result in excessive adrenal androgen production. By far, the most common cause is 21-hydroxylase deficiency; other enzyme defects (e.g., 11β-hydroxylase, 3β-hydroxysteroid dehydrogenase) are relatively rare. In all, the pathophysiology relates primarily to decreased cortisol production, which stimulates a compensatory increase in pituitary ACTH secretion, causing adrenal hyperplasia; increased levels of steroid hormones proximal to the enzyme block seek an alternative metabolic pathway, resulting in increased production of androgens. The disorder is inherited in an autosomal recessive fashion and is discussed in detail in Chapters 9 and 10.

Females with classical CAH (both salt-wasting and simple virilizing forms) present at birth with ambiguous genitalia (adrenogenital syndrome).[166–168] ***Those with the nonclassical ("late-onset") form of CAH have normal external genitalia and present later, during childhood or early adolescence, with precocious puberty or as young adults with other signs of hyperandrogenism, such as acne, hirsutism, and menstrual irregularity, very much like those with PCOS.***

Whereas it would seem that nonclassical CAH should be excluded specifically in all women with hirsutism, the yield from routine testing is quite low, because the disorder is

uncommon.[31, 36, 169] In the United States, the prevalence of the disorder among white women who present with hirsutism is between 1% and 4%.[170] ***Therefore, specific testing for nonclassical CAH can be reserved for those having an early onset of hirsutism (pre- or perimenarcheal, including girls with premature adrenarche), women with a family history of the disorder, and those in high-risk ethnic groups (Hispanic, Mediterranean, Slavic, or Ashkenazi Jewish heritage).***

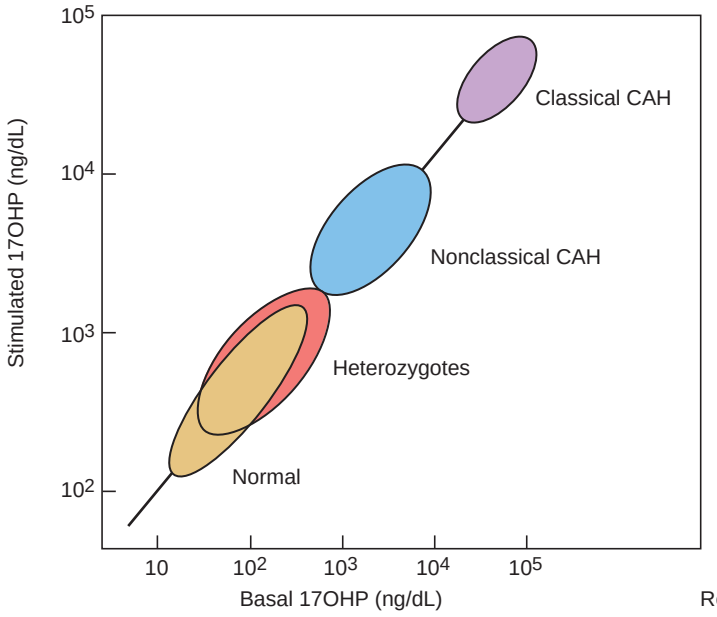


The Serum 17α-Hydroxyprogesterone Concentration

In both normal women and those with nonclassical CAH, the normal diurnal pattern of pituitary ACTH secretion is reflected in serum 17α-hydroxyprogesterone (17OHP) concentrations, which peak in the morning and nadir late in the day. Morning serum 17OHP levels during the follicular phase of the menstrual cycle are clearly higher in women with nonclassical CAH than in normal women, whereas evening concentrations overlap significantly. ***A follicular-phase morning serum 17OHP concentration less than 200 ng/dL effectively excludes the diagnosis of nonclassical CAH.***[67, 171, 172] ***Levels greater than 800 ng/dL are virtually diagnostic for 21-hydroxylase deficiency and concentrations between 200 and 800 ng/dL strongly suggest the diagnosis, which should be confirmed by performing an ACTH stimulation test.***

The ACTH Stimulation Test

The ACTH stimulation test is performed by obtaining blood samples before and 60 minutes after administering cosyntropin (synthetic ACTH 1–24; 0.25 mg). ***In most affected women, the response to ACTH stimulation is exaggerated and the 17OHP level rises above 1,500 ng/dL.***[169, 173, 174] Heterozygote carriers exhibit lesser responses to ACTH stimulation that overlap with those of normal subjects.[175, 176]



Redrawn with permission of M. New

Cushing Syndrome

Some women with hirsutism have symptoms and signs of hypercortisolism or Cushing syndrome, which vary with the duration and extent of excess cortisol secretion. In addition to hirsutism, the classical features of Cushing syndrome include progressive central obesity, excess fat accumulation in the cheeks ("moon face") or at the back of the neck ("buffalo hump"), severe fatigue and muscle weakness, hypertension, atrophy of the skin and subcutaneous tissue (easy bruising and purple striae on the abdomen and flanks), hyperpigmentation (caused by excess secretion of α-melanocyte-stimulating hormone, as a byproduct of ACTH synthesis from pro-opiomelanocortin, the common precursor molecule) in areas most exposed to light (the face, neck, and back of the hands) or chronic mild trauma, friction, or pressure (the elbows, knees, knuckles, and shoulders), diabetes, cognitive impairment, and menstrual disorders.

Cushing syndrome can be caused by: (1) ingestion of prescribed glucocorticoids, (2) a pituitary ACTH-secreting corticotroph adenoma (Cushing disease, accounting for most cases), (3) cortisol-secreting adrenal adenomas and carcinomas, or (4) ectopic corticotrophin-releasing hormone (CRH) or ACTH secretion by bronchial carcinoids and other rare tumors. Whereas the ingestion of prescribed glucocorticoids (oral, rectal, inhaled, topical, or injected) is the most common cause of Cushing syndrome overall, most patients with Cushing syndrome relating to treatment with glucocorticoids are not hirsute. ***Nonetheless, the first step in the evaluation of suspected Cushing syndrome is to exclude exogenous glucocorticoid exposure.***

Diagnosis of Cushing Syndrome

When hirsutism is accompanied by distinct symptoms and signs of hypercortisolism, screening for Cushing syndrome is indicated. However, because Cushing syndrome is rare, the risk of false-positive diagnostic tests is high. Widespread screening of overweight and obese women has a negligible yield and results in false-positive tests and needless anxiety.[177] Therefore, screening is best limited to individuals having a relatively high pretest probability of having the disorder. The currently recommended testing strategy is aimed at reducing the

number of false-positive tests while using threshold values with high sensitivity to minimize false-negative results, and emphasizes the most convenient and least costly tests.[178]

There are three methods of screening for Cushing syndrome that have comparable diagnostic accuracy:[179] the 24-hour urinary free cortisol excretion (measured twice), the late-night (11:00 P.M.–midnight) salivary cortisol level (measured twice), and the overnight dexamethasone suppression test. ***Whereas all clinicians who care for patients with hirsutism should be able to identify individuals having features of Cushing syndrome who merit screening, those unfamiliar with the disorder and its causes should consult with a medical or reproductive endocrinologist when screening yields an abnormal or equivocal result.***

The ***24-hour urinary free cortisol excretion*** provides a direct and reliable integrated measure of the serum free cortisol concentration.[180, 181] Because a reliable 24-hour urine collection can be difficult to obtain, creatinine excretion should be measured in the same specimen to judge compliance with instructions. A systematic review and meta-analysis of diagnostic tests for Cushing syndrome reported a likelihood ratio of 10.6 (95% CI 5.5–20.5) for an abnormal result, and 0.16 (95% CI 0.08–0.33) for a normal result (reflecting how the odds of the disease increase when the test is abnormal, and decrease when the test is normal).[179] ***Whereas values for 24-hour urinary free cortisol excretion greater than 3 times the upper normal limit are clearly abnormal, lower abnormal values are equivocal and more likely to represent a false positive result.***[182] Patients must be instructed to avoid excessive fluid intake and the use of any products containing glucocorticoids during the collection. Because the hypercortisolism of Cushing syndrome can wax and wane, the test should be performed at least twice before judging the result.

The ***late-night salivary cortisol level*** is based on the fact that whereas serum cortisol levels normally peak at 7:00–9:00 A.M. and fall throughout the day to very low levels late at night,[183] the normal circadian rhythm is lost in patients with Cushing syndrome.[184, 185] Moreover, because free cortisol in the blood is in equilibrium with cortisol in saliva,[186] the late-night salivary cortisol (11:00 P.M.–midnight) level can be used to establish a diagnosis of Cushing syndrome.[187–191] Saliva is easy to collect by the patient at home, by passive drooling or after chewing on a cotton pledget for 1–2 minutes. Specimens are stable at room or refrigerator temperature for several weeks.[178] In contrast to salivary sex steroid hormone assays, salivary cortisol assays yield reliable results. Results are interpreted by comparison to established normal ranges, which vary among studies, probably due to differences in assays. The meta-analysis of diagnostic tests for Cushing syndrome reported a likelihood ratio of 9.5 (95% CI 1.7–54.1) for an abnormal result, and 0.09 (95% CI 0.08–0.33) for a normal result.[179] Again, because the hypercortisolism of Cushing syndrome can fluctuate, the late-night salivary cortisol level should be performed at least twice. Patients should be advised to avoid using licorice, chewing tobacco, and smoking, which may falsely elevate salivary cortisol levels. The method also probably is not the best choice for shift workers and those with bedtimes well after midnight.

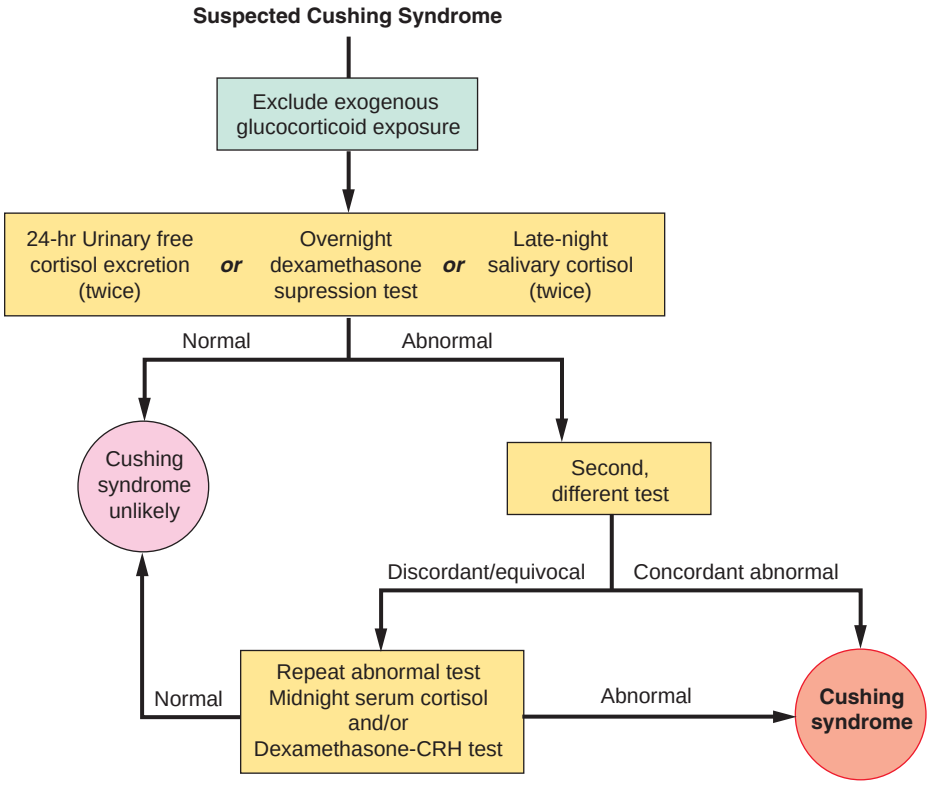
The ***overnight dexamethasone suppression test*** is based on the concept that dexamethasone (a potent synthetic glucocorticoid) should suppress ACTH secretion by the normal pituitary gland, thereby also suppressing cortisol secretion. The test is performed by administering 1.0 mg of dexamethasone between 11:00 P.M. and midnight and measuring the serum cortisol at 8:00 a.m. the next morning; values less than 1.8 μg/dL are normal.[178] The aforementioned systematic review and meta-analysis reported a likelihood ratio of 16.4 (95% CI 9.3–28.8) for an abnormal result, and 0.06 (95% CI 0.03–0.14) for a normal result.[179] A 2-day low-dose dexamethasone suppression test also can be used (0.5 mg dexamethasone every 6 hours for a total of 8 doses beginning at 9:00 A.M. on day 1, with measurement of serum cortisol 6 hours after the last dose), applying the same criteria for normal suppression (<1.8 μg/dL). However, the 2-day test has lower diagnostic accuracy than the overnight test, with a likelihood ratio of 7.3 (95% CI 3.6–15.2) for an abnormal result, and 0.18 (95% CI 0.06–0.52) for a normal result.[179] ***Given its simplicity and ability to discriminate, the overnight dexamethasone suppression test is the best first test to perform when***

Cushing syndrome is suspected in patients with hirsutism. Because estrogens increase the cortisol-binding globulin concentration and serum assays measure total cortisol, false-positive overnight dexamethasone test results are common in women using oral contraceptives.[192] Therefore, whenever possible, estrogen-containing drugs should be discontinued for 6 weeks before testing or retesting.[193]

If the first screening test is normal, Cushing syndrome is excluded and no further testing is necessary, unless clinical suspicion (based on clinical presentation) is strong; re-evaluation in 6 months also is indicated if symptoms or signs of Cushing syndrome progress. If the first test yields an abnormal result, a second, different, test should be performed. The diagnosis of Cushing syndrome is confirmed when two different tests are unequivocally abnormal. Patients with discordant or equivocal test results require further evaluation.[178]

When necessary, additional evaluation should begin by repeating the abnormal test. Expanded evaluation also should include a midnight serum cortisol concentration and/or the dexamethasone-CRH test.

The ***midnight serum cortisol concentration*** has the same rationale as the late-night salivary cortisol level. Although the test is difficult to perform, it can be useful when 24-hour urinary free cortisol excretion or the overnight dexamethasone suppression test is normal but clinical suspicion for Cushing syndrome is high. A sleeping midnight serum cortisol greater than 1.8 μg/dL, or an awake midnight value greater than 7.5 μg/dL increases the probability of Cushing syndrome.[194, 195] Conversely, when the clinical suspicion for Cushing syndrome is low, as in simple obesity, but 24-hour urinary free cortisol excretion or the overnight dexamethasone suppression test is mildly abnormal, a sleeping midnight serum cortisol less than 1.8 μg/dL or awake value less than 7.5 μg/dL effectively excludes Cushing syndrome.[195] The test has similar utility in patients receiving anticonvulsant medications, which can accelerate metabolism of dexamethasone, causing a false-positive overnight dexamethasone suppression test.[196]



The ***dexamethasone-CRH test*** helps to differentiate patients with Cushing syndrome from those whose hypercortisolism relates to physical or psychological stress or depression (pseudo-Cushing syndrome).[182] The test is performed by administering dexamethasone 0.5 mg every 6 hours over 2 days for a total of 8 doses, administering CRH (1 μg/kg, intravenously) 2 hours after the last dose of dexamethasone, and measuring the serum cortisol 15 minutes later; in patients with pseudo-Cushing syndrome, values generally are less than 1.4 μg/dL.

Establishing the Cause of Cushing Syndrome

Once the diagnosis of Cushing syndrome is made, its cause must be determined. The tests used for diagnosis of Cushing syndrome cannot distinguish between pituitary, adrenal, and other causes of the disorder, because normal pituitary corticotrophs behave and respond in much the same way as a corticotroph adenoma. ***The evaluation required to establish the cause of Cushing syndrome can be challenging and complicated and is best conducted by an endocrinologist having the necessary training and experience.***

The first step in the process is to measure the ***plasma ACTH concentration***, to determine whether the hypercortisolism is ACTH-dependent (due to an ACTH-secreting tumor), or ACTH-independent (due to a primary adrenal source). The best ACTH test is a two-site irmmunoradiometric assay.[197] Although plasma ACTH levels normally exhibit a circadian rhythm (20–80 pg/mL at 8:00 A.M., falling to <20 pg/mL at 4:00 P.M., and to <10 pg/mL within an hour after the usual time of falling asleep), ACTH can be measured at any time in patients with hypercortisolism, because the normal circadian rhythm is lost. At least two measurements should be obtained. Values less than 5 pg/mL indicate ACTH-independent disease,[198] and values greater than 20 pg/mL indicate ACTH-dependent disease. Values in between are less easily interpreted, but usually are associated with ACTH-dependent disease; in such unusual cases, a CRH stimulation test can help steer evaluation in the right direction.

The ***CRH stimulation test*** (described below) is based on the fact that plasma ACTH and serum cortisol concentrations increase promptly after CRH stimulation in most patients with an ACTH-secreting pituitary adenoma, but not in those with a primary adrenal source, because pituitary ACTH secretion is suppressed.[199–203] An ACTH response to CRH stimulation thus indicates ACTH-dependent Cushing syndrome, and the lack of response indicates ACTH-independent disease.[129, 130, 204]

ACTH-Independent Cushing Syndrome

For patients with ACTH-independent Cushing syndrome, the next diagnostic step is ***thin-section computed tomography*** (CT) imaging of the adrenal glands, seeking to identify an adrenal mass. If imaging reveals a unilateral adrenal adenoma, no further testing is necessary. Findings suggesting a possible carcinoma require additional evaluation to stage the suspected cancer. Findings of bilateral disease also require further evaluation to determine whether bilateral masses are both functional or if one is a nonfunctional incidentaloma, or to distinguish among the various causes of bilateral adrenal hyperplasia, which include long-term adrenal stimulation by pituitary or ectopic sources of ACTH and rare forms of ACTH-independent macro- and micronodular disease.[129, 130]

Adrenal adenomas typically are smaller (<4 cm in diameter) than carcinomas and have characteristically low unenhanced CT attenuation values. Irregular margins, necrosis, hemorrhage, or calcification suggest a carcinoma.[127] When needed, additional information to help define the nature of an adrenal mass lesion can be obtained by magnetic resonance imaging (MRI), by functional nuclear imaging with scintigraphy using a labeled cholesterol

analog (^{131}I-6-iodomethyl norcholesterol),[128] or positron emission tomography (PET) scanning. Whereas most benign cortisol-secreting adenomas produce relatively little androgen, most androgen-secreting adrenal tumors are malignant.[125, 126]

ACTH-Dependent Cushing Syndrome

For patients with ACTH-dependent Cushing syndrome, evaluation is aimed at identifying the source of ACTH secretion. By far, most such patients will have a pituitary corticotroph adenoma (Cushing disease). Ectopic ACTH-secreting or CRH-secreting tumors are rare.

The ***CRH stimulation test*** helps to distinguish pituitary from ectopic sources of ACTH, for the same reason it helps distinguish ACTH-dependent from ACTH-independent Cushing syndrome in women with equivocal plasma ACTH concentrations. ACTH and cortisol levels increase promptly after CRH stimulation in most patients with an ACTH-secreting pituitary adenoma, but not in those with ectopic sources of ACTH because pituitary ACTH secretion is suppressed.[199–203] After a period of fasting for 4 hours or more, blood samples are obtained 15 minutes and immediately before administering an intravenous bolus of synthetic ovine or human CRH (1 μg/kg or 100 μg total dose), and every 15 minutes for 60 minutes thereafter; samples are assayed for both ACTH and cortisol.[200, 205] There are no established uniform criteria for interpretation of the CRH stimulation test in patients with Cushing syndrome. In various studies, a 35–50% increase in ACTH and a 20–50% increase in cortisol over basal concentrations excluded all patients with ectopic ACTH secretion and correctly identified over 90% of patients with Cushing disease.[199, 200, 203] In approximately 8–10% of patients with Cushing disease, ACTH levels do not rise appreciably in response to CRH.[206]

The ***high-dose dexamethasone suppression test*** also helps to distinguish pituitary from ectopic sources of ACTH. The test is based on the fact that ACTH-secreting pituitary adenomas are only relatively resistant to negative feedback by glucocorticoids; low doses of dexamethasone do not suppress their ACTH secretion, but high doses typically do.[207] In contrast, ectopic sources of ACTH are completely resistant to glucocorticoid suppression because they are not controlled by feedback inhibitory mechanisms.[208] The overnight high-dose dexamethasone suppression test is easier to perform than the standard 2-day test (2 mg every 6 hours for a total of 8 doses) and has comparable sensitivity and specificity.[209] The test is performed by administering 8 mg of dexamethasone between 11:00 P.M. and midnight and measuring the serum cortisol level at 8:00 A.M. the next morning; the cortisol concentration is less than 5 μg/dL in most, but not all, patients with an ACTH-secreting pituitary adenoma (Cushing disease).[210] Alternatively, the morning serum cortisol on the days before and after dexamethasone treatment can be compared, with 50% or greater suppression indicating Cushing disease.[209, 211, 212]

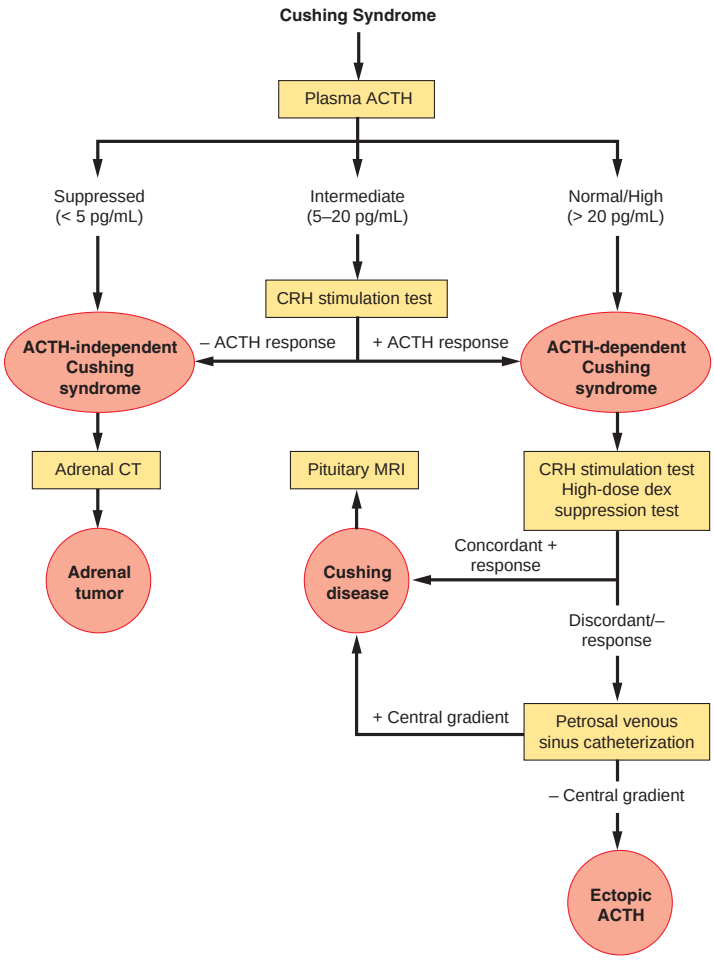
Combined, results obtained with the CRH stimulation test and the high-dose dexamethasone suppression test point to the source of ACTH in most patients with ACTH-dependent Cushing syndrome. Those whose cortisol levels are suppressed by high-dose dexamethasone (positive test) and stimulated by CRH (positive test) almost certainly have a pituitary ACTH-secreting adenoma.[213, 214] Patients with discordant or negative test results are still more likely to have Cushing disease than ectopic ACTH secretion, but require additional evaluation to establish the source of ACTH.

Petrosal venous sinus catheterization is the most direct way to demonstrate the presence or absence of excessive pituitary ACTH secretion, by comparing the ACTH concentrations in the petrosal venous sinuses (draining the pituitary via the cavernous sinus) to that in the peripheral blood.[215–217] Blood samples are obtained from the petrosal sinuses and from a peripheral vein before and within 10 minutes after administration of CRH. A petrosal venous ACTH concentration 2-fold or greater than the peripheral venous concentration,

or 3-fold or greater after CRH stimulation, distinguishes pituitary from ectopic sources of ACTH, with rare exceptions.[206, 218–220] Some have reported that an ACTH concentration gradient between the two petrosal sinuses of 1.4 or greater can predict the location of the pituitary adenoma,[206] but others have found the measure unreliable.[221, 222] Petrosal venous sinus catheterization also has significant potential risks, including cerebrovascular accident, cranial nerve palsies, pulmonary embolus, and deep vein thromboses and hematomas. Therefore, the procedure is difficult to justify solely for the purpose of localizing a tumor.

Pituitary MRI (unenhanced and gadolinium-enhanced) is more sensitive than CT for detecting corticotroph adenomas, but still identifies only about half of such tumors.[223] The imaging protocol (field of view and repetition time/echo time value) should be optimized for detection of corticotroph adenomas.[224] When results obtained with CRH stimulation and high-dose dexamethasone suppression support the diagnosis of Cushing disease and imaging demonstrates an obvious tumor (>6 mm), no further evaluation is required. When endocrine test results are discordant or negative, petrosal venous sinus catheterization generally is required to isolate the source of ACTH, although preliminary pituitary imaging to exclude an obvious tumor still is prudent, so as to avoid unnecessary petrosal sinus sampling and its attendant risks.

Ectopic ACTH-secreting tumors can be difficult to localize. Chest CT or MRI is the logical first step because that is where most such tumors (e.g., bronchial carcinoids) are found.[225] Others may be detected with PET scanning or scintigraphy using ^{111}In-pentetreotide (an octreotide analog) because, like other neuroendocrine tumors, ACTH-secreting tumors have somatostatin cell-surface receptors.[226, 227]



Summary of Key Points and Recommendations for the Evaluation of Hirsutism

1. Laboratory evaluation is recommended for women with moderate or severe hirsutism, or hirsutism that is sudden in onset, rapidly progressive, or associated with symptoms or signs of virilization. Routine laboratory evaluation of women with mild hirsutism is unnecessary.

2. The serum total testosterone concentration is the best overall measure of androgen production and is the only hormone that need be measured in most women with hirsutism that merit evaluation.

3. An androgen-secreting tumor should be suspected, and excluded, in women with rapidly progressive hirsutism, symptoms or signs of virilization, or a serum testosterone concentration 150 ng/dL or greater. However, most such patients will not have a tumor.

4. Nonclassical congenital adrenal hyperplasia should be suspected, and excluded, in patients with an early onset of hirsutism (pre- or peri-menarcheal, including those with a premature adrenarche), women with a family history of the disorder, and those in high-risk ethnic groups (Hispanic, Mediterranean, Slavic, and Ashkenazi Jewish heritage).

5. Cushing syndrome should be suspected, and excluded, in women with symptoms and signs of hypercortisolism.

Treatment of Hirsutism

The treatment of hirsutism should be directed towards its cause, whenever possible, but also must consider the extent to which the patient views it as a problem, and her therapeutic and reproductive goals. Whereas laboratory evaluation is recommended only for women with moderate or severe hirsutism, treatment should be considered for all women who judge themselves hirsute; many with mild hirsutism are worried or bothered by their hair growth and seek treatment.[72]

Although hirsutism can be managed using cosmetic measures such as shaving, plucking, waxing, and depilatory agents, most women with hirsutism have increased androgen production, and hair growth recurs if treated only by removal; most already are using one or more such methods. Consequently, almost all who seek treatment for hirsutism require drug therapy.

The severity of hirsutism should be defined before treatment begins to provide the means for monitoring response; the methods and frequency of hair removal provide the most practical and clinically relevant measure. ***Serial measurements of serum androgen levels during treatment are neither necessary nor helpful, but repeated evaluation is indicated when hirsutism progresses despite treatment.***

Before treatment begins, it also is important to foster reasonable expectations regarding its likely impact. Finer, lighter and slower hair growth, and the prevention of new terminal hair growth, all can be expected; a complete cessation or elimination of hair growth cannot.

No significant reduction in hair growth may occur for up to 6 months, which approximates the half-life of a hair follicle growth cycle. After 6 months, a change in dose, drug, or the addition of a second drug should be considered if the patient judges her response inadequate. In general, treatment should be continued indefinitely because the problem rarely goes away and almost always recurs when treatment is discontinued.[228] Patients planning to attempt pregnancy are the obvious exception, because most treatments prevent pregnancy or are contraindicated during pregnancy due to the risk of adverse impact on sexual development in a male fetus.

Treatments for hirsutism are aimed at reducing the production, increasing the binding, and/or blocking the action of androgens, and estrogen-progestin contraceptives and antiandrogens are the primary weapons in the therapeutic arsenal. It is important to emphasize that even women with idiopathic hirsutism relating to increased end organ sensitivity to androgens can benefit from treatments that lower free/active androgen concentrations or block the androgen receptor;[229] clinical response correlates with circulating levels of 3α-androstanediol glucuronide (the peripheral metabolite of dihydrotestosterone), supporting increased peripheral 5α-reductase activity as the cause of idiopathic hirsutism.[230]

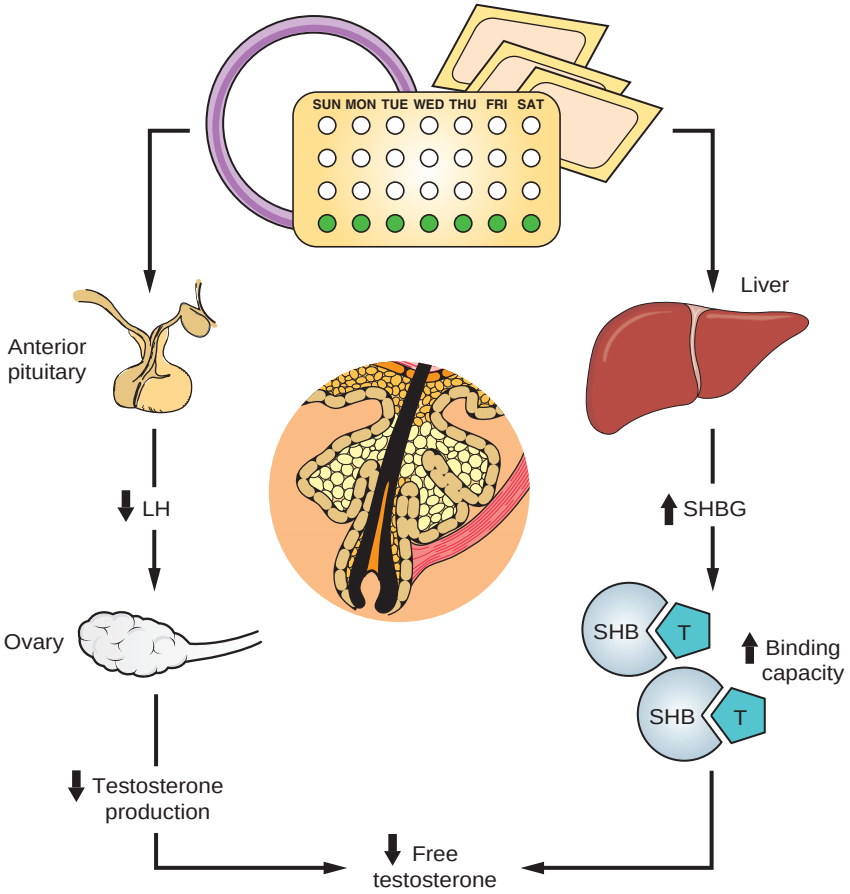
Estrogen-Progestin Contraceptives

Estrogen-progestin contraceptives have a number of complementary non-contraceptive actions that make them a logical and effective treatment for hirsutism:

- Androgen production in hirsute women usually is an LH-dependent process. Estrogen-progestin contraceptives suppress pituitary LH secretion and thus also suppress ovarian androgen production.[231–234]
- The high level of estrogen in combination contraceptives stimulates hepatic SHBG production, thereby increasing binding capacity for circulating androgens and decreasing the amount of free/active androgen.[234–236]
- Directly or indirectly, estrogen-progestin contraceptives can decrease adrenal DHEA-S secretion.[237–240]
- Contraceptive progestins inhibit 5α-reductase activity in skin,[241] which decreases the production of dihydrotestosterone (DHT), the major nuclear androgen in hair follicles and sebaceous glands.

The large majority of the benefits resulting from treatment with estrogen-progestin contraceptives derive from the first two actions. In addition to these specific actions on androgen production, binding, and metabolism, combination contraceptives have other effects that often are equally important in the clinical management of women with hirsutism. Most hirsutism results from chronic anovulation, which frequently causes menstrual irregularity and episodic dysfunctional bleeding, and also predisposes to abnormal patterns of endometrial growth. Treatment with estrogen-progestin contraceptives induces regular, predictable menses and attenuates endometrial growth, thereby eliminating the risk for developing endometrial hyperplasia and neoplasia.

Current oral contraceptives contain ethinyl estradiol, in doses ranging from 20 μg to 50 μg daily, and one of a variety of progestins. ***All low-dose oral contraceptives (containing 20–35 μg ethinyl estradiol) have similar effectiveness in the treatment of acne and hirsutism.*** Although estrogen induces a dose-dependent increase in serum SHBG concentrations,[234, 236] low- and higher-dose pills suppress free testosterone levels to a comparable extent.[242, 243] Similarly, although contraceptive progestins have varying impact on SHBG levels, there are no detectable differences in their overall clinical effectiveness.[242–246] Drospirenone,



a derivative of spironolactone, has some intrinsic anti-androgenic properties,[247, 248] but the dose (3 mg) is too small (equivalent to approximately 25 mg spironolactone)[85] to have any significant impact beyond that of other oral contraceptives.[249]

The transdermal contraceptive patch (delivering 20 μg ethinyl estradiol and 150 μg norelgestromin daily) and the vaginal contraceptive ring (releasing 15 μg ethinyl estradiol and 120 μg etonogestrel daily) also can be used for the treatment of hirsutism, although data relating to their effects on androgens are limited. The average circulating ethinyl estradiol concentration in patch users is approximately 60% higher than in women using an oral contraceptive containing 35 μg ethinyl estradiol, resulting in a greater increase in SHBG, but the overall decrease in androgen levels is no different.[250]

For patients with contraindications to the use of estrogen-progestin contraceptives, treatment with medroxyprogesterone acetate (150 mg intramuscularly every 3 months, or 10–20 mg orally daily) is an alternative. Although the progestin suppresses gonadotropins secretion to a lesser extent than estrogen-progestin regimens, LH still is suppressed sufficiently to cause a significant decrease in ovarian androgen production. In addition, testosterone clearance increases during treatment with medroxyprogesterone acetate,[251] due to induction of hepatic enzyme activity. Although SHBG levels are decreased during treatment, the decrease in androgen production is so great that free testosterone levels still are decreased overall.[252]

Subjectively, 60–100% of women report improvement in their hirsutism during treatment with oral contraceptives,[231–234, 253] which agrees with observations in studies using objective measures of hirsutism.[234, 236, 253–255] Oral contraceptives have similar effectiveness for

improving acne and seborrhea.[256, 257] Clinical improvement reflects the decrease in free/active androgen during treatment: new terminal hair growth decreases or stops, terminal hairs already present grow more slowly and produce finer hair, and acne gradually improves or disappears. ***Hormone therapy must be continued for at least 6 months before judging its effectiveness.*** In the meantime, the patient can continue to use her preferred method of hair removal (e.g., shaving, plucking, waxing). After 1–2 years, or when pregnancy becomes the goal, treatment can be discontinued and the patient observed for a return of ovulatory cycles, although most again will exhibit chronic anovulation. ***Permanent hair removal by electrolysis or laser methods (discussed below) may be required ultimately, at least in some patients, but is best postponed until hormonal suppression has achieved its maximum benefits.***

Although some studies have suggested that treatment with oral contraceptives may cause a modest increase in insulin resistance in patients with PCOS,[258–260] most have observed no significant changes.[261–267] Treatment with low-dose oral contraceptives does not adversely affect lipid and biochemical markers for cardiovascular disease, retinopathy, or nephropathy in women with insulin-dependent diabetes,[268–271] and has very limited impact on glucose tolerance, even in obese women with severe insulin resistance.[261] In a long-term study (6–18 years follow-up) of women with PCOS, metabolic parameters (body weight, glucose tolerance, insulin and high-density lipoprotein cholesterol levels) improved in those using oral contraceptives, whereas they worsened in nonusers.[272] Taken together, these observations support the safety of treatment with oral contraceptives in women with PCOS and insulin resistance. Also notably, in one study involving five women with the HAIR-AN syndrome, serum androgen concentrations decreased into the normal range in four of the five patients during combined treatment with oral contraceptives and spironolactone.[273]

Antiandrogens

Antiandrogens are an effective treatment for hirsutism, but are best used in combination with oral contraceptives or another means of highly effective contraception, because they have the potential to adversely affect sexual development in a male fetus if the patient conceives during treatment. In patients with contraindications to oral contraceptives, an alternative means of reliable contraception (e.g., an intrauterine device) should be provided during treatment with antiandrogens. Combined treatment with oral contraceptives and antiandrogens also is a logical choice for patients who respond inadequately to oral contraceptives alone.

Spironolactone

Spironolactone is an aldosterone antagonist having structural similarity to progestins. The drug also acts as an androgen receptor antagonist, competing with dihydrotestosterone (DHT) for binding to the androgen receptor, and to varying extent, also inhibits ovarian and adrenal androgen synthesis.[274] Although serum androstenedione levels decrease, those of DHEA, DHEA-S and cortisol do not significantly change during treatment with spironolactone.

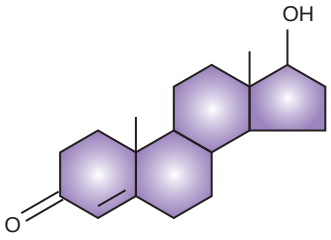
The effects of spironolactone are dose-dependent and best results are achieved with doses of 50–100 mg twice daily.[275–278] In two clinical trials comparing spironolactone (100 mg daily) with placebo, active treatment resulted in significantly greater subjective improvement in

hirsutism.[279, 280] As with all treatments for hirsutism, maximal effects are observed only after approximately 6 months of therapy. Side effects are relatively few, including diuresis in the early days of treatment and occasional complaints of fatigue and dysfunctional uterine bleeding. Although the drug can cause hyperkalemia, the effect is rare and monitoring of potassium levels is not necessary in women with normal renal function.

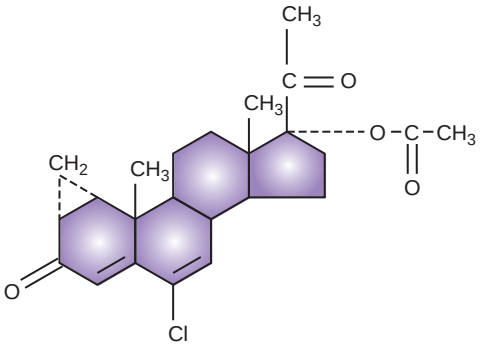
The action of spironolactone, peripheral androgen receptor blockade, nicely complements those of oral contraceptives and may thus provide additional benefit for those who fail to achieve adequate results from oral contraceptives alone. However, the results achieved with combined treatment are not greatly different.[281–283]

Cyproterone Acetate

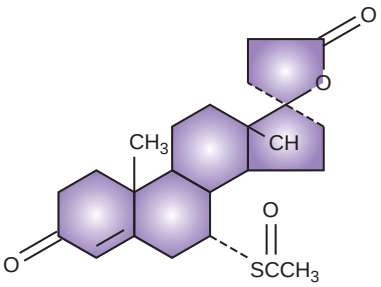
Cyproterone is a derivative of 17α-hydroxyprogesterone (17OHP) having potent progestational activity that inhibits gonadotropin secretion, but also acts as a competitive androgen receptor antagonist and inhibits enzymes involved in androgen synthesis, like spironolactone. Cyproterone acetate is the progestin in the combined estrogen-progestin oral contraceptive called "Diane" (2 mg cyproterone acetate and 50 μg ethinyl estradiol) in common use in many parts of the world, but not available in the United States; "Dianette" or "Diane 35" contains 2 mg cyproterone acetate and 35 μg ethinyl estradiol. The drug also has been used in higher doses (12.5–100 mg), alone or in combination with estrogen.[284, 285] A systematic review including data from 9 clinical trials concluded that combined treatment with cyproterone acetate and ethinyl estradiol is more effective than placebo and yields results comparable to those achieved with oral contraceptives, spironolactone, and other treatments.[286] The most common side effects associated with cyproterone treatment are fatigue, edema, loss of libido, weight gain, and mastalgia.



Testosterone



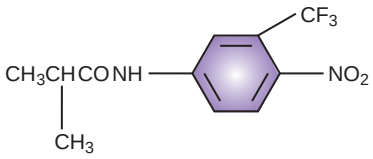
Cyproterone acetate



Spironolactone

Flutamide

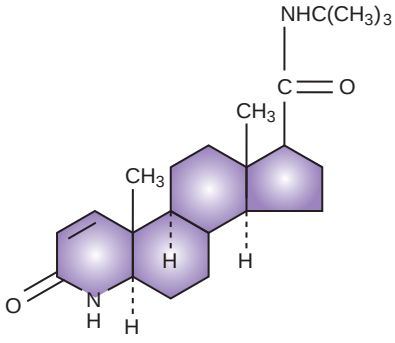
Flutamide is a nonsteroidal androgen receptor antagonist used primarily in the treatment of prostate cancer. The drug (250–750 mg daily) inhibits hair growth directly and is as effective as spironolactone,[255, 280, 283, 287, 288] but its higher cost and potential for causing severe hepatotoxicity make it an unattractive therapeutic choice, by comparison.[289, 290]



Flutamide

Finasteride

Finasteride inhibits 5α–reductase and thus blocks the conversion of testosterone into DHT. The enzyme exists in two forms, with type 1 most prevalent in skin and type 2 predominating in reproductive tissues.[291] Although finasteride inhibits the type 1 enzyme only to a limited extent, evidence from clinical trials indicates its efficacy is comparable to those of spironolactone and flutamide.[280, 283, 292] Because external male genital development requires the action of DHT, the risks of inadvertent finasteride treatment during pregnancy are a particular concern and finasteride should not be used without a highly effective method of contraception.



Finasteride

Insulin-Sensitizing Drugs

Given that PCOS is the most common cause of hirsutism and that insulin resistance is a common feature of the disorder, insulin-sensitizing drugs offer another potential useful approach to the treatment of hirsutism.[85, 293] Indeed, treatment with metformin and thiazolidinediones (rosiglitazone, pioglitazone) decreases circulating insulin and androgen levels in women with PCOS.[294–301] However, a recent systematic review and meta-analysis including 9 placebo-controlled trials concluded that insulin-sensitizing drugs have no important

benefits for the treatment of hirsutism.[162] Accordingly, guidelines issued by the Endocrine Society suggest against their use for the treatment of hirsutism.[85]

Other Treatments

The mainstay of medical treatment for hirsutism has been, and remains, estrogen-progestin contraceptives, with the addition of an antiandrogen after approximately 6 months if the desired cosmetic result has not yet been achieved. When conventional treatment is contraindicated or proves inadequate, other treatments can be considered.

Gonadotropin-Releasing Hormone Agonists

In women with severe hyperandrogenism who fail to respond to or cannot tolerate treatment with estrogen-progestin contraceptives and antiandrogens, GnRH agonist therapy can be considered. GnRH agonists (e.g., leuprolide, nafarelin, goserelin) are not recommended for routine use, primarily because they induce a severe hypoestrogenism, but also because they are more costly and inconvenient to use.[85]

Serum androgen levels decrease dramatically during GnRH agonist treatment, typically falling to near castrate levels within as little as a month.[302–305] The addition of estrogen to GnRH agonist therapy to eliminate estrogen deficiency symptoms and prevent bone loss does not diminish its efficacy and can even increase it. Cyclic or continuous treatment with estrogen (e.g., 0.3–0.625 mg conjugated estrogens daily, or equivalent) and progestin (e.g., 5–10 mg medroxyprogesterone), or an estrogen-progestin contraceptive, can be used. Combined treatment decreases free testosterone concentrations to lower levels than GnRH therapy alone, due to the added benefit of increased SHBG concentrations induced by estrogen.[306–309] ***Nevertheless, combined treatment with a GnRH agonist and oral contraceptives is no more effective than treatment with a GnRH agonist alone, and somewhat less effective than combined treatment with oral contraceptives and an antiandrogen.***[310]

The effectiveness of GnRH agonist therapy relates directly to the suppression of LH-dependent ovarian androgen production. Adequate suppression may not be achieved in obese women, as suggested by the absence of expected estrogen deficiency symptoms. When suspected, the possibility can be confirmed by measuring the serum estradiol concentration. If results indicate inadequate suppression, the dose of GnRH agonist therapy should be increased.

GnRH agonist therapy should be an effective treatment for women with ovarian hyperthecosis who typically have severe hyperandrogenism. However, the impact of treatment on their hirsutism can be less than expected, even when gonadotropin secretion is suppressed profoundly, because most also have severe insulin resistance, with hyperinsulinemia driving their androgen production.[311]

Glucocorticoids

Glucocorticoids are used to suppress endogenous ACTH secretion in the long-term management of women with classical congenital adrenal hyperplasia (CAH). They also have been used for the treatment of hirsutism in women with the nonclassical, late-onset, form of the disorder, but with limited benefit. ***Although glucocorticoids suppress serum***

adrenal androgen levels effectively in women with nonclassical CAH, they are less effective than oral contraceptives or antiandrogens for the treatment of hirsutism.[312, 313] Consequently, glucocorticoid treatment has even less to offer women with other causes for hirsutism.[303, 314, 315]

Eflornithine Hydrochloride

Eflornithine hydrochloride (13.9% cream) is a topically applied inhibitor of ornithine decarboxylase, an enzyme active in the dermal papilla that is essential for hair growth; it is not a depilatory agent. In clinical trials, twice daily application produced noticeable improvement in facial hair growth within a few weeks in the majority of patients. However, the drug must be used continuously, because hair growth reverts to pretreatment characteristics within approximately 8 weeks after treatment is discontinued.[316] When used in conjunction with laser hair removal, eflornithine produces a more rapid response than laser treatment alone.[317, 318] Treatment with topical eflornithine hydrochloride is perhaps best suited for patients with mild facial hirsutism, such as occurs after menopause.

Permanent Hair Removal

Removal of hair by plucking, waxing, shaving, or use of depilatory agents is common in women with hirsutism, but the results achieved are only temporary. Hairs that are plucked again become apparent after approximately 6–8 weeks. Waxing, using melted wax ("hot waxing") or a liquid wax ("cold waxing") can be used on larger areas of the body, but results last no longer. Both methods remove the entire hair, but typically not the dermal papilla. Because shaving removes hair to a level only slightly below the skin, its results are short-lived and most women will need to shave again within 1–3 days. Electrolysis and photoepilation (laser and pulsed light therapies) are aimed at permanent hair removal.

Electrolysis

Electrolysis has been used as a method of permanent hair removal for more than 100 years.[319] The earliest method, called galvanic electrolysis, used direct current applied to a fine needle inserted into the hair follicle, which produced sodium hydroxide from saline in tissues, causing a chemical destruction of the dermal papilla. Modern "thermolytic" techniques use high frequency alternating current, causing thermal destruction of the hair follicle, or a "blend" of the two methods.[320, 321]

Although electrolysis is an effective method of "permanent" hair removal,[322] hair growth recurs in up to 25% of women by 6 months after treatment is discontinued.[320, 321] In sensitive areas, topical anesthetic creams typically are applied first, because electrolysis can be painful. Electrolysis also can cause inflammation and erythema and, in some, pigmentation changes and scarring.

Unfortunately, there are no governing standards of practice for electrolysists and no formal training is required in many jurisdictions before starting a practice. In experienced hands, electrolysis can produce satisfying results, but the quality of care, and results, can vary considerably.

Laser and Pulsed Light Therapies

Photoepilation therapies use laser or intense pulsed light to destroy hair follicles.[323, 324] Both methods attempt to selectively target the hair bulb by using wavelengths absorbed specifically by melanin, but absorbsion by pigment in the epidermis also can occur. Consequently, they are best suited for light-skinned individuals with dark hair, in whom most of the energy will be absorbed by melanin in the hair bulb; the risk of burns and other complications such as postinflammatory pigment changes increases with the amount of skin pigmentation, although advances in technology are improving options for patients with darker skin.[325–327] Photoepilation therapies can be used for removal of any color of hair but, as might be expected, are more effective in patients with black or brown hair than in those with red or blonde hair.

Success rates achieved with laser and pulsed light therapies vary with hair growth phase, skin and hair color, location, the type of laser, and the number of treatments.[328, 329] Most patients require a series of 4–6 treatments at 4–6 week intervals to achieve the desired result,[326] followed by maintenance treatments every 6–12 months to remove any hairs that grow back.

Summary of Key Points and Recommendations for the Treatment of Hirsutism

1. The response to all medical treatments for hirsutism is relatively slow, generally requiring 6 months to achieve significant benefits, which approximates the duration of the life cycle of a hair follicle.

2. The first treatment of choice for hirsutism is a low-dose estrogen-progestin contraceptive.

3. In patients having an inadequate response to treatment with estrogen-progestin contraceptives alone, an antiandrogen should be added, with spironolactone generally being the best choice.

4. The use of GnRH agonists should be reserved for patients who fail to respond to or cannot tolerate more traditional treatments and should be combined with sex steroid add-back therapy, which prevents the consequences of hypoestrogenism and does not diminish the efficacy of GnRH agonist treatment.

5. Permanent hair removal using electrolysis or photoepilation therapies (laser, pulsed light), when necessary, is best postponed until hormonal suppression has achieved its maximum benefits.

All references are available online at: http://www.clinicalgynendoandinfertility.com

14 Menstrual Disorders

	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31			

Since antiquity, the temporal relationship between menses and the lunar phases has inspired names for menstruation, such as the "period." The regularity of menses was easily appreciated by the ancients, even if they had no understanding of its cause or purpose. Ancient physicians viewed menstruation as a process of detoxification and, throughout history, myths and superstitions have perpetuated negative attitudes toward menses.[1]

The health care profession has an obligation to provide and to promote education on menstruation and related subjects, which must start with itself. Clinicians must have a thorough understanding of reproductive physiology before they can impart that knowledge to their patients, and must be sensitive to the need to present the information in a positive context that fosters healthy attitudes toward sexual and reproductive functions. An educated understanding of normal reproductive processes is a powerful tool for addressing the symptoms and disorders of menstruation.

Some menstrual disorders, such as dysmenorrhea, can be explained in a physiologic framework that both educates and provides the foundation for appropriate treatment. Unfortunately, others, such as the premenstrual syndrome, remain poorly understood. This chapter considers several medical problems that are temporally linked with menstruation, and their pathophysiology when that is known.

Historical Views of Menstruation and Menstruating Women

Recorded history includes a wide variety of myths regarding menstruation and menstruating women. In ancient times, menstruating women commonly were thought possessed

by an evil spirit. Aristotle (384–322 B.C.), the Greek philosopher, student of Plato, and teacher of Alexander the Great, said that a menstruous woman could dull a mirror with a look, and that the next person to peer into it would be bewitched. Pliny, born in 23 A.D., consulted approximately 2,000 available books by physicians while writing his treatise, *Historia Naturalis*, a resource used throughout the Dark Ages; more than a hundred copies, all 37 volumes, still exist. Pliny wrote extensively on menstruation, including the following[2]:

> Contact with it turns new wine sour, crops touched by it become barren, grafts die, seeds in gardens are dried up, the fruit of trees falls off, the edge of steel and the gleam of ivory are dulled, hives of bees die, even bronze and iron are at once seized by rust, and a horrible smell fills the air; to taste it drives dogs mad and infects their bites with an incurable poison. If a women strips herself naked while she is menstruating and walks around a field of wheat, the caterpillars, worms, beetles, and other vermin will fall off from the ears of corn. All plants will turn of a yellow complexion on the approach of a woman who has the menstrual discharge upon her. Bees will forsake their hives at her touch, for they have a special aversion to a thief and a menstruous woman, and a glance of her eyes suffices to kill a swarm of bees.

Throughout early history, the fear of blood spawned many ancient taboos. Almost universally, menstruating women were isolated and prevented from handling food. Most primitive peoples regarded women as unclean during menstruation and subjected them to segregation and special rituals. It is therefore not surprising that, even with growing sophistication, negative attitudes toward menstruation persisted into modern times.

In 19th and early 20th century Europe, menstruation was commonly associated with antisocial behavior.[3] In 1845, a domestic servant who murdered one of her employer's children was acquitted on the grounds of insanity due to obstructed menstruation. In 1851, a woman was acquitted of murdering her baby niece due to insanity arising from disordered menstruation. As recently as 1984, Dalton argued that the premenstrual phase of the cycle was associated with an increased incidence of crime, jailing for alcoholism, poor academic performance, sickness in industry, and hospitalization for accidents.[4] However, careful studies have found no significant variations in cognitive or motor functions across the menstrual cycle,[5–8] suggesting that, to a large extent, behaviors merely reflect societal expectations. Unfortunately, even today, expectations and attitudes toward menstruation are influenced heavily by old traditions and social and cultural beliefs.

The Premenstrual Syndrome and Premenstrual Dysphoric Disorder

The simplest definition of the premenstrual syndrome (PMS) is a common sense one: cyclic physical and behavioral symptoms that appear in the days preceding menses and interfere with work or lifestyle, followed by a symptom-free interval. Premenstrual dysphoric disorder (PMDD) describes a severe form of PMS that some consider a distinct

clinical entity, characterized by prominent symptoms of irritability, anger, internal tension, dysphoria, and mood lability.[9]

Historically, the phrase "premenstrual syndrome" was used first by Greene and Dalton in their report of 84 cases in 1953.[10] However, R.T. Frank, then the chief of obstetrics and gynecology at Mt. Sinai Hospital in New York City, generally is credited with having first described PMS, in 1931:[11]

> The group of women to whom I refer especially complain of a feeling of indescribable tension from 10 to 7 days preceding menstruation which in most instances continues until the time that the menstrual flow occurs. The patients complain of unrest, irritability, like jumping out of their skin and a desire to find relief by foolish and ill considered actions. Their personal suffering is intense and manifests itself in many reckless and sometimes reprehensible actions. Not only do they realize their own suffering, but they feel conscience-stricken toward their husbands and families, knowing well that they are unbearable in their attitude and reactions. Within an hour or two after the onset of the menstrual flow complete relief from both physical and mental tension occurs.

An extraordinary array of different physical and behavioral symptoms has been attributed to PMS. The most common physical symptoms include abdominal bloating, extreme fatigue, breast tenderness, and headaches, all occurring in 50–90% of cases. The most prevalent behavioral symptoms of PMS are mood lability, irritability, depressed mood, increased appetite, forgetfulness, and difficulty with concentration, occurring in 50–80% of cases. Other less common symptoms include anxiety or tension, easy crying, thirst, acne, gastrointestinal upset, hot flushes, palpitations, dizziness, and lower extremity edema. Symptoms of PMS typically arise during the last 7–10 days of the cycle.[12]

Premenstrual symptoms are very common, reported by up to 75% of women with regular menstrual cycles. However, because women naturally relate symptoms and behaviors to menstruation retrospectively, estimates of their frequency are subjective and inherently biased.[13] Moreover, both men and women have been conditioned to expect symptoms during the premenstrual phase of the cycle, such as fluid retention, pain, and emotional lability, and not surprisingly, they report such symptoms when asked in retrospect.[14] The power of conditioned response was illustrated in a classic study by Ruble, in which 44 undergraduates at Princeton University were deliberately deceived about the phase of their menstrual cycle.[15] A mock electroencephalogram was performed, complete with electrodes attached to the head, after being described as a new technique that could predict the onset of menstruation. Subjects were informed that they were premenstrual (expected menses in 1–2 days) or intermenstrual (expected menses in 7–10 days), and only those led to believe they were premenstrual reported increased symptoms of pain, water retention and changes in eating habits—a self-fulfilling prophesy. Subjects tend to comply with what they believe is the investigator's hypothesis, and in studies of PMS, no differences in symptoms can be demonstrated when the purpose of the study is disguised or expectations are manipulated.[16–18] Carefully designed prospective studies have revealed that some women who have no demonstrable cyclic premenstrual symptoms or changes in cognitive function nonetheless believe that they do,[13] even including some diagnosed with PMDD.[5, 19, 20] ***When strictly defined based on prospective symptom diaries, clinically significant PMS occurs in 20–30% and PMDD affects 2–8% of women.***[9, 21–24]

Diagnostic Criteria

The diagnosis of both PMS and PMDD depends on the presence of typical symptoms, their timing, severity, and the exclusion of other diagnoses. ***Both diagnoses require a prospective symptom diary documenting specific cyclic symptoms associated with the luteal and menstrual phases of the cycle and evidence of socioeconomic dysfunction.***[25] The specific collection of symptoms in a given individual is much less important than the cyclic nature of the symptom complex and its temporal relationship with menses. When symptoms are charted accurately, as much as 40% of women presenting with presumed PMS do not exhibit the distinctly cyclic pattern required for diagnosis and actually have another mood or anxiety disorder.[26]

The most commonly used criteria for diagnosis of PMS are those proposed by investigators at the University of California at San Diego.[12, 27] The diagnostic critera were based on prospective symptom diaries of women in whom underlying medical conditions and psychiatric disorders were carefully excluded, and on an analysis of cycle phase defined by daily urinary steroid metabolites.[27] A symptom survey instrument known as the Calendar of Premenstrual Experiences (COPE) was constructed, including the 10 most commonly reported physical symptoms and the 12 most commonly reported behavioral symptoms, with each ranked on a 4-point Likert scale of severity across the menstrual cycle.[12]

Calendar of Premenstrual Experiences[12, 28]

Name _____ **Month/Year** _____ **Age** _____

Begin your calendar on the first day of your menstrual cycle. Enter the date below the cycle day. Day 1 is your first day of bleeding. Shade the box above the cycle day if you have bleeding. ☐ Enter an X for spotting ☒ .

If more than one symptom is listed in a category (ie, nausea, diarrhea, constipation), you do not need to experience all of these. Rate the most disturbing of the symptoms on the 0–3 scale below.

Weight: Weigh yourself before breakfast. Record your weight in the box below date. Symptoms: Indicate the severity of your symptoms by using the 1–3 scale below. Rate each symptom at about the same time each evening.

0 = None (symptoms not present)

1 = Mild (noticeabe but not troublesome)

2 = Moderate (interferes with normal activities)

3 = Severe (intolerable, unable to perform normal activities)

Other Symptoms: If there are other symptoms you experience, list and indicate severity. Medications: List any medications taken and enter an X on the corresponding day(s).

Bleeding																													
Cycle Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Date																														
Weight																														
Symptoms																														
Acne																														
Bloating																														
Breast tenderness																														
Dizziness																														
Fatigue																														
Headache																														
Hot flashes																														
Nausea, diarrhea, constipation																														
Palpitations																														
Swelling (hands, ankles, feet)																														
Angry outbursts, arguments, violent tendencies																														
Anxiety, tension, nervousness																														
Confusion, poor concentration																														
Easy crying																														
Depression																														
Food cravings (salt, sweets)																														
Forgetfulness																														
Irritability																														
Increased appetite																														
Mood swings																														
Overly sensitive																														
Wish to be alone																														
Other symptoms																														
1.																														
2.																														
Medications																														
1.																														
2.																														

The COPE survey instrument yields reliably reproducible scores that correlate well with those derived from administering the Profile of Mood States[29] and the Beck Depression Inventory.[30] A number of other scales can be used for diagnosis of PMS, including the Moos Menstrual Distress Questionnaire (MDQ),[31] the Premenstrual Assessment Form (PAF),[32, 33] and the Prospective Record of the Severity of Menstruation (PRISM).[34] The COPE survey remains among the most popular, primarily because analysis of data derived from its use has revealed that virtually all women with PMS can be identified using a simplified list of six behavioral symptoms and four physical symptoms, yielding a set of diagnostic criteria that can be applied easily in a patient interview, as follows:[12]

1. Self-report of 1 or more affective symptoms ***and*** 1 or more somatic symptoms during the 5 days preceding menses in each of three menstrual cycles:

Affective Symptoms	*Somatic Symptoms*
Depression	Breast tenderness
Angry outbursts	Abdominal bloating
Irritability	Headache
Confusion	Swollen extremities
Social withdrawal	
Fatigue	

2. Relief from symptoms within 4 days after the onset of menses, without recurrence before cycle day 12.

3. Absence of any medications, hormone therapy, drug or alcohol use.

4. Socioeconomic dysfunction, as indicated by one of the following:

   Discord in the relationship with a partner, confirmed by the partner
   Parenting difficulties
   Poor work/school performance or attendance
   Increased social isolation
   Legal problems
   Suicidal ideation
   Seeking medical care for somatic symptoms

According to guidelines from the National Institute of Mental Health (NIMH),[35] the diagnosis of PMS also should require at least a 30% increase in the severity of symptoms over the 5 days before menses, compared with the 5 days after onset of menses. Based on the UCSD and NIMH criteria, it is estimated that approximately 5% of women of reproductive age can be diagnosed with disruptive PMS.[36–38]

The most commonly used criteria for the diagnosis of PMDD are those proposed by the American Psychiatric Association, as they appear in the current Diagnostic and Statistical Manual of Mental Disorders (DSM-IV):[39]

A. Symptoms occur regularly during the last week of the luteal phase in most menstrual cycles during the past year, remit within a few days after the onset of menses, and always are absent in the week after menses.

B. Five or more of the following symptoms must be present, including at least one among the first four:

1. Feeling sad, hopeless, or self-deprecating
2. Feeling tense, anxious, or "on edge"
3. Marked lability of mood, interspersed with frequent tearfulness
4. Persistent irritability, anger, and increased interpersonal conflicts
5. Decreased interest in usual activities that may be associated with withdrawal from social relationships
6. Difficulty concentrating
7. Feeling fatigued, lethargic, or lacking in energy
8. Marked changes in appetite that may be associated with binge eating or certain food cravings
9. Hypersomnia or insomnia
10. A subjective feeling of being overwhelmed or out of control
11. Other physical symptoms such as breast tenderness or swelling, headaches, sensations of "bloating" or weight gain with tightness of fit of clothing, shoes, or rings, or joint or muscle pain

C. The symptoms are of comparable severity (but not duration) to those of a mental disorder, such as a major depressive episode or a generalized anxiety disorder, and cause obvious and marked interference with work, usual activities, or relationships.

D. The symptoms may be superimposed on another disorder but are not merely an exacerbation of the symptoms of another disorder.

It is important to note that whereas the diagnosis of PMS requires both affective and somatic symptoms, the diagnosis of PMDD can include, but does not require, somatic symptoms. Another distinction between the two disorders is that PMDD may be superimposed on another psychiatric disorder, whereas the diagnosis of PMS can only be made in their absence.

The diagnoses of PMS and PMDD must be differentiated from other underlying psychiatric disorders, which are common among women with similar symptoms.[26, 40, 41] Medical conditions, such as hyperthyroidism and hypothyroidism, also should be excluded. A study involving a group of women evaluated in a specialty PMS clinic found that 13% had a distinct affective psychiatric disorder, 38% had premenstrual exacerbation of an underlying depressive or anxiety disorder, and only 44% actually met strict diagnostic criteria for PMS.[42] Women with PMS often have a history of a previous major depressive episode and also are at increased risk for major depression in the future.[43, 44] ***Women who have no demonstrable symptom-free interval during the follicular phase of the cycle merit careful evaluation for a mood or anxiety disorder.***

Migraine headache and symptoms of chronic fatigue syndrome and irritable bowel syndrome frequently are more pronounced during the premenstrual phase of the cycle. However, in women with these syndromes, symptoms also occur at other times in the cycle.[45]

Pathophysiology

Scientific evidence for the mechanism(s) involved in PMS and PMDD has been difficult to produce, but there has been no shortage of theories; the list is impressive:

Low progesterone levels
High estrogen levels
Falling estrogen levels
Changes in the estrogen/progesterone ratio
Increased aldosterone activity

Increased rennin-angiotensin activity
Increased adrenal activity
Endogenous opiate withdrawal
Subclinical hypoglycemia
Central changes in catecholamines
Responsiveness to prostaglandins
Vitamin deficiencies
Excess prolactin secretion

In his original description of PMS, R.T. Frank summarized 15 cases, theorized that the problem resulted from an excess of female sex hormones due to inadequate excretion, and reported that he could provide relief by withdrawing blood from his patients. Accordingly, he applied treatments designed to increase excretion, such as calcium lactate, caffeine, and laxatives. For severe cases, he prescribed pelvic irradiation to cause ovarian failure. In 1934, S. Leon Israel proposed the opposite theory—that PMS was caused by defective luteinization, progesterone deficiency, and relative hyperestrogenism.[46]

The role of ovarian steroid hormones in PMS is suggested strongly by the lasting reponse to oophorectomy in women unresponsive to medical therapy[47, 48] and by the dramatic decrease in symptoms after suppression of the hypothalamic-pituitary-ovarian axis by treatment with a long-acting gonadotropin-releasing hormone (GnRH) agonist.[49–52] However, studies comparing serum estrogen and progesterone levels have failed to identify any consistent differences between women with and without PMS/PMDD.[53–55] Moreover, an early onset of menses induced by treatment with a progesterone antagonist during the luteal phase does not decrease or otherwise change the symptoms of PMS, even when progesterone levels are maintained by simultaneous treatment with hCG.[56, 57] In women with PMS treated with a GnRH agonist, symptoms recur when exogenous estrogen or progesterone is added to the treatment regimen, but not in those receiving add-back placebo treatment, and not in normal women receiving the same treatment.[58] Studies involving well-defined patient populations also have failed to demonstrate any differences in the levels of testosterone, follicle-stimulating hormone (FSH), luteinizing hormone (LH), prolactin, sex hormone-binding globulin, and aldosterone between women with and without symptoms of PMS, in any phase of the menstrual cycle.[59] ***Taken together, these observations suggest that the symptoms of PMS are not caused directly by endocrine events during the luteal phase, but reflect an abnormal response to normal cyclic changes in ovarian steroid hormone levels.***

The menstrual cycle is associated with significant changes in the opioid,[60] gamma-aminobutyric acid (GABA),[61] and serotonin neurotransmitter systems,[62] suggesting a possible pathophysiologic mechanism for PMS. Lower midcycle and luteal phase serum β-endorphin levels have been observed in women with PMS, compared to normal women.[63–65] The anxiolytic actions of certain progesterone metabolites that act as ligands for the GABA-A receptor[61] and the effectiveness of alprazolam (a short-acting benzodiazepine) in relieving symptoms of PMS[66] suggest the disorder might involve a disturbance in the GABA-ergic system. However, studies comparing the levels of anxiolytic progesterone metabolites in women with and without PMS have observed no consistent differences.[55, 67] Several lines of evidence suggest that the mood symptoms of PMS may relate to serotonin depletion. Whole blood serotonin concentrations and platelet serotonin uptake and imipramine binding (a marker of CNS binding and serotonergic activity) during the luteal phase are lower in women with PMS than in asymptomatic women.[68–72] Symptoms of PMS are aggravated by acute dietary tryptophan depletion, which suppresses brain serotonin synthesis,[73] and relieved by treatment with fenfluramine, a serotonin agonist,[74] or fluoxetine, a serotonin reuptake inhibitor (SRI). Moreover, mood symptoms in women with PMDD receiving treatment with fluoxetine promptly return after treatment with metergoline, a serotonin antagonist.[75] ***The weight of current evidence thus suggests strongly that PMS and PMDD result from an abnormal or exaggerated effect of cyclic changes in***

ovarian steroid hormones on central neurotransmitter mechanisms and that serotonin, in particular, plays an important role in their pathophysiology.

In general, thyroid function is normal in women with PMS.[76] Approximately 10% of women with PMS have demonstrably abnormal thyroid function, but the prevalence is not significantly different from that of subclinical hypothyroidism in the general population. Overall, the thyroid-stimulating hormone (TSH) response to thyroid-releasing hormone (TRH) is normal. Although abnormal responses (both exaggerated and blunted) are observed more often in women with PMS,[77] they occur just as often during the follicular phase as in the luteal phase. Moreover, the effect of treatment with thyroxine is no different from that of placebo, even in patients with an abnormal response to TRH.

Several studies have found evidence to suggest that genetic factors might predispose to PMS and PMDD. A large twin study found that PMS was highly heritable, with environmental factors also contributing.[78] The correlation between menstrual symptoms in mothers and daughters and between sisters suggests a genetic influence, but also might merely reflect a learned or conditioned response, at least to some extent.[79, 80] Whereas many have speculated that differences in personality, stress levels, or coping mechanisms may play a role in PMS, there is little or no evidence to support the hypothesis.[81–83]

Some evidence suggests that women with PMS ingest more alcohol than others and that women with a family history of alcoholism exhibit more premenstrual anxiety and other behavioral symptoms, but a link between alcoholism and PMS has not been established.[84, 85] Efforts to identify vitamin deficiencies in women with PMS have failed. Studies comparing serum levels of vitamin A and vitamin E in women with and without PMS have observed no significant differences.[86, 87] The results of treatment with vitamin B6 generally are unimpressive and inconsistent.[88, 89] A number of studies have found that intracellular magnesium concentrations are lower in women with PMS than in asymptomatic control women, but the significance of the observation remains unclear.[90–93]

Treatment

The key to effective treatment of PMS and PMDD is accurate diagnosis, which rests primarily on the collection of objective evidence that the patient's symptoms are clearly cyclic, as documented by use of the COPE survey or other similar calendar-based screening tool over the interval spanning three menstrual cycles.

A number of medications have proven beneficial to women with PMS or PMDD, including SRIs,[94, 95] alprazolam (a benzodiazepine),[66, 96, 97] and GnRH agonists.[51, 98] Accumulated evidence indicates that certain oral contraceptives,[99–101] exercise,[102, 103] relaxation techniques,[104, 105] and spironolactone[106–109] also have some value. Older treatment regimens involving progesterone or progestins,[110–114] tricyclic antidepressants, monoamine oxidase inhibitors, lithium, evening primrose oil,[115] essential free fatty acids,[116] dietary restrictions, vitamin supplements, and ginkgo biloba generally are not effective.

The Placebo Response

The strange sounding word placebo derives from the Latin verb meaning "I shall please." Clinicians and patients have been conditioned to observe a prescription ritual. Many patients seem to feel that their complaints are not taken seriously unless a medication is prescribed. However, placebo is more than a pill; it is a process.[117]

The process begins with confidence in the clinician and extends to the patient's own healing system. Interaction with a clinician provides a better understanding of a patient's symptoms, eliminates some unfounded fears, and offers hope for improvement. Many of the treatments for PMS that provide women with a greater sense of control, even simple measures such as changes in diet and lifestyle, can yield benefits. The very process of making detailed, prospective observations of events in life can improve one's sense of self-control, which is intrinsically therapeutic.

Leon Eisenberg penned some very insightful thoughts regarding the placebo response, which plays a prominent role in the therapeutics of PMS and PMDD[118]:

> So emphatically does the phrase "placebo response" discredit the psychosocial aspects of the therapeutic encounter that it may be time to eradicate it from our language. Let us replace it by some such term as "the response to care," "the response to the doctor," or "the healing response" in order to emphasize that it is (a) powerful, (b) no less "real" than drug actions, and (c) embedded in every therapeutic transaction… Its mechanisms are some compound of the arousal of hope, the comfort of reassurance, taking an active rather than a passive role in managing the illness experience, and reinterpreting the meaning of the illness… It is perverse that "placebo" has almost become an epithet implying charlatanism rather than a descriptor of a fundamental characteristic of medical practice… We ought equally to seek an understanding of the healing response rather than disdaining it, as the "hard" scientist does, or being deceived by it, as practitioners often are.

Treatments for Premenstrual Syndrome and Premenstrual Dysphoric Disorder

Demonstrated Effective	*Possibly Effective*	*Ineffective*
Serotonin reuptake inhibitors	Oral contraceptives	Progesterone
Alprazolam	Diuretics	Vitamins
GnRH agonists	Exercise	Dietary restrictions

Serotonin Reuptake Inhibitors

There is substantial evidence for the effectiveness of SRIs in the treatment of PMS and PMDD.[94, 95, 119] Fluoxetine, in a daily dose of 20 mg, has demonstrated sustained efficacy for relieving both somatic and mood symptoms, and generally is well-tolerated.[120–123] Other SRIs also are effective, including sertraline (50–150 mg daily),[124, 125] paroxetine (20–30 mg daily),[126] and citalopram (20–30 mg daily).[127] Venlafaxine (50–200 mg daily), which inhibits the reuptake of both serotonin and norepinephrine, also has efficacy,[128] as do other antidepressants that inhibit serotonin reuptake or antagonize its action, such as clomipramine[129, 130] and nefazodone.[131]

Although SRIs and related drugs are administered most commonly on a continuous daily basis, an intermittent treatment regimen limited to the luteal phase, or one beginning at the onset of symptoms, can be equally or more effective while offering the potential advantages of lower cost and fewer side effects.[127, 130, 132, 133] In some clinical trials, treatment for as few as 3 days was effective.[134, 135] However, some women require higher doses or continuous treatment to achieve benefit.[136–138]

Alprazolam

Alprazolam, a benzodiazepine, is another medication that may be useful in the treatment of PMS and PMDD,[66, 96, 97] although its effectiveness may be limited to the relief of depressive symptoms. Because the drug also has addictive potential, it generally is considered a second-line agent and is best used only intermittently.

Gonadotropin-Releasing Hormone Agonists

The clinical utility of GnRH agonists in the treatment of PMS/PMDD was first demonstrated in 1984.[139] Although now an established treatment,[51, 98] GnRH agonists generally are more effective for relieving irritability and physical symptoms than for the treatment of prominent symptoms of depression or dysphoria.[51, 140] GnRH agonist treatment is associated with the well-known symptoms of hypoestrogenism (e.g., hot flushes), which can be severe, and prolonged use invites longer-term consequences (bone mineral depletion).[141] However, these limitations can be negated largely by simultaneous "add-back" treatment with low doses of estrogen or estrogen and progestin, which does not reduce the overall efficacy of GnRH agonist therapy.[49, 142–144]

Oral Contraceptives

Oral contraceptives are one of the oldest and simplest methods for treatment of PMS/PMDD, based on the idea of substituting a constant hormonal environment for the dynamic cyclic pattern of the normal menstrual cycle. Results achieved with this approach to treatment have been mixed. Early clinical studies found that oral contraceptives helped to relieve breast pain and symptoms of bloating but had no detectable benefits for relieving mood symptoms.[145] More recent studies involving the use of an oral contraceptive containing the progestin drospirenone have observed that treatment can achieve modest improvement in a wide range of symptoms, including behavioral and mood symptoms, particularly when the usual 7-day pill-free interval is shortened to 4 days.[99, 100, 146, 147] There is every reason to believe that any oral contraceptive administered in a similarly extended regimen would produce similar results. Oral contraceptives also might be used in a continuous fashion to achieve a relative hormonal steady state and to eliminate the cycle, and menses, altogether.

Exercise and Relaxation Techniques

There is some evidence to indicate that aerobic exercise,[102, 103] relaxation,[104] and reflexology[105] can help to relieve the symptoms of PMS, but the data suggesting their efficacy is not compelling and may reflect only a placebo response.

Spironolactone

Spironolactone is a potassium-sparing diuretic having structural similarity to steroid hormones and has been used widely in the treatment of PMS. In clinical trials, spironolactone has proven more effective than placebo in relieving symptoms of irritability, depression, bloating,

breast tenderness, and food craving.[107–109] In one placebo-controlled randomized trial, a significant difference in serum androgen levels from the follicular to the luteal phase of the cycle was observed in those who subsequently responded to treatment with spironolactone.[106]

Progesterone

In the past, progesterone treatment by injection or suppository was used commonly in the management of PMS, having been actively promoted by Dalton.[4] Early studies that failed to detect a benefit were criticized for the size of the study population and doses of progesterone that were employed.[148–151] In a study that attempted to negate the placebo response by eliminating any contact with the investigators or any health care providers during the trial, both progesterone and placebo failed to achieve any measurable benefits.[152] Controversies regarding the value of progesterone treatment were largely put to rest by the results of large double-blind, placebo-controlled trials finding that the effects of progesterone (400 mg, 800 mg, 1,200 mg daily) were not different from those of placebo.[110, 153] A meta-analysis including 10 trials of progesterone therapy involving 531 women and 4 trials of progestin therapy involving 378 women found that neither is effective for management of the symptoms of PMS.[111] A recent Cochrane systematic review concluded the available data do not indicate that progesterone is an effective treatment for PMS.[154]

Choice of Treatment

When symptoms are mild and evidence of significant socioeconomic dysfunction is lacking, patients can be advised to consider aerobic exercise as treatment. If symptoms of bloating and fluid retention are prominent, a trial of spironolactone can be justified. Women in need of contraception are logical candidates for treatment with oral contraceptives, with a shorter than usual pill-free interval, or in a daily continuous regimen.

Women who meet strict criteria for diagnosis of PMS or PMDD, including socioeconomic dysfunction, are candidates for treatment with an SRI (fluoxetine, sertraline, paroxetine, venlafaxine), administered daily or only during the luteal phase. Common side effects of SRIs include nausea, jitteriness, and headache. When side effects prove limiting, a trial with a lower dose or alternative drug is warranted. ***Sexual dysfunction, including anorgasmia and diminished sexual interest, is perhaps the most significant potential adverse effect of SRI treatment, and women should be advised of this specific possibility before treatment begins.*** Unfortunately, lower doses often do not eliminate this side effect. Approximately 30–40% of women may not respond to treatment with an SRI over several cycles; a switch to another drug in the class is reasonable and often effective. Those who fail to respond to intermittent luteal phase treatment or to daily therapy may do better with the alternative treatment regimen. When SRI treatment proves unsuccessful, low-dose alprazolam is a logical choice to try, although sedating side effects may limit its usefulness.

In women diagnosed with PMS or PMDD who fail to respond to the treatments usually effective, it is important to consider underlying conditions such as major depression, a generalized anxiety disorder, or substance abuse. In those with severe PMDD unassociated with such disorders, treatment with a GnRH agonist can be considered, preferably administered with low doses of add-back estrogen or estrogen/progestin. In those who respond well to this last-resort therapy, GnRH agonist treatment with add-back can be extended beyond the usual 6 months without risk for loss of bone density.

Dysmenorrhea

Dysmenorrhea is pain with menstruation, usually cramping in nature and centered in the lower abdomen. Dysmenorrhea generally is classified as primary or secondary.[155] ***Primary dysmenorrhea is associated with ovulatory cycles and results from myometrial contractions, in the absence of demonstrable disease. Secondary dysmenorrhea refers to pain during menstruation that is associated with pelvic pathology, such as endometriosis, adenomyosis, or uterine myomas.***

Epidemiology

Dysmenorrhea is one of the most common gynecologic problems in reproductive age women. Primary dysmenorrhea usually begins during adolescence, but only after ovulatory cycles are established; 20–45% of teenage girls are ovulatory by 2 years after menarche, and 80% by 4–5 years.[156]

The overall prevalence of primary dysmenorrhea among adolescent girls is between 60% and 90% and decreases as age increases.[157–160] However, only approximately 15% of adolescent girls seek medical attention for complaints of menstrual pain.[160] In a random sample of 19-year-old women in Göteborg, Sweden, 72% reported dysmenorrhea, 38% regularly used medication, 15% had to limit their daily activity despite use of medication, and 8% missed school or work during every menses.[161] The severity of dysmenorrhea related directly to the volume and duration of menstrual flow. A later survey of the same cohort of women found that the prevalence of dysmenorrhea was decreased to 67% by age 24, with 10% still reporting limitation of daily activities.[162] The severity of dysmenorrhea also was reduced in women who had delivered a child in the intervening years, but not in those who had a miscarriage or abortion, and also was decreased in oral contraceptive users. A Canadian survey in a random sample of over 1,500 menstruating women observed that the prevalence of moderate or severe dysmenorrhea was 60%, resulting in decreased activity in 50% and missed school or work in 17%.[163] In the United States, approximately 60% of menstruating adolescents report dysmenorrhea, causing 14% to miss school regularly.[160] In a longitudinal American study of college women ages 17–19 years, 13% reported severe pain in more than half of their menstrual periods and 42% indicated that dysmenorrhea interfered with daily activities at least once.[164]

Risk factors for dysmenorrhea include a body mass index less than 20, early menarche (before age 12), longer intermenstrual intervals and duration of bleeding, irregular or heavy flow, premenstrual molimina, previous sterilization or history of sexual assault, and smoking.[165] Oral contraceptives, exercise, being married or in a stable relationship, and higher parity decrease the likelihood of dysmenorrhea.[165]

Pathophysiology

Accumulated evidence indicates that primary dysmenorrhea is caused by myometrial ischemia due to frequent and prolonged uterine contractions. Studies of uterine blood flow using Doppler ultrasonography have revealed that uterine and arcuate artery resistance on the first day of menses is significantly higher in women with primary dysmenorrhea than in women without dysmenorrhea, suggesting that constriction of uterine vessels is the proximate cause for pain.[166]

The secretory endometrium contains substantial stores of arachidonic acid, which is converted to prostaglandin $F_{2\alpha}$ ($PGF_{2\alpha}$), prostaglandin E_2 (PGE_2), and leukotrienes during menses. $PGF_{2\alpha}$ always stimulates uterine contractions and is the primary mediator of dysmenorrhea.[167] Endometrial concentrations of $PGF_{2\alpha}$ and PGE_2 correlate with the severity of dysmenorrhea.[168] Treatment with cyclooxygenase (COX) inhibitors decreases prostaglandin levels in menstrual fluid and uterine contractile activity; response curves correlate closely with serum drug levels.[169, 170]

Uterine smooth muscle contractions cause the crampy, spasmodic lower abdominal and low back pain typical of dysmenorrhea and of prostaglandin-induced labor or abortion. In women with primary dysmenorrhea, uterine contractions during menses begin from an elevated level of basal tone (>10 mm Hg), generate higher intrauterine pressures that frequently reach 150–180 mm Hg and can exceed 400 mm Hg, occur more frequently (>4–5/10 min), and are not rhythmic or coordinated.[170] When intrauterine pressure exceeds arterial pressure for a sustained period of time, ischemia results in the production of anaerobic metabolites that stimulate small type C pain neurons, which contributes to the pain of dysmenorrhea. Moreover, $PGF_{2\alpha}$ and PGE_2 can stimulate bronchial, bowel, and vascular smooth muscle contractions, causing bronchoconstriction, nausea, vomiting, diarrhea and hypertension.

Classically, primary dysmenorrhea begins just before or coincident with the onset of menses and declines gradually over the subsequent 72 hours. The menstrual cramps are intermittent, vary in intensity, and usually are centered in the suprapubic region, although some women also experience pain in their thighs and lower back. Typically, the pattern is consistent across cycles. In contrast, women with secondary dysmenorrhea related to pelvic pathology, such as endometriosis, frequently report increasingly severe pain that often occurs at midcycle and during the week preceding menses, in addition to symptoms of deep dyspareunia and dyschezia (painful bowel movements). In those with secondary dysmenorrhea related to uterine myomas, pain results primarily from menorrhagia, with an intensity that correlates with the volume of menstrual flow.

Diagnosis

Primary dysmenorrhea is a clinical diagnosis, based primarily on a history of characteristic symptoms and a physical examination yielding no evidence or suspicion of specific pelvic pathology such as endometriosis, adenomyosis, uterine myomas or chronic pelvic inflammatory disease. In general, laboratory tests, imaging and laparoscopy are not necessary for diagnosis.

A careful menstrual history should include the age at menarche and at onset of dysmenorrhea, intermenstrual interval, volume and duration of flow, and note any symptoms of intermenstrual or premenstrual spotting or staining. The relationship between the onset of pain and the onset of flow, the severity and location of pain, and the presence of any associated nausea, vomiting, diarrhea, back pain, or headache should be determined. The extent to which the pain interferes with daily activities (work, school, or exercise), the use of medications and their effectiveness, any progression in severity over time, and the presence of pain at times other than during menses also should be defined. These historical features generally can distinguish reliably women with primary dysmenorrhea from those with secondary dysmenorrhea.

Whereas women with primary dysmenorrhea typically report the onset of menstrual pain before the age of 25, those with adenomyosis tend to present later, generally after age 35, and often also report noncyclic chronic pelvic pain. Women with endometriosis typically have

pain at times other than during menses, and frequently also report premenstrual spotting, dyspareunia, dyschezia, limited relief from treatment with non-steroidal anti-inflammatory drugs (NSAIDs), and increasing severity over time. Because NSAIDs generally are highly effective in relieving the pain of primary dysmenorrhea, pain that proves refractory to such treatment suggests pelvic pathology.

Women with primary dysmenorrhea usually have a normal pelvic examination. In those with secondary dysmenorrhea related to pelvic pathology, the pelvic examination can be normal but often is not, providing clues to the underlying cause. Adenomyosis often is associated with a bulky, globular, and tender uterus, whereas those with myomas frequently have an enlarged uterus with irregular contours. Women with secondary dysmenorrhea related to endometriosis often have abnormal physical examination findings, which can include thickening, nodularity, or focally tender uterosacral ligaments, lateral cervical displacement due to shortening of one uterosacral ligament, cervical stenosis, or an enlarged ovary due to an endometrioma.[171, 172] Interestingly, red hair, scoliosis, and dysplastic nevi also are observed more frequently in women with endometriosis.[173–175]

The diagnosis of primary dysmenorrhea does not require laboratory tests or imaging. However, transvaginal ultrasonography can be very helpful in identifying uterine myomas, endometriomas, and adenomyosis in women with secondary dysmenorrhea.[176] Although numerous studies have demonstrated that serum CA-125 often is elevated in women with endometriosis, the test has limited clinical utility due to its low negative predictive value.[177, 178]

Treatment

A wide variety of therapies has been proposed for the treatment of dysmenorrhea. These include the application of heat, dietary and vitamin or herbal therapies, exercise, and behavioral interventions, as well as more traditional medications such as NSAIDs and oral contraceptives.

Data from 2 clinical trials suggest that application of a heated abdominal patch or wrap for 8–12 hours daily is more effective than placebo and can be as effective as treatment with an NSAID.[179, 180] Results from a few small clinical studies have suggested that a vegetarian diet,[181] vitamin E,[182, 183] combinations of vitamins (B1, B6, E), and Asian herbal remedies are more effective than placebo, but 2 systematic reviews have concluded that evidence is insufficient to justify recommending dietary and herbal therapies for the treatment of dysmenorrhea.[184, 185] Studies of the effects of exercise on dysmenorrhea have yielded mixed results, with some reporting improvement and others finding that regular exercise worsened symptoms.[186, 187] Although behavioral interventions aimed at changing the way women think about or respond to pain appear to help some women with dysmenorrhea, evidence for their effectiveness derives from small studies in heterogeneous populations and is not compelling.[188]

For both adolescent girls and women with primary dysmenorrhea, NSAIDs are the first treatment of choice.[189–191] There are numerous NSAIDs to choose from, including the proprionic acid derivatives (e.g., naproxen, ibuprofen, and ketoprofen) and fenamates (e.g., mefenamic acid, tolfenamic acid, flufenamic acid, and meclofenamate); all are very effective. Numerous clinical trials have demonstrated that NSAIDs provide effective relief in 70–90% of patients.[192–197] Their efficacy derives from both a decrease in endometrial prostaglandin production and from decreased menstrual flow. The fenamates also block prostaglandin actions.[198] Although some selective COX-2 inhibitors have been approved for treatment of primary dysmenorrhea, their higher cost and greater potential risks suggest their use should be limited to women at high risk for serious gastrointestinal side effects.

NSAID treatment can be started at the onset of menses and continued for the usual duration of pain. Women with severe dysmenorrhea might benefit from starting treatment 1–2 days before menstrual bleeding begins. NSAIDs should be taken with food to minimize common gastrointestinal side effects. The proprionic acid derivatives are a good initial choice because they are inexpensive and available over-the-counter without a prescription. Patients do exhibit variations in response to different NSAIDs. Consequently, if one fails, substituting another in a different drug class is reasonable and appropriate (e.g., changing from ibuprofen to mefenamic acid). Because prostaglandins play a role in ovulation, treatment with NSAIDs has the potential to delay or prevent ovulation.[199–201] However, given that NSAID treatment for dysmenorrhea is remote from ovulation, women attempting to conceive generally can be reassured that treatment will not adversely affect their fertility.[202]

Common NSAID Treatment Regimens for Dysmenorrhea

Drug	*Initial Dose (mg)*	*Maintenance Dose (mg)*
Proprionic acid derivatives		
Ibuprofen	400	400 q 6 h
Naproxen	500	250 q 6–8 h
Naproxen sodium	550	275 q 6–8 h
Ketoprofen	75	75 q 8 h
Fenamates		
Mefenamic acid	500	250 q 4 h
Meclofenamate	100	50–100 q 6 h

Oral contraceptives also are effective treatment for dysmenorrhea. They can be considered a first line agent in sexually active women who require contraception and are a logical alternative for those who do not tolerate or gain sufficient relief from NSAID treatment.[203] The efficacy of oral contraceptives derives from their inhibition of ovulation, thereby decreasing endometrial prostaglandin production, and from the decrease in the volume and duration of flow that results from endometrial attenuation after months of use.[203–206] Oral contraceptives can be used in the standard cyclic fashion (21–24 active pills followed by 4–7 inactive pills), or in an "extended" cyclic manner using one of the newer formulations containing 12 weeks of active pills, followed by 7 pills that are inactive or contain a low dose of estrogen only.[207, 208] All regimens are effective. Extended cycles offer the added advantage of fewer menses, but also are associated with a higher prevalence of unscheduled spotting or bleeding.[209]

The contraceptive ring appears as effective as oral contraceptives for the treatment of dysmenorrhea.[210] Although no studies have focused on the use of depot-medroxyprogesterone acetate for dysmenorrhea, a small study in adolescents found that two-thirds of subjects reported decreased dysmenorrhea during treatment.[211] A decrease in dysmenorrhea also has been demonstrated in trials evaluating the levonorgestrel intrauterine system (device),[212] and a single-rod implantable hormonal contraceptive containing etonogestrel (Implanon).[213]

Women who fail to respond to treatment with NSAIDs and/or hormonal contraceptives and those having recurrent or worsening pain merit re-evaluation to exclude causes of secondary dysmenorrhea, such as endometriosis.[191] In a study of women with pelvic pain who failed to obtain adequate relief from treatment with NSAIDs, the large majority had demonstrable endometriosis at laparoscopy.[214] These observations suggest that women with severe dysmenorrhea who fail to respond adequately to treatment with NSAIDs or oral contraceptives are candidates for diagnostic laparoscopy. Women with endometriosis should have their disease ablated, to the extent possible. Post-operative treatment with a GnRH agonist (e.g., leuprolide acetate depot 3.75 mg every 4 weeks) is effective for the management of those (with or without

endometriosis) who have persistent dysmenorrhea. Although empirical treatment with GnRH agonists, based on clinical criteria,[214] has been promoted as an alternative to surgery in women having a high likelihood of endometriosis, surgical treatment offers the advantages of a specific diagnosis, immediate pain relief, and better informed longer-term management.

Menstrual Migraine

Headaches are very common, but the cause is rarely serious. Most headaches result from vasodilation, muscle contraction, or psychological stress. Menstrual headaches include all headaches that are temporally related to menses, beginning before or during menstruation.[215] For many women with PMS, headache is part of their cyclic symptom complex. The focus here is on pure menstrual migraine, describing headaches that occur exclusively in association with menses, and on "menstrually-related migraine," describing patients who have migraines with menses but also at other times in the cycle.[216]

Pathophysiology

The traditional "vascular theory," holding that migraine and cluster headaches result from vasodilation and the preceding aura from vasoconstriction, is no longer considered viable.[217] Instead, vasodilation is probably an epiphenomenon that results from instability in central neurovascular control. ***Current concepts of the pathophysiology of migraine center on the trigeminovascular system.*** The large cerebral vessels, those of the pia and dura mater, and the large venous sinuses are innervated by sensory afferents originating from the trigeminal ganglion and upper cervical dorsal roots.[218] The two converge at the trigeminal nucleus caudalis, explaining the distribution of migraine pain, which typically involves the front and back of the head and upper neck. When the trigeminal ganglion is stimulated, vasoactive neuropeptides (substance P, calcitonin gene-related peptide, neurokinin) are released,[219] causing neurogenic inflammation, which results in vasodilation, extravasation of plasma proteins, and pain. In turn, neurogenic inflammation may cause sensitization, describing a process in which neuronal response thresholds decrease, the magnitude of response increases, and receptive fields expand.[220–222] The sensitization phenomenon helps to explain some of the clinical symptoms of migraine, which include a worsening with coughing or bending, hyperalgesia (increased sensitivity to painful stimuli) and allodynia (pain caused by normally non-painful stimuli). The classical aura associated with migraine headaches (visual or auditory symptoms, nausea or vomiting, paresthesias) is attributed to "cortical spreading depression," describing a self-propagating wave of neuronal and glial depolarization that spreads across the cerebral cortex.[217] Migraine without aura may involve areas of the brain where depolarization is not consciously perceived.[223]

The incidence of migraine headache increases significantly between the ages of 15 and19 years, peaks in women in their late 30s to early 40s, and falls after menopause.[224–227] Up to 70% of women with migraine headaches observe an association with menses; 7–21% have pure menstrual migraine and the remainder has menstrually-related migraines.[228] Compared with headaches that occur at other times of the month, menstrual migraines usually are more resistant to treatment, generally not associated with aura, of longer duration, and associated with more functional disability.[229–232]

The association between changes in the clinical course of migraine headaches and reproductive milestones, such as menarche, pregnancy, and menopause, suggests that ovarian steroid hormones are involved in their pathophysiology. Migraine appears to be

associated with declining hormone levels, as occurs at the end of the normal menstrual cycle, postpartum, and during scheduled pill-free weeks in women using oral contraceptives.[233, 234] ***Biochemical and genetic evidence suggests that menstrual migraine is triggered primarily by declining estrogen levels.*** Estrogen has a number of actions within the central nervous system. Its effects on the serotonergic neurotransmitter system, in particular, may explain its association with migraine.[228] In women, serotonergic tone correlates with estrogen levels; when estrogen levels decline, serotonin concentrations also fall, due both to a decrease in production and to an increase in clearance. Estrogen also may modulate the balance between excitatory and inhibitory neurotransmission via its effects on other chemical mediators, such as nitric oxide, magnesium, or prostaglandins.[228, 235]

Evaluation

Although headaches are common, they also can be an indication of a serious condition, such as an intracranial space-occupying mass, vascular lesion, infection, or a metabolic disease. Chronic headaches should be characterized according to their location, quality, changes over time, and associated symptoms and signs.

Common tension-type headaches are classified as episodic (<15 per month) or chronic (≥ 15 per month).[236] Tension-type headaches were named originally for their suspected cause, excessive stress or tension, resulting in muscle-contraction, but that traditional explanation is no longer considered viable.[237] Although their pathogenesis remains unclear, current models center on heightened sensitivity of pain pathways in the central nervous system, and possibly in the peripheral nervous system,[238–240] nitric oxide,[241] and genetic factors.[242, 243] Tension-type headaches generally are mild to moderate in intensity, bilateral, and usually described as "dull," "pressure," "a tight cap," "band-like," or a sense of weight on the head and shoulders. They can vary in intensity and frequency, but generally do not exhibit progression over time.

When headaches are cyclic, completely absent for periods of time, and exhibit characteristic features, they generally can be ascribed comfortably to migraine. ***Migraines are characteristically throbbing in nature, frequently but not always are preceded by a prodrome, and typically begin slowly, rising to a crescendo over 1–2 hours. Most patients with migraines have had numerous similar headaches in the past.*** Migraines can be precipitated by stress, alcohol, or tyramine- and tryptophan-rich foods (e.g., red wine, chocolate, ripe cheeses). Classic migraine is now known as "migraine with aura" and common migraine as "migraine without aura." ***Menstrual migraine headaches typically are migraines without aura.***[227]

Headaches that are sudden in onset, persistent headaches that worsen over time, and those that become severe shortly after their onset or are associated with changes in mental status warrant careful and thorough evaluation. Focal neurologic symptoms, other than a typical visual or sensory aura, suggest a mass lesion, arteriovenous malformation, or collagen vascular disease. Fever is not a characteristic of migraine headache and suggests an intracranial, systemic, or local infection (e.g., paranasal or mastoid sinusitis). New headaches in women over 50 or in patients with cancer or human immunodeficiency virus (HIV) infection suggest pathology. Chronic nasal congestion suggests sinusitis. Blurred vision, headaches upon waking in the morning that improve after sitting or standing, double vision or loss of balance, and worsening headaches associated with chronic nausea suggest increased intracranial pressure. Sudden unilateral loss of vision suggests optic neuritis. ***Brain imaging generally is indicated for patients having a recent significant change in the pattern, frequency, or severity of headaches, worsening headaches despite treatment, focal neurologic symptoms or signs, or an onset of headaches with exertion or after the age of 40 years.***[244]

Treatment

The treatment for acute menstrual migraine is similar to that for common migraine; serotonin agonists (known as triptans), such as sumatriptan (50–100 mg), rizatriptan (10 mg), and frovatriptan (2.5 mg) are effective in aborting migraine headaches.[245–247] Triptans inhibit the release of vasoactive peptides, promote vasoconstriction, and inhibit neurotransmission in the trigeminal nucleus, which blocks afferent input to second order neurons. They also help in relieving associated nausea and photophobia. When acute abortive therapy is required repeatedly or proves inadequate, prevention strategies, both hormonal and non-hormonal, merit consideration.

Hormonal prevention therapies are aimed at minimizing or eliminating the premenstrual decline in serum estrogen levels.[248] They are a logical choice for patients with other menstrual pathology, such as irregular cycles, dysmenorrhea, or menorrhagia, and generally are preferred over non-hormonal prevention therapies. If menstrual migraines represent a reaction to cyclic changes in circulating levels of sex steroid hormones, it's logical to try to minimize or eliminate cyclicity, which can be achieved with continuous, daily oral contraceptives. For women who prefer to have menses, extended cycle oral contraceptives containing 9–12 consecutive weeks of active pills greatly reduce their frequency; standard formulations also can be used, but in either case, estrogen supplementation should be provided during any scheduled placebo week to prevent the estrogen withdrawal migraine.[249] The use of oral contraceptives in women who have migraine with aura has been controversial due to concerns about a potential increase in the risk for stroke. Whereas European studies have suggested that oral contraceptives may increase the risk,[250–252] American studies have not,[253, 254] probably because low-dose pills are more prevalent and smoking over age 35 has been regarded as a relatively strong contraindication to use of oral contraceptives in the United States (discussed in detail in Chapter 22). Nonetheless, both the World Health Organization and the American College of Obstetricians and Gynecologists currently discourage the use of oral contraceptives in women over age 35 whose migraines are associated with focal neurologic symptoms or signs.

In women with contraindications to use of oral contraceptives, estrogen supplementation still can be used to buffer the decline in estrogen levels at the end of the cycle, beginning just before the onset of menses and continuing for a total of approximately 7 days. In clinical trials, a transdermal estradiol patch (0.1 mg/day) or gel (1.5 mg/day) was effective,[255, 256] but must be continued until endogenous estrogen levels rise again or treatment merely postpones estrogen withdrawal and related migraine.[256] A long-acting GnRH agonist is another alternative. Clinical studies in women with menstrual migraine indicate that treatment with leuprolide acetate can decrease markedly the frequency of headaches, even when add-back estrogen/progestin treatment is provided.[257, 258]

The prevalence of migraine is relatively high during the perimenopause, especially in women with a history of menstrual migraine.[259] In the Women's Health Study, 11% of more than 17,000 postmenopausal female health professionals reported migraine; current users of hormone therapy were more often affected than never users (OR = 1.42; 95% CI = 1.24–1.62) and prevalence correlated with the dose of estrogen therapy.[260] The treatment of migraine in menopausal women is not different from that in premenopausal women. Again, hormone regimens that minimize changes in circulating estrogen levels are preferable; therefore, continuous treatment is a better choice than cyclic therapy. Addition of a progestin, when needed, does not affect the frequency of migraines.[260]

Non-hormonal therapies for menstrual migraines include treatment with an NSAID or a triptan. Treatment with NSAIDs, beginning 7 days before onset and continuing through the end of menses, can help to decrease the frequency, duration, and severity of menstrual migraine.[261] Triptan treatment, beginning 2–3 days before menses and continuing for a total

of 5–6 days, effectively prevents or decreases the severity of headaches in the majority of patients. Proven regimens include sumatriptan 25 mg 3 times daily,[262] naratriptan 1 mg twice daily,[263] and frovatriptan 2.5 mg daily or twice daily).[264]

Catamenial Epilepsy

In ancient times, catamenial epilepsy was attributed to the moon, giving rise to the word, "lunatic."[265] The word, "catamenial," derives from the Greek word *katomenios*, meaning "monthly." Catamenial epilepsy describes seizures that are clustered around specific points in the menstrual cycle, usually during the perimenstrual or periovulatory intervals. ***The diagnosis of catamenial epilepsy is based on demonstrating the temporal relationship between menstruation and seizure activity; a 2-fold or greater increase in seizure frequency during a particular phase of the menstrual cycle generally is considered as evidence of catamenial epilepsy.***[266, 267] Based on that definition, catamenial epilepsy affects from 30% to 60% of women with epilepsy.[268–270] Given that epilepsy affects an estimated 1.3 million women in the United States,[271, 272] approximately 400,000 women with epilepsy experience catamenial seizures, which often are quite resistant to treatment.[273]

Pathophysiology

Catamenial seizures are associated with every type of epilepsy but are more common among women with focal epilepsy (e.g., temporal lobe epilepsy) than in those with generalized epilepsy.[274] Their specific cause is still unclear. Hypotheses include fluctuations in antiepileptic drug levels, changes in water and electrolyte balance, and variations in ovarian steroid hormone secretion.[267] Cyclic changes in estrogen and progesterone across the menstrual cycle are widely regarded as playing a central role.

Estrogens generally are considered as proconvulsant, although their effects depend on the duration of treatment, dosage, the mode of administration, and the seizure model. Estrogens affect neuronal excitability via modulation of gene expression, regulation of neurotransmitter release, and direct interaction with neurotransmitter receptors. Estrogen acts on neurons within the limbic system, cerebral cortex and other regions important for seizure susceptibility.[275] Estrogen also interacts with neurotrophins, which enhance hippocampal excitability. In women with epilepsy, seizure susceptibility correlates with the estrogen/progesterone ratio, which peaks in the premenstrual and preovulatory intervals. Although serum estradiol concentrations in women with catamenial epilepsy are similar to those in normal controls across the menstrual cycle, evidence suggests that progesterone levels are lower and estrogen/progesterone ratios are higher.[270, 276] Seizure frequency decreases during the midluteal phase when progesterone levels are highest and increases in the premenstrual phase when progesterone levels fall and the estrogen/progesterone ratio increases.

Progesterone generally is considered to have an anticonvulsant action, by lowering seizure susceptibility. Changes in serum progesterone levels have been correlated directly with catamenial seizures,[270, 276] and evidence from studies in mice supports the concept that 5α-reduced metabolites of progesterone, particularly allopregnanolone, are responsible for its antiseizure activity.[267, 277] Neurosteroids like allopregnanolone are synthesized locally within the brain, both *de novo* and from circulating steroid precursors, and modulate neural excitability. Allopregnanolone potentiates the action of gamma aminobutyric

acid (GABA), the major inhibitory neurotransmitter in the brain, via modulation and direct interaction with GABA-A receptors.[278]

The interactions between gonadal steroid hormones, seizures, and antiepileptic drugs are complex. Whereas gonadal steroid hormones can affect seizure susceptibility, seizures may disrupt patterns of steroid hormone secretion, and antiepileptic drugs can affect sex steroid hormone levels by altering their metabolism. In adolescents with epilepsy, the incidence of generalized tonic-clonic seizures often increases during puberty,[279] when the levels of steroid hormones increase and menses begin. At menopause, seizure activity increases in some women, but decreases or remains unchanged in others.[280–282] Hormone therapy can increase seizure activity in postmenopausal women, particularly in those having a history of catamenial epilepsy.[283, 284] On the other hand, the prevalence of menstrual disorders such as polycystic ovary syndrome is increased,[285] and fertility is decreased among women with epilepsy,[286] suggesting that seizures may predispose to reproductive dysfunction, possibly by altering the pattern of hypothalamic GnRH release. The association between epilepsy and reproductive dysfunction also might reflect the effects of antiepileptic drugs on steroid hormone metabolism. Some, such as phenytoin, carbamazepine, and phenobarbital, induce hepatic cytochrome P450 enzymes, which can accelerate the metabolism of steroid hormones that share common metabolic pathways.[287, 288] They also can increase serum SHBG concentrations, thereby further decreasing the concentrations of free or biologically active steroids. Other antiepileptic drugs, such as sodium valproate, inhibit hepatic enzymes, which may increase bioactive steroid hormone levels. However, there is no direct evidence linking catamenial epilepsy with specific antiepileptic drugs. The extent to which all of these competing mechanisms contribute to catamenial epilepsy remains to be determined.

Some studies have suggested that oral contraceptives might increase seizure activity, but most have observed no effect,[283, 289, 290] or a decrease in seizure frequency.[291, 292] Although certain antiepileptic drugs have the potential to reduce the efficacy of oral contraceptives by accelerating their metabolism,[283, 293] there is no evidence the effect is clinically important.[294, 295, 296, 297] Conversely, oral contraceptives can decrease circulating concentrations of some antiepileptic drugs via the same mechanism, which may increase the risk of seizures;[298] evidence indicates that women receiving treatment with lamotrigine or valproic acid may need dosage adjustments if also using oral contraceptives.[297, 299–301]

Treatment

Currently there is no specific treatment for catamenial epilepsy. Antiepileptic drugs are the mainstay for its management. However, approximately one-third of women with catamenial seizures require treatment with more than one drug, partly because catamenial epilepsy often proves refractory to conventional medications. At least in theory, gabapentin, levetiracetam, tiagabine, zonisamide, and pregabalin are attractive choices, because they do not induce hepatic enzymes. Acetazolamide, a potent inhibitor of carbonic anhydrase, has been used empirically for years in the treatment of catamenial epilepsy, but there are few direct studies attesting to its effectiveness;[302] drug tolerance, requiring progressive dosage escalation, is a common problem. Benzodiazepines, such as clonazepam and clobazam, increase GABA-A receptor activity and have broad-spectrum antiseizure activity. Clobazam (20–30 mg daily), administered intermittently from 2–4 days before menses through the first 3 days of bleeding, has proven effective for the treatment of catamenial epilepsy; intermittent treatment helps to avoid problems with tolerance, which otherwise is common.[303] Limited evidence suggests that treatment with lamotrigine (25–200 mg daily) can decrease or eliminate catamenial seizures, at least in some women.[304] Ganaxolone, a synthetic neurosteroid (a 3β-methyl analog of allopregnanolone) that has shown promise in

preclinical models of catamenial epilepsy and preliminary investigations in women,[305] is now in clinical trials.[306]

Hormonal therapies for the management of catamenial seizures also deserve consideration, particularly for women who prove resistant to antiepileptic drugs. Treatment with depot-medroxyprogesterone acetate in doses that typically eliminate menses (e.g., 150 mg intramuscularly every 3 months) can improve seizure control in many women.[307] Cyclic natural progesterone also has been demonstrated effective for the treatment of catamenial seizures; 100–200 mg, administered orally or vaginally three times daily between days 15 and 28 of the cycle, decreased seizure frequency by approximately 50–75%.[308, 309] Its apparent efficacy and safety have prompted an ongoing trial sponsored by the National Institutes of Health,[310] but progesterone is not yet recognized as an approved treatment for catamenial epilepsy. Evidence from studies in animal models and clinical data suggest the antiseizure effects of progesterone derive from its metabolic conversion to neurosteroids, primarily allopregnanolone.[311, 312] Continuous oral contraceptives, the progestin-only minipill,[313] and GnRH agonists with add-back estrogen/progestin therapy also have demonstrated some effectiveness in small numbers of women with catamenial epilepsy.[314–316]

In sum, catamenial epilepsy is a complex and multifaceted condition. Ovarian hormones play a central role but the exact cause is unknown. Evidence indicates that estrogen, progesterone, and endogenous neurosteroids are involved in its pathophysiology, but no specific hormonal dynamic predisposing to seizures has been identified. Neurosteroid withdrawal may be the critical factor that increases seizure susceptibility during the perimenstrual interval, acting via changes in central GABA-A receptors. Conventional antiepileptic drugs are not effective in most patients with catamenial seizures, possibly because they can alter sex steroid concentrations and metabolism. Hormonal therapies such as progesterone, oral contraceptives, and GnRH agonists, have proven efficacy. Finally, newly developed synthetic neurosteroids have shown considerable promise in animal studies of catamenial epilepsy and may offer a specific treatment in the future.

Premenstrual Asthma

Approximately 20–40% of women with asthma have an increase in symptoms associated with menstruation.[317–319] Even asthmatics not aware of a link to menses demonstrate a worsening of pulmonary function during menstruation.[320, 321] Women with hormonally-triggered asthma generally tend to have more severe asthma than those whose asthma is not affected by changes in hormones. The mechanism is unknown, but prostaglandin release, changes in the immune system, and a direct effect of declining estrogen and progesterone levels on bronchial smooth muscle all have been suggested.

The best method for treating premenstrual asthma has not been established. Treatment with estrogen (micronized estradiol, 2 mg orally daily) has been reported to improve symptoms and measures of pulmonary function.[321] A randomized clinical trial found no difference between the effects of estrogen and placebo, although the subjects all had mild asthma generally under good control.[322] Another study found that administration of intramuscular progesterone helped to ameliorate premenstrual asthma.[323] The logic of eliminating menstrual periods by daily treatment with oral contraceptives, intramuscular depot-medroxyprogesterone acetate, or treatment with a long acting GnRH agonist is tempting, but there are no available data demonstrating their efficacy. Alternatives include a rescue inhaler, when necessary, or a leukotriene modifier, such as montelukast, zafirlukast, or zileuton.

Catamenial Pneumothorax, Hemothorax, and Hemoptysis

Thoracic endometriosis syndromes are closely linked to the presence of pelvic endometriosis. In a retrospective analysis of 110 cases, the mean age at presentation of thoracic endometriosis was 35 ± 0.6 years, with a range from 15 to 54 years.[324] Catamenial pneumothorax was the most common presentation, occurring in 80/110 patients (73%); catamenial hemothorax occurred in 15 (14%), catamenial hemoptysis in 8 (7%) and lung nodules were observed in 7 (6%). Fifty-one of 61 patients who underwent laparoscopy or laparotomy had evidence of pelvic endometriosis. Pleural implants were observed in less than 15% of those who had thoracostomy or thoracotomy, and diaphragmatic defects, parenchymal cysts or blebs were observed in approximately 25% of cases.[324]

The most plausible explanations for catamenial pneumothorax, hemothorax, and hemoptysis are peritoneal to pleural transfer of endometrial tissue through diaphragmatic defects[325, 326] and microembolization through pelvic veins.[324] A review of 154 cases of catamenial pneumothorax treated surgically found that 16% of patients had demonstrable diaphragmatic perforations in the absence of thoracic endometriosis and 12% had visible diaphragmatic endometriosis associated with one or more perforations.[327]

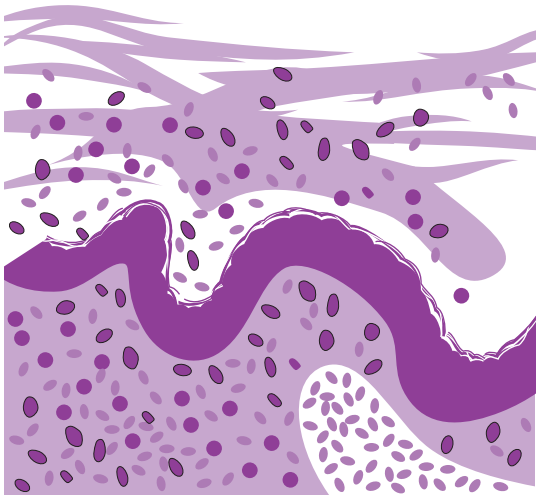
In affected women, symptoms of thoracic endometriosis typically arise within 24–48 hours after the onset of menses. Chest pain is the most common symptom, occurring in 90% of patients, and one-third have dyspnea. Pneumothorax occurs most commonly on the right side and usually is small to moderate in size; hemothorax also usually is right sided.[328, 329] Women with endobronchial or pulmonary parenchymal endometriosis usually present with catamenial hemoptysis.

The diagnosis of thoracic endometriosis should be suspected in reproductive aged women presenting with recurrent chest pain, pneumothorax or hemoptysis during menses. Although not required, diagnosis can be established by pleural fluid cytology,[330] needle aspiration of lung masses,[331] cytology performed on bronchoscopic aspirations,[332] or video thoracoscopy.[333] In women with catamenial pneumothorax, chest computerized tomography (CT) can reveal bullae, cavities, or fibrosis, and in those with hemoptysis, can demonstrate small parenchymal nodules, which may be visible only during menstruation.[334, 335]

The initial treatment of patients with symptomatic thoracic endometriosis is the same as for others with pneumothorax, hemothorax, or hemoptysis. Longer-term, successful treatment requires suppression or excision of thoracic endometrial implants, prevention of re-seeding from the pelvis, and prevention of air leakage across diaphragmatic perforations. Although hormonal suppressive treatments (e.g., oral contraceptives, progestins, danazol, GnRH analogs) generally are considered first line therapy, the recurrence rate is greater than 50%.[324] When they fail, surgical treatment is indicated. Direct inspection of the pleura via video thoracoscopy or thoracotomy can identify diaphragmatic endometrial implants or perforations, which can be excised or closed, usually followed by chemical pleurodesis using talc poudrage (blowing powder or an aerosol into the pleural cavity) or pleural abrasion.[327, 329] Surgical treatments generally are very effective for preventing recurrent catamenial pneumothorax or hemothorax, but some who continue to have cyclic chest pain due to pleuropulmonary implants may require longer-term medical suppressive therapy.[336]

All references are available online at: http://www.clinicalgynendoandinfertility.com

15 Abnormal Uterine Bleeding



Abnormal uterine bleeding is the single most common complaint that reproductive age women bring to their clinicians. All clinicians who provide primary care for women must therefore be familiar with its causes and have an organized, logical approach to the evaluation and treatment of the problem.

Anovulatory or dysfunctional uterine bleeding describes the spectrum of abnormal menstrual bleeding patterns that can occur in anovulatory women who have no medical illness or pelvic pathology. The mechanisms involved in anovulatory bleeding vary, but each reflects an abnormal pattern of steroid hormone stimulation that deviates from the sequence characterizing the normal ovulatory menstrual cycle. The key to successful clinical management of dysfunctional bleeding is to recognize or identify which mechanism is operating or responsible. Anovulatory bleeding can be effectively and confidently managed with medical treatment regimens based on sound physiologic concepts. The treatment regimens described in this chapter are time-tested and designed to achieve two specific but interrelated goals. The first is to reverse the abnormalities of endometrial growth and development that result from chronic anovulation and predispose to excessive and prolonged menstrual flow. The second is to induce or restore cyclic predictable menses of normal volume and duration.

Bleeding related to a wide assortment of pathology inside and outside of the reproductive tract can masquerade as anovulatory bleeding. A careful menstrual history and physical examination usually provide most of the information needed to distinguish anovulation from other causes of abnormal bleeding. When pathology is strongly suspected or treatment for presumed anovulatory bleeding fails, additional evaluation is indicated but is also straightforward.

Terminology

Clinicians use a wide variety of terms to describe abnormal patterns of menstrual bleeding that do not always mean or communicate the same thing to others. Traditional terms having Greek or Latin roots still are used widely to describe different abnormalities relating to the frequency, regularity, duration and volume of menses.

Traditional Terms Describing Abnormalities of Menstrual Bleeding

- **Amenorrhea** — **absent menses**
- **Oligomenorrhea** — **infrequent menses, occurring at intervals > 35 days**
- **Polymenorrhea** — **frequent menses, occurring at intervals < 24 days**
- **Metrorrhagia** — **menses occurring at irregular intervals**
- **Menorrhagia *or* Hypermenorrhea** — **abnormally long or heavy menses, lasting > 7 days or involving blood loss > 80 mL**

Although the definitions above are reasonably well established, the terms are not always used or understood accurately.[1, 2] For example, in the United States, the term ***abnormal uterine bleeding*** generally describes all abnormal patterns of bleeding that may result from a wide variety of causes, including anovulation, pregnancy, uterine pathology, and coagulopathies.[3] The term ***dysfunctional uterine bleeding*** is synonymous with anovulatory bleeding, in the absence of pregnancy or any demonstrable pathology (a diagnosis of exclusion), and the term ***menorrhagia*** describes regular, heavy or prolonged bleeding. However, in other countries, dysfunctional uterine bleeding and menorrhagia often are used to describe both ovulatory (regular) or anovulatory (irregular) bleeding that is heavy or prolonged.[1] The confusion surrounding the exact meaning of the traditional terms has spurred a call to abandon them, in favor of simple terms that can be understood by patients and translated easily into languages other than English, with the ultimate goal of improving communication among health care providers, investigators, and patients. To that end, recommendations arising from an international consensus conference proposed terms to describe the most important features of menstrual bleeding during the reproductive years, as follows[1]:

Characteristic	*Descriptive Terms*	*Normal Limits*
Frequency of menses	Frequent	<24 days
	Normal	24–38 days
	Infrequent	>38 days
Regularity (cycle to cycle variation)	Absent	———
	Regular	± 2–20 days
	Irregular	> 20 days
Duration of flow	Prolonged	> 8 days
	Normal	4-8 days
	Shortened	< 4 days
Volume of monthly blood loss	Heavy	>80 mL
	Normal	5–80 mL
	Light	< 5 mL

The suggested normal limits for frequency, regularity, and duration of menstrual flow were based on the 5th and 95th percentiles for data drawn from population studies.[4–6] As such, they are influenced by the prevalence of common anovulatory disorders, such as the polycystic ovary syndrome, in a given population. Consequently, the population-based norms are

wider than the generally accepted norms for menstrual frequency (24–35 days), regularity (± 5 days variation), and duration (2–7 days) among ovulatory women. The normal limits for the volume of menstrual blood loss were based primarily on measurements of hemoglobin loss in a Swedish community.[4] The expectation is that a structured menstrual history can clarify the details needed to categorize a patient's complaint in clear and simple terms (e.g., irregular, heavy menstrual bleeding).[7, 8]

Although the effort to simplify and standardize the terminology used to describe menstrual abnormalities is reasoned and laudable, the adoption of a new nomenclature likely will be slow because the traditional terms, however confused, are firmly entrenched.

Normal Menstrual Bleeding

It is ovulation or, more specifically, the organized sequence of endocrine signals that characterizes the ovulatory cycle, that gives menses regularity, predictability, and consistency. The endocrinology of the normal menstrual cycle is discussed in detail in Chapter 6. Only the most basic concepts and characteristics are summarized here, with a focus on the major events and mechanisms that control the endometrial cycle and the volume and duration of menstrual flow.

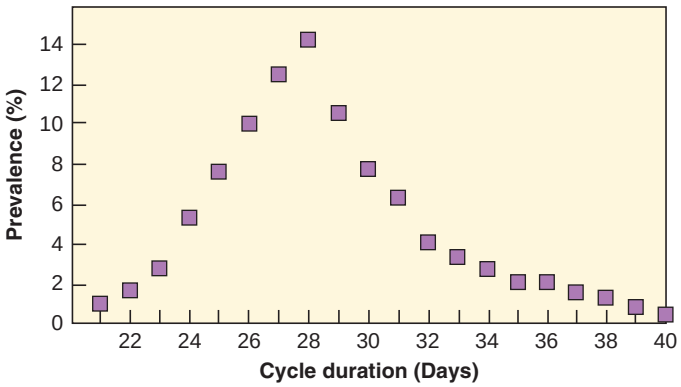
During the follicular phase of the normal ovarian cycle (corresponding to the proliferative phase of the endometrial cycle), estrogen levels rise, slowly at first and then more rapidly, as the dominant ovarian follicle emerges, grows, and matures. In response to that estrogen, the functional layer of the endometrium regrows, after having been shed during the preceding menses. After ovulation, the corpus luteum derived from the ovulatory follicle continues to produce estrogen, but now and more importantly, also progesterone. During the luteal phase of the ovarian cycle (corresponding to the secretory phase of the endometrial cycle), estrogen and progesterone levels rise together as the corpus luteum grows to maturity. In response to the combined actions of estrogen and progesterone, the endometrium transforms and organizes in preparation for the anticipated arrival and implantation of a conceptus. If pregnancy and rapidly rising levels of human chorionic gonadotropin (hCG) do not come to its "rescue," the corpus luteum regresses spontaneously in a form of preprogrammed cell death. As it does, estrogen and progesterone levels fall steadily, eventually withdrawing the functional support for the endometrium. Menses begin, marking the end of one endometrial cycle and the beginning of another.

From the endometrial perspective, the endocrine features of the ovarian cycle are quite simple; the quantities of hormones produced are not nearly as important as the sequence in which they appear: estrogen, followed by estrogen and progesterone, followed by withdrawal of both hormones. Of all the different hormone effects on the endometrium, estrogen-progesterone stimulation produces the most stable endometrium, and their combined withdrawal yields the most consistent menstrual characteristics. ***The sequence is so controlling that most ovulatory women have a pattern, volume, and duration of menstrual flow they recognize as their own and come to expect, very often accompanied by an equally consistent and predictable pattern of premenstrual molimina (bloating, breast tenderness, mood swings).*** Even slight deviations from the usual pattern in the timing, amount, or length of flow can cause concern. Careful attention to the finer details of the menstrual history can be very helpful in distinguishing anovulatory bleeding from other causes.

Variations in menstrual flow and cycle length are common at the extremes of reproductive age, during the early teenage years and those preceding the menopause. Menstrual cycles often are irregular for the first 12–18 months after menarche, due to immaturity

of the hypothalamic-pituitary-ovarian axis.[9, 10] In a study conducted by the World Health Organization, the median length of the first cycle after menarche was 34 days; almost 40% of cycles were longer than 40 days and fewer than 10% were less than 20 days.[11] Cycles remain relatively long for the first 5–7 years after menarche, thereafter decreasing gradually in length and becoming more regular.[11] The prevalence of anovulatory cycles is higher in women under age 20 and over age 40.[12, 13] Menstrual cycle characteristics generally do not change appreciably during the reproductive years,[6] although overall cycle length and variability slowly decrease. On average, mean cycle length and range reach their lows at about age 40–42.[6, 14] Over the subsequent 8–10 years before the menopause, the trend is reversed; both average cycle length and variability increase steadily as ovulation becomes less regular and frequent.[5, 14–16] Mean cycle length is greater in women at the extremes of body mass and composition; both high and low body mass index (BMI), body fat mass, and body lean mass are associated with an increased mean cycle length.[17, 18]

In general, variations in cycle length reflect differences in the length of the follicular phase of the ovarian cycle. Women who have a 25-day cycle ovulate on or about cycle day 10–12 and those with a 35-day cycle ovulate approximately 10 days later. Within a few years after menarche, the luteal phase becomes extremely consistent (13–15 days in duration) and remains so until the perimenopause.[5, 14] At age 25, over 40% of cycles are between 25 and 28 days in length, and between age 25 and 35, over 60% are. ***Although 28 days is the most commonly reported intermenstrual interval, only approximately 15% of cycles among reproductive aged women actually are 28 days in length.*** Less than 1% of women have a regular cycle lasting less than 24 days or more than 35 days.[19] Most women have cycles that last from 24 to 35 days, but at least 20% of women experience irregular cycles.[6]



The usual duration of menstrual flow is 4–6 days, but for some women (approximately 3%) menses may last as few as 2 days or as many as 7 days.[20] The average volume of menstrual blood loss is approximately 30 mL;[4] greater than 80 mL is abnormal. Flow can be excessive without being abnormally long because most of menstrual blood loss occurs during the first 3 days.[21, 22]

Women who menstruate more often than every 24 days or less often than every 35 days deserve evaluation,[5, 6] as do those who consistently flow for more than 7 days and women with monthly menstrual blood loss exceeding 80 mL. Any of these abnormal patterns can result in anemia that also requires treatment.[23, 24] The intermenstrual interval and duration of menses are relatively easy to determine, but the volume of menstrual blood loss is difficult to measure. The correlation between perceived and actual blood loss is relatively poor.[25] In population-based studies, one-fourth to one-third of women with normal periods considered their menstrual blood loss excessive, and 40% of those with documented menorrhagia (blood loss > 80 mL) described their menses as light or moderate.[4, 26] Complaints of heavy menstrual bleeding appear to relate more to perceived interference with daily function than to actual blood loss,[27] and evidence indicates that psychosocial factors may

have significant influence on those perceptions; the incidence of depression and anxiety is increased among women with complaints of heavy menstrual bleeding.[28–30]

Mechanisms Controlling Onset and Cessation of Normal Menstruation

A conceptual understanding of the mechanisms involved in the onset and cessation of normal menstrual bleeding provides both the foundation and the context for understanding the pathophysiology of anovulatory bleeding.

The classic concepts of normal menstruation derived primarily from direct observations of the cyclic changes in endometrium transplanted from the uterus to the anterior chamber of the eye in nonhuman primates; vascular events played the key role in the explanation for how menses both began and ended.[31, 32] Basically, menstruation was envisioned as ischemic necrosis of the endometrium caused by vasoconstriction of the spiral arterioles in the basal layer, triggered by withdrawal of estrogen and progesterone. Similarly, the end of menses was explained by longer and more intense waves of vasoconstriction, combined with coagulation mechanisms activated by vascular stasis and endometrial collapse, aided by rapid re-epithelialization mediated by estrogen derived from the emerging new follicular cohort.

The results of more contemporary investigations do not support the classic hypoxia theory of menstruation. Perfusion studies in women have failed to demonstrate reduced endometrial blood flow just before menses.[33] Hypoxia-inducible factor (HIF)-1, a nuclear protein that activates gene transcription in response to reduced cellular oxygen (the earliest known marker of response to hypoxia), is barely detectable and not widely distributed in human premenstrual endometrium cultured under hypoxic conditions.[34] Histologically, early menstrual endometrium exhibits focal necrosis, inflammation, and coagulation rather than the diffuse hyalinization or coagulation necrosis that would be expected to result from vasoconstriction and hypoxia.[35] Slowly but surely over the last decade or so, the operational paradigm for menstruation has shifted. ***Instead of vascular events, the central theme of the new model of the initiation of menstruation is an enzymatic autodigestion of the functional layer of the endometrium and its subsurface capillary plexus, possibly extending to the spiral arteriolar system in the basal layer.[35] The classic concept of the mechanisms that end normal menstruation is essentially unchanged; coagulation mechanisms, local vasoconstriction, and re-epithelialization all contribute to hemostasis in the menstrual endometrium with vascular events playing the key role.***

The enzymatic degradation of the endometrium triggered by estrogen-progesterone withdrawal involves a number of different but interrelated mechanisms including the release of intracellular lysosomal enzymes, proteases from infiltrating inflammatory cells, and the actions of matrix metalloproteinases. In the first half of the secretory phase, acid phosphatase and other potent lytic enzymes are confined to intracellular lysosomes, their release inhibited by progesterone via stabilization of lysosomal membranes. As estrogen and progesterone levels fall in the days preceding menses, lysosomal membranes destabilize and the enzymes within are released into the cytoplasm of epithelial, stromal, and endothelial cells, and eventually, into the intercellular space. These proteolytic enzymes digest their cellular constraints as well as surface membranes and desmosomes (intercellular bridges). In the vascular endothelium, their actions result in platelet deposition, prostaglandin release, vascular thrombosis, extravasation of red blood cells, and tissue necrosis.[35, 36]

Progesterone withdrawal also stimulates an inflammatory response in the endometrium. Just before menstruation, the total number of leukocytes in the endometrium increases

markedly to as much as 40% of the stroma.[37, 38] The inflammatory infiltrate (including neutrophils, eosinophils, and macrophages or monocytes) is drawn by chemo-attractive molecules (chemokines) synthesized by endometrial cells, some of which are down-regulated by progesterone (interleukin 8; IL-8).[37] When activated, the leukocytes produce a wide assortment of regulatory molecules including cytokines, chemokines, and a range of enzymes that contribute to degradation of the extracellular matrix, directly, or indirectly, via activation of other proteases.

Matrix metalloproteinases are a family a proteolytic enzymes that degrade components of the extracellular matrix and basement membrane.[39] The metalloproteinases include collagenases that degrade interstitial and basement membrane collagens, gelatinases that further digest collagens, and stromelysins that attack fibronectin, laminin, and glycoproteins. Each member of the family is substrate specific and secreted as an inactive zymogen that requires activation by plasmin, leukocyte proteases, or other metalloproteinases. The expression, secretion, and activation of endometrial matrix metalloproteinases is cycle dependent and increases markedly in the late secretory phase just before menstruation.[40, 41] Overall, progesterone inhibits endometrial metalloproteinase expression, an action mediated by transforming growth factor (TFG)-β.[42] Progesterone withdrawal has the opposite effect—increased metalloproteinase secretion and activation, followed by dissolution of the extracellular matrix.[43] Local modulators (predominantly cytokines), derived from endometrial epithelial, stromal, and endothelial cells, and natural tissue inhibitors of matrix metalloproteinases that bind the active form of the enzymes also play an important role in their regulation.[44] In cycles of conception wherein elevated progesterone levels are sustained, matrix metalloproteinase activity remains effectively suppressed. In the normal menstrual cycle, metalloproteinase expression is suppressed again after menses, presumably by increasing estrogen levels.

Progressive enzymatic degradation of the endometrium eventually disrupts the subsurface capillary and venous vascular system, causing interstitial hemorrhage; dissolution of the surface membrane allows blood to escape into the endometrial cavity. Ultimately, degeneration extends to the deepest extent of the functional layer where rupture of the basal arterioles contributes to bleeding. A natural cleavage plane develops at the junction of the loose, vascular, edematous stroma with the basal layer. Desquamation begins in the fundus and gradually extends toward the isthmus. The end result is the typical deflated and shallow but dense menstrual endometrium.[45, 46]

The menstrual fluid is comprised of an autolysed endometrium rich in inflammatory exudates, red blood cells, and proteolytic enzymes.[35, 46] One of those enzymes, plasmin, formed by activation of its inactive precursor, plasminogen, has potent fibrinolytic actions that help to prevent clotting of menstrual fluid and to facilitate the expulsion of degenerated tissue. Plasminogen activators that mediate the conversion of plasminogen to plasmin are found in late secretory and menstrual endometrium and are released from degenerated endometrial vascular endothelium.[35, 46] The volume of menstrual bleeding is controlled, at least to some extent, by the local balance between fibrinolysis and clotting. Endometrial stromal cell tissue factor and plasminogen activator inhibitor (PAI)-1 promote clotting and help to balance fibrinolytic processes.[47–49] Early in menstruation, intravascular platelet plugs, and later, thrombi form at the shedding surface, helping to limit blood loss. Their importance to hemostasis in the menstrual endometrium can be inferred from the increased volumes of menstrual blood loss observed in women with thrombocytopenia and von Willebrand disease. Ultimately, however, the cessation of menstrual bleeding depends on vasoconstriction in the denuded spiral arterioles in the basal layer of the endometrium, and also possibly in the radial arteries of the superficial myometrium. Endothelins are potent long-acting vasoconstrictors of vascular smooth muscle produced by endometrial glandular, stromal, and endothelial cells. Menstrual endometrium contains high concentrations of endothelins and prostaglandins, which together cause intense vasoconstriction in the spiral arterioles.[35]

The myometrial contractions associated with menstrual events very likely reflect the actions of prostaglandin $F_{2\alpha}$, but in contrast to postpartum bleeding, myometrial contractions are not important for control of menstrual bleeding.

Surface re-epithelialization also contributes to hemostasis in the menstrual endometrium. The process occurs very rapidly, beginning at the mouth of the basal portions of residual glands in areas otherwise completely denuded, and spreading outward. The peripheral regions of the cavity at the isthmus and near the tubal ostia (which do not shed during menses) also contribute to its resurfacing.[38, 46] Generally, by cycle day 5, these scattered areas of epithelial proliferation converge and fuse; bleeding stops completely only when the new epithelial surface is complete. The mechanisms that govern this initial phase of tissue repair and the role that estrogen has, if any, are uncertain. In the first few days of the new cycle, circulating estrogen levels and endometrial estrogen and progesterone receptor concentrations are low and unchanged from premenstrual levels.[50, 51] Moreover, even after oophorectomy and vigorous endometrial denudation, the endometrium heals, suggesting that the initial phase of tissue repair is largely independent of estrogen.[38, 46]

The stroma regenerates from stem cells located in the basal layer of the endometrium, but only after a confluent surface epithelium has been restored. Damaged endometrial vessels are quickly repaired. New vessel growth and mitotic activity in all parts of the regenerated human endometrium coincide with increasing serum estrogen levels and rising endometrial estrogen and progesterone receptor concentrations.[38, 46] Matrix metalloproteinases present in the menstrual endometrium and other proteases may be important mediators of the release and activation of growth factors needed for endometrial repair. Vascular endothelial growth factor is an important promoter of endometrial mitosis and can be induced by tumor necrosis factor (TNF)-α, TGF-β, and insulin-like growth factor-1.[35, 52, 53] Experimental evidence derived from model systems suggests that activins and other members of the TGF-β superfamily also may play a role.[38, 54]

There are two basic reasons why normal menstrual bleeding is self-limited.

1. In response to a simultaneous estrogen-progesterone withdrawal, endometrial shedding is universal. Because the onset and end of menses relate to organized cyclic hormonal events, menstrual changes occur uniformly, throughout the endometrial cavity. Shedding of the functional layer and exposure of the basal regenerative layer of the endometrium stimulates coagulation, vasoconstriction, and epithelial reconstruction mechanisms that effectively limit the volume and duration of bleeding.

2. In response to cyclic sequential estrogen-progesterone stimulation, growth and development of the endometrial epithelium, stroma, and microvasculature is structurally stable and random breakdown is avoided. The sequence of events leading to the enzymatic disintegration of the endometrium proceeds in an orderly and synchronous fashion. The endometrium is not just repaired, but is completely remodeled, at regular intervals.

Endometrial Responses to Steroid Hormones: Physiologic and Pharmacologic

The normal menstrual bleeding that occurs at the end of an ovulatory cycle results from estrogen-progesterone withdrawal. The same mechanism operates when the corpus luteum

is removed or when its gonadotropin support is suddenly interrupted during the luteal phase, such as by treatment with a gonadotropin-releasing hormone (GnRH) antagonist. Other examples include the bleeding that follows discontinuation of both estrogen and progesterone in women receiving cyclic postmenopausal hormone therapy and the bleeding that comes at the end of a standard cycle of treatment with an estrogen-progestin contraceptive. The bleeding that follows estrogen-progesterone withdrawal generally is regular, predictable, and consistent in volume and duration. However, estrogen-progesterone withdrawal is not the only pattern of steroid hormone signals that can provoke endometrial bleeding. Bleeding also can result from estrogen withdrawal, estrogen breakthrough, progestogen withdrawal, and progestogen breakthrough.

Estrogen Withdrawal Bleeding

One clinical example of estrogen withdrawal bleeding is that which may follow bilateral oophorectomy during the follicular phase of the cycle. The bleeding that occurs after removal of the ovaries can be delayed by exogenous estrogen therapy, but will occur when treatment stops. Other examples include cyclic estrogen-only hormone therapy in castrate or postmenopausal women and the midcycle bleeding that can accompany the transient but abrupt fall in estrogen levels immediately preceding ovulation.

Estrogen Breakthrough Bleeding

The best clinical examples of estrogen breakthrough bleeding are the different patterns of bleeding observed in women with chronic anovulation. The amount and duration of estrogen breakthrough bleeding can vary widely, depending on the amount and duration of unopposed estrogen stimulation that the endometrium has received. ***Relatively low levels of chronic estrogen exposure typically result in intermittent spotting or staining that is generally light in volume but may be prolonged. In contrast, sustained high level estrogen stimulation commonly results in long intervals of amenorrhea punctuated by acute episodes of often profuse bleeding that vary in duration.***

Progestogen Withdrawal Bleeding

Progestogen withdrawal bleeding is observed when treatment with exogenous progesterone or a synthetic progestin is discontinued. Progestogen withdrawal bleeding usually occurs only when the endometrium has first been primed with endogenous or exogenous estrogen. The amount and duration of bleeding can vary widely and generally correlates with the level and duration of previous estrogen-stimulated endometrial proliferation. In women with marginal to frankly low estrogen levels or short intervals of amenorrhea, bleeding is generally light to scant and may not occur at all. In those with sustained high estrogen levels or long intervals of amenorrhea, bleeding can be heavy and somewhat prolonged, but still is self-limited. Between the extremes, the amount and duration of bleeding induced by progestogen withdrawal is typically similar to that observed at the end of a normal ovulatory cycle. In women receiving cyclic hormone therapy with exogenous estrogen and progestin, bleeding follows withdrawal of progestin even if estrogen treatment continues; progestin withdrawal bleeding can be delayed, but only if estrogen levels are increased by 10–20 fold.[55]

Progestogen Breakthrough Bleeding

Progestogen breakthrough bleeding occurs when the ratio of progestogen to estrogen is unfavorably high. Unless there is sufficient estrogen to balance its action, continuous treatment with exogenous progesterone or synthetic progestins will result in intermittent bleeding of varying duration that is generally light, a pattern very similar to low level estrogen breakthrough bleeding described above. Clinical examples of progestogen breakthrough bleeding are the bleeding observed in women using the progestin-only contraceptive "minipill" or other long-acting progestin-only contraceptive methods (progestin implants, depot medroxyprogesterone acetate).[56] The breakthrough bleeding observed in women using combination estrogen-progestin contraceptives also is a form of progestogen breakthrough bleeding. ***Although all estrogen-progestin contraceptive regimens contain pharmacologic quantities of both estrogen and progestin, the progestin component is always the dominant hormone and the net effect on the endometrium is profoundly progestational.***

Pathophysiology of Anovulatory Bleeding

Anovulatory bleeding can result from estrogen withdrawal bleeding, reflecting the transient fall in estrogen levels that accompanies regression of a follicular cohort, or from estrogen breakthrough bleeding, due to focal break down of an overgrown and structurally fragile endometrium under continuous estrogen stimulation. The heaviest episodes of anovulatory bleeding tend to occur in women with sustained high levels of estrogen; women with polycystic ovary syndrome, obese women, postmenarcheal adolescents, and perimenopausal women are common clinical examples. The clinical presentation spans the spectrum from the pale frightened teenager who has bled for weeks to the older woman who is deeply concerned that she may have cancer.

In contrast to the organized predictable pattern of sequential estrogen-progesterone stimulation and withdrawal that characterizes the normal ovulatory menstrual cycle, the patterns of ovarian steroid hormone production and endometrial stimulation in anovulatory women are disorganized and unpredictable. By definition, the anovulatory woman is always in the follicular phase of the ovarian cycle and in the proliferative phase of the endometrial cycle. There is no luteal or secretory phase because there is no ovulation or cycle. The only ovarian steroid signal the endometrium receives is estrogen, levels of which constantly fluctuate, rising and falling as each new cohort of follicles begins to grow but ultimately loses its developmental momentum and, sooner or later, lapses into atresia. Although the amplitude of the signal may vary, the message, growth, stays the same.

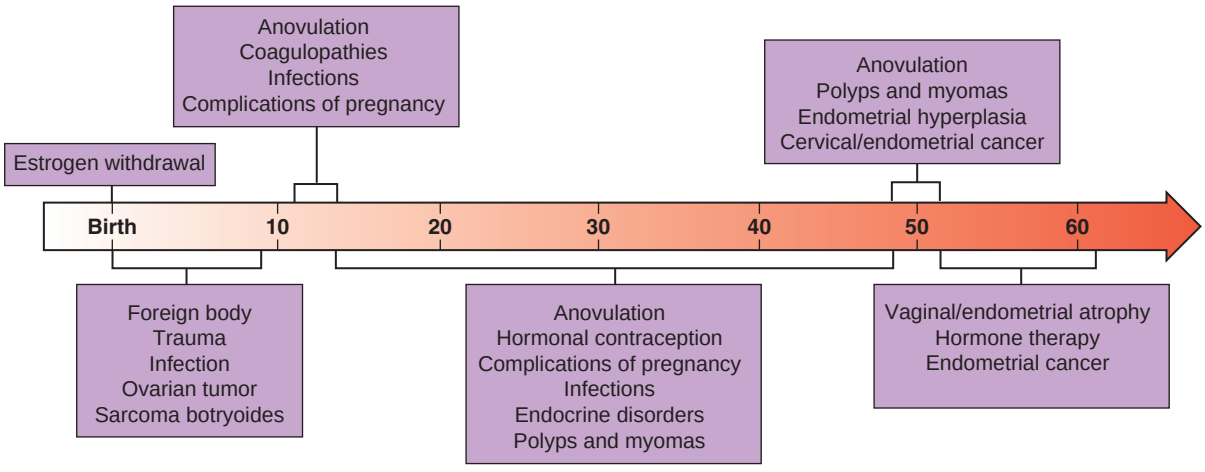
Over a period of time, an unrelenting, uninterrupted estrogen growth stimulus can stimulate the endometrium to proliferate to abnormal heights where it becomes fragile. Without the growth limiting and organizing effects of progesterone, the endometrium lacks the stromal support structure to maintain stability. Focal areas breakdown and bleed and, as those areas heal under the influence of continued estrogen stimulation, others break down and bleed. Persistent proliferative and hyperplastic endometrium characteristically exhibits numerous discrete foci of stromal breakdown near the epithelial surface, associated with pools of extravasated red blood cells, capillary platelet/fibrin thrombi, and repair-related changes recognized as ball-like aggregates of tightly packed stromal cells beneath a cap of intact but hypertrophied epithelium.[35] The cause for the focal breakdowns in persistent proliferative endometrium is not entirely clear. However, abnormal endometrial growth involves not only epithelial and stromal cells but also the microvasculature.

Venous capillaries in persistent proliferative and hyperplastic endometrium are increased, dilated, and often form abnormal irregular channels; ultrastructural studies have revealed a number of abnormal structural elements that predispose to fragility.[57, 58] The abnormal microvasculature could be the result, but is more likely the proximate cause, of abnormal bleeding. ***The weight of available evidence from histologic and molecular studies indicates that anovulatory bleeding results from an increased density of abnormal vessels having a fragile structure prone to focal rupture, followed by release of lysosomal proteolytic enzymes from surrounding epithelial and stromal cells and migratory leukocytes and macrophages.*** Once initiated, the process is further aggravated by local release of prostaglandins, with greater sensitivity to those that vasodilate (PGE_2) than to those that vasoconstrict ($PGF_{2\alpha}$).[59] Other molecules (perforins) inhibit capillary plug formation and further degrade the capillary venous network. Vasoconstriction of basal endometrial and superficial myometrial vessels does not occur because tissue loss is only focal and superficial, and does not typically reach the basal layer where denudation triggers an intense vasoconstrictive response. The final mechanism that normally controls menstrual bleeding, surface epithelial reconstruction, operates in persistent proliferative endometrium, but not in a normal way. Epithelial repair is focal, in the areas of breakdown, not universal; the result is a constantly changing patchwork of small repairs instead of an organized and well structured remodeling.[35]

Differential Diagnosis of Abnormal Uterine Bleeding

Anovulatory dysfunctional uterine bleeding is a diagnosis made by exclusion. The differential diagnosis includes problems relating to pregnancy, infection, vaginal and cervical abnormalities, benign and malignant uterine neoplasia, coagulopathies, endocrine disorders, trauma, foreign bodies, systemic disease, and bleeding relating to medications. The most common causes vary with age. In premenarcheal girls, foreign bodies, trauma, and infection are the most common. In postmenarcheal adolescents, anovulatory bleeding, coagulopathies, infections, and complications of pregnancy head the list. During the reproductive years, most abnormal bleeding results from anovulation, hormonal contraception, complications of pregnancy, infections, endocrine disorders, and polyps and myomas. In perimenopausal women, anovulation, benign uterine neoplasia, and endometrial hyperplasia cause the majority of problems, and in postmenopausal women, vaginal/endometrial atrophy and hormone therapy are the most common causes of abnormal bleeding; only about 10% of postmenopausal bleeding results from endometrial cancer.

Usual Causes of Abnormal Bleeding By Age



Complications of pregnancy always should be considered and excluded, particularly in adolescents who may be reluctant to reveal their sexual history. ***It is important to emphasize that the most common cause of a sudden departure from a well-established pattern of regular and predictable menses is a complication of pregnancy; threatened or spontaneous abortion and ectopic pregnancy are the most common, but possibilities also include retained products of conception and gestational trophoblastic disease.***

Although abnormal bleeding is a relatively common problem in women using ***hormonal contraception*** or receiving physiologic combined ***continuous estrogen-progestin hormone therapy***, the possibility of underlying pathology must not be forgotten. ***Infections*** such as cervicitis, endometritis, and salpingitis may be associated with abnormal bleeding. Bleeding relating to benign uterine neoplasia, chiefly ***cervical and endometrial polyps and uterine myomas***, is confused frequently with anovulatory bleeding. Other pathology of the reproductive tract associated with abnormal bleeding includes ***adenomyosis*** and ***malignancies of the cervix and endometrium***. Abnormal menstrual cycles occasionally are one of the earliest signs of a ***thyroid disorder*** (hypothyroidism or hyperthyroidism).[60]

The possibility of a ***coagulopathy*** also should be kept in mind, especially in adolescents whose menstrual history is short and not yet well defined. The most common cause of abnormal uterine bleeding in adolescents is anovulation, but up to a third may have a coagulation defect,[61–64] including von Willebrand disease, Glanzmann thrombasthenia, idiopathic thrombocytopenic purpura, platelet dysfunction, and thrombocytopenia related to malignancy or treatment for malignancy. Bleeding disorders usually are associated with cyclic, regular, heavy or prolonged bleeding (menorrhagia). The same pattern may be observed in women receiving treatment with anticoagulants.[65] ***Previous history of postpartum hemorrhage or excessive bleeding with surgery, dental procedures, or trauma should raise suspicion, but menorrhagia since menarche may be the only clue.[66] Coagulation defects are not as rare as is generally perceived and may be found in 10–20% of women with unexplained menorrhagia.[66–69]***

A variety of different ***medications*** can predispose to abnormal bleeding, by interfering with hemostasis (usually resulting in menorrhagia), by affecting the concentrations of endogenous or exogenous hormones (causing fluctuations in circulating levels), or by disrupting the hypothalamic-pituitary-ovarian axis. Drugs associated with abnormal menstrual bleeding include hormonal contraceptives, those used for postmenopausal hormone therapy, digitalis, anticonvulsants, anticoagulants, and psychopharmacologic medications. Some common herbs have estrogenic activity (e.g., ginseng) and may be associated with abnormal bleeding.[70]

Although uncommon, other diagnostic possibilities include ***systemic illnesses*** that predispose to anovulation or coagulation abnormalities; examples include diabetes mellitus, systemic lupus erythematosus, malignancy, and myelodysplasia. Chronic renal disease is associated with both ovulatory and platelet dysfunction. Liver disease can result in abnormal bleeding by adversely affecting estrogen metabolism (predisposing to anovulation) or the synthesis of clotting factors. ***In adolescents, genital trauma, sexual abuse, cervicitis relating to sexually-transmitted infections (Chlamydia trachomatis), and foreign bodies (e.g., retained tampons) merit specific consideration.***

The existence of a ***post-tubal ligation syndrome*** of menstrual abnormalities has been debated for decades. Numerous studies have addressed the question with conflicting results. Some have examined the prevalence of menstrual complaints before and after sterilization.[71, 72] Others have compared the incidence of hospitalization or hysterectomy for abnormal uterine bleeding in women with and without a previous tubal sterilization procedure.[73–75] The popular theory that extensive tubal electrocoagulation adversely affects ovarian blood supply and steroid hormone production was supported by data suggesting that the incidence of menstrual problems increased with time after sterilization by electrocautery but not in women sterilized with rings or clips.[76–78] However, no correlation has been found

between poststerilization menstrual changes and the amount of tissue destroyed.[76, 78] Analysis of data from the U.S. Collaborative Review of Sterilization, a multicenter prospective cohort study that followed almost 10,000 women for up to 5 years after a tubal sterilization procedure, revealed that sterilized women were no more likely than women with sterilized male partners to report persistent changes in intermenstrual bleeding or cycle length.[79] Sterilized women were more likely to have decreased menstrual duration, volume, and pain, and among women with heavy bleeding at baseline, those sterilized were more likely to report decreased menstrual bleeding after the procedure.[79] Another more recent study of menstrual patterns and ovarian function before and 3 months after bipolar electrocauterization of the fallopian tubes found no evidence for an adverse effect on menstrual characteristics or ovarian reserve (as assessed by basal FSH concentrations).[80] ***These data suggest strongly that women who have a tubal sterilization procedure are no more likely than other women to have menstrual abnormalities.***

Diagnostic Evaluation of Abnormal Uterine Bleeding

A careful history and physical examination are the most useful tools for differentiating anovulatory bleeding from other causes. The details of the history and the physical findings narrow the number of possibilities meriting serious consideration and define the scope and content of the evaluation required to establish a diagnosis. The history should seek to define each of the following characteristics:

- **Intermenstrual interval (number of days, regularity)**
- **Volume (heavy, light, or variable)**
- **Duration (normal or prolonged, consistent or variable)**
- **Onset of abnormal menses (perimenarcheal, sudden, gradual)**
- **Temporal associations (postcoital, postpartum, post-pill, weight gain or loss)**
- **Associated symptoms (premenstrual molimina, dysmenorrhea, dyspareunia, galactorrhea, hirsutism)**
- **Underlying systemic illness (renal, hepatic, hematopoietic, thyroid)**
- **Medications (hormonal, anticoagulants)**

In the majority of women with true anovulatory bleeding, the menstrual history alone can establish the diagnosis with sufficient confidence that treatment can begin without additional laboratory evaluation or imaging. Infrequent, irregular, unpredictable menstrual bleeding that varies in amount, duration, and character and is not preceded by any recognizable or consistent pattern of premenstrual molimina or accompanied by any visible or palpable genital tract abnormality is not difficult to interpret. ***Conversely, regular monthly periods that are heavy or prolonged are more likely related to an anatomical lesion or a bleeding disorder than to anovulation.***

Objective methods for measuring menstrual blood loss include the photometric alkaline hematin test (the gold standard for research purposes),[81, 82] and menstrual pictograms (illustrations of blood stains of different size on feminine hygiene products),[83] both of which provide an accurate means of quantifying menstrual blood loss.[84] However, the most practical approach is the menstrual history. Although subjective, a history of changing pads or tampons more often than every 3 hours, use of more than 20 over a single menses, the need to change protection during the night, the passage of clots larger than an inch in diameter, menses lasting longer than 7 days, and diagnosis of anemia indicate abnormally heavy menstrual bleeding.[26] ***Regardless of the actual amount of blood loss, menstrual bleeding that interferes with daily activities or causes anxiety and concern merits evaluation.***

Midcycle bleeding may be an occasional consequence of the transient but abrupt fall in estrogen levels that occurs at the time of ovulation, but women who have recurrent episodes of intermenstrual bleeding often have intrauterine pathology and deserve evaluation.

Physical examination should aim first at establishing the source of bleeding when that is uncertain. Although most abnormal genital bleeding comes from the uterine corpus, other sources should be excluded, particularly in women whose bleeding is unrelated to the menstrual cycle. Extrauterine sources for abnormal bleeding include the urethra (urethritis), bladder (urinary tract infections, cancers), the vagina (vaginitis and ulcerative lesions), the cervix (ectropion, cervicitis, polyps, focal lesions), the vulva (trauma, skin lesions), and the anus and rectum (anal fissures, hemorrhoids, inflammatory bowel disease, cancers). Examination also should define uterine size (normal or enlarged), contour (smooth and symmetrical or irregular), consistency (firm or soft), and tenderness.

Laboratory Evaluation

Laboratory tests can be very helpful but are not always necessary. A sensitive urine or serum ***pregnancy test*** can quickly exclude the possibility that abnormal bleeding relates to a complication of pregnancy; a positive test leaves only a few diagnostic possibilities, which generally are not difficult to distinguish. A ***complete blood count*** to exclude anemia and thrombocytopenia is prudent in all women with complaints of abnormal bleeding, especially when heavy or prolonged.

After excluding pregnancy, the most important question to answer is whether the patient is ovulating, because the causes and clinical management of ovulatory and anovulatory uterine bleeding are quite different. When the menstrual history alone does not allow a confident conclusion, a well timed ***serum progesterone*** determination during the putative luteal phase of the cycle can help to document ovulation or anovulation. A logical strategy is to obtain the test between cycle day 22 and 24, after ovulation in the longest normal cycle and before the end of the shortest normal cycle; any value greater than 3 ng/mL provides reliable evidence that ovulation has occurred recently.[85] However, when bleeding episodes are frequent or poorly documented, proper timing for a progesterone measurement can be difficult to determine. It also is important to remember that many women with abnormal bleeding, especially perimenarcheal and perimenopausal women, ovulate at least occasionally. Although most commonly applied to assess ovulatory function in women with infertility and now seldom used even for that purpose, basal body temperature recordings can be very informative in women with a confusing pattern of bleeding. Endometrial biopsy also can be used to assess ovulatory function (proliferative vs. secretory endometrium) but cannot be justified for that purpose alone when a less costly and less invasive serum progesterone measurement provides the same qualitative information; endometrial sampling should be reserved for those at risk for endometrial hyperplasia or neoplasia, as discussed below.

In sexually active women, a nucleic acid based test for ***chlamydia and gonorrhea*** and a ***wet prep*** to exclude trichomonas infection merit consideration, particularly in those with evidence of vaginitis and/or cervicitis. In presumed or proven anovulatory women, a ***serum thyroid-stimulating hormone (TSH)*** level excludes any associated thyroid disorder. ***Liver or renal function tests*** are indicated only for those with known or strongly suspected disease.

Adolescents, women with a suspicious personal or family history of bleeding symptoms (easy bruising, frequent gum bleeding when flossing or brushing teeth, epistaxis), and women with unexplained menorrhagia warrant evaluation with ***coagulation studies*** to exclude coagulopathies, such as von Willebrand disease, factor deficiencies, and platelet function abnormalities.[66, 68, 86, 87] ***In addition to a platelet count, screening should include***

both a prothrombin (PT), which evaluates the extrinsic and final common clotting pathways, and an activated partial thromboplastin time (aPTT), which tests the intrinsic and common pathways of coagulation. Although the PT and aPTT have relatively low positive and negative predictive value for detecting underlying bleeding disorders,[88] they are adequate screens for severe factor deficiencies.[89] ***The high prevalence of von Willebrand disease among women with menorrhagia (approximately 13%) warrants specific exclusion of the diagnosis and justifies measurement of von Willebrand factor, ristocetin cofactor activity (von Willebrand factor activity), the factor VIII level, and blood typing.***[87, 90, 91] It is important to note that test results can fluctuate over time,[92] and also may vary across the menstrual cycle; repeated testing, ideally during the first few days of the cycle, may be required to establish the diagnosis of von Willebrand disease.[90,93] The blood type is helpful because von Willebrand factor and factor VIII levels are 25% lower in patients with type O blood than in those with other blood types.[94] ***Although the bleeding time is the traditional method for evaluating platelet function, an automated laboratory test (Platelet Function Analyzer, PFA-100) is taking its place because it has greater sensitivity and reproducibility and is less invasive.***[95, 96] The instrument exposes platelets in citrated whole blood to high shear inside a capillary tube and monitors the drop in flow rate as the platelets form a plug in the center of a membrane coated with collagen and either adenosine diphosphate or epinephrine. For patients with abnormal coagulation studies, consultation with a hematologist is recommended.[97, 98]

Endometrial Sampling

An endometrial biopsy can exclude endometrial hyperplasia or cancer. Age over 35 or 40 years is widely considered a risk factor for endometrial disease and cited as an indication for biopsy in women with abnormal bleeding. ***Endometrial hyperplasia and cancer are more commonly detected in older than in younger women, but the duration of exposure to unopposed estrogen stimulation is the more critical risk factor.*** Long-term exposure is more likely in older than in younger women, but women under age 30, and even teenagers, can develop endometrial cancer.[99–102] In premenopausal women, the likelihood of abnormal endometrial histology is relatively high (14%) when menses are irregular, but very low (< 1%) when cycles are regular.[103] The small flexible suction cannulas now widely available cause less discomfort than older traditional biopsy instruments and yield comparable results.[104–106] Unfortunately, hospital-based curettage without hysteroscopy is still commonly performed, even though it is no longer the gold standard.

In addition to revealing any intrinsic endometrial disease, such as chronic endometritis, hyperplasia, or adenocarcinoma, biopsy can help to direct further evaluation or to guide the choice of treatment in women with a confusing history of abnormal bleeding. An inactive or atrophic endometrium identifies women unlikely to respond to progestational therapy. ***In women with no recent exposure to exogenous progestins, a secretory endometrium provides reliable evidence of recent ovulation and signals the need to search for an anatomical cause.***

Imaging

Imaging can help to differentiate anovulatory bleeding from anatomical causes, myomas and endometrial polyps being the most common examples. Standard transvaginal ultrasonography can provide accurate information about the size and location of any uterine fibroids that may explain abnormal bleeding or exaggerate the bleeding due to other causes.[107]

Ultrasonography may reveal an obvious cavitary lesion or an abnormally thin or thick endometrium. A very thin endometrial "stripe" (<5 mm), like a biopsy that yields minimal tissue, suggests an attenuated or denuded endometrium best treated first with estrogen rather than with a progestin or an estrogen-progestin combination (discussed below). ***In perimenopausal and postmenopausal women with abnormal bleeding, endometrial biopsy generally is considered unnecessary when the endometrial thickness is less than 4 or 5 mm because the risk of endometrial hyperplasia or cancer is remote.***[108–110] It seems logical to apply the same criterion for the same reason in premenopausal women with abnormal bleeding, although there is no substantial direct evidence to support the extrapolation. Otherwise, the decision to biopsy or not should be based primarily on clinical suspicion and risk factors rather than on ultrasonographic measurements of endometrial thickness. That does not mean that endometrial thickness has no bearing on the decision whether to perform a biopsy; a grossly increased endometrial thickness (>12 mm) increases the risk of disease and is an indication for sampling, even when clinical suspicion of pathology is otherwise low.[111] ***In summary, we believe that biopsy is unnecessary when the endometrial thickness is less than 5 mm, that biopsy is indicated when the clinical history suggests long-term unopposed estrogen exposure even when the endometrial thickness is "normal" (5–12 mm), and that biopsy should be performed when endometrial thickness is greater than 12 mm even when clinical suspicion of disease is low.***

Sonohysterography, involving transvaginal ultrasonography during or after introduction of sterile saline using any of a variety of available catheters (also known as hydrosonography and saline infusion sonography) sharply defines cavity contours and readily demonstrates even small intrauterine lesions. The sensitivity and specificity of sonohysterography exceed that of standard transvaginal ultrasonography and compare favorably with hysteroscopy.[112–115] ***The combination of sonohysterography and endometrial biopsy offers a high sensitivity and high negative predictive value for detection of endometrial and uterine pathology in women with abnormal bleeding.***[116] One disadvantage of the technique is that minor cavity contour abnormalities or blood clots may be misinterpreted as polyps.

Hysteroscopy is the definitive method for both diagnosis and treatment of symptomatic intrauterine pathology, but also is the most invasive. Traditionally, hysteroscopy has been reserved for treatment of disease identified by other less invasive methods, but modern hysteroscopes having an outer diameter of 2 or 3 mm now permit diagnostic and minor operative procedures to be performed in the office setting with minimal anesthesia.[117] For clinicians having the necessary training and experience, office hysteroscopy has a very low incidence of complications, which may include uterine perforation, infections, and excessive bleeding. Major intrauterine pathology generally requires more traditional operative hysteroscopy using instruments having a larger caliber and greater capabilities.

Magnetic resonance imaging (MRI) is gaining acceptance in the evaluation of abnormal uterine bleeding. It can reliably define uterine anatomy, distinguish between adenomyosis and leiomyomata, and demonstrate the proximity of myomas to the uterine cavity.[118] MRI can be very helpful in women who cannot be imaged adequately with ultrasonography, but its cost is otherwise difficult to justify.

In general, diagnostic uterine imaging can be reserved for women in whom the menstrual history or the results of other evaluation provide strong evidence for an anatomical cause of abnormal bleeding, including any of the following:

- **Regular monthly cycles with increasing volume or duration of bleeding**
- **Regular monthly cycles complicated by intermenstrual bleeding in the absence of a vaginal or cervical lesion.**
- **Abnormal bleeding despite objective evidence of ovulation from measurement of serum progesterone (>3 ng/mL) or from endometrial sampling (secretory endometrium)**
- **Failed empirical medical management**

As in all aspects of clinical medicine, the success of treatment hinges on an accurate diagnosis. When there is good reason to suspect a coagulopathy or uterine pathology as the cause of abnormal bleeding, diagnostic laboratory tests, endometrial biopsy, or uterine imaging should be considered carefully before beginning empiric medical management. ***However, when there is every reason to believe that anovulation is the cause, empiric medical management based on that premise is entirely reasonable; a prompt resolution of the problem also should be expected. When bleeding persists despite appropriate empiric medical management, further diagnostic evaluation is more productive than a higher dose or otherwise different medical treatment regimen.***

Treatment of Anovulatory Bleeding

The primary objective of treatment in women with anovulatory bleeding is to induce or restore the natural control mechanisms that are not operating—orderly, synchronous growth, development, and shedding of a structurally stable endometrium. Without treatment or correction of the cause of chronic anovulation, recurrent episodes of heavy or prolonged bleeding can be expected.

Although most women with anovulatory bleeding can be managed effectively on an outpatient basis, acute bleeding occasionally may be severe enough to require hospitalization and emergency treatment. ***Hospitalization is indicated for women with active hemorrhage who are hemodynamically unstable and those with symptomatic anemia or a serious underlying medical illness.*** The most effective initial strategy in emergent circumstances is to insert a Foley catheter with a 30 mL balloon into the uterus to ***tamponade*** the bleeding while establishing intravenous access for fluid administration and, if necessary, transfusion.[119] Once the patient is stabilized, the diagnostic evaluation can proceed to determine the cause of the bleeding and the most appropriate treatment strategy. Acute anovulatory bleeding can be treated with estrogen, estrogen-progestin, or progestin alone. The best choice in a given patient depends primarily on the condition of the endometrium at the time. ***It is important to emphasize that progestin treatment (remembering that the net effect of all estrogen-progestin contraceptives is progestational) is unlikely to be effective in patients with a thin, attenuated, or denuded endometrium***. Ideally, given the importance of making the correct choice in women with acute bleeding, transvaginal ultrasonography should be performed before treatment begins, to identify any obvious pathology that may dictate management and to assess the endometrial thickness.[107]

Progestin Therapy

Just as progesterone is the dominant and controlling influence in normal menstrual cycles, progestins are the mainstay of treatment for anovulatory bleeding. Progestins are powerful anti-estrogens. Progestins stimulate 17β-hydroxysteroid dehydrogenase and sulfotransferase activity, the enzymes that work in concert to convert estradiol to estrone sulfate (which is rapidly cleared from the body).[120] Progestins further antagonize estrogen action by inhibiting estrogen's induction of its own receptor (estrogen receptor replenishment). Progestins also suppress estrogen-mediated transcription of oncogenes.[121] Together, these actions explain the anti-mitotic, growth limiting effects of progesterone and progestins on the endometrium (arrest of growth during the secretory phase of the cycle, prevention and reversal of hyperplasia, and marked attenuation during pregnancy or treatment with estrogen-progestin contraceptives).

In most circumstances, progestin therapy will control anovulatory bleeding once uterine pathology has been excluded. ***In oligomenorrheic anovulatory women with episodic abnormal bleeding, orderly, predictable, self-limited progestogen withdrawal bleeding can be induced by cyclic treatment with an orally active progestin (e.g., medroxyprogesterone acetate 5–10 mg daily for 12–14 days each month).*** Cyclic progestin therapy restores the normal sequence of endometrial steroid hormone stimulation—estrogen, followed by estrogen plus progestogen, followed by withdrawal. The interval of progestin therapy can be fixed to the calendar (beginning on the first of every month) or to the onset of menses (beginning 15–16 days after onset of the last progestin-induced menses); both regimens work well. ***Failed progestin treatment suggests strongly that other pathology is causing or contributing to the problem and signals the need for additional diagnostic evaluation.***

Although cyclic progestin therapy generally works well in women who are completely anovulatory and not sexually active, treatment with an estrogen-progestin contraceptive is the better choice for those who likely still ovulate (albeit infrequently) or want to avoid pregnancy. Inevitably, programmed cyclic progestin treatment will not coincide with endogenous progesterone production in random ovulatory cycles like those that occur in aging women. When that happens, bleeding may deviate from the predicted pattern and be misinterpreted or cause alarm. ***Moreover, standard cyclic progestin treatment regimens do not reliably suppress the hypothalamic-pituitary-ovarian axis, will not prevent random ovulation, and are not contraceptive.*** In contrast, contraceptive doses of exogenous steroids effectively suppress endogenous function and prevent such confusion.

Acute severe anovulatory bleeding also can be treated effectively with high-dose progestin alone (medroxyprogesterone acetate 10–20 mg twice daily; megestrol acetate 20–40 mg twice daily; norethindrone 5 mg twice daily), provided that the endometrium is normal or increased in thickness.[122–124] Treatment should continue for approximately 3 weeks, decreasing to once daily treatment after 7–10 days. High-dose progestin treatment induces stabilizing predecidual changes in a thickened, vascular, and fragile endometrium. However, a substantial amount of tissue remains to be shed upon progestin withdrawal, resulting in a so-called "medical curettage." ***If not warned to expect the heavy menses and increased dysmenorrhea likely to arrive within 2–4 days after treatment stops, most women will interpret the experience as more of the same and treatment failure.*** Thereafter, standard cyclic progestin treatment or an estrogen-progestin contraceptive may be offered for longer term management.

Depot-medroxyprogesterone acetate (150 mg intramuscularly every 3 months) can be a useful option for maintenance therapy in women who have difficulty with or cannot take estrogen-progestin contraceptives. ***However, depot progestin treatment has no place in the acute management of abnormal bleeding. Once given, it cannot be withdrawn, and if unsuccessful, its effects can be difficult to overcome.*** Episodic breakthrough bleeding is relatively common and can be treated with estrogen as discussed below.

Estrogen-Progestin Therapy

Women with anovulatory bleeding who are sexually active and not immediately prepared to pursue pregnancy generally are best managed by treatment with an estrogen-progestin contraceptive. A gradual but progressive decrease in the volume and duration of flow and associated dysmenorrhea can be expected and is reassuring. In women with normal uteri, estrogen-progestin contraceptives reduce menstrual flow by at least 60% from that in natural cycles.[125] Longer cycles of treatment offer the advantage of fewer and lighter menses, but increase the incidence of episodic breakthrough bleeding.

Acute prolonged episodes of heavy anovulatory bleeding also can be treated effectively with high-dose estrogen-progestin therapy, provided that the endometrium is normal or increased in thickness. Ideally, transvaginal ultrasonography should be performed as or before treatment begins to minimize the risk of unsuccessful treatment with continued heavy blood loss, as can occur in women with a denuded endometrium. In women with a thickened, vascular, and fragile endometrium, estrogen-progestin treatment inhibits further growth and induces structural changes that organize and stabilize the endometrium, thereby preventing further random breakdown. ***Any monophasic combination oral contraceptive can be used, beginning with one pill twice daily, and decreasing to one pill daily thereafter. Treatment should continue for a total of at least 2 weeks, even when bleeding markedly slows or stops, which generally can be expected within 24–48 hours.*** Attention can then turn to evaluation to determine the cause of anovulation and to treatment for any associated anemia. To provide a longer respite from the heavy bleeding that only recently stopped and the opportunity for hemoglobin levels to increase in anemic patients, estrogen-progestin treatment can continue (one pill per day) for a longer interval of time. ***Failed estrogen-progestin indicates the need for additional diagnostic evaluation.***

Estrogen Therapy

Intermittent vaginal spotting frequently is associated with marginal or frankly low levels of estrogen stimulation (estrogen breakthrough bleeding) and a very thin, unstable endometrium. In this setting, the usual beneficial effect of progestin treatment cannot be achieved because estrogen levels are insufficient to stimulate the growth that serves as the foundation for the actions of progestin; the growth-limiting effects of progestin are not helpful and may further aggravate the problem, as in women whose endometrium becomes denuded after prolonged heavy bleeding. Logically, estrogen therapy is the most effective initial treatment strategy.

Estrogen also is the obvious and best choice for management of episodic progestogen breakthrough bleeding, as commonly observed in women receiving low dose estrogen-progestin contraceptives, depot medroxyprogesterone acetate, or other forms of continuous progestin treatment for contraception (the progestin-only "minipill," progestin implants) or in the management of endometriosis.[126] A common clinical story involves long-term estrogen-progestin contraception, markedly decreased or absent menstrual flow during the pill-free week, and episodes of breakthrough bleeding at other times. Over time, unless there is sufficient endogenous or exogenous estrogen to effectively balance its effects, progestin attenuates the endometrium, inducing pseudo-atrophy. Histologically, the endometrium has little height and is composed almost entirely of pseudodecidualized stroma and blood vessels with relatively few glands. Although the mechanism is different, the light spotting, staining, or bleeding that may occur is in many ways similar to the type of estrogen breakthrough bleeding observed in women with marginal or very low levels of circulating estrogen. In all such scenarios, a short interval of added estrogen (conjugated estrogens 1.25 mg or micronized estradiol 2.0 mg daily for 7–10 days) is generally highly effective. In some women, the problem recurs frequently or persists. Higher doses, longer durations, or repeated courses of estrogen treatment are sometimes needed. When treatment fails, further evaluation with transvaginal ultrasound or sonohysterography can exclude the possibility of a previously unrecognized endometrial polyp or submucous myoma. Rarely, only a brief hiatus in estrogen-progestin or progestin treatment will resolve the problem.

When acute, heavy bleeding results in a thin, denuded endometrium, high-dose estrogen therapy is the best initial treatment; progestin or estrogen-progestin therapy is unlikely to succeed and may aggravate the problem. Estrogen stimulates endometrial re-epithelialization and proliferation and stabilizes lysosomal enzymes. Evidence suggests that high-dose estrogen therapy also stimulates clotting at the capillary level.[127, 128]

In patients who are hemodynamically unstable, intravenous estrogen therapy (25 mg conjugated equine estrogens every 4 hours intravenously until bleeding subsides, for up to 24 hours) is very effective. Treatment with an antiemetic is recommended (e.g., promethazine, 12.5–25 mg intramuscularly or rectally), because high-dose estrogen results in nausea and vomiting in up to 40% of patients.[129] The regimen controls acute bleeding effectively in more than 70% of patients, usually within 4–8 hours.[129] Thereafter, high-dose estrogen therapy should continue orally (2.5 mg conjugated estrogens or 2.0 mg micronized estradiol every 6 hours), tapering to a once daily dose after bleeding is controlled, and adding a progestin (e.g., medroxyprogesterone acetate 5–10 mg daily × 7–10 days) or changing to an estrogen-progestin contraceptive after 14–21 days to stabilize the estrogen-stimulated endometrial growth. In the hemodynamically stable patient with a denuded endometrium whose bleeding is less emergent but still acute and quite heavy, the same high-dose oral estrogen and antiemetic treatment regimen generally is effective.

High-dose intravenous or oral estrogen treatment may increase the risk of thromboembolism. There are no data that quantify the risk, but venous and pulmonary embolism are a potential complication and have been reported.[130] As with any therapeutic decision, the benefits of treatment must be weighed against its potential risks and those of alternative methods for the management of abnormal uterine bleeding. In women with a past episode or family history of thromboembolism, high-dose estrogen treatment should be avoided, if at all possible.

Curettage

In women with acute bleeding, ***dilation and curettage*** can be performed as both a therapeutic and diagnostic procedure. Curettage is an expeditious and effective way to stop acute uncontrollable uterine bleeding in the absence of any apparent pathology.[131] The mechanism responsible for the therapeutic effects of curettage is not entirely clear, but surgical denudation of the basal layer of the endometrium is presumed to acutely stimulate all of the normal processes involved in cessation of normal menstrual bleeding—local clotting mechanisms, vasoconstriction of basal arterioles, and re-epithelialization.[35] Blind curettage easily can miss focal lesions and, in most cases, does not treat the underlying cause of bleeding.[132] Consequently, curettage ideally should be combined with hysteroscopy to improve diagnostic accuracy and the effectiveness of postoperative treatment, thereby minimizing the risk for recurrence.[119, 133]

Endometrial Hyperplasia

Endometrial hyperplasia is a histologic diagnosis, based on findings of proliferating glands of varying size and shape and a greater gland-to-stroma ratio than is observed in normal endometrium.[134] Endometrial hyperplasia results almost exclusively from unopposed chronic estrogen stimulation.

Currently, endometrial hyperplasia is classified as simple or complex (reflecting the architectural pattern), with or without nuclear atypia (enlargement, rounding, pleomorphism, and aneuploidy). Simple hyperplasia is characterized by dilated glands with no or only occasional branching, lined by glandular cells that may or may not exhibit mitoses. In complex hyperplasia, the endometrial glands are crowded (“back-to-back”) with minimal intervening stroma, exhibit branching, and are lined by cells that, again, may or may not exhibit mitoses. Lesions without atypia basically represent only exaggerated forms of persistent proliferative

endometrium; they regress spontaneously, after curettage, or with progestin treatment, and are associated with little risk (1–3%) for progression to adenocarcinoma.[135–138] In contrast, atypical endometrial hyperplasia exhibits an entirely different behavior; it does not often spontaneously regress, can be quite resistant to even repeated curettage or prolonged high-dose progestational therapy, has significant risk (10–30%) of progression to adenocarcinoma if left untreated, and must therefore be regarded as a precancerous lesion.[135–138] Atypical lesions are distinguished from invasive carcinomas by the absence of stromal invasion. It is important to note that, despite concerted efforts to standardize classification criteria, there is significant inter-observer variability in the classification assigned by pathologists, even within the same institution.[139, 140] Other classifications systems have been proposed,[141, 142] but none has gained wide acceptance.

When an office biopsy reveals endometrial hyperplasia, further evaluation is required to exclude atypia or a coexisting cancer that was not represented in the tissue specimen, and if subsequent curettage reveals atypical endometrial hyperplasia, there is significant risk of an unrecognized adenocarcinoma. In a retrospective study involving 824 women with a diagnosis of complex atypical hyperplasia after office sampling, 100 were diagnosed with cancer after further evaluation with curettage, but 298 of the remaining 724 (41%) had unexpected cancer in a hysterectomy specimen obtained within 6 months of the original diagnosis; among those having an occult cancer, 30% had been further evaluated with curettage and 45% had not.[143]

Simple and complex endometrial hyperplasia without atypia has a low risk for progression to endometrial cancer and can be corrected using progestin treatment regimens similar to those recommended for management of anovulatory bleeding in oligomenorrheic anovulatory women. Cyclic progestin therapy (medroxyprogesterone acetate 5–10 mg daily or norethindrone acetate 5 mg daily for 14 days/month × 3–6 months) induces regression in at least 80–90% of patients;[136, 144] continuous progestin treatment for a similar interval also is effective. Another option for women interested in longer-term contraception is to insert a levonorgestrel-releasing intrauterine system (LNG-IUS).[145–147] Repeat biopsy to confirm regression is recommended and, in those with a LNG-IUS, can be performed without removing the device.

Endometrial hyperplasia with atypia is best treated by hysterectomy. Women intent on preserving their reproductive potential may be treated with progestins, but more potent and longer durations of treatment (megestrol acetate 80 mg twice daily for 3–6 months) are required and repeated biopsies to monitor response and confirm resolution of the lesion are essential. Insertion of an LNG-IUS is another effective treatment option.[148] Most, but not all, will respond to medical treatment.[149–151] The median time to regression is approximately 9 months, and persistent disease after 7–9 months of treatment predicts failure.[149, 151] Resistant lesions in women who remain adamantly opposed to surgery may require even higher and longer durations of progestational therapy. The resistance of atypical lesions to progestin therapy should not be surprising since nuclear atypia reflects a degree of cellular dedifferentiation. ***Women who respond to medical management should be encouraged to pursue pregnancy at the earliest possible time and must be carefully monitored because recurrence is common. Those who fail to respond to medical treatment will require hysterectomy.***

Treatment of Abnormal Bleeding from Other Causes

Not all irregular or abnormally heavy or prolonged menstrual bleeding relates to anovulation. History and physical examination generally are all that is needed to exclude trauma and foreign bodies. With few exceptions (retained nonviable products of conception),

complications of pregnancy are easily excluded with a simple pregnancy test. Cervical biopsy of any suspicious lesion and endometrial biopsy in women with risk factors for endometrial cancer eliminate reproductive tract malignancies as a possibility. The differential diagnosis of abnormal bleeding in ovulatory women and in anovulatory women who fail appropriate hormonal treatment centers on a few major possibilities—chronic endometritis, uterine leiomyomata, endometrial polyps, adenomyosis, and bleeding disorders.

Chronic Endometritis

Chronic endometritis is a histologic diagnosis, based on the finding of plasma cells in the endometrial stroma. The condition can result from infections (chlamydia, tuberculosis, mycoplasma), intrauterine foreign bodies or growths (intrauterine contraceptive device, submucous myoma), and radiation exposure. No cause can be identified in about one-third of affected patients.[152] ***Women with symptomatic chronic endometritis typically present with abnormal uterine bleeding, which can vary from intermenstrual spotting and postcoital bleeding to menorrhagia.*** Some may have vague, crampy lower abdominal pain. The most common physical manifestation is uterine or cervical motion tenderness, but many or most have no symptoms at all.

Chronic endometritis is seldom the direct cause, but often may be an indirect or contributing cause of abnormal bleeding. Inflammatory cells release proteolytic enzymes that damage the subepithelial capillary plexus and surface epithelium, rendering them fragile and prone to breaks and micro-erosions. Proteases also interfere with both repair processes and new vessel formation. In addition, leukocytes and macrophages release platelet-activating factor and prostaglandins which are potent vasodilators.

Chronic inflammation related to a foreign body reaction is almost certainly directly responsible for the increased menstrual bleeding associated with a copper intrauterine device (IUD) and one mechanism that may cause abnormal bleeding in women with retained products of conception. Histologic studies suggest that chronic endometritis also contributes to abnormal bleeding relating to submucous and deep intramural myomas and endometrial polyps (discussed below).[153]

Myomas and Polyps

Uterine leiomyomas are extremely common, and abnormal uterine bleeding is the most common clinical problem they cause. However, most women with uterine fibroids do not experience abnormal bleeding. The high prevalence of uterine myomas guarantees that they will often be identified in women who are also anovulatory or have other causes of bleeding. ***Myomas cannot, therefore, be regarded as the cause of abnormal bleeding before other obvious possibilities have been excluded, particularly when they do not protrude into or displace the uterine cavity.*** Fibroids may be the cause of heavy or prolonged bleeding in ovulatory women, they may aggravate the bleeding that results from anovulation or other causes, or they may represent only an incidental finding. Transvaginal ultrasonography generally provides accurate information regarding the size, number, and location of myomas, but images can be difficult to interpret when fibroids are multiple and large. Sonohysterography more clearly defines the proximity of myomas to the uterine cavity and can thus help to differentiate clinically relevant myomas from those that are not.[112]

The mechanisms by which uterine myomas may cause abnormal bleeding are not entirely clear but seem closely related to their location. Histologic studies suggest that

submucous and large deep intramural myomas cause the overlying endometrium to stretch. Compression from below and trauma from intracavitary friction at the epithelial surface combine to cause focal chronic inflammation or even ulceration, resulting in bleeding.[153] In compressed or damaged endometrium, other hemostatic mechanisms like platelet plug formation also may be impaired. Erosion and rupture of the larger caliber surface vessels observed on some myomas can further contribute to prolonged or heavy bleeding.[35, 154] The greater surface area of a grossly enlarged uterine cavity probably explains menorrhagia in women with fibroids that are numerous and large but distant from the endometrium.

In some women, medical treatment can be helpful in the management of abnormal bleeding directly related to uterine myomas. Estrogen-progestin contraceptives can decrease the volume and duration of blood loss in the same way they do in women without fibroids; benefits are less likely in women with submucous myomas. Non-steroidal anti-inflammatory drugs and gonadotropin-releasing hormone agonists also have benefits to offer and are discussed below.

The surgical management of abnormal bleeding resulting from or aggravated by uterine myomas must be individualized after considering the size, number, and location of the fibroids, the relative risks, benefits, and consequences of different surgical treatments, age, and desire for future fertility. In general, hysteroscopic myomectomy is a logical choice for single small submucous myomas, regardless of age and future reproductive goals. Hysteroscopic surgery for large and multiple submucous myomas requires greater technical expertise and poses greater risks, including sterility resulting from severe postoperative intrauterine adhesions—an important consideration in women who hope to preserve their fertility. Submucous myomas that extend deeply into the myometrium can be managed by subtotal hysteroscopic myomectomy, abdominal myomectomy, or hysterectomy, depending on surgical skill and the need to preserve fertility. For those experienced with the procedure, laparoscopic myomectomy offers another option for women who have not yet completed childbearing, but the laparoscopic approach does not eliminate the risk of pelvic adhesions or the need for caesarean delivery. Hysterectomy is certainly an option for women with abnormal uterine bleeding, multiple large fibroids, and no interest in future pregnancy.

Endometrial polyps often cause abnormal bleeding, most likely due to vascular fragility, chronic inflammation, and surface erosions. Larger pedunculated polyps may develop ischemic necrosis at their apex that extends to the subsurface capillaries as a consequence of intermittent torsion and related thrombosis. ***When polyps are identified by transvaginal ultrasonography or sonohysterography, hysteroscopic surgery offers a simple and highly effective remedy.***[35]

Adenomyosis

Adenomyosis is a disorder characterized by the extension of endometrial glands and stroma into the myometrium, a relatively common finding in hysterectomy specimens from women with menorrhagia unrelated to uterine myomas or endometrial pathology. Hypertrophy and hyperplasia in the surrounding myometrium generally results in diffuse uterine enlargement.[155] However, some women develop focal nodular lesions called adenomyomas (exaggerated myometrial proliferation around foci of ectopic endometrium), which resemble leiomyomas clinically. The pathogenesis of adenomyosis and associated menorrhagia are unknown. The disease may develop from endomyometrial invagination of the endometrium or develop *de novo* from müllerian rests.[156–158]

In symptomatic women, transvaginal ultrasonography can suggest the diagnosis; myometrial cysts are the most specific diagnostic criterion.[159] A meta-analysis including 14 studies involving women who had ultrasonography performed before hysterectomy found that

sonography has 83% sensitivity and 85% specificity for diagnosis of adenomyosis.[160] MRI is a more sensitive diagnostic technique,[161] particularly in the presence of uterine myomas; thickening of the junctional zone on T2 weighted imaging is characteristic.[155, 159, 162] However, the costs of MRI are difficult to justify when results will not affect clinical management.

There have been no large or controlled studies of medical or limited surgical treatment for adenomyosis. In individual patients, continuous treatment with progestins, suppression by treatment with a GnRH agonist, and aromatase inhibitors can be effective, as they are in patients with endometriosis.[163] ***Growing evidence indicates that insertion of a LNG-IUS can be very effective in providing relief from both menorrhagia and dysmenorrhea in women with adenomyosis.***[164–168] Conservative surgery can be technically challenging because, unlike the pseudocapsule that typically surrounds leiomyomas, there is no clear plane separating adenomyotic nodules from normal myometrium. Uterine artery embolization has proven effective in some women, but is not appropriate for women who have not completed childbearing.[169–171]

Bleeding Disorders

Numerous studies have documented the association between menorrhagia and inherited coagulation defects.[66–68, 90, 97, 172] Combined, they provide ample justification for performing screening coagulation studies in women with unexplained menorrhagia.

Von Willebrand disease is the most common inherited bleeding abnormality affecting women. The disease is the consequence of quantitative or qualitative defects of von Willebrand factor, a protein that plays an important role in primary hemostasis by binding to both platelets and the vascular endothelium, forming a bridge between them and adjacent platelets at the site of injury.[173, 174] By serving as a carrier protein for factor VIII and thereby prolonging its half-life in the circulation, Von Willebrand factor also contributes to fibrin clot formation. There are three types of inherited von Willebrand disease.[175–177] Type I is an autosomal dominant disease, the most common, and results in a quantitative deficiency of von Willebrand factor. Type 2, also usually autosomal dominant, has four subtypes, all of which involve qualitative abnormalities of von Willebrand factor. Type 3 is an autosomal recessive disorder resulting in a total deficiency of von Willebrand factor and severe disease. The tendency to excess bleeding can vary widely, even within individuals and their families. There are also acquired forms of von Willebrand disease associated with a variety of different diseases.

Desmopressin (dDAVP) is a synthetic analog of arginine vasopressin that has been used to treat abnormal uterine bleeding in women with coagulation disorders, especially those with von Willebrand disease.[178–180] The drug promotes the release of von Willebrand factor from endothelial cell storage sites and also may have other actions.[181] Desmopressin can be administered intravenously, subcutaneously, or by intranasal spray. The nasal spray formulation generally is recommended for home and prophylactic treatment of von Willebrand disease. Treatment induces a rapid increase in coagulation factor VIII and von Willebrand factor that lasts 6–12 hours. ***Although its effects may be only modest, desmopressin has been used successfully in the management of heavy menstrual bleeding in women with von Willebrand disease, beginning treatment with the onset of menses.***[179, 180, 182–184]

Antifibrinolytic therapy is an alternative to treatment with desmopressin in women with menorrhagia relating to von Willebrand disease. Tranexamic acid prevents clot dissolution, particularly in mucous membranes having naturally high fibrinolytic activity such as the endometrium.[184, 185] Estrogen-progestin contraceptives or the LNG-IUS also help to reduce the volume and duration of menses in women with von Willebrand disease.[186, 187]

Other Treatments for Heavy Menstrual Bleeding

A specific cause for heavy or prolonged menstrual bleeding in ovulatory women cannot always be identified; local defects in endometrial hemostasis are presumed responsible.[188] Nevertheless, the problem still can be effectively managed using a variety of nonspecific medical and surgical therapies.

Nonsteroidal Anti-Inflammatory Drugs

There is little question that prostaglandins have important actions on the endometrial vasculature and in endometrial hemostasis. The concentrations of PGE_2 and $PGF_{2\alpha}$ increase progressively in human endometrium during the menstrual cycle and are found in high concentrations in menstrual endometrium.[35] Nonsteroidal anti-inflammatory drugs (NSAIDs) inhibit prostaglandin synthesis and decrease menstrual blood loss. NSAIDs may also alter the balance between thromboxane A_2 (a vasoconstrictor and promoter of platelet aggregation) and prostacyclin (a vasodilator and inhibitor of platelet aggregation).[189]

Although the exact mechanism involved is unclear, NSAIDs decrease both normal menstrual bleeding and the increased bleeding associated with an intrauterine device.[190, 191] ***In general, NSAID treatment reduces blood loss by approximately 20–40% and to a greater extent in those with excessive bleeding.***[192–194] Ibuprofen (400 mg, 3 times daily) and mefenamic acid (500 mg 3 times daily) have been studied most extensively, but no NSAID offers any clear advantage.[194] Treatment with NSAIDs might be considered the first line therapy for ovulatory women with heavy menstrual bleeding and no demonstrable pathology. Side effects are few because treatment is limited, usually beginning with the onset of bleeding and continuing for 3–5 days as necessary. NSAIDs have the added advantage of providing relief from dysmenorrhea, even when menses are normal.

Estrogen-Progestin Contraceptives

Estrogen-progestin contraceptives can be used to reduce menstrual blood loss in ovulatory women with heavy menstrual bleeding, regardless whether menorrhagia is associated with pathology (myomas, adenomyosis) or is unexplained. ***In women with unexplained menorrhagia, estrogen-progestin contraceptives can be expected to decrease bleeding by up to 40%.***[195, 196]

The Levonorgestrel-Releasing Intrauterine System

The levonorgestrel-releasing intrauterine system (LNG-IUS; Mirena) has a reservoir containing 52 mg levonorgestrel mixed with polydimethylsiloxane, which controls the rate of hormone release. For contraceptive purposes, the device is approved for 5 years, but lasts for 7 years, and perhaps up to 10 years. ***Menstrual blood loss in women with heavy menstrual bleeding can be reduced by 75–95%, due to progestin-induced decidualization of the endometrium.***[197, 198] Data from randomized trials indicate that the decrease in menstrual blood loss achieved with the LNG-IUS is greater than with cyclic administration of norethindrone (5 mg 3 times daily, cycle days 5 to 26),[198, 199] NSAIDs,[200, 201] or tranexamic acid,[202] and

approaches or even equals that achieved with endometrial ablation.[203, 204] Patient satisfaction with the LVG-IUS also compares favorably to that with ablation or hysterectomy.[203] ***The LNG-IUS is an attractive option for ovulatory women with heavy menstrual bleeding and for women with intractable bleeding associated with chronic illnesses (renal failure).***

Gonadotropin-Releasing Hormone Agonists

Treatment with a long-acting gonadotropin-releasing hormone agonist (GnRHa) can achieve short-term relief from a bleeding problem and has been used effectively as a preoperative adjunct in women awaiting conservative (myomectomy, endometrial ablation) or definitive surgery (hysterectomy) for abnormal bleeding.

In women with severe anemia resulting from menorrhagia, preoperative GnRHa-induced amenorrhea can provide temporary relief from further bleeding, allow hemoglobin levels to return to normal, and decrease the probability of transfusion with surgery. GnRHa treatment also will often decrease the size of myomas and overall uterine mass. In women with large fibroids awaiting hysterectomy, the effect can provide an added benefit by allowing vaginal surgery when an abdominal operation might otherwise have been required. In women awaiting myomectomy, a GnRHa-induced decrease in the size and firmness of myomas can make the identification and removal of fibroids more difficult. As a method for thinning the endometrium before ablation, GnRHa treatment improves operating conditions and outcomes.[205]

GnRHa treatment is also useful in the management of abnormal menstrual bleeding that may follow organ transplantation where the toxicity of immunosuppressive drugs makes use of sex steroids less desirable. However, the expense and side effects resulting from estrogen deficiency (hot flashes, bone mineral depletion) make GnRHa an unattractive long-term strategy for treatment of abnormal bleeding.

Tranexamic Acid

Tranexamic acid is an antifibrinolytic agent that has been used widely in Europe for the treatment of menorrhagia. The drug reversibly blocks lysine binding sites on plasminogen, thereby preventing fibrin degradation. An oral form of the drug was approved by the U.S. Food and Drug Administration in 2009 for the treatment of heavy menstrual bleeding. The drug is administered for 4–7 days during menses (1.0–1.5 gm 3–4 times daily) and decreases menstrual blood loss by 35–60%.[206, 207] The risk for thrombosis associated with tranexamic acid is controversial.[208] Consequently, it has limited value in women with contraindications to hormone therapy, because most relate to the risk of thrombosis.

Endometrial Ablation

Persistent bleeding despite treatment is both frustrating and concerning. Myomas and polyps usually can be removed, with improvement or resolution of abnormal bleeding. Hysterectomy is an appropriate choice for some, but many prefer to avoid a major operation if possible, and still others have conditions that make them poor candidates for major surgery. Endometrial ablation is another, increasingly popular option for the management of unexplained menorrhagia when medical treatments are rejected, unsuccessful, or poorly tolerated.[209]

A wide variety of methods have been developed for endometrial ablation. The first method described was hysteroscopic nd:—YAG (neodymium: yttrium, aluminum, garnet) laser photovaporization, almost 30 years ago.[210] Soon thereafter, less costly techniques were developed using electrosurgical instruments (resectoscopic loop, roller ball).[211, 212] A number of randomized controlled trials have compared hysteroscopic electrosurgical endometrial ablation to hysterectomy as treatment for heavy menstrual bleeding. Overall, hysterectomy involves longer operating and recovery time, a higher risk for complications, and greater expense, but provides a permanent solution; the need for retreatment of many women after ablation narrows the cost difference over time. Satisfaction rates with both procedures are high.[213]

Several additional techniques for endometrial ablation have been developed; most do not require hysteroscopy.[214] Hysteroscopic approaches include a bipolar vaporizing electrode[215] and a hydrothermal technique.[216] Two different balloon devices are available, one that circulates heated water (87°±5°C) inside the balloon[217, 218] and the other using electrodes on the outer surface and radiofrequency-induced thermal destruction.[219] Another method involves a gold-plated mesh electrode that conforms to the uterine cavity and bipolar radiofrequency thermal ablation.[220, 221] Still others employ microwave,[222] laser,[223] and cryosurgical technologies.[224] ***Compared to traditional hysteroscopic methods, the "blind" techniques for ablation are technically easier to perform, take less time, are more likely to require only local anesthesia, and achieve similar results, but equipment problems are more common.***[188, 225]

Although all methods are effective, there are reasons for choosing one method over another in individual women. Cryoablation may be the best option for women who prefer an office-based procedure using minimal or no anesthesia.[225] Radiofrequency electrosurgical ablation is another excellent office-based method and does not require medical pretreatment to thin the endometrium. Hydrothermal ablation is the procedure of choice for women having an abnormally shaped uterine cavity, unrelated to uterine myomas. For women with submucous fibroids smaller than 3 cm, microwave ablation may be ideal, although hysteroscopic myomectomy and rollerball electrosurgical ablation also is appropriate and is preferred for women with larger submucous myomas.

Method	*Advantages*	*Disadvantages*
Cryoablation	Not completely blind	No outcomes data for women with intracavitary lesions
	Less pain than methods using heat energy	
	Requires minimal or no anesthesia	
Thermal balloon ablation	First global technique approved for use	Not recommended for women with an abnormal uterine cavity (anomaly, enlarged, polyps, myomas, adhesions)
	Easy to learn	
Hydrothermal ablation	Circulating hot water contacts all endometrial surfaces, regardless of shape	Not recommended for women with a uterus > 10 cm
	Direct visualization of uterine cavity	Requires 8 mm hysteroscope
		Hot water stimulates pain
		Risk for burns to vagina and perineum
Bipolar radiofrequency ablation	Short procedure time	Not recommended for women with an enlarged or abnormal uterine cavity
	Easy to perform	
	Requires no endometrial pretreatment	
Microwave ablation	Applicable in women with large cavity or small myomas (<3 cm)	Requires pretreatment ultrasonography to document minimum 1 cm myometrial thickness in all areas
		Contraindicated for women with previous transmural myomectomy or classical cesarean section

Preoperative evaluation before ablation should include thorough examination of the uterine cavity by saline sonography or office hysteroscopy to exclude focal lesions such as polyps and myomas that can be resected, and to identify women not having a normally shaped uterine cavity who may not be appropriate candidates for some global ablation methods like the thermal balloon. Best results with ablation can be achieved if the endometrium also is first rendered thin and inactive, to increase the likelihood that ablation will include the basal layer of the endometrium, which is 4–6 mm beneath the surface, depending on the phase of the cycle. Several methods to achieve that purpose have been described, including curettage immediately prior to performing the ablation and preoperative treatment with progestins, estrogen-progestin contraceptives, danazol, and GnRH agonists.[205, 226] Bipolar radiofrequency ablation is an exception, in that the technique is equally effective with and without endometrial pretreatment.

Among women with menorrhagia who undergo an endometrial ablation procedure, 80–90% report reduced bleeding, 25–50% develop amenorrhea, 70–80% report less menstrual pain, 75–90% are satisfied with the surgical outcome, and 80% require no additional surgery up to 5 years after ablation.[216, 218, 220, 222–225, 227–229] In a recent study of outcomes 10 years after endometrial ablation, 94% of women indicated they would recommend the procedure to a friend.[230] Approximately 10% of women who have an endometrial ablation will later have a hysterectomy. Overall, despite the lower risks, fewer complications, and more rapid recovery associated with endometrial ablation, women treated with hysterectomy tend to be more satisfied with the outcome.[231–234]

For obvious reasons, endometrial ablation is not an appropriate treatment for women who have not completed childbearing. Conversely, endometrial ablation is not a sterilization procedure. Although uncommon, pregnancy is still possible after ablation and is association with an increased risk for complications, including miscarriage, antepartum hemorrhage, preterm delivery, and abnormal placentation.[235, 236] Consequently, sexually active women still require contraception after endometrial ablation.

There are legitimate concerns that endometrial carcinomas might be inadvertently treated by endometrial ablation[237, 238] or that the procedure might obliterate portions of the uterine cavity leaving isolated, residual islands of endometrium in which adenocarcinoma could develop and go unrecognized in the absence of bleeding.[239, 240] These observations emphasize the importance of thorough preoperative evaluation, to include endometrial biopsy, and proper patient selection for ablation procedures. ***Although the risk can never be completely avoided, endometrial ablation is not recommended for women at increased risk for endometrial cancer (obesity, diabetes, hypertension, smoking, family history, chronic anovulation).***[241] ***Importantly, in women who have had an endometrial ablation and receive postmenopausal hormone therapy, treatment also should include a progestin.***

Other complications of endometrial ablation include hematometra, cervical stenosis, and uterine perforation. Hematometra develops when active islands of endometrium remain above ablated areas that adhere. Thorough ablation of the upper limits of the cavity, including the cornua and tubal ostia but excluding the cervix and cervical-uterine junction, decreases risk for the complication. Uterine perforation complicates 1% or less of endometrial ablation procedures.

Endometrial ablation can be an effective treatment for women with acute or prolonged bleeding who are hemodynamically stable when medical treatment fails or is contraindicated.[242–245] In one reported series of 26 women with acute severe bleeding treated by hysteroscopic endometrial resection, no further medical or surgical treatment was required for 24 of the women over 19 months of follow-up; one with endometrial cancer detected in the surgical specimen and another with uterine myomas required hysterectomy.[242]

Summary of Clinical Principles

Considering that abnormal menstrual bleeding is the single most common complaint that reproductive age women bring to their physician, all clinicians who provide primary care for women must have an organized, logical approach to the evaluation and treatment of the problem. The following summarizes the key elements of the clinical evaluation and treatment of abnormal menstrual bleeding in premenopausal women.

Diagnostic Evaluation

- Anovulatory bleeding is usually irregular, infrequent, and unpredictable, variable in amount, duration, and character, and most often observed in adolescents and aging women, the obese, and women with the clinical features of polycystic ovary syndrome.
- Regular and predictable but increasingly heavy or prolonged periods or new onset of episodic intermenstrual bleeding more often result from an anatomic abnormality than from anovulation.
- Recurrent episodes of intermenstrual bleeding often result from intrauterine pathology and warrant evaluation.
- The most common cause of a sudden departure from a well established pattern of predictable menses is a complication of pregnancy.
- The evaluation of women with a complaint of abnormal uterine bleeding should include a pregnancy test and complete blood count to exclude the possibility of pregnancy and to identify those with anemia and thrombocytopenia.
- When the clinical history and examination clearly point to anovulatory bleeding, empiric medical treatment can be offered without additional laboratory evaluation or imaging.
- A well timed serum progesterone determination can help to confirm the diagnosis of anovulatory bleeding when doubt exists. A serum TSH can exclude thyroid disorders in anovulatory women. Liver or renal function tests are indicated only for those with known or strongly suspected disease.
- Bleeding disorders are more common than is generally perceived. Coagulation tests are indicated for adolescents with menorrhagia from menarche, women with past episodes of excessive bleeding from trauma or surgery, and those with unexplained heavy or prolonged menstrual bleeding.
- Endometrial biopsy should be seriously considered before treatment begins when the clinical history suggests long-term unopposed estrogen exposure, regardless of age, but is unnecessary when the endometrium is very thin (<5 mm). Endometrial biopsy should be performed when the endometrium is abnormally thick (>12 mm), even when clinical suspicion of disease is low.
- Uterine imaging with ultrasound or sonohysterography should be performed when examination reveals abnormal uterine size or contours, when history (regular cycles of increasing volume or duration, new onset intermenstrual bleeding), laboratory tests (serum progesterone >3 ng/mL), or biopsy results (secretory endometrium) provide objective evidence of ovulation, and when empiric medical treatment fails.
- The combination of sonohysterography and endometrial biopsy has high sensitivity and high negative predictive value for detection of endometrial and uterine pathology in women with abnormal bleeding.

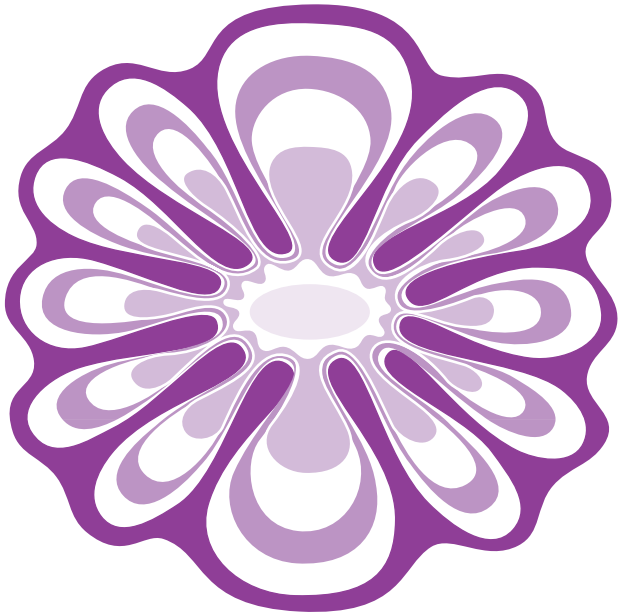
Treatment

- Cyclic progestin therapy is appropriate treatment for oligomenorrheic anovulatory women with episodic abnormal bleeding who do not need contraception, but treatment with an estrogen-progestin contraceptive is otherwise the better choice. Standard cyclic progestin treatments do not reliably suppress the hypothalamic-pituitary-ovarian axis, will not prevent random ovulation, and are not contraceptive.
- Estrogen therapy is the best initial treatment when a denuded or attenuated endometrium is strongly suspected or demonstrated. Clinical examples include women in whom biopsy yields minimal tissue, women receiving chronic progestin treatment, and women with prolonged heavy bleeding. Progestin or estrogen-progestin therapy is unlikely to succeed and may aggravate the problem.
- Failed medical management for presumed anovulatory bleeding suggests strongly that other pathology is causing or contributing to the bleeding and signals the need for additional diagnostic evaluation.
- In women with acute heavy bleeding, imaging with transvaginal ultrasonography helps to guide the choice of treatment by defining the endometrial thickness and revealing anatomic abnormalities not otherwise suspected.
- Acute prolonged episodes of heavy anovulatory bleeding can be treated effectively with high-dose estrogen-progestin therapy, or with high-dose progestin alone (when estrogen is contraindicated), provided that the endometrium is normal or increased in thickness.
- Treatment with depot medroxyprogesterone acetate has no place in the acute management of abnormal bleeding. Once administered, it cannot be withdrawn, and if unsuccessful, its effects can be difficult to overcome.
- Endometrial curettage should be performed when bleeding is acute and demands immediate action or fails to respond promptly to intensive medical therapy. Hysteroscopy at time of curettage helps to ensure an accurate diagnosis.
- Endometrial hyperplasia without cytologic atypia is an exaggerated form of persistent proliferative endometrium resulting from long-term unopposed estrogen stimulation in women with chronic anovulation. With few exceptions, the lesion can be treated effectively with cyclic or continuous progestin therapy or by insertion of a levonorgestrel-releasing intrauterine system.
- Endometrial hyperplasia with cytologic atypia is a precancerous lesion best treated surgically except in women intent on preserving reproductive potential. Medical management of atypical endometrial hyperplasia requires high doses and longer durations of progestin treatment or insertion of a levonorgestrel-releasing intrauterine system, serial endometrial biopsies to monitor response, and longer-term close surveillance.
- Uterine myomas are extremely common and cannot be regarded as the cause of abnormal bleeding before other possibilities have been excluded, particularly when they do not protrude into or displace the uterine cavity. Sonohysterography clearly defines the proximity of myomas to the uterine cavity and helps to differentiate clinically relevant myomas from those that are not.
- Desmopressin is very effective for the management of heavy menstrual bleeding in women with von Willebrand disease, beginning treatment with the onset of menses. Tranexamic acid, estrogen-progestin contraceptives, or insertion of a levonorgestrel-releasing intrauterine system also help to reduce the volume and duration of menses in women with coagulation disorders.
- Nonsteroidal anti-inflammatory drugs, estrogen-progestin contraceptives, the levonorgestrel-releasing intrauterine system, and tranexamic acid are effective medical treatment options for the management of heavy menstrual bleeding in ovulatory women with adenomyosis, global cavity enlargement related to multiple intramural leiomyomata, and otherwise unexplained menorrhagia.

- Endometrial ablation using hysteroscopic or non-hysteroscopic techniques is an effective alternative to hysterectomy for management of abnormally heavy menstrual bleeding when medical treatments are rejected, unsuccessful, or poorly tolerated.

All references are available online at: http://www.clinicalgynendoandinfertility.com

16 The Breast



The form, function, and pathology of the human female breast are major concerns of medicine and society. As mammals, we define our biologic class by the function of the breast in nourishing our young. Breast contours occupy our attention. As obstetricians, we seek to enhance or diminish function, and as gynecologists, the appearance of inappropriate lactation (galactorrhea) may signify serious disease. Cancer of the breast is the most prevalent cancer in women.

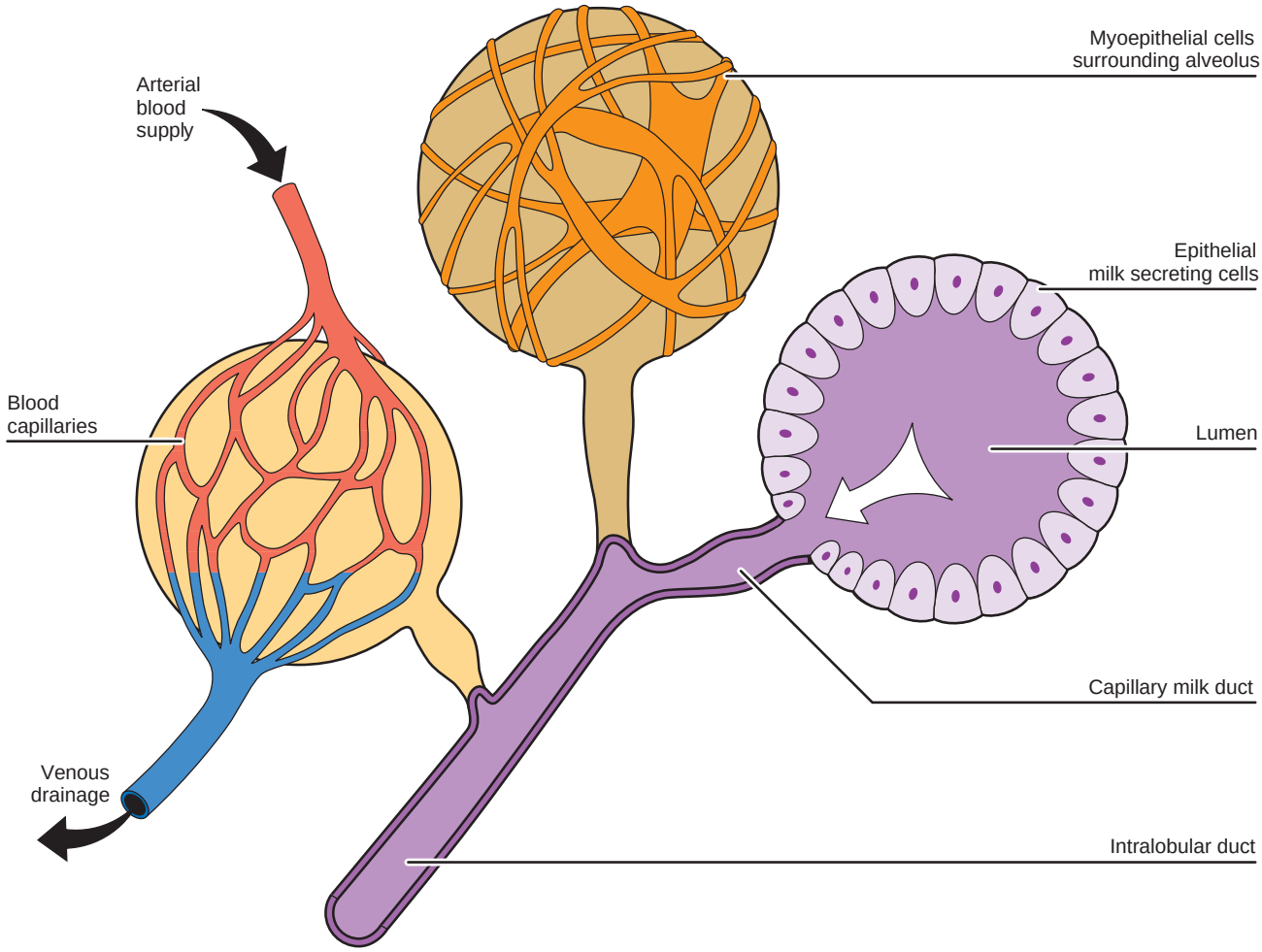
This chapter reviews the factors involved in normal growth and development of the breast, including the physiology of normal lactation, describes the numerous factors leading to inappropriate lactation, and, finally, discusses the endocrine aspects of breast cancer.

Growth and Development

The basic component of the breast lobule is the hollow alveolus or milk gland lined by a single layer of milk-secreting epithelial cells, derived from an ingrowth of epidermis into the underlying mesenchyme at 10–12 weeks of gestation. Each alveolus is encased in a crisscrossing mantle of contractile myoepithelial strands. Also surrounding the milk gland is a rich capillary network.

The lumen of the alveolus connects to a collecting intralobular duct by means of a thin non-muscular duct. Contractile muscle cells line the intralobular ducts that eventually reach the exterior via 15–20 collecting ducts in a radial arrangement, corresponding to the 15–20 distinct mammary lobules in the breast, each of which contains many alveoli.

Growth of this milk-producing system is dependent on numerous hormonal factors that occur in two sequences, first at puberty and then in pregnancy. Although there is considerable overlapping of hormonal influences, the differences in quantities of the stimuli in each circumstance and the availability of entirely unique inciting factors (human placental lactogen and prolactin) during pregnancy permit this chronologic distinction. The strength of the hormonal stimulus to breast tissue during pregnancy is responsible for the fact that nearly half of male and female newborns have breast secretions.

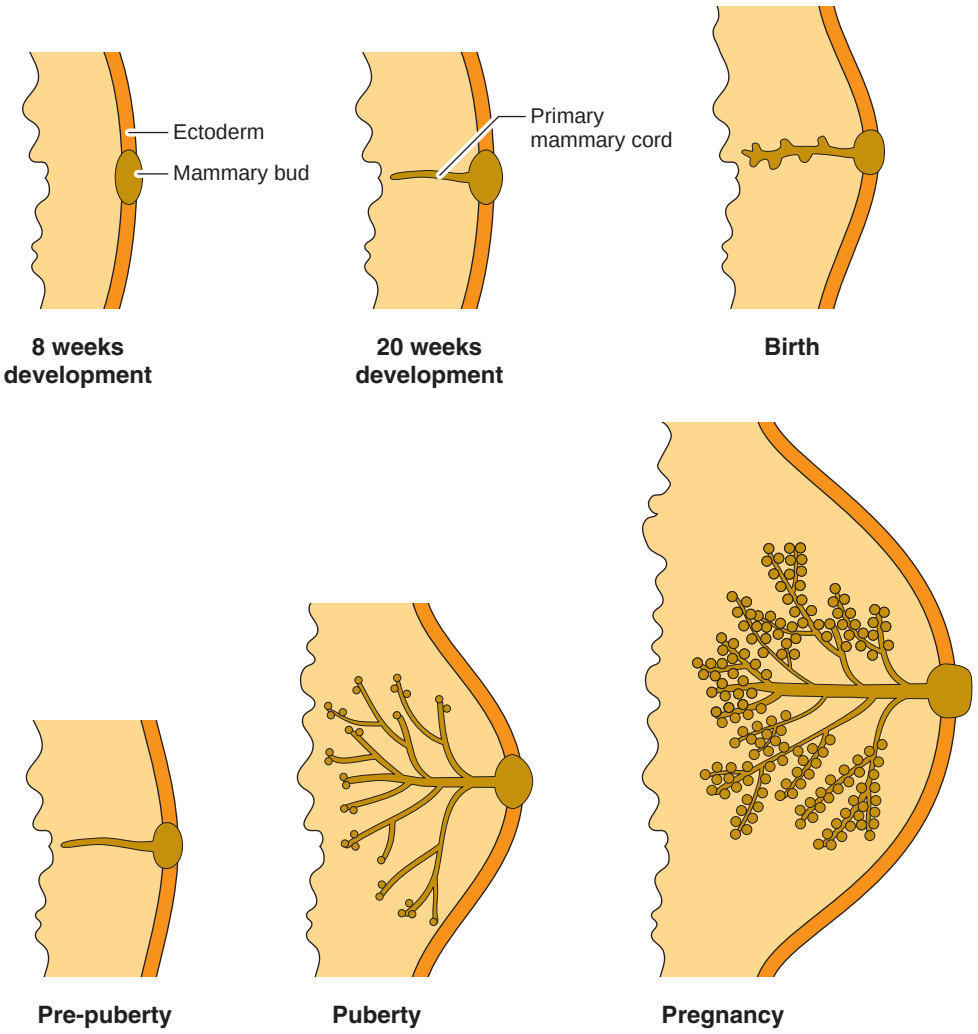


The major influence on breast growth at puberty is estrogen. In most girls, the first response to the increasing levels of estrogen is an increase in size and pigmentation of the areola and the formation of a mass of breast tissue just underneath the areola. Breast tissue binds estrogen in a manner similar to the uterus and vagina. The human breast expresses both estrogen receptors, ER-α and ER-β.[1] The development of estrogen receptors in the breast does not occur in the absence of prolactin. The primary effect of estrogen in subprimate mammals is to stimulate growth of the ductal portion of the gland system. Progesterone in these animals, in the presence of estrogen, influences growth of the alveolar components

of the lobule that later become the milk-producing structures.[2] However, neither hormone alone, or in combination, is capable of yielding optimal breast growth and development. Full differentiation of the gland requires insulin, cortisol, thyroxine, prolactin, and especially, growth hormone-induced insulin-like growth factor-I.[3, 4] Experimental evidence in mice knock-out models supports the combined actions of estrogen and progesterone, mediated primarily by estrogen receptor-α and progesterone receptor-B, but dependent on epidermal growth factor and IGF-I.[5–7] Estrogen and progesterone receptors in normal breast tissue are located in non-dividing epithelial cells and in stromal cells adjacent to proliferating epithelial cells, indicating the importance of paracrine communication using growth factors. Growth hormone-induced IGF-I is essential in both mammary development and function.[4]

The pubertal response is a manifestation of closely synchronized central (hypothalamus-pituitary) and peripheral (ovary-breast) events. For example, gonadotropin-releasing hormone (GnRH) is known to stimulate prolactin release, and this action is potentiated by estrogen.[8] This suggests a paracrine interaction between gonadotrophs and lactotrophs, linked by estrogen, ultimately with an impact on the breast.

Changes occur routinely in response to the estrogen-progesterone sequence of a normal menstrual cycle. Maximal size of the breast occurs late in the luteal phase. Fluid secretion,



mitotic activity, and DNA production of nonglandular tissue and glandular epithelium peak during the luteal phase.[9–11] This accounts for cystic and tender premenstrual changes.

During the normal menstrual cycle, estrogen receptors in mammary gland epithelium decrease in number during the luteal phase, whereas progesterone receptors remain at a high level throughout the cycle.[12] Studies using tissue from reduction mammoplasties or from breast tissue near a benign or malignant lesion have demonstrated a peak in mitotic activity during the luteal phase.[10, 13, 14] Using fine-needle biopsy tissue, an immunocytochemical marker of proliferation was higher in the luteal phase than in the proliferative phase.[12] And in this study there was a direct correlation with serum progesterone levels. However, important studies indicate that with increasing duration of exposure, progesterone imposes a limitation on breast cell proliferation.[15–17] Therefore, breast and endometrium epithelial cells may be more similar than conventionally proposed.

Final differentiation of the alveolar epithelial cell into a mature milk cell is accomplished by the gestational increase in estrogen and progesterone, combined with the presence of prolactin, but only after prior exposure to cortisol and insulin. The complete reaction depends on the availability of minimal quantities of thyroid hormone. Thus, the endocrinologically intact individual in whom estrogen, progesterone, thyroxine, cortisol, insulin, prolactin, human placental lactogen, and growth hormone are available can have appropriate breast growth and function. During the first trimester of pregnancy, growth and proliferation are maximal, changing to differentiation and secretory activity as pregnancy progresses.

Breast tissue changes with aging. During teenage years the breasts are dense and predominantly glandular. As the years go by, the breasts contain progressively more fat, but after menopause, this process accelerates so that soon into the postmenopausal years, the breast glandular tissue is mostly replaced by fat.

Abnormal Shapes and Sizes

Early differentiation of the mammary gland anlage is under fetal hormonal control. Abnormalities in adult size or shape may reflect the impact of hormones (especially the presence or absence of testosterone) during this early period of development. This prenatal hormonal influence programs the breast development that will occur in response to the increase in hormones at puberty. Occasionally, the breast bud begins to develop on one side first. Similarly, one breast may grow faster than the other. These inequalities usually disappear by the time development is complete. However, exact equivalence in size may not be attained. Significant asymmetry is correctable only by a plastic surgeon. Likewise hypoplasia and hypertrophy can be treated only by corrective surgery. Hormone therapy is totally ineffective in producing a permanent change in breast shape or size, with one exception, in patients with primary amenorrhea due to deficient ovarian function, estrogen treatment induces significant and gratifying breast growth. Breast size can be increased in current users of oral contraceptives, but there is no lasting effect associated with past use.[18]

Accessory nipples (almost always without underlying breast tissue) can be found anywhere from the groin to the neck, remnants of the mammary line that extends early in embryonic life (sixth week) along the ventral, lateral body wall. They occur in approximately 1% of women (sporadic or familial) and require no therapy. The presence of ***polythelia*** has been reported to be associated with a variety of renal and urinary tract malformations.[19, 20] However, in three series, each with a large number of children, the presence of supernumerary nipples was not associated with a higher prevalence of kidney and urinary tract malformations.[21–23] Nevertheless, it is prudent to investigate the renal-urinary tract in the presence of polythelia.[24]

Accessory breast tissue occurs because of incomplete embryologic regression of the mammary ridges, and for this reason, the location is along the mammary line that extends from the axilla to the pubic area. Ectopic breast tissue is usually detected during puberty, pregnancy, or lactation, a consequence of hormonally-induced enlargement. Accessory breasts are commonly bilateral, and occasionally are found in unusual locations such as the axilla, scapula, thigh, or labia majora, and when nipple and areola are absent, the mass can be a diagnostic dilemma.[25] Even when the diagnosis is obvious, surgical excision is indicated for cosmetic and comfort reasons.[26] Accessory breast tissue is subject to the same risk of cancer as normal breasts.

Pregnancy and Lactation

Prolactin Secretion

In most mammalian species, prolactin is a single-chain polypeptide of 199 amino acids, 40% similar in structure to growth hormone and placental lactogen. All three hormones are believed to have originated from a common ancestral protein about 400 million years ago.

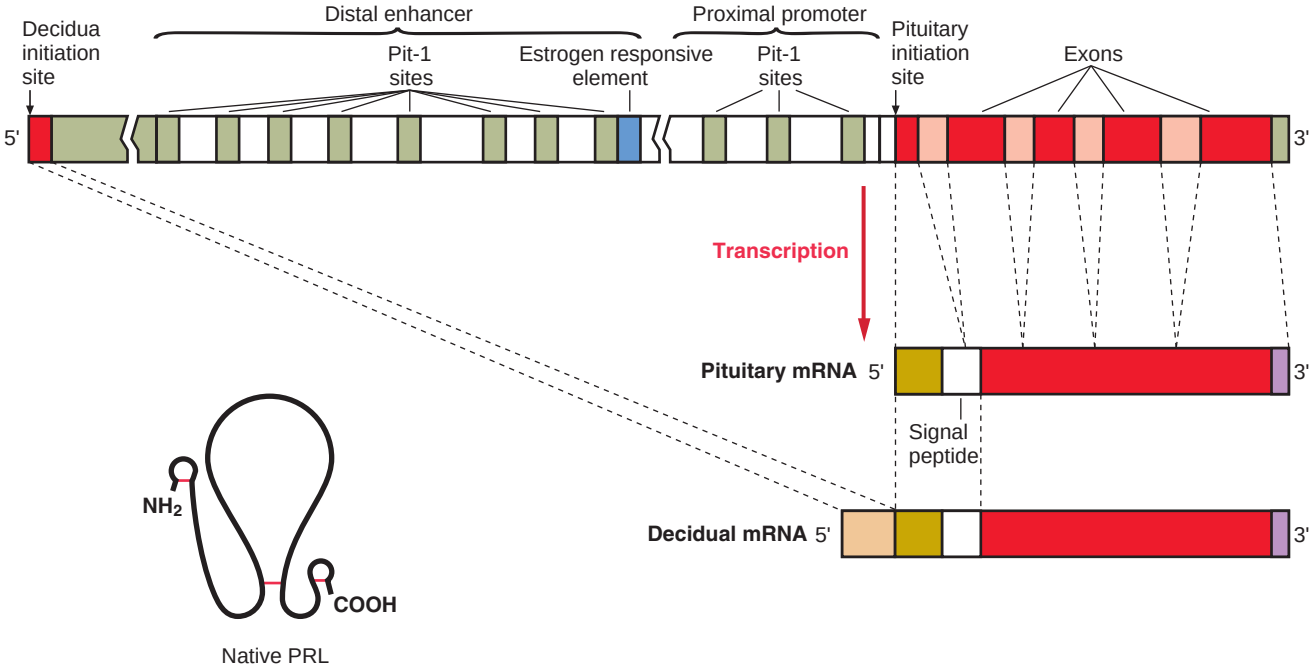
Prolactin is encoded by a single gene on chromosome 6, producing a molecule that in its major form is maintained in three loops by disulfide bonds.[27] Most, if not all, variants of prolactin are the result of posttranslational modifications. Little prolactin represents a splicing variant resulting from the proteolytic deletion of amino acids. Big prolactin can result from the failure to remove introns; it has little biologic activity and does not cross-react with antibodies to the major form of prolactin. The so-called big big variants of prolactin are due to separate molecules of prolactin binding to each other, either noncovalently or by interchain disulfide bonding. Some of the apparently larger forms of prolactin are prolactin molecules complexed to binding proteins. High levels of relatively inactive prolactin in the absence of a tumor can be due to the creation of macromolecules of prolactin by antiprolactin autoantibodies.[28, 29] Overall, big prolactins account for somewhere between 10% and 25% of the hyperprolactinemia reported by commercial assays.[30]

Other variations exist. Enzymatic cleavage of the prolactin molecule yields fragments that may be capable of biologic activity, and prolactin that has been glycosylated continues to exert activity. Differences in the carbohydrate moieties can produce differences in biologic activity and immunoreactivity. However, the nonglycosylated form of prolactin is the predominant form of prolactin secreted into the circulation.[31] Modification of prolactin also includes phosphorylation, deamidation, and sulfation.

At any one point of time, the bioactivity (e.g., galactorrhea) and the immunoreactivity (circulating levels by immunoassay) of prolactin represent the cumulative effect of the family of structural variants. Remember, immunoassays do not always reflect the biologic situation (e.g., a normal prolactin level in a woman with galactorrhea). Nevertheless, the routine radioimmunoassay of prolactin is generally clinically reliable, especially at extremely high levels associated with prolactin-secreting pituitary tumors.

The anterior pituitary cells that produce prolactin, growth hormone, and thyroid-stimulating hormone (lactotrophs, somatotrophs, and thyrotrophs) require the presence of Pit-1, a transcription factor, for transactivation. Pit-1 binds to the prolactin gene in multiple sites in both the promoter region and in an adjacent region, designated as a distal enhancer; Pit-1 binding is a requirement for prolactin promoter activity and gene transcription. Many hormones, neurotransmitters, and growth factors influence the prolactin gene, involved in a level of function beyond that allowed by Pit-1. Fundamental modulation of prolactin

secretion is exerted by estrogen, producing both differentiation of lactotrophs and direct stimulation of prolactin production.[32, 33] An estrogen response element is adjacent to one of the Pit-1 binding sites in the distal enhancer region, and estrogen stimulation of the prolactin gene involves interaction with this Pit-1 binding site. Estrogen additionally influences prolactin production by suppressing dopamine secretion.[34] Prolactin is also synthesized in extrapituitary tissues, including breast tissue and endometrial decidua.[35] In extrapituitary sites, the active promoter site is upstream of the pituitary initiation site, and is not regulated by Pit-1, estrogens, or dopamine. Progesterone increases prolactin secretion in the decidua, but has no effect in the pituitary.



Prolactin is involved in many biochemical events during pregnancy. Surfactant synthesis in the fetal lung is influenced by prolactin, and decidual prolactin modulates prostaglandin-mediated uterine muscle contractility.[36, 37] Prolactin also contributes to the prevention of the immunologic rejection of the conceptus by suppressing the maternal immune response. Prolactin is both produced and processed in breast cells. The mechanisms and purpose for mammary production of prolactin remain to be determined, but prolactin in milk is believed to be derived from local synthesis. Transmission of this prolactin to the newborn may be important for immune functions.

Prolactin-Inhibiting Factor

The hypothalamus maintains suppression of pituitary prolactin secretion by delivering a prolactin-inhibiting factor (PIF) to the pituitary via the portal circulation. Suckling suppresses the formation of this hypothalamic substance, which is believed to be dopamine (as discussed in Chapter 5).[38] Dopamine is secreted by the basal hypothalamus into the portal system and conducted to the anterior pituitary. Dopamine binds specifically to lactotroph cells and suppresses the secretion of prolactin into the general circulation; in its absence, prolactin is secreted. Dopamine binds to a G-protein-coupled receptor (Chapter 2) that

exists in a long form and a short form, but only the D_2 (long form) is present on lactotrophs. The molecular mechanism for dopamine's inhibitory action is still not known. There are several other PIFs, but a specific role has been established only for dopamine.

Prolactin Releasing Factor

Prolactin secretion may also be influenced by a positive hypothalamic factor, prolactin-releasing factor (PRF). PRF does exist in various fowl (e.g., pigeon, chicken, duck, turkey, and the tricolored blackbird). While the identity of this material has not been elucidated, or its function substantiated in normal human physiology, it is possible that thyrotropin-releasing hormone (TRH) is a potent stimulant of prolactin secretion in humans. The smallest doses of TRH that are capable of producing an increase in TSH also increase prolactin levels, a finding that supports a physiologic role for TRH in the control of prolactin secretion, at least in response to suckling.[39] TRH stimulation of prolactin release involves calcium mechanisms (both internal release and influx via calcium channels) in response to the TRH receptor, also a member of the G-protein family. However, except in hypothyroidism, normal physiologic changes as well as abnormal prolactin secretion are easily explained and understood in terms of variations in the prolactin-inhibiting factor, dopamine. A large collection of peptides has been reported to stimulate the release of prolactin in vitro. These include growth factors, angiotensin II, GnRH, vasopressin, and others. But it is unknown whether these peptides participate in the normal physiologic regulation of prolactin secretion.

The Prolactin Receptor

The prolactin receptor is encoded by a gene on chromosome 5p13–14 that is near the gene for the growth hormone receptor. The prolactin receptor belongs to a receptor family that includes many cytokines and some growth factors, supporting a dual role for prolactin as a classic hormone and as a cytokine.[27]

Prolactin receptors exist in more than one form, all containing an extracellular region, a single transmembrane region, and a relatively long cytoplasmic domain. There is evidence for more than one receptor, depending on the site of action (e.g., decidua and placenta).[40] The similar amino acid identity between prolactin and growth hormone receptors is approximately 30%, with certain regions having up to 70% homology.[41] Prolactin receptors are expressed in many tissues throughout the body. Because of the various forms and functions of prolactin, it is likely that multiple signal mechanisms are involved, and for that reason, no single second messenger for prolactin's intracellular action has been identified. A protein also exists that functions as a receptor/transporter, translocating prolactin from the blood into the cerebrospinal fluid, the amniotic fluid, and milk.

Amniotic Fluid Prolactin

Amniotic fluid concentrations of prolactin parallel maternal serum concentrations until the 10th week of pregnancy, rise markedly until the 20th week, and then decrease.

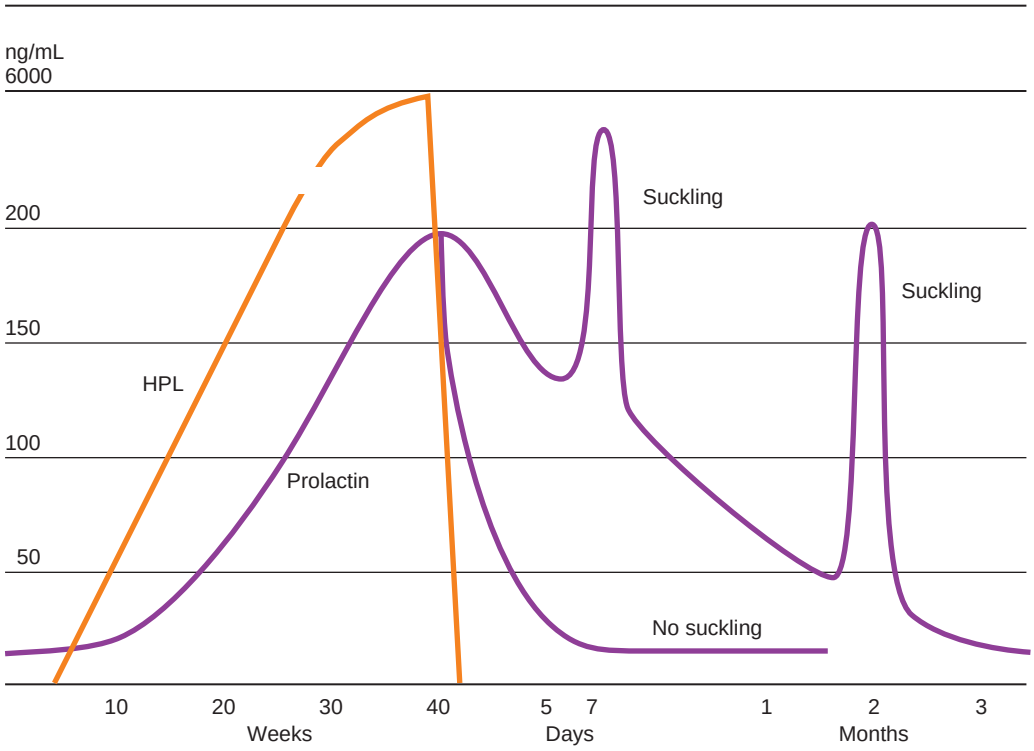
Maternal prolactin does not pass to the fetus in significant amounts. Indeed, the source of amniotic fluid prolactin is neither the maternal pituitary nor the fetal pituitary. The failure of dopamine agonist treatment to suppress amniotic fluid prolactin levels, and studies with in vitro culture systems indicate a primary decidual source with transfer via amnion receptors to the amniotic fluid, requiring the intactness of amnion, chorion, and adherent decidua. This decidual synthesis of prolactin is initiated by progesterone, but once decidualization is established, prolactin secretion continues in the absence of both progesterone and estradiol.[42] Various decidual factors regulate prolactin synthesis and release, including relaxin, insulin, and insulin-like growth factor-I. Prolactin produced in extrapituitary sites involves an alternative exon upstream of the pituitary start site, generating a slightly larger RNA transcript compared with the pituitary product. However, the amino acid sequence and the chemical and biologic properties of decidual prolactin are identical to those of pituitary prolactin. It is hypothesized that amniotic fluid prolactin plays a role in modulating electrolyte economy not unlike its ability to regulate sodium transport and water movement across the gills in fish (allowing the ocean-dwelling salmon and steelhead to return to freshwater streams for reproduction). Thus prolactin would protect the human fetus from dehydration by control of salt and water transport across the amnion. Prolactin reduces the permeability of the human amnion in the fetal to maternal direction by a receptor-mediated action on the epithelium lining the fetal surface.[43] Decidual and amniotic fluid prolactin levels are lower in hypertensive pregnancies and in patients with polyhydramnios.[44, 45] Prolactin receptors are present in the chorion laeve, and their concentration is lower in patients with polyhydramnios.[46] Thus, idiopathic polyhydramnios may be a consequence of impaired prolactin regulation of amniotic fluid.

Lactation

During pregnancy, prolactin levels rise from the normal level of 10–25 ng/mL to high concentrations, beginning about 8 weeks and reaching a peak of 200–400 ng/mL at term.[47, 48] The increase in prolactin parallels the increase in estrogen beginning at 7–8 weeks' gestation, and the mechanism for increasing prolactin secretion (discussed in Chapter 5) is believed to be estrogen suppression of the hypothalamic prolactin-inhibiting factor, dopamine, and direct stimulation of prolactin gene transcription in the pituitary.[49, 50] There is marked variability in maternal prolactin levels in pregnancy, with pulsatile secretion and a diurnal variation similar to that found in nonpregnant subjects. The peak level occurs 4–5 hours after the onset of sleep.[51]

Made by the placenta and actively secreted into the maternal circulation from the sixth week of pregnancy, human placental lactogen (hPL) rises progressively, reaching a level of approximately 6,000 ng/mL at term. hPL, though displaying less activity than prolactin, is produced in such large amounts that it may exert a lactogenic effect.

Although prolactin stimulates significant breast growth, and is available for lactation, only colostrum (composed of desquamated epithelial cells and transudate) is produced during gestation. Full lactation is inhibited by progesterone, which interferes with prolactin action at the alveolar cell prolactin receptor level. Both estrogen and progesterone are necessary for the expression of the lactogenic receptor, but progesterone antagonizes the positive action of prolactin on its own receptor while progesterone and pharmacologic amounts of androgens reduce prolactin binding.[41, 52, 53] In the mouse, inhibition of milk protein production is due to progesterone suppression of prolactin receptor expression.[54] The effective use of high doses of estrogen to suppress postpartum lactation indicates that pharmacologic amounts of estrogen also block prolactin action.



Progesterone can directly suppress milk production. A nuclear peptide (a corepressor) has been identified that binds to specific sites in the promoter region of the casein gene, thus inhibiting transcription.[55] Progesterone stimulates the generation of this corepressor. After delivery, the loss of progesterone leads to a decrease in this inhibitory peptide.

The principal hormone involved in milk biosynthesis is prolactin. Without prolactin, synthesis of lactose, lipids and the primary protein, casein, will not occur, and true milk secretion will be impossible. The hormonal trigger for initiation of milk production within the alveolar cell and its secretion into the lumen of the gland is the rapid disappearance of estrogen and progesterone from the circulation after delivery. The clearance of prolactin is much slower, requiring 7 days to reach nonpregnant levels in a nonbreastfeeding woman. These discordant hormonal events result in removal of the estrogen and progesterone inhibition of prolactin action on the breast. Breast engorgement and milk secretion begin 3–4 days postpartum when the sex steroids have been sufficiently cleared. Maintenance of steroidal inhibition or rapid reduction of prolactin secretion (with a dopamine agonist) are effective in preventing postpartum milk synthesis and secretion. Augmentation of prolactin (by TRH or sulpiride, a dopamine receptor blocker) results in increased milk yield.

In the first postpartum week, prolactin levels in breastfeeding women decline approximately 50% (to about 100 ng/mL). Suckling elicits increases in prolactin, which are important in initiating milk production. Until 2–3 months postpartum, basal levels are approximately 40–50 ng/mL, and there are large (about 10–20-fold) increases after suckling. Throughout breastfeeding, baseline prolactin levels remain elevated, and suckling produces a 2-fold increase that is essential for continuing milk production.[56, 57] The pattern or values of prolactin levels do not predict the postpartum duration of amenorrhea or infertility.[58] The failure to lactate within the first 7 days postpartum may be the first sign of Sheehan's syndrome (hypopituitarism following intrapartum infarction of the pituitary gland).

Maintenance of milk production at high levels is dependent on the joint action of both anterior and posterior pituitary factors. By mechanisms to be described in detail shortly, suckling causes the release of both prolactin and oxytocin as well as thyroid-stimulating hormone (TSH).[59, 60] Prolactin sustains the secretion of casein, fatty acids, lactose, and the volume of secretion, while oxytocin contracts myoepithelial cells and empties the alveolar lumen, thus enhancing further milk secretion and alveolar refilling. The increase in TSH with suckling suggests that thyrotropin-releasing hormone (TRH) may play a role in the prolactin response to suckling. The optimal quantity and quality of milk are dependent upon the availability of thyroid, insulin and the insulin-like growth factors, cortisol, and the dietary intake of nutrients and fluids.

Secretion of calcium into the milk of lactating women approximately doubles the daily loss of calcium.[61, 62] In women who breastfeed for 6 months or more, this is accompanied by significant bone loss even in the presence of a high calcium intake.[63] However, bone density rapidly returns to baseline levels in the 6 months after weaning.[64, 65] The bone loss is due to increased bone resorption, probably secondary to the relatively low estrogen levels associated with lactation. It is possible that recovery is impaired in women with inadequate calcium intake; total calcium intake during lactation should be at least 1,500 mg per day. Nevertheless, calcium supplementation has no effect on the calcium content of breast milk or on bone loss in lactating women who have normal diets.[66] In addition, fetuses and lactating mothers, except in unusual circumstances, do not suffer from a significant deficiency in vitamin D.[67] Furthermore, studies indicate that any loss of calcium and bone associated with lactation is rapidly restored, and, therefore, there is no impact on the risk of postmenopausal osteoporosis.[68–72] Rarely, a pregnant woman can present with osteoporosis and vertebral fractures, probably a consequence of very inadequate calcium intake and severe vitamin D deficiency.[73] Case reports of pregnancy-associated osteoporosis indicate that this acute condition can be successfully treated with either bisphosphonates or teriparatide, the parathyroid hormone fragment.[74, 75]

Antibodies are present in breast milk and contribute to the health of an infant. Besides the proteins, carbohydrates, and fats that provide a complete and balanced diet, human milk prevents infections in infants both by transmission of immunoglobulins and by modifying the bacterial flora of the infant's gastrointestinal tract. Viruses are transmitted in breast milk, and although the actual risks are unknown, women infected with cytomegalovirus, hepatitis B, or human immunodeficiency viruses are advised not to breastfeed. Vitamin A, vitamin B_{12}, and folic acid are significantly reduced in the breast milk of women with poor dietary intake. As a general rule approximately 1% of any drug ingested by the mother appears in breast milk. In a study of Pima Indians, exclusive breastfeeding for at least 2 months was associated with a lower rate of adult-onset noninsulin-dependent diabetes mellitus, probably because overfeeding and excess weight gain are more common with bottle-feeding.[76]

Frequent emptying of the lumen is important for maintaining an adequate level of secretion. Indeed, after the fourth postpartum month, suckling appears to be the only stimulant required; however, environmental and emotional states also are important for continued alveolar activity. Vigorous aerobic exercise does not affect the volume or composition of breast milk, and therefore infant weight gain is normal.[77] Maternal diet and hydration have little impact on lactation; the primary control of milk output is under the control of the infant's suckling.[78]

Suckling studied with ultrasonography indicates that the infant's instinctive attachment to a nipple immediately establishes a vacuum seal.[79] The tongue moves up and down, increasing the vacuum and producing milk flow during the downward motion. However, the ejection of milk from the breast does not occur only as the result of a mechanically induced negative pressure produced by suckling. Tactile sensors concentrated in the areola activate, via thoracic sensory nerve roots 4, 5, and 6, an afferent sensory neural arc that stimulates

the paraventricular and supraoptic nuclei of the hypothalamus to synthesize and transport oxytocin to the posterior pituitary. The efferent arc (oxytocin) is blood-borne to the breast alveolus-ductal systems to contract myoepithelial cells and empty the alveolar lumen. Milk contained in major ductal repositories is ejected from 15 to 20 openings in the nipple. This rapid release of milk is called "let-down." This important role for oxytocin is evident in knockout mice lacking oxytocin who undergo normal parturition, but fail to nurse their offspring.[80] The milk ejection reflex involving oxytocin is present in all species of mammals. Oxytocin-like peptides exist in fish, reptiles, and birds, and a role for oxytocin in maternal behavior may have existed before lactation evolved.[78]

In many instances, the activation of oxytocin release leading to let-down does not require initiation by tactile stimuli. The central nervous system can be conditioned to respond to the presence of the infant, or to the sound of the infant's cry, by inducing activation of the efferent arc. These messages are the result of many stimulating and inhibiting neurotransmitters. Suckling, therefore, acts to refill the breast by activating both portions of the pituitary (anterior and posterior) causing the breast to produce new milk and to eject milk. The release of oxytocin is also important for uterine contractions that contribute to involution of the uterus.

The oxytocin effect is a release phenomenon acting on secreted and stored milk. Prolactin must be available in sufficient quantities for continued secretory replacement of ejected milk. This requires the transient increase in prolactin associated with suckling. The amount of milk produced correlates with the amount removed by suckling. The breast can store milk for a maximum of 48 hours before production diminishes.

Breastfeeding by Adopting Mothers

Adopting mothers occasionally request assistance in initiating lactation.[81] Successful breastfeeding can be achieved by ingestion of 25 mg chlorpromazine t.i.d. together with vigorous nipple stimulation every 3–4 hours. Milk production will not appear for several weeks. This preparation ideally should be begin about a month before the expected baby is due. An electric breast pump should be used, again preferably beginning about a month before the expected baby is due to deliver. Stasis of milk within the breast, without stimulation, will lead to cessation of lactation. Metoclopramide, 10 mg t.i.d., is another drug that has produced success in increasing prolactin levels and inducing lactation.[82] Metoclopramide can also be used when nursing mothers have an inadequate milk supply. Once adequate lactation is established (usually in 7 to 10 days), drug treatment should be discontinued, tapering the dose over 3 weeks.

Cessation of Lactation

Lactation can be terminated by discontinuing suckling. The primary effect of this cessation is loss of milk let-down via the neural evocation of oxytocin. With passage of a few days, the swollen alveoli depress milk formation probably via a local pressure effect (although milk itself may contain inhibitory factors). With resorption of fluid and solute, the swollen engorged breast diminishes in size in a few days. In addition to the loss of milk let-down the absence of suckling reactivates dopamine (PIF) production so that there is less prolactin stimulation of milk secretion. Routine use of a dopamine agonist for suppression of lactation is not recommended because of reports of hypertension, seizures, myocardial infarctions, and strokes associated with its postpartum use.

Contraceptive Effect of Lactation

A moderate contraceptive effect accompanies lactation and produces child-spacing, which is very important in the developing world as a means of limiting family size. The contraceptive effectiveness of lactation, i.e., the length of the interval between births, depends on the intensity of suckling, the extent to which supplemental food is added to the infant diet, and the level of nutrition of the mother (if low, the longer the contraceptive interval; however well-nourished and undernourished women resume ovulating at the same time postpartum.[83]) If suckling intensity and/or frequency is diminished, contraceptive effect is reduced. Only amenorrheic women who exclusively breastfeed (full breastfeeding) at regular intervals, including nighttime, during the first 6 months have the contraceptive protection equivalent to that provided by oral contraception (98% efficacy); with menstruation or after 6 months, the chance of ovulation increases.[84, 85] With full or nearly full breastfeeding, approximately 70% of women remain amenorrheic through 6 months and only 37% through 1 year; nevertheless with exclusive breastfeeding, the contraceptive efficacy at 1 year is high, at 92%.[85] Fully breastfeeding women commonly have some vaginal bleeding or spotting in the first 8 postpartum weeks, but this bleeding is not due to ovulation.[86]

Supplemental feeding increases the chance of ovulation (and pregnancy) even in amenorrheic women.[87] Total protection is achieved by the exclusively breastfeeding woman for a duration of only 10 weeks.[86] Half of women studied who are not fully breastfeeding ovulate before the sixth week, the time of the traditional postpartum visit; a visit during the third postpartum week is strongly recommended for contraceptive counseling.

Rule of 3′S for Postpartum Initiation of Contraception

Full breastfeeding;	**Begin in *3rd postpartum month*.**
Partial or no breastfeeding:	**Begin in *3rd postpartum week*.**

In non-breastfeeding women, gonadotropin levels remain low during the early puerperium and return to normal concentrations during the third to fifth week when prolactin levels have returned to normal. In an assessment of this important physiologic event (in terms of the need for contraception), the mean delay before first ovulation was found to be approximately 45 days, while no woman ovulated before 25 days after delivery.[84] Of the 22 women, however, 11 ovulated before the sixth postpartum week, underscoring the need to move the traditional postpartum medical visit to the third week after delivery. In women who do receive dopamine agonist treatment at or immediately after delivery, return of ovulation is slightly accelerated, and contraception is required a week earlier, in the second week postpartum.[88, 89]

Prolactin concentrations are increased in response to the repeated suckling stimulus of breastfeeding. Given sufficient intensity and frequency, prolactin levels remain elevated. Under these conditions, follicle-stimulating hormone (FSH) concentrations are in the low normal range (having risen from extremely low concentrations at delivery to follicular range in the 3 weeks postpartum) and luteinizing hormone (LH) values are also in the low normal range. These low levels of gonadotropins do not allow the ovary, during lactational hyperprolactinemia, to display follicular development and secrete estrogen. Therefore, vaginal dryness and dyspareunia are commonly reported by breastfeeding women. ***The use of vaginal estrogen preparations is discouraged because absorption of the estrogen can lead to inhibition of milk production. Vaginal lubricants should be used until ovarian function and estrogen production return.***

The mechanism of the contraceptive effect is of interest because a similar interference with normal pituitary-gonadal function is seen with elevated prolactin levels in nonpregnant women, the syndrome of galactorrhea and amenorrhea. Earlier experimental evidence suggested that the ovaries might be refractory to gonadotropin stimulation during lactation, and, in addition, the anterior pituitary might be less responsive to GnRH stimulation. Other studies, done later in the course of lactation, indicated, however, that the ovaries as well as the pituitary were responsive to adequate tropic hormone stimulation.[90]

These observations suggest that high concentrations of prolactin can work at both central and ovarian sites to produce lactational amenorrhea and anovulation. Prolactin appears to affect granulosa cell function in vitro by inhibiting the synthesis of progesterone. It also may change the testosterone/dihydrotestosterone ratio, thereby reducing aromatizable substrate and increasing local antiestrogen concentrations. Nevertheless, a direct effect of prolactin on ovarian follicular development does not appear to be a major factor. The central action predominates.

Elevated levels of prolactin inhibit the pulsatile secretion of GnRH.[91, 92] Prolactin excess has short-loop positive feedback effects on dopamine. Increased dopamine reduces GnRH by suppressing arcuate nucleus function, perhaps in a mechanism mediated by endogenous opioid activity.[93, 94] However, blockade of dopamine receptors with a dopamine antagonist or the administration of an opioid antagonist in breastfeeding women does not always affect gonadotropin secretion.[95] The exact mechanism for the suppression of GnRH secretion remains to be unraveled. The principle of GnRH suppression by prolactin is reinforced by the demonstration that treatment of amenorrheic, lactating women with pulsatile GnRH fully restores pituitary secretion and normal ovarian cyclic activity.[96]

At weaning, as prolactin blood concentrations fall to normal, gonadotropin levels increase, and estradiol secretion rises. This prompt resumption of ovarian function is followed by the occurrence of ovulation within 14–30 days of weaning.

Inappropriate Lactation—Galactorrheic Syndromes

Galactorrhea refers to the mammary secretion of a milky fluid, which is nonphysiologic in that it is inappropriate (not immediately related to pregnancy or the needs of a child), persistent, and sometimes excessive. Although usually white or clear, the color may be yellow or even green. In the latter circumstance, local breast disease should be considered. To elicit breast secretion, pressure should be applied to all sections of the breast beginning at the base of the breast and working up toward the nipple. ***Hormonally-induced secretions usually come from multiple duct openings in contrast to pathologic discharge that usually comes from a single duct.*** A bloody discharge is more typical of cancer. The quantity of secretion is not an important criterion. Amenorrhea does not necessarily accompany galactorrhea, even in the most serious provocative disorders. Any galactorrhea demands evaluation in a nulliparous woman and if at least 12 months have elapsed since the last pregnancy or weaning in a parous woman. Galactorrhea can involve both breasts or just 1 breast. This recommendation has evolved empirically, knowing that many women have the persistence of galactorrhea for many months after breastfeeding, and therefore the rule is a soft one. The exact numbers have never been established by appropriate studies. Thus, there is room for clinical judgment with this clinical problem.

Differential Diagnosis of Galactorrhea

The differential diagnosis of galactorrhea is a difficult and complex clinical challenge. The difficulty arises from the multiple factors involved in the control of prolactin release. In most pathophysiologic states the final common pathway leading to galactorrhea is an inappropriate augmentation of prolactin release. The following considerations are important:

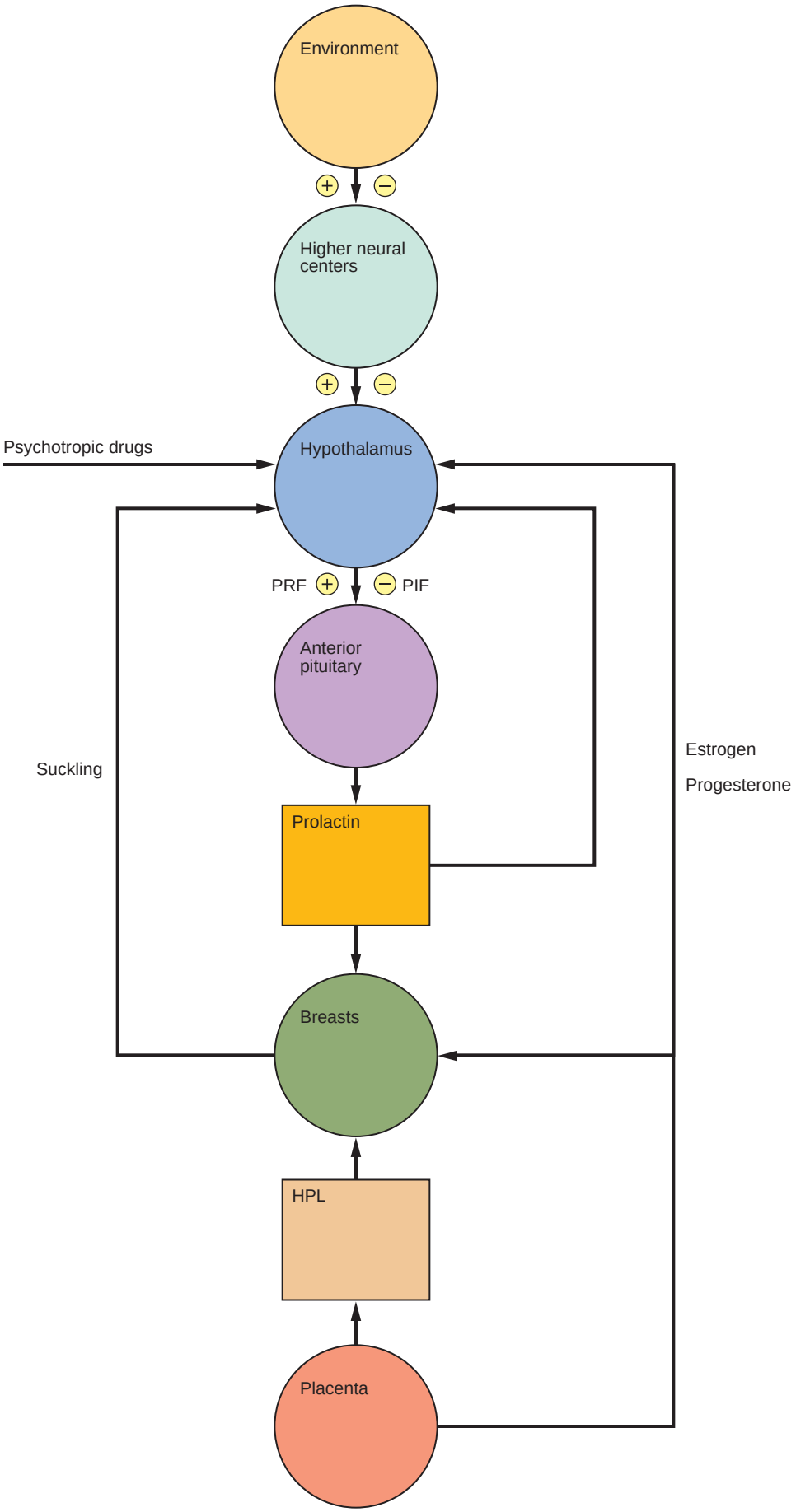
1. Increased prolactin release can be a consequence of prolactin elaboration and secretion from pituitary tumors (discussed in Chapter 11), which function independently of the otherwise appropriate restraints exerted by PIF from a normally functioning hypothalamus. This infrequent but potentially dangerous tumor, which has endocrine, neurologic, and ophthalmologic liabilities that can be disabling, makes the differential diagnosis of persistent galactorrhea a major clinical challenge. Beyond producing prolactin, the tumor may also suppress pituitary parenchyma by expansion and compression, interfering with the secretion of other tropic hormones. Other pituitary tumors may be associated with lactotroph hyperplasia and present with the characteristic syndrome of hyperprolactinemia and amenorrhea.

2. A variety of drugs can inhibit hypothalamic dopamine.[97] There are nearly 100 phenothiazine derivatives with indirect mammotropic activity. In addition, there are many phenothiazine-like compounds, reserpine derivatives, amphetamines, and an unknown variety of other drugs (opiates, diazepams, butyrophenones, verapamil, α-methyldopa, and tricyclic antidepressants) that can initiate galactorrhea via hypothalamic suppression. The final action of these compounds is either to deplete dopamine levels or to block dopamine receptors. Chemical features common to many of these drugs are an aromatic ring with a polar substituent as in estrogen and at least two additional rings or structural attributes making spatial arrangements similar to estrogen. Thus, these compounds may act in a manner similar to estrogens to decrease PIF or to act directly on the pituitary. In support of this conclusion, it has been demonstrated that estrogen and phenothiazine derivatives compete for the same receptors in the median eminence. Prolactin is uniformly elevated in patients on therapeutic amounts of these drugs, but essentially never as high as 100 ng/mL. Approximately 30–50% exhibit galactorrhea that should not persist beyond 3–6 months after drug treatment is discontinued.

3. Hypothyroidism (juvenile or adult) can be associated with galactorrhea. With diminished circulating levels of thyroid hormone, hypothalamic TRH is produced in excess and acts as a PRF to release prolactin from the pituitary. Reversal with thyroid hormone is strong circumstantial evidence to support the conclusion that TRH stimulates prolactin.

4. Excessive estrogen (e.g., oral contraceptives) can lead to milk secretion via hypothalamic suppression, causing reduction of dopamine and release of pituitary prolactin, and direct stimulation of the pituitary lactotrophs. Galactorrhea developing during oral contraceptive administration may be most noticeable in the traditional dosing regimen during the 7 days free of medication (when the steroids are cleared from the body and the prolactin interfering action of the estrogen and progestin on the breast wanes). Galactorrhea caused by excessive estrogen disappears within 3–6 months after discontinuing medication. This is now a rare occurrence with the lower-dose pills.[98] A longitudinal study of 126 women did demonstrate a 22% increase in prolactin values over mean control levels, but the response to low-dose oral contraceptives was not out of the normal range.[99]

5. Prolonged intensive suckling can also release prolactin, via hypothalamic reduction of dopamine. Similarly, thoracotomy scars, cervical spinal lesions, and herpes


Environment
Higher neural centers
Psychotropic drugs
Hypothalamus
PRF
PIF
Anterior pituitary
Suckling
Estrogen
Progesterone
Prolactin
Breasts
HPL
Placenta

zoster can induce prolactin release by activating the afferent sensory neural arc, thereby simulating suckling. Galactorrhea and elevated prolactin levels have been observed secondary to nipple piercing.[100]

6. Stresses can inhibit hypothalamic dopamine, thereby inducing prolactin secretion and galactorrhea. Trauma, surgical procedures, and anesthesia can be seen in temporal relation to the onset of galactorrhea.

7. Hypothalamic lesions, stalk lesions, or stalk compression (events that physically reduce production or delivery of dopamine to the pituitary) allow release of excess prolactin leading to galactorrhea.

8. Increased prolactin concentrations can result from nonpituitary sources such as lung, ovarian, and renal tumors and even a uterine leiomyoma. Severe renal disease requiring hemodialysis is associated with elevated prolactin levels due to the decreased glomerular filtration rate.

The Clinical Problem of Galactorrhea

A variety of eponymic designations were applied in the past to variants of the lactation syndromes. These were based on the association of galactorrhea with intrasellar tumor (Forbes, et al. 1951), antecedent pregnancy with inappropriate persistence of galactorrhea (Chiari and Frommel 1852), and in the absence of previous pregnancy (Argonz and del Castillo 1953). In all, the association of galactorrhea with eventual amenorrhea was noted. On the basis of currently available information, categorization of individual cases according to these eponymic guidelines neither is helpful nor does it permit discrimination of patients who have serious intrasellar or suprasellar pathology.

Hyperprolactinemia may be associated with a variety of menstrual cycle disturbances: oligo-ovulation, corpus luteum insufficiency, as well as amenorrhea. About one-third of women with secondary amenorrhea have elevated prolactin concentrations. Pathologic hyperprolactinemia inhibits the pulsatile secretion of GnRH, and the reduction of circulating prolactin levels restores menstrual function.

Mild hirsutism may accompany ovulatory dysfunction caused by hyperprolactinemia. Whether excess androgen is stimulated by a direct prolactin effect on adrenal cortex synthesis of dehydroepiandrosterone (DHEA) and its sulfate (DHEAS) or is primarily related to the chronic anovulation of these patients (and hence ovarian androgen secretion) is not settled. Another possibility is hyperinsulinemia. Women with elevated prolactin levels have been reported to have an association with hyperinsulinemia because of an increase in peripheral insulin resistance.[101–108] This association is independent of obesity; however, there is considerable variation and the mechanism is uncertain. We recommend that in patients with hyperprolactinemia who have a family history of early coronary heart disease or who have an abnormal lipid profile, consideration should be given to the evaluation and management of hyperinsulinemia as described in Chapter 12.

Not all patients with hyperprolactinemia display galactorrhea. The reported incidence is about 33% (Chapter 11). The disparity may not be due entirely to the variable zeal with which the presence of nipple milk secretion is sought during physical examination. The absence of galactorrhea may be due to the usually accompanying hypoestrogenic state. A more attractive explanation focuses on the concept of heterogeneity of tropic hormones (Chapter 2). The immunoassay for prolactin may not discriminate among heterogeneous molecules of prolactin. A high circulating level of prolactin may not represent material

capable of interacting with breast prolactin receptors. On the other hand, galactorrhea can be seen in women with normal prolactin serum concentrations. Episodic fluctuations and sleep increments may account for this clinical discordance, or, in this case, bioactive prolactin may be present that is immunoreactively not detectable. Remember that at any one point in time, the bioactivity (galactorrhea) and the immunoreactivity (immunoassay result) of prolactin represent the cumulative effect of the family of structural and molecular prolactin variants present in the circulation.

In the pathophysiology of male hypogonadism, hyperprolactinemia is much less common, and the incidence of actual galactorrhea quite rare. Hyperprolactinemia in men usually presents with decreased libido and potency.

If galactorrhea has been present for 6 months to 1 year, or hyperprolactinemia is noted in the process of working up menstrual disturbances, infertility, or hirsutism, the probability of a pituitary tumor must be recognized. The evaluation and management of hyperprolactinemia are presented in detail in Chapter 11.

Galactorrhea as an isolated symptom of hypothalamic dysfunction existing in an otherwise healthy woman does not require treatment. Periodic prolactin levels, if within normal range, confirm the stability of the underlying process. However, some patients find the presence or amount of galactorrhea sexually, cosmetically, and emotionally burdensome. Treatment with combined oral contraceptives, androgens, danazol, and progestins has met with minimal success. Dopamine agonist treatment, as described in Chapter 11, therefore, is the therapy of choice. Even with normal prolactin concentrations and normal imaging, treatment with a dopamine agonist can eliminate galactorrhea.

The Management of Mastalgia

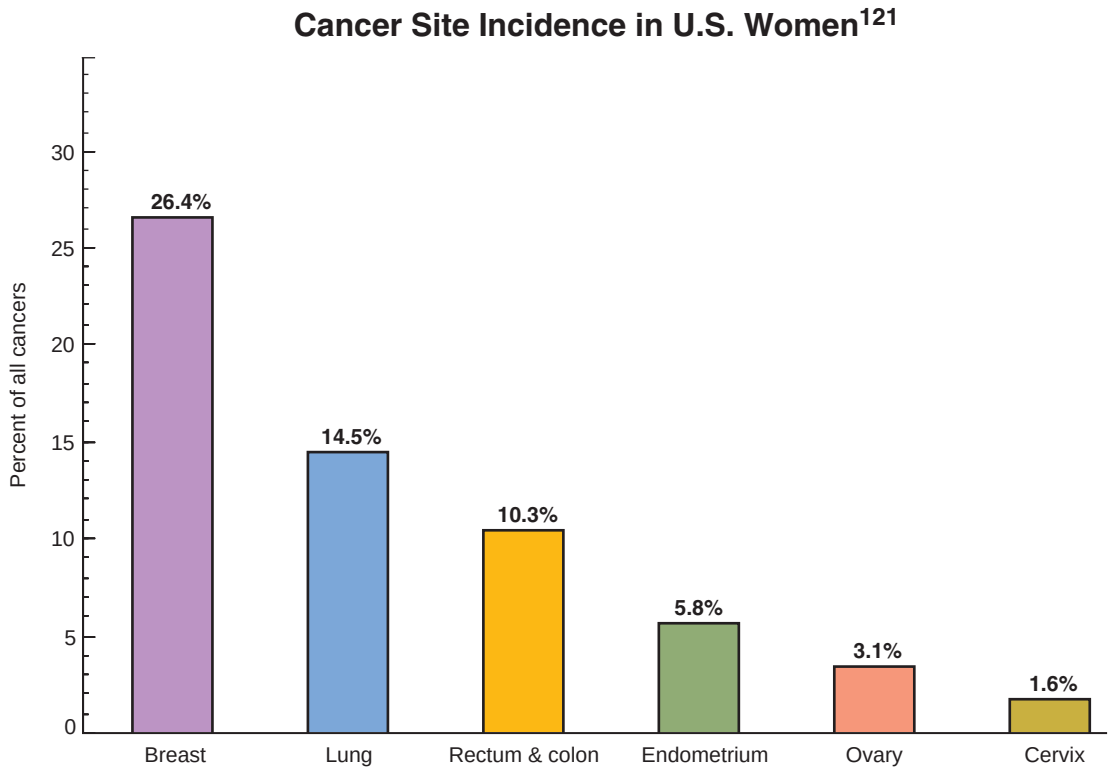
The cyclic premenstrual occurrence of breast discomfort is a common problem and is occasionally associated with dysplastic, benign histologic changes in the breast. Neither a specific etiology (although the response is probably secondary to the hormonal stimulation of the luteal phase) nor an adverse consequence (such as an increased risk of breast cancer) has been established.[109] Approximately 70% of women report premenstrual breast discomfort in surveys, and interference with activities is recorded in 10–30%.[109]

Medical treatment of mastalgia has historically included a bewildering array of options. Several are of questionable value. Diuretics have little impact, and thyroid hormone treatment is indicated only when hypothyroidism is documented. Steroid hormone treatment has been tried in many combinations, mostly unsupported by controlled studies. An old favorite, with many years of clinical experience testifying to its effectiveness, is testosterone. One must be careful, however, to avoid virilizing doses. In recent years, these methods have been supplanted by several new approaches.

Danazol in a dose of 100–200 mg/day is effective in relieving discomfort as well as decreasing nodularity of the breast.[110, 111] A daily dose is recommended for a period of 6 months. This treatment may achieve long-term resolution of histologic changes in addition to the clinical improvement. Doses below 400 mg daily do not assure inhibition of ovulation, and a method of effective contraception is necessary because of possible teratogenic effects of the drug. Significant improvement has been noted with vitamin E, 600 units/day of the synthetic tocopherol acetate. No side effects have been noted, and the mechanism of action is unknown. Bromocriptine (2.5 mg/day, which can be administered vaginally if side effects are a problem) and antiestrogens such as tamoxifen (10 or 20 mg daily) are also

effective for treating mammary discomfort and benign disease.[111–113] In a comparison study, tamoxifen was more effective than danazol.[111]

Clinical observations suggested that abstinence from methylxanthines leads to resolution of symptoms. Methylxanthines (caffeine, theophylline, and theobromine) are present in coffee, tea, chocolate, and cola drinks. In controlled studies, however, a significant placebo response rate (30–40%) has been observed. Careful assessments of this relationship in controlled studies failed to demonstrate a link between methylxanthine use and mastalgia, mammographic changes, or atypia (premalignant tissue changes).[114, 115] In addition, studies have consistently failed to detect a convincing link between methylxanthine-containing beverages and the risk of breast cancer.[116–120]



Cancer of the Breast

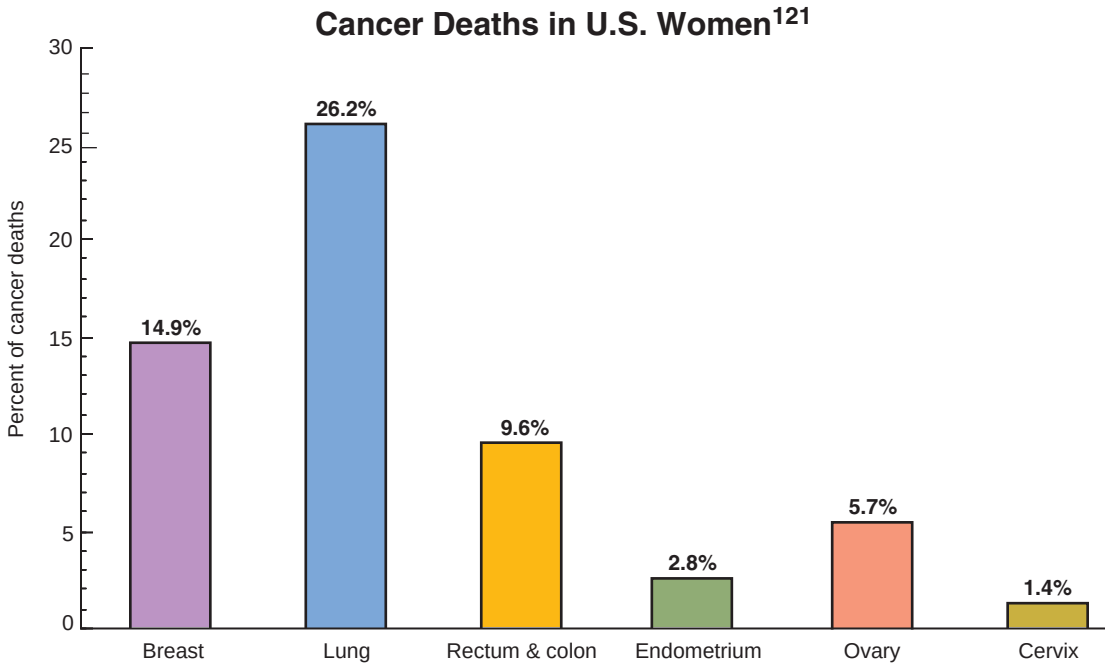
Scope of the Problem

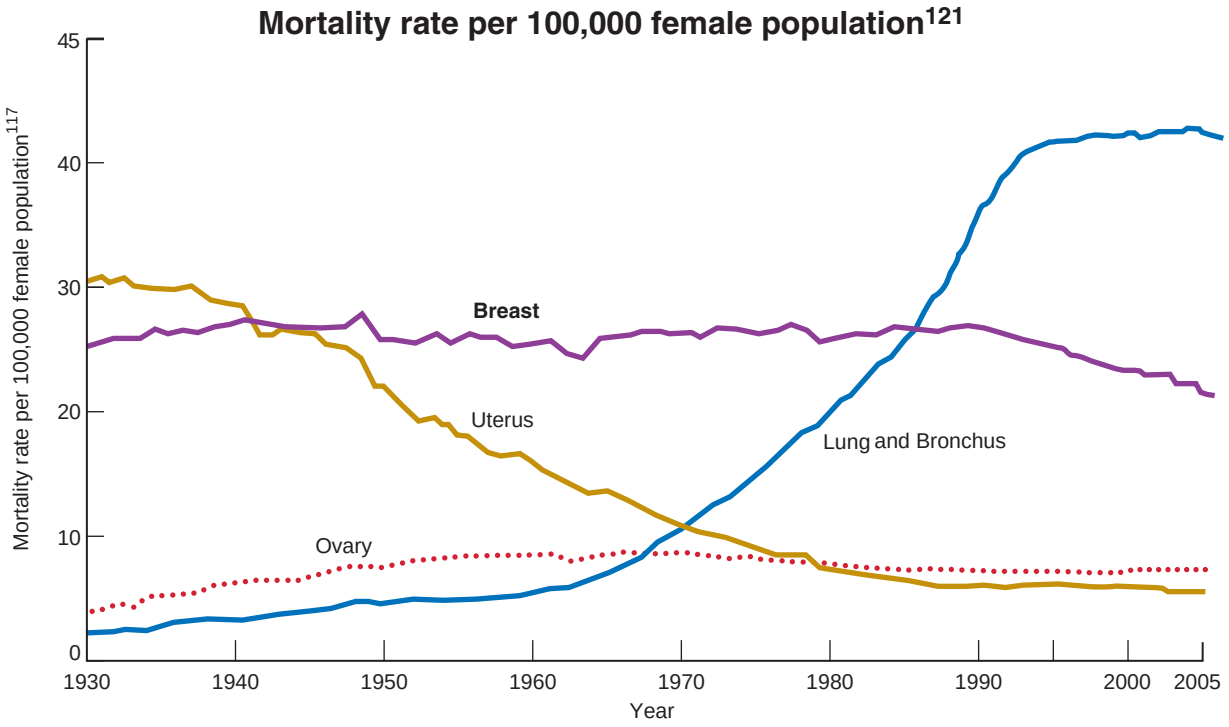
Currently, female American newborns have a lifetime probability of developing breast cancer of 12%, about one in eight, double the risk in 1940.[121,122] There are about 182,000 new cases of invasive breast cancer and 68,000 new cases of in situ breast cancer per year in the U.S. Since 1990, breast cancer incidence has decreased, by about 3% per year.[123]

This decrease is believed to reflect a reduction in the use of postmenopausal hormone therapy following the publicized results of the Women's Health Initiative and a decrease in the utilization of mammography; discussed in Chapter 18. About 87% of all breast cancers in the U.S. occur in women over age 44; only 1.9% of all cases occur under age 35, 12.5% under age 45, and 97% of breast cancer deaths in the U.S. occur in women over age 40.[122]

Mortality rates remained disappointingly constant until a decline began in the 1990s. The 5-year survival rate for localized breast cancer (about 61% of breast cancers) has risen from 72% in the 1940s to 98%.[121] This is attributed to better therapy and earlier diagnosis because of the greater utilization of screening mammography. With regional spread, the 5-year survival rate for breast cancer is 84%; with distant metastases, the rate is 27%. The breast is the leading site of cancer in U.S. women and is now, unfortunately (because smoking is obviously the reason), exceeded by lung and bronchus cancer as the leading cause of death from cancer in women.[121]

The Chances of Developing Breast Cancer in the U.S. according to Age[124]	
Birth to age 39	1 in 228
Age 40 to 59	1 in 24
Age 60 to 79	1 in 14
Birth to death	1 in 8





Over the years, breast cancer continued to have a deadly impact despite advances in surgical and diagnostic techniques. Classically, the single most useful prognostic information in women with operable breast cancer has been the histologic status of the axillary lymph nodes.[127, 128] The survival rate is higher with axillary lymph nodes negative for disease compared with positive nodes. Because of this recognition of the importance of the axillary nodes, the traditional surgical approach to breast cancer was based on the concept that breast cancer is a disease of stepwise progression. ***There is an important change in concept. Breast cancer is now viewed as a systemic disease, with spread to local and distant sites at the same time. Breast cancer is best viewed as occultly metastatic at the time of presentation.*** Therefore, dissemination of tumor cells has occurred by the time of surgery in many patients. However, this is not the story for all patients. Surely, some (if not many) cancers prior to invasion (and perhaps even some small invasive cancers) are not systemic at the time of diagnosis. For this reason, surgery is curative for many early cases of breast cancer.

Because we have been dealing with a disease that has already reached the point of dissemination in many patients, we must move the diagnosis forward several years in order to have an impact on breast cancer mortality. Earlier diagnosis requires that we be aware of what it is that makes a high-risk patient. ***However, keep in mind that the great majority of women (85%) who develop breast cancer do not have an identifiable risk factor other than age, and, therefore, every woman must be considered at risk.***

Risk Factors

A constellation of factors influences the risk for breast cancer. These include reproductive experience, ovarian activity, benign breast disease, familial tendency, genetic differences, dietary considerations, and specific endocrine factors. Clinicians can calculate the risk for an individual patient at the National Cancer Institute Internet site: http://www.cancer.gov/bcrisktool/.

Risk Factors for Breast Cancer [121]	
Relative risk greater than 4.0:	Over age 65 Inherited mutations Two or more first-degree relatives with early disease Postmenopausal breasts that are at least 75% dense on mammography
Relative risk 2.1–4.0 :	One first-degree relative with breast cancer Atypical hyperplasia on breast biopsy High-dose radiation to the chest High postmenopausal bone density
Relative risk 1.1–2.0 :	First full-term pregnancy after age 30 Menarche before age 12 Menopause after age 55 Nulliparity Never breastfed Postmenopausal obesity Previous cancer of endometrium, ovary, or colon Alcohol consumption, 2 to 5 drinks daily

Reproductive Experience

The risk of breast cancer increases with the increase in age at which a woman bears her first full-term child. A woman pregnant before the age of 18 has about one-third the risk of one who first delivers after the age of 35. To be protective, pregnancy must occur before the age of 30. Age at first birth and multiparity in women who experience their first birth before age 25 reduce the risk of breast cancer that is positive for estrogen and progesterone receptors.[127, 128] Women over the age of 30 years at the time of their first birth have a greater risk than women who never become pregnant.[129] Indeed, there is reason to believe that the age at the time of birth of the last child is the most important influence (an increasing risk with increasing age).[130] There is, however, a significant protective effect with increasing parity, present even when adjusted for age at first birth and other risk factors.[131, 132] Delayed childbearing and fewer children in modern times are believed to have contributed significantly to the increased incidence of breast cancer observed over the last decades.

Although pregnancy at an early age produces an overall lifetime reduction in risk, there is evidence that the first few years after delivery are associated with a transient increase in risk.[133] This increase probably reflects accelerated growth of an already present malignancy by the hormones of pregnancy. A very large case-control study concluded that pregnancy transiently increases the risk (perhaps for up to 3 years) after a woman's first childbirth, and this is followed by a lifetime reduction in risk.[134] And some have found that a concurrent or recent pregnancy (3–4 years previously) adversely affects survival (even after adjustment for size of tumor and number of nodes).[135, 136] It is argued that breast cells that have already begun malignant transformation are adversely affected by the hormones of pregnancy, while normal stem cells become more differentiated and resistant, reducing the number of stem cells capable of malignant change. The number of breast stem cells available for this beneficial response diminishes with age and succeeding pregnancies.[137] Although it is likely this effect is mediated by estrogen and progesterone, experimental evidence indicates the presence of LH receptors in breast tissue, and it is possible that human chorionic gonadotropin (hCG) contributes to the protective differentiation of breast cells.[138–140] Another possibility is an antiproliferative action of alpha-fetoprotein, a peptide that is secreted in the fetal liver and stimulated by the hormones of pregnancy.[141]

Initially, conflicting results were reported in over 20 studies examining the risk of breast cancer associated with the number of abortions (both spontaneous and induced abortions) experienced by individual patients.[142, 143] Concern for an adverse effect was based on the theoretical suggestion that a full-term pregnancy protects against breast cancer by invoking complete differentiation of breast cells, but abortion increases the risk by allowing breast cell proliferation in the first trimester of pregnancy, but not allowing the full differentiation that occurs in later pregnancy. In these studies there was a major problem of recall bias; women who develop breast cancer are more likely to truthfully reveal their history of induced abortion than healthy women. In studies that avoided recall bias (e.g., by deriving data from national registries instead of personal interviews), the risk of breast cancer was identical in women with and without induced abortions.[144, 145] More careful case-control studies failed to link a risk of breast cancer with either induced or spontaneous abortions.[146, 147] Similarly, newer prospective cohort studies, including the Nurses' Health Study, also reported no association between the incidence of breast cancer and induced or spontaneous abortions.[148–150]

The fact that pregnancy early in life is associated with a reduction in the risk of breast cancer implies that etiologic factors are operating during that period of life. The protection afforded only by the first pregnancy suggests that the first full-term pregnancy has a trigger effect that either produces a permanent change in the factors responsible for breast cancer or changes the breast tissue and makes it less susceptible to malignant transformation. There is evidence for a lasting impact of a first pregnancy on a woman's hormonal milieu. A small but significant elevation of estriol, a decrease in dehydroepiandrosterone and dehydroepiandrosterone sulfate, and lower prolactin levels all persist for many years after delivery.[151, 152] These changes take on significance when viewed in terms of the endocrine factors considered below.

Lactation may offer a weak to moderate protective effect (20% reduction) on the risk of breast cancer, both estrogen receptor-positive and receptor-negative tumors.[127, 128, 153–160] The same beneficial effect has been reported in *BRCA* mutation carriers in one study, but not in another.[161, 162] The Nurses' Health Study could not detect a protective effect of lactation, and a Norwegian prospective study, including a high percentage of women with long durations of breastfeeding, found no benefit on either premenopausal or postmenopausal breast cancer incidence.[163, 164] The impact of lactation, if significant, must be small. However, an analysis of the worldwide available data concluded that breastfeeding would reduce the risk of breast cancer by 4.3% per year of breastfeeding, and potentially could reduce the cumulative incidence by age 70 by more than 50%.[165] A meta-analysis indicated that breastfeeding reduced the risk of breast cancer by about 10–20%, and the impact was limited to premenopausal women.[166] There is a unique and helpful study of the Chinese Tanka, who are boat people living on the coast of southern China.[167] The women of the Chinese Tanka wear clothing with an opening only on the right side, and they breastfeed only with the right breast. All breast cancers were in postmenopausal women, and the cancers were equally distributed between the two sides, suggesting a protective effect only for premenopausal breast cancer.

In both cohort and case-control studies, there is good evidence that cosmetic breast augmentation does not increase the risk of breast cancer.[168–170] Specifically, studies have failed to indicate an increased risk of breast cancer in women who have had cosmetic breast implants.[171–174]

Ovarian Activity

Women who have a premenopausal oophorectomy have a lower risk of breast cancer, and the lowered risk is greater the younger a woman is when ovariectomized. There is a 70%

risk reduction in women who have oophorectomy before age 35. There is a small decrease in risk with late menarche and a moderate increase in risk with late natural menopause, indicating that ovarian activity plays a continuing role throughout reproductive life.[175]

Observational studies indicated that anovulatory and infertile women (exposed to less progesterone) have a small increased risk of breast cancer later in life.[176–179] However, the statistical power of these observational studies was limited by small numbers (all fewer than 15 cases). Larger numbers are available in the Nurses' Health Study, where the opposite result was apparent, a reduction in the incidence of breast cancer in women with infertility attributed to ovulatory disorders.[180]

Benign Breast Disease

Women with prior benign breast disease form only a small proportion of breast cancer patients, approximately 5%. With obstruction of ducts (probably by stromal fibrosis), ductule-alveolar secretion persists, the secretory material is retained, and cysts form from the dilation of terminal ducts (duct ectasia) and alveoli. There is good reason to eliminate the term "fibrocystic disease of the breast." In a review of over 10,000 breast biopsies in Nashville, Tennessee, 70% of the women were found to not have a lesion associated with an increased risk for cancer.[181] The most important variable on biopsies is the degree and character of the epithelial proliferation. Women with atypical hyperplasia had a relative risk of 5.3, while women with atypia and a family history of breast cancer had a relative risk of 11. In the Nurses' Health Study, biopsies with proliferative disease had a relative risk of breast cancer of 1.6, and with atypical hyperplasia, the relative risk was 3.7.[182] Only 4–10% of benign biopsies have atypical hyperplasia. The point is that we needlessly frighten patients with the use of the term fibrocystic disease. For most women, this is not a disease, but a physiologic change brought about by cyclic hormonal activity. ***Let's call this problem FIBROCYSTIC CHANGE OR CONDITION.***

The College of American Pathologists supports this position and has offered this classification.[183]

Classification of Breast Biopsy Tissue According to Risk for Breast Cancer

No increased risk:
- **Adenosis**
- **Duct ectasia**
- **Fibroadenoma without complex features**
- **Fibrosis**
- **Mild hyperplasia (3–4 cells deep)**
- **Mastitis**
- **Periductal mastitis**
- **Squamous metaplasia**
- **Ordinary cysts (fibrocystic disease)**

Slightly increased risk (1.5–2 times):
- **Fibroadenoma with complex features**
- **Moderate or florid hyperplasia**
- **Several papillomas**
- **Sclerosing adenosis**

Moderately increased (4–5 times):
- **Atypical ductal hyperplasia**
- **Atypical lobular hyperplasia**

Markedly increased (8–10 times):
- **Ductal carcinoma in situ**
- **Lobular carcinoma in situ**

Familial Tendency

Most breast cancers are sporadic; i.e., they arise in individuals without a family history of breast cancer. However, female relatives of women with breast cancer have about twice the rate of the general population. There is an excess of bilateral disease among patients with a family history of breast cancer. Relatives of women with bilateral disease have about a 45% lifetime chance of developing breast cancer. The relative risks associated with first-degree relatives are:

Relative Risk with Affected First-degree Relatives[184]

One relative	**—**	**1.80**
Two relatives	**—**	**2.93**
Three relatives	**—**	**3.90**

It is worth emphasizing that only one of nine women who develop breast cancer has an affected first-degree relative, and most women with an affected relative will never have breast cancer.

The breast and ovarian tumor suppressor gene (*BRCA1*) associated with familial cancer is on the long arm of chromosome 17, localized to 17q12–q21.[185] Although other genetic alterations have been observed in breast tumors, multiple, different mutations in *BRCA1* are believed to be responsible for approximately 20% of familial breast cancer and 80% of families with both early-onset breast and ovarian cancer. ***Overall, no more than 5–10% of breast cancers in the general population can be attributed to inherited mutations***.[127, 186] Autosomal dominant inheritance of mutations in this gene can be either maternal or paternal; male carriers are at increased risk for colon and prostate cancers.[187] A second autosomal dominant locus of multiple mutations, *BRCA2*, on chromosome 13q12–q13, accounts for up to 35% of families with early-onset breast cancer (but a lower rate of ovarian cancer), and in males, for prostate cancer, pancreatic cancer, and male breast cancer.[188, 189] Together, *BRCA1* and *BRCA2* account for 80% of families with multiple cases of early-onset breast cancer.[190] ***About 5–10% of women who develop ovarian cancer have mutations in BRCA1.***[191, 192]

BRCA1 encodes a 1,863-amino-acid protein with a zinc finger domain that is a tumor suppressor important in DNA transcription. Mutations in many different regions of the *BRCA1* gene cause a loss or reduction in its function.[193, 194] Because not every individual with a mutation in this gene develops cancer, other factors are involved, making the accuracy of prediction more difficult and arguing against widespread screening for mutations of this gene. Providing accurate numbers is a difficult task, because breast cancer has a multifactorial etiology with both genetic and environmental factors. The *BRCA1* gene could play a role in sporadic breast and ovarian cancer, but analysis of tumors has failed to find mutations in sporadic cancers that occur later in life.[195]

High-risk families have a high probability of harboring a mutation in a dominant breast cancer susceptibility gene. It is estimated that approximately 0.04% to 0.2% of women in the U.S. carry the *BRCA1* susceptibility (and *BRCA2* is less common).[196] Among women of Ashkenazi Jewish descent, the prevalence of *BRCA1* and *BRCA2* mutations is about 2%.[197] The percentage of breast cancer cases in the general population associated with a family history accounts for only a minor part of the overall prevalence. The best estimates initially ranged from 6% to 19% at most.[198] Later more representative studies revealed a lower prevalence, as low as 3% in the general population.[199, 200] In addition, there appears to be great variability in different parts of the world, and the prevalence in minority populations has not been adequately measured.

The presence of ovarian cancer within a family and three or more cases of breast cancer within a family are strong predictors of *BRCA* mutations. Genetic screening should be reserved for patients from high-risk families.

Family History Characteristics Associated with the Presence of BRCA Mutations

Early age of onset of breast cancer within a family.
Relatives with ovarian, primary peritoneal, or fallopian tube cancer.
Male relatives with breast cancer.
Three or more close relatives with breast cancer.
Close relatives with bilateral breast cancer.
Ashkenazi (Eastern European Jewish), French Canadian, or Icelandic ancestry.

Moderate-risk families are characterized by a less striking family history, the absence of ovarian cancer, and an age of onset at the time of diagnosis that is older. High-risk families have the presence of multiple cases of breast cancer in close relatives (usually at least three cases) that follows an autosomal-dominant pattern of inheritance; breast cancer is usually diagnosed before age 45; there may be cases of ovarian cancer in the family as well. Many of the cases, but not all, can be attributed to the susceptibility genes, *BRCA1* and *BRCA2*.

High-risk families have the following cumulative breast cancer risk by the age of 80 as determined by the analysis of family histories[198]:

Affected Relative	*Age of Affected Relative*	*Cumulative Breast Cancer Risk by Age 80*
One first-degree relative	<50 years old	13–21%
	50 or more years old	9–11%
One second-degree relative	<50 years old	10–14%
	50 or more years old	8–9%
Two first-degree relatives	Both <50 years old	35–48%
	Both 50 years or older	11–24%
Two second-degree relativtes but both paternal or maternal	Both <50 years old	21–26%
	Both 50 years or older	9–16%

Each child of a *BRCA* mutation carrier has a 50% chance of inheriting the mutation. In the United States, women who are carrying the *BRCA1* mutation have a 46% cumulative risk of developing breast cancer by age 70, and a 39% risk for ovarian cancer.[201] There is also a small increase in risk for other cancers, specifically of the pancreas, colon, uterus, and cervix.[202] The male relatives who are carrying this mutation have an increased risk of prostate cancer and colon cancer in addition to a cumulative risk of breast cancer of 1.2%.[203] The cancer risk for women with *BRCA2* mutations is 43% for breast cancer and 22% for ovarian cancer by age 70.[201] Male *BRCA2* mutation carriers have a higher cumulative risk of breast cancer, 6.8%, compared with male *BRCA1* carriers.[203] In addition, *BRCA2* mutation carriers have increased risks of cancers originating in the pancreas, prostate, gallbladder and bile duct, stomach, and skin.[204] Breast cancer associated with *BRCA1* mutations is histologically different (more often aneuploid and receptor-negative) compared to *BRCA2* mutations and sporadic cancers, and appears to grow faster, but paradoxically, has a better survival in response to treatment.[205] Outcome results, however, have not been consistent. A well-done Dutch study could not detect a difference in disease-free and overall survival comparing breast cancer cases from families with proven *BRCA1* mutations to patients with sporadic breast cancer.[206]

Summary of Breast and Ovarian Cancer Risk in BRCA Carriers[201]		
	Breast Cancer Risk by Age 70(%)	*Ovarian Cancer Risk by Age 70(%)*
BRCA1	46	39
BRCA2	43	22

Because not all families with breast cancer carry mutations of *BRCA1* or *BRCA2*, these families probably have breast cancer susceptibility genes yet to be identified. In addition, the current screening methods do not detect all *BRCA* mutations. For example, a mutation in a gene involved in the recognition and repair of damaged DNA, *CHEK2*, is prevalent in families with hereditary breast and colorectal cancer.[207] Other genes that infrequently cause inherited breast cancer include the *ATM* gene, the *p53* tumor suppressor gene, and the *PTEN* gene.[127] **When *three or more closely related individuals within a family have been diagnosed with breast cancer, the likelihood that an inherited dominant genetic mutation is present is very high.*** The affected women need not be first-degree relatives, but they must be related either all on the mother's side or the father's side. Identifying the families that carry the *BRCA2* gene uses the same historical criteria as that for the *BRCA1* gene. ***The family presence of just one case of ovarian cancer further increases the likelihood of the BRCA1 mutation.*** In contrast to *BRCA1* families, *BRCA2* families have only a moderately increased incidence of ovarian cancer.

Screening and counseling for families who have the appropriate history but fail to demonstrate BRCA1 or BRCA2 mutations should be exactly the same as when the mutations are found.[208]

Once it has been determined that a family is at high risk for a breast cancer gene mutation, it is recommended that this family be referred to an appropriate laboratory and service that can be identified through the medical genetics department at a regional referral institution. Although blood samples can be mailed by overnight mail, involvement with an appropriate center is highly urged because of the importance of accurate informed consent, counseling, and follow-up care. The way in which information is communicated to patients has a profound impact on decision-making and compliance with surveillance.

High-risk women who have undergone prophylactic mastectomy experience a major reduction (more than 90%) in the number of breast cancers, although total prevention is not achieved.[209–211] Because the mutation is present in every cell, and prophylactic mastectomy does not remove all tissue, there is no guarantee that breast cancer will be totally prevented. The same situation applies with prophylactic oophorectomy in that a carcinoma can arise from peritoneal cells. However, prophylactic salpingo-oophorectomy reduces the risk of ovarian cancer by about 90% and the risk of breast cancer by about 50%.[212, 213]

A growing story indicates that serous ovarian cancer originates in the fimbriae of the fallopian tubes.[214, 215] Evidence consistently indicates that tubal sterilization is associated with a major reduction in the risk of ovarian cancer.[216–220] A case-control study of *BRCA1* and *BRCA2* carriers indicated that tubal ligation reduced the risk of ovarian cancer by 60% in *BRCA1* carriers, but no protective effect was observed among *BRCA2* carriers.[221] A prospective cohort study also detected differences between *BRCA1* and *BRCA2* carriers after prophylactic salpingo-oophorectomy: an 85% reduction in ovarian cancer in *BRCA1* carriers but no significant effect in *BRCA2* carriers, and a 72% reduction in breast cancer in *BRCA2* carriers with a reduction that was not statistically significant in *BRCA1* carriers.[222] In addition, early carcinomas are found in the fallopian tube fimbriae of *BRCA1* and *BRCA2* mutation carriers.[223, 224] ***Prophylactic surgery should include bilateral salpingectomy.***

Current recommendations from experts in this field are as follows[186, 198, 225–227]: For an individual identified to be at high risk, clinical breast examination is recommended every

6 months and annual mammography beginning at age 25. An annual evaluation by magnetic resonance imaging is also recommended because there is some evidence of a higher false-negative rate with mammography in these patients, and breast cancers detected in* BRCA *mutation carriers who undergo annual MRI surveillance are of lower stage disease.[228] Clinical evaluation every 6 months is appropriate because the BRCA1-related tumors have been demonstrated to be faster growing tumors. Support should be provided for those women who choose prophylactic mastectomy. Pelvic examination, serum CA-125 levels, and transvaginal ultrasonography with color Doppler are recommended annually for women under age 40, although it has not been demonstrated that this screening will detect tumors early enough to influence prognosis. Prophylactic salpingo-oophorectomy and hysterectomy are recommended at the completion of childbearing, preferably before age 35 and certainly by age 40. In our view, estrogen-only therapy is appropriate and acceptable following surgery, as discussed below.

The epidemiologic evidence indicates that oral contraceptive use can lower the risk of ovarian cancer in *BRCA* mutation carriers. A case-control study indicated that the use of oral contraceptives in women with *BRCA1* or *BRCA2* mutations was associated with a 50% reduction in the risk of ovarian cancer (increasing with duration of use, from 20% for less than 3 years of use, up to 60% with 6 or more years of use).[229] In a large case-control study, the use of oral contraceptives reduced the risk of ovarian cancer by 44% in carriers of *BRCA1* mutations and by 61% in carriers of *BRCA2* mutations.[230] Another case-control study concluded that the use of oral contraceptives reduced the risk of ovarian cancer by 5% with each year of use in both *BRCA1* and *BRCA2* mutation carriers.[231] There is only one case-control study that found no indication of protection.[232]

In contrast to the effect on ovarian cancer risk, the impact of oral contraceptives on the risk of breast cancer is not clear at all. A cohort study from Minnesota concluded that women with a first-degree relative with breast cancer had an increased risk of breast cancer with oral contraception; however, this association was present only with oral contraceptives used prior to 1976 (high-dose formulations), and the confidence intervals were wide because of small numbers (13 ever users).[233] In a study of women with *BRCA1* and *BRCA2* mutations, an elevated risk of breast cancer associated with oral contraception was based on only a few cases and did not achieve statistical significance.[234] A larger case-control study concluded that *BRCA1* (but not *BRCA2*) mutation carriers had small increases in the risk of breast cancer in users for at least 5 years (OR=1.33, CI=1.11–1.60), in users before age 30 (OR=1.29, CI=1.09–1.52), and in those who developed breast cancer before age 40 (OR=1.38, CI=1.11–1.72).[235] In contrast, another case-control study concluded that oral contraceptive use for at least 5 years doubled the risk of breast cancer before age 50 in *BRCA2* carriers, but not in *BRCA1* carriers.[236] A retrospective analysis of an international cohort of *BRCA* carriers indicated that an increased risk of breast cancer with both *BRCA1* and *BRCA2* carriers was present only with 4 or more years of use before a first full-term pregnancy.[237] A study that focused on low-dose oral contraceptives could detect no association with breast cancer risk in *BRCA* mutation carriers.[162] Another case-control study found no increase in the risk of breast cancer diagnosed before age 40 in either *BRCA1* or *BRCA2* carriers.[238] And finally, a case-control study could detect no significant increase in the risk of contralateral breast cancer among BRCA1 and BRCA2 carriers or in noncarriers with the use of oral contraceptives or postmenopausal hormones.[239]

The data with oral contraceptives in BRCA mutation carriers are all observational and not robust. Until better information is forthcoming, it seems reasonable to inform carriers of BRCA mutations that the use of oral contraceptives is likely to reduce the risk of ovarian cancer, but the effect on breast cancer risk is uncertain.

The effect of chemoprevention by tamoxifen, raloxifene, or aromatase inhibitors has not been tested in *BRCA* mutation carriers by randomized trials. However, in subgroup analyses of the American trial assessing the effect of tamoxifen for prevention, tamoxifen

reduced the risk of breast cancer by 62% in *BRCA2* carriers, but had no impact in *BRCA1* carriers.[240, 241] This is consistent with the fact that women with *BRCA2* mutations have predominately estrogen receptor-positive tumors and women with *BRCA1* mutations have mostly estrogen receptor-negative tumors. Although no data are available, it is likely that raloxifene and aromatase inhibitors would yield results similar to those with tamoxifen. Given the side effects associated with these drugs, the decision to use one of these agents for chemoprevention is a difficult one for both clinician and patient. Prophylactic bilateral salpingo-oophorectomy remains as the superior choice for risk protection, a procedure that can in most cases, even with thorough inspection of peritoneal surfaces and peritoneal washings, be easily performed by laparoscopy. Serial sectioning of the ovaries and tubes is mandatory to detect microscopic cancers. Although concurrent hysterectomy is an individual choice, it is recommended to gain the theoretical advantage of removing the cornual portions of the fallopian tubes.

In a cohort of women with *BRCA1/2* who had oophorectomy and a 60% reduction in the risk of developing breast cancer, hormone therapy of any type did not alter the reduction in breast cancer experienced by the women undergoing oophorectomy.[242] The average length of follow-up was 2.6 years (more than 5 years in 16%) in the surgically treated group and 4.1 years (more than 5 years in 33%) in the non-oophorectomized group. There was no hint of a difference in breast cancer reduction comparing hormone users and nonusers. The findings were similar in 34 women who used a combination of estrogen and progestin, but the power of this finding was limited by the small number.

A case-control study of 472 postmenopausal women with a *BRCA1* mutation found that women who used hormone therapy after prophylactic oophorectomy, either estrogen only or combined estrogen-progestin, not only did not have an increased risk of breast cancer, but hormone use was actually associated with a decreased risk.[243] The findings were the same regardless of duration of use or current or past use. The conclusion is encouraging, but limited by the fact that 68% of the tumors in the study were estrogen receptor-negative, making the estrogen receptor-positive tumors (that are more likely to be influenced by hormone use) relatively small in number.

Women who are *BRCA* carriers face difficult decisions regarding hormonal treatment for menopausal symptoms. The experience thus far indicates that hormone therapy can be used safely for several years. Continuing follow-up of these patients may extend this period of safety even longer.

Dietary Factors

The geographic variation in incidence rates of breast cancer is considerable (the United States has the highest rates and Japan the lowest), and it has been correlated with the amount of animal fat in the diet.[244] Lean women, however, have an increased incidence of breast cancer, although this increase is limited to small, localized, and well-differentiated tumors.[245] Furthermore, studies have failed to find evidence for a positive relationship between breast cancer and dietary total or saturated fat or cholesterol intake.[246–249] One study found that dietary fat is a stronger risk factor for postmenopausal breast cancer than for premenopausal breast cancer, but another study had the opposite conclusion.[250, 251] Although a cohort study concluded that dietary fat is a determinant of postmenopausal breast cancer, the association did not achieve statistical significance.[252] And another very large cohort study in Europe demonstrated only a very weak link between saturated fat intake and the risk of breast cancer, only in nonusers of hormone therapy.[253] Thus, the epidemiologic literature provides little support for a major contribution of dietary fat to the risk of breast cancer. Nevertheless, there

is a correlation between intraabdominal fat (android obesity) and the risk of breast cancer, a consequence of excessive caloric consumption, however, not a specific dietary component.[254] Presumably, the connection between android obesity and breast cancer is through the metabolic perturbations, especially hyperinsulinemia, associated with excessive body weight.

There is no argument that the incidence of breast cancer is increased in countries associated with affluent, unfavorable diets (high fat content) and a lack of physical exercise. Indeed, increased physical activity in postmenopausal women reduces the risk of breast cancer.[255] The common denominator may be the peripheral insulin resistance and hyperinsulinemia that become prevalent with aging and weight gain in affluent, modern societies. This specific metabolic change is becoming a common theme in various clinical conditions, particularly noninsulin-dependent diabetes mellitus, anovulation and polycystic ovaries, hypertension, and dyslipidemia. Hyperinsulinemia is found more often in women with breast cancer.[256] There are, indeed, many reasons to avoid excess body weight. The risk of breast cancer is reduced in women who exercise regularly.[257]

The increased circulating levels of insulin that are a consequence of obesity-induced insulin resistance can directly stimulate breast tissue growth and can also increase levels of biologically active estradiol by lowering sex hormone-binding globulin synthesis in the liver. In a cohort of women enrolled in the Women's Health Initiative, an increase in risk of breast cancer in obese women who were *not* using hormone therapy correlated with hyperinsulinemia and elevated estradiol levels, but not with IGF-I levels.[258] An adjustment for estrogen levels indicated that the hyperinsulinemia acted independently and was the more robust factor. The inability to demonstrate this association with hyperinsulinemia in hormone users may be a consequence of the lower insulin levels caused by estrogen treatment.

In the parts of the world where soy intake is high, there is a lower incidence of breast, endometrial, and prostate cancers. For example, a case-control study concluded that there was a 54% reduced risk of endometrial cancer, and other case-control studies found a reduction in the risk of breast cancer, in women with a high consumption of soy and other legumes.[259–261] It is by no means certain, however, that there is a direct effect of soy intake.[262] Soy intake may be a marker for other factors in lifestyle or diet that are protective. Short-term studies on breast secretions have actually indicated that soy intake produces an estrogenic response.[263–265] The effect of soy intake on the risk of breast cancer is discussed in greater detail in Chapter 18.

It is well recognized that the incidence of breast cancer is higher in the U.S. than in China or Japan. It has been further observed that after migration to the U.S., Asian women gradually increase (6-fold) their risk of breast cancer over several generations, eventually reaching the level of white women.[266] Evidence indicates that this reflects a change in diet and lifestyle, with an increase in risk associated with a gain in height and weight.[267, 268] Recent weight gain is especially associated with increased risk. A reduced risk, however, is observed in heavy, younger women.

The effect of body weight on the risk of breast cancer differs in premenopausal and postmenopausal women. In premenopausal women who are overweight, the risk of breast cancer is lower compared with normal-weight individuals, and in postmenopausal women, especially in nonusers of hormone therapy, excess weight is associated with either an unchanged or slightly increased risk.[268–272] This is attributed to a more marked increase in total and free estrogen levels in overweight postmenopausal women, in contrast to lower levels with increasing weight in premenopausal women. Postmenopausal obese women have later menopause, higher estrone production rates and higher free estradiol levels (because of insulin-induced lower sex hormone-binding globulin levels), and a slightly greater risk for breast cancer.[273] A large Swedish case-control study and an American prospective

cohort study suggested that the principal factor is weight gain during adulthood, and that the impact on breast cancer emerges 10 years after menopause.[272, 274] As noted, this weight gain may be the important determinant in the increasing risk experienced by migrants from low-risk parts of the world who move to high-risk areas.

Evidence indicates that the intake of vitamins A, C, and E has no effect on the risk of breast cancer.[275]

Alcohol in the Diet

There is a modest increase in the risk for estrogen receptor-positive breast cancer with the consumption of one or more alcoholic drinks of all forms per day.[276, 277] Almost all of many studies conclude that 2 drinks daily increase the risk by about 20%.[278, 279] It is speculated that breast cancer and alcohol are linked through estrogen, either a direct or an indirect effect (e.g., on hepatic enzymes) on estrogen metabolism. An effect of alcohol ingestion by premenopausal women was not demonstrated on circulating levels of estrone, estradiol, dehydroepiandrosterone sulfate (DHEAS), or sex hormone-binding globulin in a cross-sectional study that depended upon a questionnaire to assess alcohol intake.[280] However, when alcohol is administered under experimental conditions, circulating estrogen concentrations are increased.[281–283] And in a prospective cohort study of premenopausal women in Italy, higher estradiol levels were correlated with an increased alcohol intake over a 1-year period of time.[284]

Specific Endocrine Factors

Adrenal Steroids

Subnormal levels of etiocholanolone (a urinary excretion product of androstenedione) were found from 5 months to 9 years before the diagnosis of breast cancer in women living on the island of Guernsey, off the English coast.[285] A subnormal excretion of this 17-ketosteroid was also found in sisters of patients with breast cancer. A 6-fold increase in the incidence of breast cancer was found between women excreting less than 0.4 mg of etiocholanolone and those excreting over 1 mg/24 hours. After 37 years of follow-up, low levels of androsterone and etiocholanolone were observed to correlate with an increase in breast cancer only in women under age 50; over age 50, the reverse was true.[286] Measurement of these 17-ketosteroids might be a useful screening procedure to detect a high-risk group of patients because approximately 25% of the population excretes less than 1 mg/24 hours, but these early results have never been pursued.

Endogenous Estrogens and Androgens

Epidemiologic and other information continue to suggest some estrogen-related promoter function. These include the following: (1) the condition is 100 times more common in women than in men; (2) breast cancer invariably occurs after puberty; (3) untreated gonadal dysgenesis and breast cancer are mutually exclusive; (4) a 65% excess rate of breast cancer

has been observed among women who have had an endometrial cancer; and (5) breast tumors contain estrogen receptors, which are biologically active as indicated by the presence of progesterone receptors in tumor tissue. Taken together, these data suggest an element of estrogen dependence, if not provocation, in many breast cancers.

Estriol generally has failed to produce breast cancer in rodents, and in fact, estriol protects the rat against breast tumors induced by various chemical carcinogens (but so did estradiol).[287] The hypothesis is that a higher estriol level protects against the more potent effects of estrone and estradiol. This might explain the protective effect of early pregnancies. Women having had an early pregnancy continue to excrete more estriol than nulliparous women. Premenopausal healthy Asiatic women have a lower breast cancer risk than Caucasians and also have a higher rate of urinary estriol excretion.[288] When Asiatic women migrate to the United States, however, the risk of breast cancer increases, and their urinary excretion of estriol decreases, perhaps a consequence of dietary changes as noted above. A study of Westernized Asian women also documented a reduction in estrogen 2-hydroxy metabolites, with decreasing values linked to increasing breast cancer risk associated with measures of Westernization.[289]

A major factor in the potency differences among the various estrogens (estradiol, estrone, estriol) is the length of time the estrogen-receptor complex occupies the nucleus. The higher rate of dissociation with the weak estrogen (estriol) can be compensated for by continuous application to allow prolonged nuclear binding and activity. Estriol has only 20–30% affinity for the estrogen receptor compared to estradiol; therefore, it is rapidly cleared from a cell. But if the effective concentration is kept equivalent to that of estradiol, it can produce a similar biologic response.[290]

In pregnancy, where the concentration of estriol is very great, it can be an important hormone, not just a metabolite. Thus, higher estriol levels are not necessarily protective. Indeed, antagonism of estradiol occurs only within a vary narrow range of the ratio of estradiol to estriol, a range rarely encountered either physiologically or pharmacologically.[291] Below this range, estradiol is unimpeded; above this range, estriol itself exerts estrogenic activity. Indeed, no inhibition of mammary tissue proliferation markers could be detected in women administered estriol in the presence of an estrogen-progestin oral contraceptive.[292] There have been no epidemiologic studies of breast cancer risk in women treated with estriol, and, therefore, the contention that estriol protects against breast cancer remains speculative.

There have been many studies assessing the relationship between endogenous hormone levels and the risk of breast cancer. A pooled analysis of nine prospective studies concluded that the risk of breast cancer, especially estrogen receptor-positive tumors, increases with increasing concentrations of all endogenous estrogens and androgens, including estradiol, estrone, estrone sulfate, androstenedione, dehydroepiandrosterone, dehydroepiandrosterone sulfate, and testosterone.[293] The overall increase in breast cancer risk was about 2-fold comparing the lowest endogenous levels in postmenopausal women with the highest levels. This relationship is seen with both estrogens and androgens.[294, 295] Postmenopausal women who are overweight have an increased risk of breast cancer, and analyses that adjusted for the increase in circulating estrogens associated with obesity concluded that the increasing risk with increasing body weight is the result of the increase in estrogens and a decrease in sex hormone-binding globulin.[296, 297] The increased risk of breast cancer in a cohort of obese women not using hormone therapy in the Women's Health Initiative was attributed to elevated circulating levels of insulin and biologically active estradiol, highlighting the critical role for hyperinsulinemia as discussed above.[258]

Bone mass is generally regarded as a marker of estrogen exposure, and women with the highest bone densities have a greater risk of breast cancer compared with women who have low bone densities.[298–300] Another attempt to link the risk of breast cancer to

the endogenous estrogen level focused on prenatal exposure. A reduced risk for breast cancer is observed for women born to mothers with pregnancy-induced hypertension, suggesting that this finding is due to the lower estrogen levels associated with preeclampsia.[301, 302]

The biologic plausibility and epidemiologic support for an estrogen link are impressive arguments. Whether the important factor is the total amount of estrogen, the amount of estrogen unopposed by progesterone, the amount of free (unbound) estradiol, the duration of exposure to estrogen, or some other combination is not known.

Endogenous Progesterone

Because mitotic activity in the breast reaches its peak during the progesterone-dominant luteal phase of the menstrual cycle,[303–305] it is argued that progesterone is the key to influencing the risk of breast cancer. This would be consistent with experimental demonstrations in mice that progesterone is the primary hormonal stimulus for mammary growth and differentiation.[2] However, studies do not support a major role for a progestational influence. Indeed, evidence indicates that with increasing duration of exposure, progesterone can limit breast epithelial growth as it does with endometrial epithelium.[15–17] In vitro studies of normal breast epithelial cells reveal that progestins inhibit proliferation.[306] Human breast tissue specimens removed after the patients were treated with estradiol and progesterone indicate that progesterone inhibits in vivo estradiol-induced proliferation.[15, 17] Women who ultimately develop breast cancer do not have different blood levels of progesterone.[295, 307] In addition, several clinical observations would argue against progesterone as a key factor. Although there is some disagreement, most studies indicate that the high levels of estrogen and progesterone during pregnancy have no adverse impact on the course of breast cancer diagnosed during pregnancy or when pregnancy occurs subsequent to diagnosis and treatment. Medroxyprogesterone acetate is not associated with an increased risk of breast cancer when used for contraception over long durations (Chapter 24).

Exogenous Estrogen and Progestin

Epidemiologic studies have indicated a small increase in risk of breast cancer associated with postmenopausal estrogen-progestin therapy. The most important unanswered question is whether postmenopausal hormone therapy initiates the growth of new breast cancers or whether the epidemiologic results reflect an impact on pre-existing tumors. This important question is discussed in Chapter 18.

Women with a greater mammographic breast density have a higher risk of breast cancer.[308] If more than 75% of the breast is dense, the risk is four to five times increased. Mammographic density is associated with epithelial and stromal cell proliferation.[309] Studies in twins and families have indicated that there is a strong genetic determinant of an individual's breast density, and that this genetic influence is shared with the other genetic factors that increase the risk of breast cancer.[309–311] In general, breast density declines with age and increasing body weight and numbers of pregnancies.

About 25% of women on estrogen-progestin therapy have an increase in their breast density. However, it is not certain that the short-term increase in density with hormone therapy changes an individual's risk of breast cancer. The increase in breast density associated with postmenopausal hormone therapy appears in some studies to be a transient,

reversible change, a change not consistent with a persistent effect on cellular proliferation. After discontinuing hormone therapy, some reports indicated that breast density rapidly decreases.[312–315] However, in a large randomized trial of 1,704 women age 45 to 80, although suspension of hormone therapy for 1 or 2 months produced small but significant decreases in density, mammography recall rates of 10% to 12% were not affected.[316] In one small screening population of 47 women, a 4-week period without hormones before mammography had no measurable effect on density readings.[317] Therefore, the evidence is mixed regarding a recommendation to discontinue hormone therapy for 2 to 4 weeks prior to mammography in women who have dense breasts. Another approach is to consider lower doses of hormone therapy; there is some evidence that low-dose treatment has little effect on breast density.[318]

Thyroid, Prolactin, and Various Nonestrogen Drugs

Despite isolated suggestions of increased risk, hypothyroidism, reserpine, and prolactin excess, whether spontaneous or drug-induced, are not associated with an enhanced risk of breast cancer.[319, 320]

Oral Contraception and Breast Cancer

The large number of women taking or having taken oral contraceptive steroids, combined with the belief that sex steroids provoke or promote abnormal breast growth and possibly cancer, has provided a source of major concern for years. The Royal College of General Practitioners, Oxford Family Planning Association, and Walnut Creek studies have indicated no significant differences in breast cancer rates between users and nonusers. However, patients were enrolled in these studies at a time when oral contraceptives were used primarily by married couples spacing out their children. Because this population did not reflect use by younger women for long durations to delay their first pregnancy, case-control studies in the last decade have focused on the contemporary use of oral contraceptives. This subject is reviewed in detail with complete references in Chapter 22.

The largest case-control study by far on this subject is that performed by the Centers for Disease Control and Prevention, involving 4,575 American women with breast cancer, aged 35 to 64.[321] Initiation at a younger age had no impact. The risk of breast cancer was not increased in current users or past users of oral contraception. There was no adverse effect of increasing duration of use or higher doses of estrogen, with no differences in current or recent users, and no increase in risk in women with a family history of breast cancer. This large American study had consistently negative results. The next largest study, involving women from California, Canada, and Australia, focused on breast cancer diagnosed before age 40, and could not detect an increase in current or past users of oral contraceptives.[238] A multicenter, large case-control study of women younger than 55 years with breast cancer concluded that the use of oral contraceptives or postmenopausal hormone therapy either before or after diagnosis did not increase the risk of the first breast cancer or recurrent breast cancer.[322] This negative finding was not changed by duration of use or age of use. Furthermore, no increase in breast cancer mortality can be detected in women who have used oral contraceptives.[323, 324]

A team of epidemiologists from several institutions in the U.S. performed a case-control study of the association between oral contraceptive use and lobular and ductal breast cancer occurring in young women (under age 44), concluding that the use of oral contraceptives has no meaningful effects on breast cancer risk according to histologic subtype.[325]

This is very reassuring because it is well-recognized that lobular cancer is more hormonally sensitive than ductal breast cancer.

Some studies have reported small increases in premenopausal breast cancer, but these studies were unable to avoid being confounded by a very likely possibility: early and recent use of oral contraceptives may affect the growth of a pre-existing malignancy. This is supported by the fact that those studies with positive findings find an increase limited to current and recent use, and the increase has been largely localized disease (in many studies, only localized disease). Even if there is a small increase in premenopausal breast cancer associated with oral contraceptives, this would be a very small number of cases because most cases of breast cancer occur after age 40. ***Well-done and large case-control studies of modern low-dose oral contraceptives have been consistently negative and reassuring. Older positive studies cannot escape the possibility of detection/surveillance bias because of an effect on pre-existing tumors.***

The use of oral contraception does not further increase the risk of breast cancer in women with positive family histories of breast cancer or in women with proven benign breast diseases. There is no evidence that the use of oral contraceptives prior to the diagnosis of breast cancer has an adverse impact on prognosis.[326] ***Overall, the epidemiological data derived from the studies with the largest numbers of cases indicate that oral contraceptives should continue to be offered as an appropriate choice for women with family histories of breast cancer, but the impact on BRCA carriers is unsettled.***

Higher-dose oral contraception, used for 2 or more years, protected against benign breast disease, but this protection was limited to current and recent users. It is still uncertain whether this same protection is provided by the lower-dose products. A French case-control study indicated a reduction of nonproliferative benign breast disease associated with low-dose oral contraceptives used before a first full-term pregnancy, but no effect on proliferative disease or with use after a pregnancy.[327] A Canadian cohort study that almost certainly reflected the use of modern low-dose oral contraceptives concluded that oral contraceptives do protect against proliferative benign disease, with an increasing reduction in risk with increasing duration of use.[328]

Breast Cancer in Diethylstilbestrol (DES)-Exposed Women

From 1940 to 1970, diethylstilbestrol (DES), a potent synthetic estrogen, was prescribed in high doses in the mistaken belief that it would reduce the risk of pregnancy-related complications. Exposure to DES occurred in association with 2 million live births; therefore, the risk for induction of breast cancer during a period of breast differentiation could be significant if DES were a true breast carcinogen. The first study on this subject reported on the follow-up of women who participated in a controlled trial of DES in pregnancy between 1950 and 1952 at the University of Chicago. In this study, an increase in breast cancer risk that did not reach significance was observed with DES exposure.[329] A large collaborative study, involving approximately 6,000 women, concluded that there is a small but significant increase in the risk of breast cancer many years later in life in women exposed to DES during pregnancy.[330] In a longer follow-up (more than 30 years) of this large cohort of DES-exposed women, exposure to DES was associated with a significant, but modest (about 2-fold), increase in the risk of breast cancer.[331] Importantly, the relative risk did not increase with duration of follow-up and remained stable over time. This conclusion was confirmed in a prospective study by the American Cancer Society and in a national cohort of women followed since the 1970s.[332, 333] Certainly it is wise to recommend to DES-exposed women that they adhere religiously to screening for breast cancer, including mammography as

discussed later. Thus far, an increased risk of cancer has not been detected in daughters or sons of DES-exposed women.[334]

Receptors and Clinical Prognosis

There is a strong correlation between the presence of estrogen receptors and certain clinical characteristics of breast cancer.[335] Premenopausal, younger patients are more frequently receptor negative. Patients with receptor-positive tumors survive longer and have longer disease-free intervals after mastectomy than those with receptor-negative tumors. The presence of estrogen receptors correlates with increased disease-free interval regardless of the presence of positive axillary nodes or the size and location of the tumors. Similarly, patients without axillary lymph node metastases, but with estradiol receptor-negative tumors, have the same high rate of recurrence as do patients with axillary lymph node metastases. Patients with tumors that are positive for estrogen receptors are more likely to respond to endocrine treatment. Estrogen receptor status correlates with the degree of differentiation of the primary tumor. A large proportion of highly differentiated Grade I carcinomas are receptor-positive, while the reverse is true of Grade III tumors.

Remember that it takes estrogen to make progesterone receptors. Therefore, the presence of progesterone receptors proves that the estrogen receptor in the tumor is biologically active. Thus, the presence of progesterone receptors has a correlation with disease-free survival of patients only second to the number of positive nodes.[335] Overall, about 80% of breast cancers are positive for estrogen receptors, and of these, about 70% are positive for progesterone receptors, thus about 58% of all breast cancers express both estrogen and progesterone receptors.[336] The best prognosis is seen in patients with positive progesterone receptors, even with subsequent disease if the recurrent disease is still progesterone receptor-positive. The loss of progesterone receptors is an ominous sign. Tumors that are estrogen receptor-positive, but progesterone receptor-negative, express higher levels of the epidermal growth factor receptors, HER-1 and HER-2, and are more aggressive and tamoxifen-resistant.[336] The total absence of estrogen and progesterone receptors indicates a very different disease, one that should be treated aggressively with chemotherapy.

Hormone Therapy of Breast Cancer

Tamoxifen

The purpose of adjuvant therapy of breast cancer is to provide treatment in the absence of recognized active disease in order to reduce the risk of future recurrence or to minimize systemic recurrence in the presence of metastatic disease. Tamoxifen is very similar to clomiphene (in structure and actions), both being nonsteroidal compounds structurally related to diethylstilbestrol. In vitro, the estrogen binding affinity for its receptor is 100–1,000 times greater than that of tamoxifen. Thus, tamoxifen must be present in a concentration 100–1,000 times greater than estrogen to maintain inhibition of breast cancer cells. Dose-response studies with tamoxifen have failed to demonstrate an increase in activity with doses larger than the standard, 20 mg daily. When bound to the estrogen receptor, tamoxifen prevents gene transcription by the TAF-2 pathway. In vitro studies demonstrate that these actions are not cytocidal, but rather cytostatic (and thus tamoxifen use must be long-term). The mechanism of tamoxifen action is discussed in detail in Chapter 2.

We have available a remarkable worldwide overview of 37,000 women involved in tamoxifen randomized trials.[337, 338] Adjuvant treatment with the antiestrogen tamoxifen achieved highly significant reductions in recurrence and increases in survival. The beneficial effect of tamoxifen was evident no matter what the age of the patient, in both premenopausal and postmenopausal women, in node-positive and node-negative disease, and in both estrogen receptor-positive and -negative tumors (however, the effect of tamoxifen on estrogen receptor-negative tumors is small). The impact on recurrence occurred in the first 5 years, but continued impact on survival occurred throughout 15 years.[338] Hormonal adjuvant treatment yields worldwide an extra 100,000 10-year survivors. With tamoxifen, there is an increased survival at 5 years of approximately 25%, most evident in women over age 50. Response rates in advanced breast cancer are 30–35%, most marked in patients with tumors that are positive for estrogen receptors, reaching 75% in tumors highly positive for estrogen receptors. There is a lower rate (a 47% reduction with 5 years of treatment) of a second primary breast cancer in the contralateral breast in women treated with tamoxifen.

Data from randomized clinical trials document that a treatment duration of 5 years is superior to 2 years.[337, 339] However, the results indicated that there is little reason to extend tamoxifen treatment of breast cancer patients beyond 5 years.[340, 341] Indeed, the data suggested that survival and recurrence rates worsened with longer therapy, probably due to the emergence of tamoxifen-resistant tumors. There are several possible explanations for resistance (discussed in Chapter 2), and whichever of these are operative, it is believed that a subpopulation resistant to tamoxifen is present from the beginning, and over time grows to be clinically apparent.[342] Indeed, tamoxifen may be stimulating the growth of these "resistant" tumors, operating through pathways not involving the estrogen receptor, such as growth factor mechanisms. This may be the explanation for the observation that even 5 years of tamoxifen treatment is associated with an increase in estrogen receptor-negative cancer in the contralateral breast.[343]

The efficacy of tamoxifen is significantly dependent on the formation of active metabolites, 4-hydroxytamoxifen and endoxifen, which have a greater affinity for the estrogen receptor than tamoxifen. A cytochrome enzyme, P450 2D6, is involved in this metabolism, and genetic variants in the enzyme can account for lower activity leading to reduced efficacy for tamoxifen. Genotyping of *CYP2D6* could allow better selection of patients for tamoxifen treatment.[344] However, others contend that the metabolite levels are sufficient for good efficacy even in the presence of reduced enzyme activity, and that the studies linking *CYP2D6* genotype and breast cancer recurrence have yielded heterogeneous results.[345] It is not certain that *CYP2D6* variation is the explanation for recurrence of resistant tumors.

Tamoxifen has many important side effects, attributed to both its estrogen agonist action and its antiestrogen impact in different target tissues. The major disturbing side effect is an increase in hot flushing. The serious side effects of tamoxifen include endometrial cancer (discussed later), venous thrombosis, and cataracts. In a report from the prevention trial in England and in the U.S. preventive trial, tamoxifen treatment of postmenopausal women prevented bone loss, but premenopausal women treated with tamoxifen had significant reductions in bone mineral density.[240, 346] Blurred and decreased vision has been reported associated with retinal changes.[347] In a prospective study of 63 patients in Greece, 6.3% developed retinopathy, which was reversible except for retinal opacities.[348] In the 2,673 patients in the protocols of the Eastern Cooperative Oncology Group, premenopausal women who received tamoxifen and chemotherapy had significantly more venous and arterial thrombosis than those who received chemotherapy without tamoxifen, and in postmenopausal women, tamoxifen alone was associated with more venous thrombosis.[349]

Serum protein changes reflect the estrogenic (agonistic) action of tamoxifen. This includes decreases in antithrombin III, cholesterol, and LDL-cholesterol, while HDL-cholesterol

and sex hormone-binding globulin (SHBG) levels increase (as do other binding globulins). Because of the significant impact on sex hormone-binding globulin, a marked increase in circulating estrogens has been observed in premenopausal women; however, unbound, free estrogen is actually reduced. For example, in a clinical study of premenopausal women receiving tamoxifen, 20 mg daily, the percent free estradiol *decreased* from 1.72% to 1.47% after 3 months because of the increase in SHBG.[350]

The estrogenic activity of tamoxifen, 20 mg daily, is nearly as potent as 2 mg estradiol in lowering FSH levels in postmenopausal women, 26% vs. 34% with estradiol.[351] The estrogenic actions of tamoxifen include the stimulation of progesterone receptor synthesis, an estrogen-like maintenance of bone and the cardiovascular system, and estrogenic effects on the vaginal mucosa and the endometrium. Indeed, patients with breast cancer who have been treated with tamoxifen were reported to have less coronary heart disease in some studies, but not all.[240, 337, 352, 353] Tamoxifen increases the frequency of hepatic carcinoma in rats at very large doses. This is consistent with its estrogenic, agonistic action, but this effect is unlikely to be a clinical problem, and it has not been observed at doses used clinically.[337]

Gynecologic Problems with Tamoxifen

Tamoxifen is both an estrogen antagonist and an estrogen agonist. A tissue that is highly sensitive to estrogen, the endometrium, responds to the weak estrogenic action of tamoxifen, which is present in high doses for long durations in women receiving adjuvant treatment for breast cancer.

The National Surgical Adjuvant Breast and Bowel Project compared the rates of endometrial cancers in tamoxifen and non-tamoxifen-treated patients who had breast cancer.[354] The rate of endometrial cancer in the tamoxifen-treated group equaled an increased relative risk of 7.5. Although 88% of the endometrial tumors were stage I, four patients died of advanced endometrial cancer. It is worth noting that the incidence of endometrial cancer in the tamoxifen-treated group was estimated to be 6.3 per 1,000 patients after 5 years of treatment. This incidence is very similar to what would be expected with unopposed estrogen treatment, a similarity to be expected in that the agonistic estrogenic action of tamoxifen over the long-term should be similar to the relatively low doses of estrogen used for postmenopausal hormone therapy. Similar results were reported from the Stockholm tamoxifen trial, and an increased rate of postmenopausal endometrial cancer was confirmed in the U.S. Breast Cancer Prevention Trial.[240, 355] In the world overview of randomized trials, the incidence of endometrial cancer quadrupled with 5 years of tamoxifen treatment.[337] In addition, women being treated with tamoxifen were reported to develop atypical hyperplasia of the endometrium, endometrial polyps, ovarian cysts, growth of fibroids, adenomyosis, and rapid exacerbation of endometriosis.[356–359] The proper surveillance and management of women being treated with tamoxifen are critical problems.

It is inappropriate to advocate oral progestational treatment to prevent the endometrial response to tamoxifen. The progestational impact (at the low doses currently used for endometrial protection) on the risk of breast cancer recurrence and the interaction with tamoxifen are not known. Indeed, a relatively high dose of norethindrone (2.5 mg daily for 3 months) was ***unable*** to exert a protective effect on the endometrium in healthy women participating in the U.K. tamoxifen prevention trial, and a high dose of megestrol acetate failed to reverse endometrial hyperplasia.[360, 361] Periodic endometrial aspiration biopsy, of course, would be sufficient for surveillance, but this procedure carries with it the potential

for a very significant negative effect on patient compliance (with her tamoxifen and with her clinician), and a low rate of positive results.

It is argued that endometrial assessment should be limited to tamoxifen-treated women who report vaginal bleeding.[362, 363] However, in the U.K. tamoxifen prevention trial, a greater endometrial response to tamoxifen was observed in those women who developed amenorrhea.[364] To be sure, most tamoxifen-treated women who have developed endometrial cancer have been symptomatic with vaginal bleeding, but not all. Furthermore, some of these women have had advanced, invasive disease at the time of presentation. Stage III and stage IV endometrial cancers with a poor prognosis were reported more frequently in long-term tamoxifen users.[365] In addition, tamoxifen is associated with a higher rate of mixed mesodermal tumors and sarcomas of the endometrium.[365] It makes sense to detect abnormal changes as early as possible. The progestin challenge test (discussed in Chapters 11 and 18) would be a cost-effective method to detect the presence of stimulated endometrium, and a pilot study documented its use in tamoxifen-treated women.[366] However, until data are available documenting the reliability of this approach, we favor the use of ultrasonographic measurement of endometrial thickness, with saline instillation sonohysterography when the appearance is not totally benign (also discussed in Chapter 18).[367, 368] Tamoxifen is associated with an ultrasonographic image that is unique, characterized by sonolucent changes that are subepithelial in the presence of atrophic epithelium, thus the usefulness of saline instillation to discriminate epithelial thickness from combined endometrial changes.[369] Investigators who have concluded that screening with ultrasonography is not useful because of low specificity and predictive value failed to utilize saline instillation sonohysterography to avoid unnecessary endometrial biopsies.

It is also logical to expect these patients to be at increased risk for the development and progression of endometriosis. There are case reports of women being treated with tamoxifen, 20 mg daily, who required hysterectomy and oophorectomy for severe endometriosis.[370–373] In addition, women receiving tamoxifen develop adenomyosis, ovarian cysts, and endometrioid cancer of the ovary.[374–376] In our view, an annual pelvic examination is not sufficient; every 6 months is best.

The levonorgestrel IUS effectively protects the endometrium against hyperplasia and polyps in women using tamoxifen or postmenopausal estrogen therapy.[377–385] Breakthrough bleeding is a problem in the early months, but this method is suitable for both premenopausal and postmenopausal women using tamoxifen. This IUD can also be used to treat endometrial hyperplasia.[386–391] Comparison studies indicate that the levonorgestrel IUS is as effective, and probably better than standard treatment with an oral progestin.[387, 392, 393] However, the persistence of atypia at biopsy follow-up after 6 months is an indication that regression is unlikely to occur. Although the levonorgestrel IUD confidently provides good protection against endometrial hyperplasia, clinicians should maintain a high degree of suspicion of unusual bleeding (bleeding that occurs after a substantial period of amenorrhea) and aggressively assess the endometrium.

A woman being treated for breast cancer will naturally focus her attention and energy on the cancer itself, especially in the early years of treatment. The same can be said for the specialist who is monitoring the treatment. It falls to the patient's health care manager, her primary clinician, to look at the broader picture. A clinician interacting with patients being treated for breast cancer has an obligation to consider the impact of the patient's treatment on other body systems and functions. Tamoxifen offers the hope of adding many years to a woman's life. Medical intervention by a clinician can help make those years better with good preventive health care.

We recommend the following program for monitoring women during and after long-term tamoxifen treatment:

All women: Careful pelvic examination every 6 months to detect the emergence of endometriosis, ovarian cysts, uterine leiomyomas.

Tamoxifen should be discontinued before major surgery, and appropriate antithrombotic measures used during and after major surgery, and during immobility.

Postmenopausal women: Annual measurement of endometrial thickness by transvaginal ultrasonography. Endometrial biopsy of all women with a 2-layer thickness of 5 mm or greater. Saline instillation (sonohysterography) when appearance is not totally benign.

The use of the levonorgestrel-releasing IUD is highly recommended as prophylactic treatment.

Premenopausal women: Periodic assessment for ovulation; if ovulatory, no further intervention is necessary; however, contraceptive counseling should not be ignored.

If anovulatory, an annual endometrial aspiration biopsy; interpretation of endometrial thicknessmeasurements by ultrasonography is uncertain in premenopausal women, although a thickness less than 5 mm makes hyperplasia very unlikely. Consider the use of the progestin-releasing IUD for both contraception and protection against endometrial change.

An increased risk of endometrial cancer lingers for up to 10 years after discontinuing therapy with estrogen (without the addition of a progestin).[394, 395] It is not known whether a similar persistent increased risk is present in the years after tamoxifen treatment. It would be prudent to investigate any unexpected vaginal bleeding in women who have been previously exposed to tamoxifen.

Aromatase Inhibitors for the Treatment of Breast Cancer

Aromatase inhibitors block the conversion of androgen precursors to estrogen at all target tissue sites, nearly completely inhibiting total body estrogen production in postmenopausal women. This inhibition is not as complete in premenopausal women. The modern aromatase inhibitors include two nonsteroidal inhibitors, anastrozole (Armidex) and letrozole (Femara), and one steroidal inactivator, exemestane (Aromasin).

The aromatase enzyme is present in the stromal tissue of normal and abnormal breast tissue, and in breast epithelial cells. Aromatase activity is increased in breast cancer tissues, associated with a switch from a promoter controlled primarily by glucocorticoids and cytokines to a promoter regulated through cyclic AMP pathways.[396] However, aromatase activity in malignant breast epithelial cells is either undetectable or very low. Thus, growth stimulation of hormonally sensitive breast cancer is presumed to be influenced by local estrogen synthesis in adjacent stromal cells that is increased in a paracrine fashion by malignant cells activating alternative aromatase gene promoters.[397] A relationship between

the aromatase and prostaglandin cyclooxygenase systems may explain the beneficial effects of nonsteroidal anti-inflammatory drugs in epidemiologic reports on the risk of breast cancer.[398] Cyclooxygenase is overexpressed in breast cancer, and the treatment combination of aromatase and cyclooxygenase inhibitors is being evaluated in clinical trials.[399]

The specific inhibitors of P450arom that have been developed produce intense blockage of estrogen production, and importantly, reduce estrogen biosynthesis in the cells adjacent to breast tumors. The initial development of drugs that reduce estrogen production focused on alterations of the androstenedione molecule to produce competitive inhibitors. The preparation of a large number of altered steroid agents yielded exemestane, a steroidal inactivator of the aromatase enzyme, that clinically is grouped in the family of aromatase inhibitors.

The first nonsteroidal inhibitors, such as aminoglutethimide, affected other CYP-450 enzymes, producing unwanted toxic effects. The current generation of nonsteroidal aromatase inhibitors containing a triazole ring, anastrozole and letrozole, are highly specific with no effect on the biosynthesis of other steroids. These agents are 100 to 3,000 times more potent than aminoglutethimide and reduce total body aromatization by 97–99%.[400]

The first studies with the nonsteroidal aromatase inhibitors demonstrated that anastrozole (1 mg daily) and letrozole (2.5 mg daily) were more effective than tamoxifen in women with advanced breast cancer.[401–403] Similar results were reported with exemestane.[404] More recent clinical trials focused on the treatment of early breast cancer.

The ATAC Trial. [405, 406, 407] The Arimidex, Tamoxifen, Alone or in Combination trial included 9,366 patients in 380 sites in 23 countries. Eighty four percent had estrogen receptor-positive tumors and one-third had positive lymph nodes. The patients were randomized to daily treatment with anastrozole, 1 mg; tamoxifen, 20 mg; or a combination of both for 5 years.

Compared with tamoxifen, anastrozole increased the disease-free survival by 14%, decreased the incidence of new contralateral primary tumors by 38% (although this difference did not reach statistical significance), and increased the time to recurrence by 17%. The combination treatment was not better than anastrozole alone. In patients negative for estrogen and progesterone receptors, the small effect of anastrozole was equivalent to that of tamoxifen. After long-term follow-up, it was apparent that the carryover effect after 5 years of treatment was greater with anastrozole compared with tamoxifen.[407]

There were significant differences comparing the adverse effects of the two drugs:

	Anastrozole (%)	*Tamoxifen (%)*
Hot flushes:	34.3	39.7
Vaginal bleeding	4.5	8.2
Endometrial cancer	0.1	0.5
Venus thromboembolism	2.1	3.5
Joint complaints	27.8	21.3
Fractures	5.9	3.7

Adverse gynecologic events occurred less frequently with anastrozole compared with tamoxifen in the ATAC trial.[408] These events included lower incidences of vaginal hemorrhage, vaginal discharge, endometrial polyps, endometrial hyperplasia, and endometrial cancer. As a result there was a 4-fold increase in hysterectomy in the women treated with tamoxifen. Importantly, more women remained adherent to treatment with anastrozole.

The BIG Trial. The Breast International Group trial randomized 8,028 women to either tamoxifen or letrozole for 5 years and reported an improved disease-free survival in the letrozole group.[409, 410]

The ITA Trial. The Italian Tamoxifen-Anastrozole Trial of 448 women compared 5 years of tamoxifen to a group switched to anastrozole after 2–3 years of tamoxifen.[411] The disease-free survival was significantly increased in the sequentially treated group.

The IES Trial. The Intergroup Exemestane Study included 4,742 patients and compared 5 years of tamoxifen with a sequential group switched to exemestane, 25 mg daily, after 2–3 years of tamoxifen.[412, 413] There was a 32% reduction in risk with exemestane for recurrence, contralateral breast cancer, or death that equaled an improvement of 4.7% in disease-free survival. The risk of contralateral breast cancer was reduced by 56%.

The TEAM Trial. The Tamoxifen Exemestane Adjuvant Multinational Study randomized 9,775 women to either exemestane, 25 mg daily, or tamoxifen, 20 mg daily.[414] In 2004, the trial was modified, switching tamoxifen patients to exemestane, and an additional 2,500 patients were recruited. The results were not yet published by 2010.

The MA.17 Trial. After 5 years of tamoxifen treatment, 5,187 women were randomized to 5 years of letrozole or placebo.[415, 416] The analysis indicated an improvement in disease-free survival with letrozole, 94.3% compared to 91.4% in the placebo group. This impact was present in both patients with positive nodes and negative nodes. Overall the letrozole group experienced a 39% reduction in contralateral primary breast cancer, a 42% reduction in recurrences, and a 38% reduction in distant metastases. Because of these benefits, the trial was unblinded and patients were given the option of switching from placebo to letrozole. In an updated analysis, disease-free survival was improved with letrozole treatment.[417, 418] This study reported significantly more joint complaints and more hot flushing compared with placebo, and an increase in fractures and cardiovascular events.

Aromatase inhibitors are more effective than tamoxifen for the treatment of estrogen-sensitive breast cancers in postmenopausal women, either for early disease or for metastatic breast cancer. The three aromatase inhibitors have similar side effect profiles. The major problem has been an increase in fractures due to the bone loss associated with the profoundly low estrogen levels (nearly a 99% decrease), an effect that can be prevented with bisphosphonate treatment. Besides hot flushing, other major side effects are joint arthralgias, reduced sexual function, and myalgia.[419] Compared with tamoxifen, there is less, if any, endometrial stimulation, and less venous thromboembolism. Anecdotal experience has suggested that anastrozole users have an increased prevalence of retinal hemorrhages, presumably due to vascular fragility secondary to estrogen depletion.[420]

A decision analysis using computer modeling suggested that a modestly improved outcome is associated with sequential therapy (tamoxifen for 2.5 years followed by an aromatase inhibitor) compared with 5 years of an aromatase inhibitor alone. The actual increase was only 1–2%.[421] This small difference was not supported by a phase 3 clinical trial, a continuation of the BIG trial, comparing letrozole monotherapy with tamoxifen-letrozole sequential therapy, in which the overall survival in the treatment groups did not differ.[422]

The American Society of Clinical Oncology and the National Comprehensive Cancer Network, based on the results of the clinical trials, now make the following recommendations[417, 423]:

- Postmenopausal women with hormone-positive breast cancers should be treated with an aromatase inhibitor for 5 years.
- Premenopausal women with hormone-positive breast cancers should be treated with tamoxifen for 5 years, to be followed by 5 years of an aromatase inhibitor if the patient becomes postmenopausal during treatment.

- Treatment options include 5 years of aromatase inhibitor treatment alone or sequential therapy with 2–3 years of tamoxifen followed by aromatase inhibitor treatment for 5 years.
- Aromatase inhibitors are appropriate as initial treatment for women with contraindications to tamoxifen.
- Aromatase inhibitor treatment has been associated with better response rates compared with tamoxifen in postmenopausal women with tumors overexpressing HER-2. This evidence is not strong, but should be considered.
- Postmenopausal women finishing 5 years of treatment with tamoxifen should consider treatment with an aromatase inhibitor for 5 years.
- There is insufficient evidence available to support the use of tamoxifen after treatment with an aromatase inhibitor.

A reasonable and important addition to these recommendations is to promote adequate calcium and vitamin D supplementation and to consider prophylactic bisphosphonate treatment to prevent bone loss and fractures. Treatment with zoledronic acid (4 mg intravenously every 6 months) increased bone density in women being treated with letrozole, and *even improved disease-free survival rates*.[424, 425] Similar bone density results were obtained with the once-a-month oral bisphosphonate, ibandronate.[426] ***More effective protection against bone loss and fractures is achieved by starting bisphosphonate treatment simultaneously with an aromatase inhibitor.[427] Bisphosphonates appear to have an independent beneficial impact on the incidence of breast cancer. In the Women's Health Initiative, oral bisphosphonate users had a significant 32% reduction in estrogen receptor-positive breast cancer incidence.[428] A case-control study in Israel found a 29% reduction in the risk of postmenopausal breast cancer with the use of bisphosphonates for more than 1 year.***

A meta-analysis of randomized trials comparing aromatase inhibitors with tamoxifen in early breast cancer, focused on cardiovascular risk.[429] The meta-analysis included seven randomized trials with a total of 19,818 patients. The relative risk for cardiovascular adverse events with aromatase inhibitors was 1.31 (CI=1.07–1.60). The number of patients needed to harm 1 patient was 189. There was a 47% reduced risk of thromboembolic events with aromatase inhibitors, RR=0.53 (CI=0.42–0.65).

The increase in cardiovascular disease reflects the absence of a beneficial influence of estrogen on the lipid profile and on important vascular epithelial functions such as nitric oxide synthesis. The meta-analysis suggested that this is a relatively low risk, but the actual risk will not be known until the ongoing trials comparing aromatase inhibitors to placebo treatment are completed with long-term follow-up data. Furthermore, it is inadvisable to consider only one of the estrogen deficiency side effects. The overall impact on a patient will be determined by the additive effects on all estrogen target tissues. The effects on cognition and the risk of Alzheimer's disease are major potential issues. Cognitive assessments were performed in a subgroup of women in the IBIS anastrozole trial for prevention of breast cancer; no significant differences were observed comparing the treatment group with the placebo group; however the duration of the study was only 2 years.[430] Another study reported worse verbal and visual learning in women treated with anastrozole compared with tamoxifen treatment.[431] Dutch women in the TEAM trial who were tamoxifen users performed worse than healthy controls on verbal memory and executive function testing whereas no adverse effects were observed in exemestane users after 1 year.[432] Continuing follow-up of treated women over a longer period of time will be necessary to acquire a better understanding of the impact of aromatase inhibitors on cognition.

Some important questions regarding aromatase inhibitors remain unanswered. The optimal duration of therapy is not established. At the present time, treatment longer than 5 years awaits appropriate clinical trial data. The long-term safety is unknown; will a low estrogen and a relatively high androgen hormonal environment lead to clinical consequences? Balancing the benefits and risks will require this information. Nevertheless,

aromatase inhibitor treatment has justifiably usurped the place of tamoxifen in the adjuvant treatment of breast cancer; the reason being that it is more effective to pharmacologically block estrogen biosynthesis, especially at the local level, than it is to interfere with estrogen action.

Tamoxifen, Raloxifene, and Aromatase Inhibitors for Prevention of Breast Cancer

Tamoxifen

Women at increased risk for breast cancer participated in a breast cancer prevention trial initiated in the U.S. in 1992. The study compared two groups of women,one treated with placebo and one with 20 mg tamoxifen daily for 5 years. Early in 1998 (after about 4 years of follow-up), the study was unblinded because there were 49% fewer cases of invasive breast cancer and 50% fewer cases of noninvasive breast cancer in the tamoxifen-treated arm of the study.[240] This outcome was not without risk. There was a 2.4-fold increase in postmenopausal endometrial cancer, a 2.8-fold increase in pulmonary embolism, a 1.6-fold increase in venous thrombosis, and a 1.6-fold increase in cataracts.

There have been four randomized placebo-controlled tamoxifen prevention trials. In the 7-year follow-up report of the American tamoxifen for prevention study, the risk for breast cancer was 0.57 (CI=0.46–0.79), a 43% reduction, not the 50% cited in the results above, and the risk for in-situ disease was 0.63 (CI=0.45–0.89), a 37% reduction.[241] Follow-up of the Italian national trial demonstrated a 23% reduction of estrogen receptor-positive cancers in the group of women considered to be at the highest risk of cancer.[433]

The Royal Marsden randomized, double-blinded tamoxifen breast cancer prevention trial began in 1986, enrolling 2,494 women with a positive first-degree family history of breast cancer.[434] The treatment group received 20 mg tamoxifen daily for 8 years. Twenty years later (median follow-up of 13 years), there were 139 estrogen receptor-positive breast cancers for a 39% reduction (HR=0.61; CI=0.43–0.86). The lowered risk did not become statistically significant until after the 8-year treatment period.

The International Breast Cancer Intervention Study (IBIS), also a randomized, double-blinded trial, began in 1992 and enrolled 7,145 women; the treatment period with tamoxifen 20 mg daily was 5 years.[435] After a median follow-up of 8 years, there was a 34% reduction in estrogen receptor-positive cancers (RR=0.66; CI=0.50–0.87). The IBIS trial found a greater reduction during the treatment period, but when the analysis was restricted to estrogen receptor-positive cancers, the IBIS and Royal Marsden trials were similar, finding a greater effect after treatment. The differences among these trials are attributed to variations in risk factors in the studied populations. The American trial enrolled women with risk assigned by the Gail model. The women in the International trial were at a lower risk than those in the Royal Marsden trial, and the women in the Italian trial were not assessed for risk.

Epidemiologists from England, Italy, and Australia reviewed the combined results of the breast cancer tamoxifen prevention trials and added updated results.[436] The combined data indicated a 48% reduction in estrogen receptor-positive cancers and no effect on the incidence of estrogen receptor-negative cancers. The overall relative risk of endometrial cancer with tamoxifen was increased 2.4-fold, and the relative risk of

venous thromboembolic events was 1.9. The length of follow-up and patient numbers do not allow data regarding breast cancer mortality. The impact of 5 years of tamoxifen treatment on 1,000 high-risk women should yield an 18% reduction in mortality within 10 years of diagnosis. Experts and organizations in the breast cancer world have agreed that tamoxifen reduces the incidence of estrogen receptor-positive cancers in high-risk women. In an assessment of the women in the American preventive trial, tamoxifen reduced the incidence of breast cancer among *BRCA2* carriers, but not in *BRCA1* carriers, perhaps reflecting the fact that most of the *BRCA2* carriers have estrogen receptor-positive tumors in contrast to the prevalence of estrogen receptor-negative tumors in *BRCA1* carriers (another problem was small numbers, 8 with *BRCA1* mutations and 11 with *BRCA2* mutations).[437]

The evidence supports tamoxifen reduction of the risk for estrogen receptor-positive breast cancer, but at the same time, tamoxifen should be recommended as a preventive agent only for women at very high risk. This conclusion is based upon the degree of reduction in risk compared with the incidence of side effects. An evaluation by the National Cancer Institute is very helpful.[438, 439] Because the risks associated with tamoxifen (endometrial cancer, stroke, pulmonary embolism, and deep vein thromboembolism) increase with age, balancing the risks and benefit indicates that tamoxifen is best for younger women with an elevated risk of breast cancer (an increased relative risk of approximately 1.7). A similar conclusion was reached by a working group of the American Society of Clinical Oncology.[440] This means that only a relatively small number of women qualify, about 5% of American white women and 0.6% of black women.[439]

There is one lingering concern. There has been a slight increase in estrogen receptor-negative cancers in the follow-up period after treatment in all of the prevention trials. It is uncertain if this is related to tamoxifen exposure; however, in the trials assessing tamoxifen treatment of breast cancers, survival and recurrence rates worsened with longer therapy, probably due to the emergence of tamoxifen-resistant tumors.

In conclusion, tamoxifen exposure for 5 to 8 years is associated with about a 30% to 50% reduction in estrogen receptor-positive breast cancers for at least 15 years after the treatment ends. An estimate of the absolute impact puts this in better perspective. The absolute reduction in cumulative *overall* incidence of breast cancer after 5 years is estimated to about 1.1% and after 10 years, 1.7%. This small impact, combined with the serious side effects, have made tamoxifen treatment an unattractive option.

Raloxifene

The MORE trial, the Multiple Outcomes of Raloxifene Evaluation trial, was a randomized, double-blind, multicenter clinical study of postmenopausal women with osteoporosis that reported a 72% reduction in estrogen receptor-positive invasive breast cancer in the treatment group after 4 years compared with placebo.[441] The CORE study, the Continuing Outcomes Relevant to Evista trial, was designed to measure the impact of 4 additional years of raloxifene (60 mg/day), to begin during the fourth year of the MORE trial.[442] Of the 7,705 participants initially randomized in the MORE trial, 3,510 women elected to continue raloxifene treatment (2,336 completed the CORE trial) and 1,703 continued on placebo (1,106 completed the trial). During the 4-year CORE study, raloxifene treatment was associated with a 66% (HR=0.34; CI=0.18–0.66) reduction of estrogen receptor-positive invasive breast cancers in the treated group. There was no difference in estrogen receptor-negative tumors. Over the entire 8-year period, the reduction in estrogen receptor-positive cancers reached 76%. In the 8-year period, there was no difference in the number of deaths in the two groups.

The Study of Tamoxifen and Raloxifene (STAR) trial enrolled 19,747 women at increased risk of breast cancer who were randomized to treatment with either raloxifene, 60 mg daily, or tamoxifen, 20 mg daily, in more than 500 centers in the U.S., Canada, and Puerto Rico.[443] The reported results after an average treatment period of almost 4 years were as follows[443]:

	Raloxifene (9,745 women)	*Tamoxifen (9,726 women)*
Invasive breast cancer	167 cases	163 cases
Breast cancer-in-situ	81	57
Deep venous thrombosis	65	87
Pulmonary embolus	35	54
Strokes	51	53
Fractures	96	104
Cataracts	313	394
Uterine cancer	23	36

The numbers of invasive breast cancers were identical in the two groups of women. It was estimated that these results were equivalent to about a 50% reduction (based on the previous results in the tamoxifen prevention trial),[240, 241] but without a placebo arm, an accurate assessment was impossible. Thus, raloxifene appears to achieve the same reduction as tamoxifen in invasive breast cancers with a lesser increase in venous thrombosis, and perhaps no increase in cataracts and uterine cancer. "Quality of life" was said to be the same for both drugs.

The fracture rates in the hip, wrist, and spine in the STAR trial were similar in the two groups. In the 7-year follow-up report of the U.S. breast cancer prevention trial with tamoxifen, osteoporotic fractures were reduced by 32%; compared with placebo, there were 11 fewer hip fractures, 13 fewer spinal fractures, and 9 fewer fractures of the radius.[241] However, even after 8 years of follow-up of the raloxifene trial involving women with osteoporosis, no effect of raloxifene has been evident on non-vertebral fractures.[444] A similar fracture rate in the STAR trial with the two treatments must reflect the incidence of spinal fractures. Neither tamoxifen nor raloxifene can achieve the efficacy in preventing all fractures well-proven with both hormone therapy and bisphosphonate treatment. Raloxifene's lack of effect on the risk of hip fractures makes it less advantageous than tamoxifen for bone protection.

The rate of strokes was equivalent in the two treatment arms of the STAR trial. The rate of stroke was increased by 42% in the tamoxifen prevention trial (coming close, but not achieving statistical significance.[241] This is a serious risk for both drugs.

Aromatase Inhibitors

Prevention trials are underway with aromatase inhibitors.[445] The International Breast Cancer Intervention Study (IBIS) compares anastrozole to placebo in 6,000 postmenopausal women. The Mammary Prevention 3 (MAP3) trial compares exemestane to placebo in 4,560 postmenopausal women. A third study (STELLAR) has been proposed, comparing letrozole and raloxifene. The MAP1 randomized trial evaluated the effect of letrozole treatment on breast density; no effect on breast density was observed after 1 year of treatment.[446] The French Onco-03/LIBER trial is assessing the use of letrozole in *BRCA* mutation carriers.

These results lead us to recommend tamoxifen prophylaxis (20 mg daily for 5 years) or raloxifene prophylaxis (60 mg daily for 5 years) for those women who are diagnosed with carcinoma in situ of the breast or who have atypical hyperplasia in a breast biopsy (especially if a positive family history of breast cancer is also present). The important positive family history criteria are at least one first-degree relative with breast cancer diagnosed before the age of 50 or two or more relatives (at least one first-degree relative) with breast cancer. For others who seek preventive treatment, we advise that the final answers are not in, and that clinical trial results from long-term follow-up will be necessary before fully informed decision-making is possible. Women at very high risk for breast cancer who choose tamoxifen or raloxifene treatment deserve support and appropriate surveillance. Bisphosphonate treatment is recommended to prevent bone loss and gain the added benefit of a further reduction in breast cancer risk.

Vasomotor Symptoms with Tamoxifen and Aromatase Inhibitors in Breast Cancer Survivors

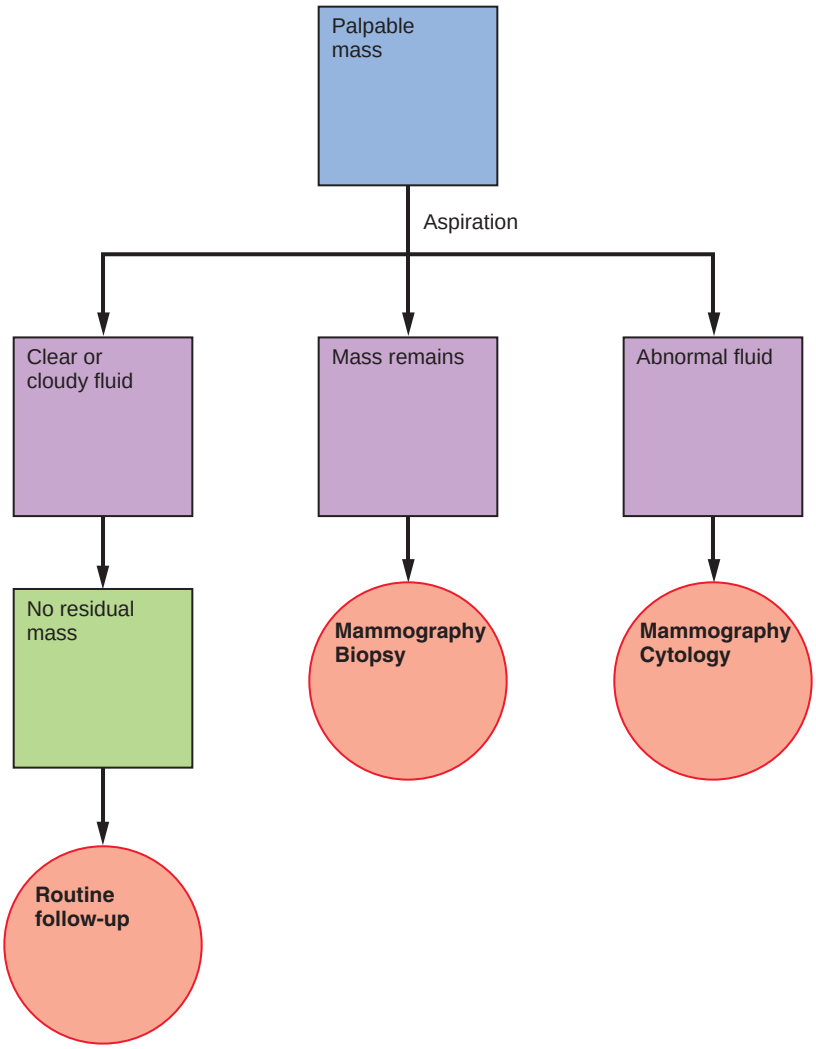
The problem of hot flushing should not be underrated. Women have vasomotor symptoms on tamoxifen, raloxifene, and aromatase inhibitors, and those that already had flushing sometimes have worse flushing. Various treatments are available, discussed in Chapter 18, under "Treatment Options for Hot Flushes" and under "Should a Women Who Has Had Breast Cancer Use Postmenopausal Hormones?"

The SSRIs are the best choice after hormone therapy. It is worth trying to titer the dose down to its lowest effective level because of a low but bothersome incidence of decreased libido. In addition, clinical experience indicates that it is best to slowly titrate upward to the recommended dose and, likewise, to wean the patient slowly when discontinuing treatment. SSRIs are effective for flushing secondary to both tamoxifen and hypoestrogenemia, and the efficacy is similar in women with and without breast cancer.[447] An added advantage of the SSRIs is the fact that the clinical studies have also reported improvements in depression, anxiety, and sleep.

There is a concern that is of potential clinical importance. Tamoxifen is converted to an active metabolite by enzymes that are inhibited by SSRIs. Paroxetine coadministration decreases plasma concentrations of the active metabolite.[448, 449] A lesser effect is associated with fluoxetine and sertraline. In a retrospective cohort study, only paroxetine use during tamoxifen therapy was associated with an increased risk of death due to breast cancer.[450] ***Paroxetine, fluoxetine, and sertraline are best avoided in women being treated with tamoxifen.***

Needle Aspiration

Needle aspiration of breast lumps should be part of the practice of everyone who cares for women.[451] The technique is easy. A small infiltrate of lidocaine is placed in the skin (many clinicians believe that local anesthesia is unnecessary). Holding the lesion between thumb and index fingers with one hand, the other hand passes a 22-gauge needle attached to a 3-finger control syringe into the lesion. Aspiration will reveal the presence of cystic fluid from a cyst. If the mass is solid, the needle should be passed at least 2–4 times (even more if nothing is being obtained) back and forth through the lesion with continuous suction on the syringe. Air is forcibly ejected through the needle on to a cytology slide for smearing and fixing. The usual Pap smear fixative can be used.

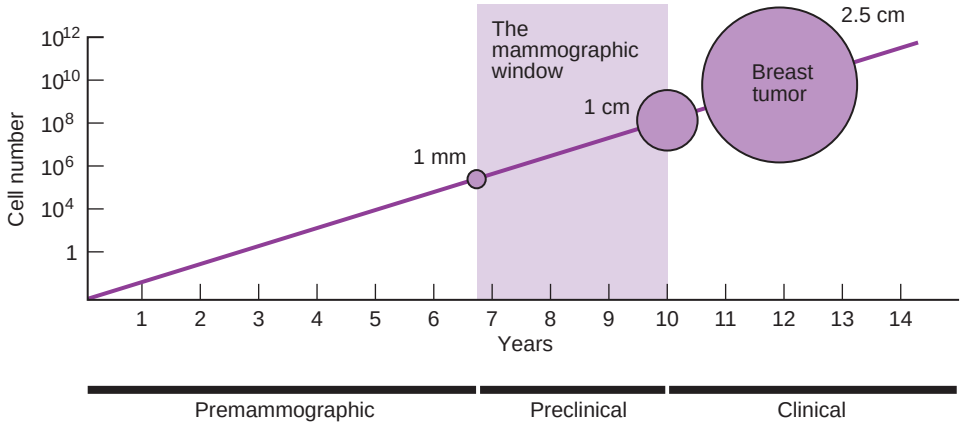


The procedure is very cost-effective. When aspiration yields clear or cloudy, green-gray or yellow fluid and the mass disappears, the procedure is both diagnostic and therapeutic. Fluid of any other nature requires cytologic assessment.[452] Failure to obtain material for cytologic evaluation or the persistence of a mass requires biopsy. The mass should not have returned at the follow-up examination 1 month after the aspiration. Locally recurrent cysts should be surgically removed for histologic diagnosis.

Screening Mammography

Mammography is a means of detecting a nonpalpable cancer. Technical advancements have significantly improved the mammographic image and reduced the radiation dose.[453] The doubling time of breast cancer is very variable, but, in general, a tumor doubles in size every 100 days. Thus, it takes a single malignant cell approximately 10 years to grow to a clinically detectable 1-cm mass, but by this time a tumor of 1 cm has already progressed through 30 of the 40 doublings in size, which is estimated to be associated with fatal disease.[454] Furthermore, the average size at which a tumor is detected has been (prior to mammography) 2.5 cm, a size that has a 50% incidence of lymph node involvement. Studies of breast self-examination have been disappointing in their failure to demonstrate an impact

on breast cancer stage of disease and mortality.[455] To decrease the mortality from breast cancer, we must utilize a technique to find the tumors when they are smaller.



After Wertheimer, et al.[454]

Mammography is the only method that detects clustered microcalcifications. These calcifications are less than 1 mm in diameter and are frequently associated with malignant lesions. More than 5 calcifications in a cluster are associated with cancer 25% of the time and require biopsy. Besides microcalcifications, the following mammographic findings usually require surgical evaluation: the appearance of a mass, calcifications associated with a mass, an area of distortion or asymmetrical density, a stellate lesion. A pattern of dysplasia on the mammogram carries with it an increased risk (2.0–3.5 times normal) of breast cancer.

Mammography has a false-negative rate of 5–10%. This means that masses are palpable but not visible. Mammography cannot and should not replace examination by patient and clinician. Breast examinations by clinicians do detect cancers that are missed by mammography.[456] Cancer commonly presents as a solitary, solid, painless (only 10% of cancers are painful), hard, unilateral, irregular nonmobile mass. A mass requires biopsy regardless of the mammographic picture.

The Effectiveness of Mammography

Mammography reduces breast cancer mortality. The clinical trial results indicate increased survival with tumors detected by screening mammography, and in addition, early detection increases the options for treatment. About a 30% reduction in mortality can be expected with screening mammography of asymptomatic women over age 50.[457, 458]

The U.S. Preventive Services Task Force recommended in 2009 against routine screening mammography in women aged 40–49 and for extending the screening interval for women aged 50–74 to every 2 years.[459] The Task Force further recommended the discontinuation of breast self examinations. The recommendations of the U.S. Preventive Services represent the consensus of a panel of clinicians, academicians, and epidemiologists after a systematic review of the literature. The various task forces are often very conservative, refusing to make clinical judgements when evidence is deemed insufficient, and focusing on a collective impact that includes not only clinical outcome but cost as well. We disagree with the Task Force recommendation on screening mammography for the following reasons.

About 17% of breast cancers occur in women 40–49, accounting for approximately 10.5% of all deaths due to breast cancer.[460] The American Breast Cancer Detection Demonstration

Project demonstrated that screening was just as effective for women in their 40s as in women over 50.[461] This program that was organized by the American Cancer Society and the National Cancer Institute began operating in 1973 in 28 locations throughout the United States, enrolling more than 280,000 women. Despite the fact that this was not an organized research study with a control group, the massive database permits many valuable conclusions. From 1977 to 1982, similar high survival rates (87%) for women in their 40s compared with women in their 50s verify that screening was just as effective in the younger women. A 5-year survival rate for patients under 50 with breast cancers detected by examination was 77% compared to 95% in those patients with breast cancers detected by mammography.[462] In a randomized trial in Gothenburg, Sweden, women ages 39–49 undergoing mammographic screening every 18 months had a 45% reduction in breast cancer mortality in an early report, and a 31% reduction after 13 years of follow-up.[463, 464] Meta-analyses of randomized clinical trials concluded that in women aged 40–49 offered mammography screening, there was about a 20% reduction in breast cancer mortality.[457, 465, 466]

It takes longer for a significant difference in mortality to appear in 40–49-year-old women compared with women over age 50. There are 2 explanations. One is that tumors grow faster in younger women, and the other is the greater difficulty in achieving accurate mammography because of the denser, more glandular breasts in younger women compared to the more fatty breasts in older women. Because the breast density changes gradually, rapid tumor growth must be the more critical factor.

Once detected by mammography, the stage of disease and survival expectations are the same comparing women aged 40–49 with women over age 50.[467] However, cancers that are detected between screenings have lower survival rates (at all ages). Therefore, another reason that it has been difficult to demonstrate an impact of screening in the age group 40–49 is that because of less than annual screening, more of the cancers are detected late (between screenings). This in turn reflects the faster tumor growth in younger women.[468] Because the randomized clinical trials have screened younger women at 2-year or longer intervals, it is not surprising that screening has been less effective for these faster growing tumors. It is logical that women aged 40–49 should have annual screening mammography.[469, 470] A randomized trial in the U.K. of annual mammographic screening beginning at age 40 indicated a 24% reduction in breast cancer mortality in the screened women.[471]

There are problems to be anticipated with extensive mammography screening. Small nonpalpable lesions have less than a 5% chance of being malignant, and overall only about 20–30% of biopsy specimens contain carcinoma. About 10% of mammograms require additional evaluation. That means there will be a large number of biopsies and mammograms performed (including the treatment of clinically irrelevant lesions), which involves costs to the health care system and cost to the individual in terms of stress and anxiety. Nevertheless mammography is the most potent weapon we possess in the battle against breast cancer. Mammography not only lowers mortality, but it also decreases morbidity because less radical surgery is necessary for smaller lesions. Most importantly, the number of unnecessary surgical procedures can be minimized by combining physical examination and mammography with needle aspiration.[472] With the so-called triple approach (examination, mammography and possibly ultrasonography in young women, and needle aspiration), the detection of a malignancy with at least one of the three diagnostic tests is very reliable; open biopsy can be avoided.[473–475]

It is appropriate to be concerned over the increased cost of annual screening. However, analysis of the increased cost, taking into account the greater efficacy of capturing early tumors comparing annual to biannual screening, reveals that the overall benefit is worthwhile, and compares favorably to the cost and benefits of Pap smear screening for cervical cancer.[476, 477]

There is a special problem with elderly women. Old women are less likely to be screened with mammography, probably due to both patient misconceptions and erroneous clinician

beliefs. Mammography reduces mortality and is cost-effective over age 65.[478] Decision analysis of available data predicts a major benefit for elderly women, and a retrospective study indicated that screening mammography in women over age 74 is as beneficial as it is in younger women.[479, 480] Older women need to be reminded that risk continues to increase with increasing age.

Digital Mammography

Digital mammography replaces the screen and x-ray film with a detection system that coverts x-ray photons to electric charge that is then converted to a digital image. This method performs better in women with dense breasts, accounting for the results in a randomized trial in which digital and film screening yielded equal results in women over age 50, but digital mammography was superior in younger women.[481, 482] Digital mammography has some important advantages: easier access to images, more efficient storage of images, the use of computer-aided reading, and rapid data transfer between clinical sites. Studies have also indicated lower recall rates because of better image quality and fewer artifacts.

Adding Ultrasound to Mammography

A prospective, multicenter, randomized trial was designed to validate the performance of screening ultrasound in conjunction with mammography in women with dense breasts and at high risk for breast cancer.[483] The study is known as ACRIN, the American College of Radiology Imaging Network 6666 trial. Each patient underwent mammography and ultrasound in a randomized sequence. Forty cases of cancer were diagnosed, 12 on ultrasound alone, 12 on mammography alone, 8 suspicious with both techniques, and 8 with negative exams. Adding ultrasound yielded an additional 4.2 cancers per 1,000 high risk women. The false-positive rate for mammography alone was 4.4%, for ultrasound alone, 8.1%, and for combined mammography plus ultrasound 10.4%. Thus, adding ultrasound to mammography screening in high-risk women with dense breasts improved the sensitivity of screening, but increased the rate of false-positive examinations. Breast cancer mortality was not an endpoint in this trial, but the fact that the cancers detected by ultrasound are usually asymptomatic, node-negative, and not detected by mammography should yield a reduction in mortality.

Ultrasound screening can detect cancers not seen on mammography and its performance is not affected by dense breast tissue. Adding ultrasound to a screening program seems straight-forward, even though its impact on mortality reduction has not been measured in a large trial. In the single center studies of screening ultrasound that have been published, cancers were found only by ultrasound, and most were small, early-stage tumors. An Italian multicenter study reported that 29 cancers were found by ultrasound in 6,449 women with dense breasts and negative mammograms.[484] Nevertheless, a majority of facilities do not offer screening ultrasound because of a lack of qualified personnel and standardized protocols.

The problem with all screening methods is a substantial rate of false positives. In the American study, 91.4% of suspicious ultrasound findings were benign.[483] The positive predictive value for ultrasound was only 8.6%, but the value for mammography was only 14.7%. Remember that ultrasound tends to find earlier tumors. The crucial question is how many false positives are worth the gain in additional cancer diagnoses. In the American study, the gain was an additional 29% (the number of cancers detected only by ultrasound). In women

with elevated risks, this seems worthwhile. ***Women at high risk probably have a greater fear of diagnosing breast cancer late than of a false positive.***

Adding MRI to Mammography

MRI is the most sensitive technique, but it is very expensive, requires the intravenous injection of contrast, and isn't always tolerated by patients. Ultrasound has the advantage of being less expensive, easily tolerated, and widely available. Thus the combination of ultrasound and mammography seems best for women of intermediate risk. Ultrasound has a disadvantage of not detecting ductal carcinoma in situ, which is detected by mammography and MRI. ***Combining MRI with mammography yields a very high sensitivity, and this is now recommended for women at very high risk for breast cancer, especially younger women.***[485–487]

MRI is more sensitive in diagnosing breast ductal carcinoma in situ.[488] Ductal carcinoma in situ is a precursor of invasive breast cancer, with progression occurring more often and more rapidly with higher grade in situ lesions, and the subsequent invasive disease is of a higher grade with a poorer prognosis. Diagnosis of higher grade ductal carcinoma in situ is, therefore, highly desirable. Mammography has led to an increase in the diagnosis of ductal carcinoma in situ from 2% of breast cancers in 1980 to 20% today. Earlier studies concluded that MRI was no better, and even worse than mammography in diagnosing ductal carcinoma in situ. However, it has been learned that diagnostic criteria differ with the two techniques, incorporating not only morphology but enhancement kinetics with contrast during MRI. MRI detects lesions without microcalcifications (a different group of tumors), whereas mammography detects cases of ductal carcinoma in situ that have microcalcifications caused by necrosis.

Both film-screen mammography and digital mammography have limited sensitivity for diagnosing ductal carcinoma in situ (determined by the size of microcalcifications). An important message is that MRI is better for the detection of the higher grade ductal carcinoma in situ associated with worse prognosis. The reason for this is the contribution of contrast enhancement. Tissues with higher grade lesions will have greater capillary permeability and an increase in microvasculature, accounting for more contrast enhancement.

The availability of MRI in general population screening is currently limited by an insufficient number of radiologists with the required level of expertise, but there are an increasing number of specialty centers with the expertise and technology to perform accurate MRIs. The full use of MRI to detect breast cancer at its earliest stage awaits the results of a large multicenter trial that is obviously now indicated.

The final protocol for the best screening use of the three modalities, mammography, ultrasound, and MRI, will also require consideration of cost. The total cost is a complex summary of the technology, the time consumed, the increase in patient anxiety and discomfort, and the expense of additional testing because of false positives. Nevertheless, the evidence now seems sufficient to individualize decision-making and to recommend more than the single technique of mammography for high-risk patients (defined as a combination of factors that produces a 3-fold increase in risk), especially in women with dense breasts. Thus far, over 90% of cancers detected only on ultrasound were in women with dense breasts.

Every woman should be regarded as at risk. Health care professionals who interact with women have the opportunity to initiate an aggressive program of preventive health care. The major deterrent to patient use of screening is the absence of a strong clinician recommendation. We urge you to follow these guidelines:

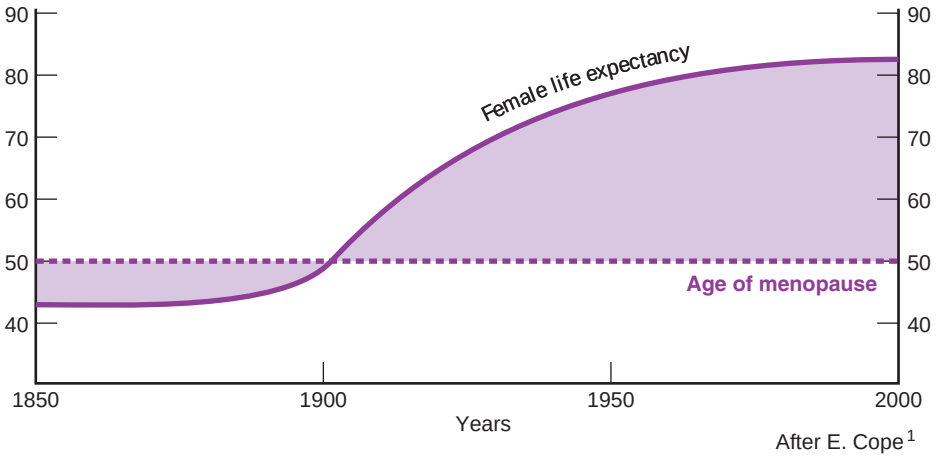
Screening for Breast Cancer

- **All women should be taught self-examination of the breast by age 20. Because of the changes that occur routinely in response to the hormonal sequence of a normal menstrual cycle, breast examination is most effective during the follicular phase of the cycle and should be performed monthly.**
- **All women over the age of 35 should have an annual breast examination.**
- **Women with a first-degree relative with premenopausal breast cancer should begin annual mammography 5 years before the age of the relative when diagnosed.**
- **Annual mammography should be performed in all women over age 39.**
- **Digital mammography is preferred for women with dense breasts.**
- **It is advisable to add ultrasonography to mammography for hormone users who develop dense breasts and the density persists despite a short period without hormone therapy.**
- **MRI should be added to mammography for women at very high risk for breast cancer (defined as a combination of factors that produces a 3-fold increase in risk), especially younger women. For an individual identified to be at high risk, especially women with inherited mutations, annual clinical breast examination is recommended every 6 months and annual mammography and MRI beginning at age 25. Clinical evaluation every 6 months is appropriate because the *BRCA1*-related tumors have been demonstrated to be faster growing tumors. Some argue that ultrasonography examination every 6 months is a useful and cost-effective addition to detect rapidly growing tumors.[489] Support should be provided for those women who choose prophylactic mastectomy. Pelvic examination, serum CA-125 levels, and transvaginal ultrasonography with color Doppler are recommended annually for women under age 40, although it has not been demonstrated that this screening will detect ovarian tumors early enough to influence prognosis. Prophylactic salpingo-oophorectomy and hysterectomy are recommended at the completion of child-bearing, preferably before age 35 and certainly by age 40.**

All references are available online at: http://www.clinicalgynendoandinfertility.com

17

Menopause and the Perimenopausal Transition



After E. Cope[1]

Throughout recorded history, multiple physical and mental conditions have been attributed to the menopause. Although medical writers often wrote colorfully in the past, unfortunately they were also less than accurate, unencumbered by scientific information and data. A good example of the stereotypical, inaccurate thinking promulgated over the years is the following written in 1887[2]:

> The ovaries, after long years of service, have not the ability of retiring in graceful old age, but become irritated, transmit their irritation to the abdominal ganglia, which in turn transmit the irritation to the brain, producing disturbances in the cerebral tissue exhibiting themselves in extreme nervousness or in an outburst of actual insanity.

The belief that behavioral disturbances are related to manifestations of the female reproductive system is an ancient one that has persisted to contemporary times. This belief regarding the menopause is not totally illogical; there is reason to associate the middle years of life with negative experiences. The events that come to mind are impressive: onset of a major illness or disability (and even death) in a spouse, relative, or friend; retirement from employment; financial insecurity; the need to provide care for very old parents and relatives; and separation from children. Thus, it is not surprising that a middle-age event, the menopause, shares in this negative outlook.

The scientific study of all aspects of menstruation has been hampered by the overpowering influence of social and cultural beliefs and traditions. Problems arising from life events have often been erroneously attributed to the menopause. But data, especially more reliable community-based longitudinal data, now establish that the increase in most symptoms and problems in middle-aged women reflects social and personal circumstances, not the endocrine events of the menopause.[3–12] The variability in menopausal reactions makes

the cross-sectional study design particularly unsuitable. Longitudinal studies are better for documenting what is normal and the variations around normal.

The Massachusetts Women's Health Study, a large and comprehensive prospective, longitudinal study of middle-aged women, provides a powerful argument that the menopause is not and should not be viewed as a negative experience by the vast majority of women.[4, 13] The cessation of menses was perceived by these women (as have the women in other longitudinal studies) as having almost no impact on subsequent physical and mental health. This was reflected by women expressing either positive or neutral feelings about menopause. An exception was the group of women who experienced surgical menopause, but there is good reason to believe that the reasons for the surgical procedure were more important than the cessation of menses.

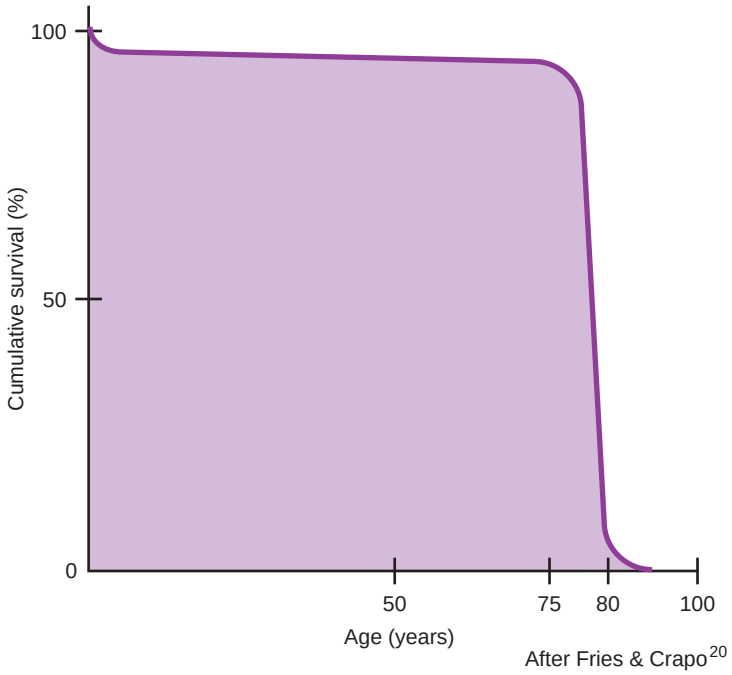
Changes in menstrual function are not symbols of some ominous "change." There are good physiologic reasons for changing menstrual function, and understanding the physiology will do much to reinforce a healthy, normal attitude. Attitude and expectations about the menopause are very important. Women who have been frequent users of health services and who expect to have difficulty do experience greater symptoms and higher levels of depression.[5, 9, 10] The symptoms that women report are related to many variables within their lives, and the hormonal change at menopause cannot be held responsible for the common psychosocial and lifestyle problems we all experience. It is important to stress the normalcy of this physiologic event. Menopausal women do not suffer from a disease (specifically a hormone deficiency disease), and postmenopausal hormone therapy should be viewed as specific treatment for symptoms in the short term and preventive pharmacology in the long term.

It can be further argued that physicians have had a biased (negative) point of view, because the majority of women, being healthy and happy, do not seek contact with physicians.[14, 15] It is vital, therefore, that clinicians not only are familiar with the facts relative to the menopause but also have an appropriate attitude and philosophy regarding this period of life. Medical intervention at this point of life should be regarded as an opportunity to provide and reinforce a program of preventive health care. The issues of preventive health care for women are familiar ones. They include family planning, cessation of smoking, control of body weight and alcohol consumption, prevention of cardiovascular disease and osteoporosis, maintenance of mental well-being (including sexuality), cancer screening, and treatment of urologic problems.

Growth of the Older Population

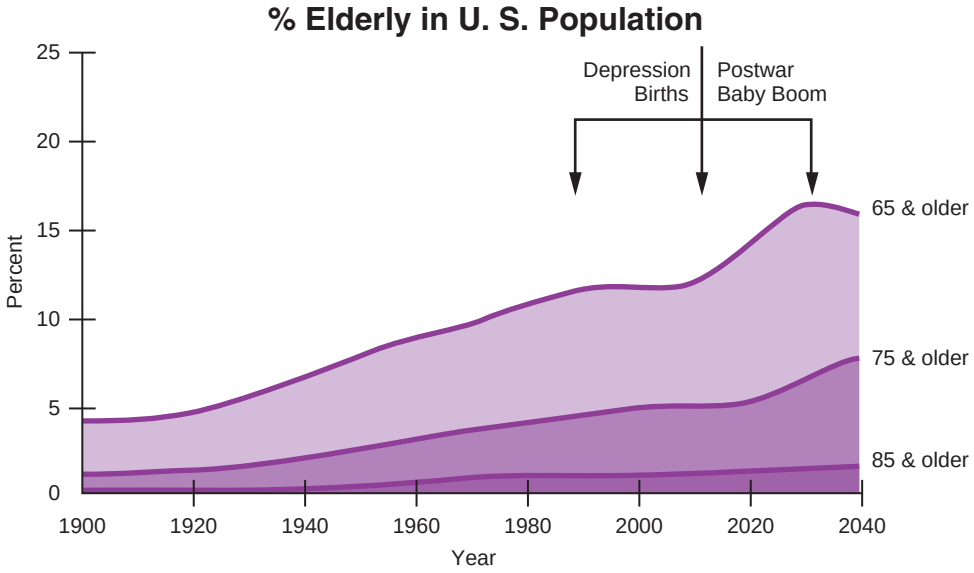
We are experiencing a relatively new phenomenon: we can expect to become old. We are on the verge of becoming a rectangular society. This is a society in which nearly all individuals survive to advanced age and then succumb rather abruptly over a narrow age range centering around the age of 85.

In 1000 B.C., life expectancy was only 18 years. By 100 B.C., the time of Julius Caesar, it had reached 25 years. In 1900, in the U.S., life expectancy still had reached only 49 years. In 2005, the average life expectancy was 80.7 years for women and 75.4 for men.[16] Today, once you reach 65, if you are a man you can expect to reach 82.2, if you are a woman, age 85.[17] We can anticipate that eventually about two-thirds of the population will survive to 85 or more, and more than 90% will live past age 65—this would be the nearly perfect rectangular society.[18, 19] Currently, Sweden and Switzerland are closest to this demographic composition.



A good general definition of elderly is 65 and older, although it is not until age 75 that a significant proportion of older people show the characteristic decline and problems. Today the elderly population is the largest contributor to illness and human need in the U.S. There are more old people (with their greater needs) than ever before.[21] In 1900, there were approximately 3 million Americans 65 and older (about 4% of the total population), and in 2000, there were 35 million (about 12% of the total population). By 2030, the elderly population in the U.S. will reach about 70 million, and about one in five Americans will be elderly.[21] The world's elderly population will more than double from 1998 to 2025, rising from 264 million in 2009 to 416 million in 2050.[22] Population aging must be added to population growth as very important social problems.

Two modern phenomena have influenced the rate of change. The first was the post–World War II baby boom (1946–1964) that temporarily postponed the aging of the population, but now is causing a faster aging of the general population. The second major influence



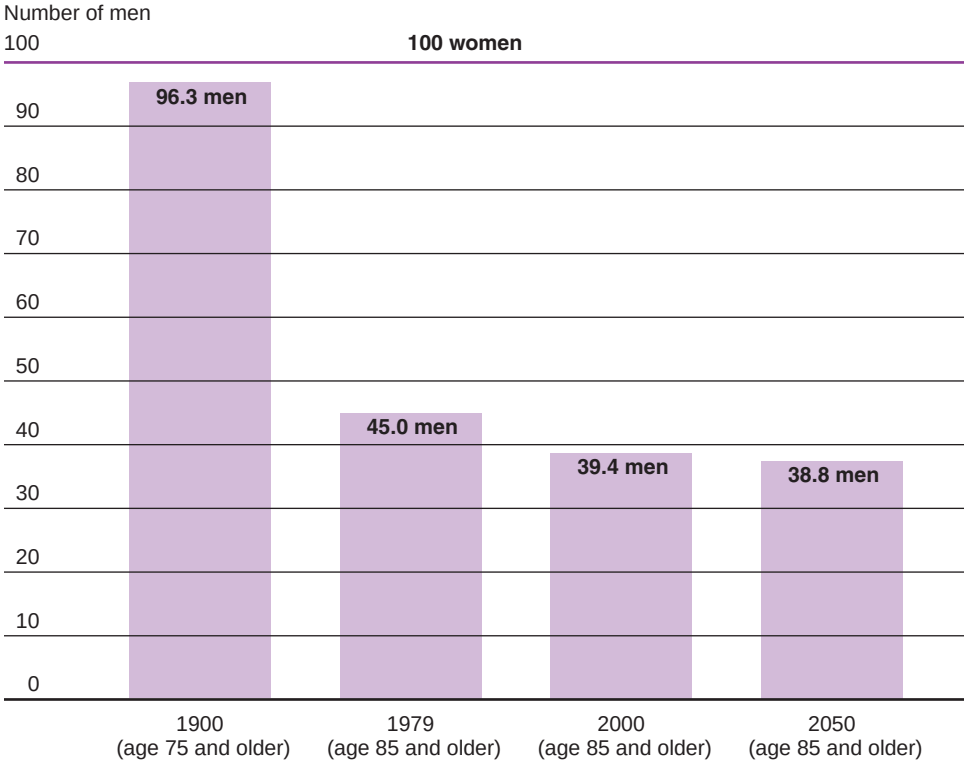
has been the modern decrease in old-age mortality. Our success in postponing death has increased the upper segment of the demographic contour. By 2050, the current developed nations will be rectangular societies. China, by 2050, will contain more people over age 65 than the number of people of all ages currently living in the U.S.

Current World Population Changes[23]

	Births	*Deaths*	*Growth*
Year	140,773,000	51,315,000	89,458,000
Month	11,731,080	4,276,250	7,454,834
Week	2,707,173	140,589	245,090
Hour	16,070	5,858	10,212
Minute	268	96	170
Second	4.5	1.6	2.8

This is a worldwide development, not limited to affluent societies.[23] The population of the earth will continue to grow until the year 2100 or 2150, when it is expected to stabilize at approximately 11 billion. After 2020, all of this growth will occur in developing countries.[22] In 2000, the poorest countries (located in Africa and Asia) accounted for 87% of the world's population. In most developing countries, the complications associated with pregnancy, abortion, and childbirth are either the first or second most common cause of death, and almost half of all deaths occur in children under age 5. Limiting family size to two children would cut the annual number of maternal deaths by 50% and infant and child mortality also by 50%.[24] Thus, it is essential to focus attention on population control, but declining fertility will increase population aging. In 1950, only 40% of people 60 and older lived in developing countries. By 2050, about 80% will live in those countries, as fertility in the developing regions is expected to drop from 2.73 children per woman in 2005–2010 to 2.05 by 2050.[22]

Men per 100 U.S. Women[25]



In 1900, men over age 65 in the U.S. outnumbered women 102 to 100. Now, there are only 70 men for every 100 women over age 65.[25] By age 85, only 39 men are alive for every 100 women. Nearly 90% of white American women can expect to live to age 70. Vital statistics data indicate that this gender difference is similar in both the black and white populations in the U.S.[26] Approximately 55% of girls, but only 35% of boys, live long enough to celebrate their 85th birthday.[27] One in 5,600 individuals can expect to live to be 100.[25]

The 15 leading causes of death in the U.S. in 2006 were as follows:[16]

1. Diseases of the heart — 26%
2. Malignant neoplasms — 23%
3. Cerebrovascular diseases — 5.7%
4. Chronic lower respiratory diseases
5. Accidents
6. Alzheimer's disease
7. Diabetes mellitus
8. Influenza and pneumonia
9. Renal diseases
10. Septicemia
11. Suicides
12. Liver diseases
13. Hypertensive diseases
14. Parkinson's disease
15. Homicides

Men and women reach old age with different prospects for older age, a sex differential that (it can be argued) is due in significant part to the sex hormone-induced differences in the cholesterol-lipoprotein profile and other cardiovascular factors, and thus the greater incidence of atherosclerosis and earlier death in men. From a public health point of view, the greatest impact on the sex differential in mortality would be gained by concentrating on lifestyle changes designed to diminish atherosclerosis: low-cholesterol diet, no smoking, optimal body weight, and active exercise. The death rate is higher for men at all ages, and coronary heart disease accounts for 40% of the mortality difference between men and women. Another one-third is from lung cancer, emphysema, cirrhosis, accidents, and suicides. It is interesting to note that in our society the mortality difference between men and women is largely a difference in lifestyle. Smoking, drinking, coronary-prone behavior, and accidents account for most of the higher male mortality rate over age 65. It has been estimated that perhaps two-thirds of the difference has been due to cigarettes alone. But we should emphasize that this is due to a greater prevalence of smoking in men. Women whose smoking patterns are similar to those of men have a similar increased risk of morbidity and mortality.[28]

The Older U.S. Female Population[25]								
Age	*1990*		*2000*		*2010*		*2020*	
55–64	10.8 mill.	(8.6%)	12.1 mill.	(9.0%)	17.1 mill.	(12.1%)	19.3 mill.	(12.9%)
65–74	10.1	(8.1%)	9.8	(7.3%)	11.0	(7.8%)	15.6	(10.4%)
> 75	7.8	(6.2%)	9.3	(7.0%)	9.8	(6.9%)	11.0	(7.3%)
Total	28.7		31.2		37.9		45.9	

The mortality sex difference has been decreasing since 1979. The U.S. Census Bureau projects that the difference in life expectancy between men and women will increase until the year 2050, and then level off. In 2050, life expectancy for women will be 82 years and for men, 76.7 years.[29] There will be 33.4 million women 65 and older, compared with 22.1 million men.

In addition to the growing numbers of elderly people, the older population itself is getting older. For example, in 1984, the 65–74 age group in the U.S. was over 7 times larger than in 1900, but the 75–84 group was 11 times larger and the 85 and older group was 21 times larger. In the 1990s, the population 85 years and older increased by 38%.[25] The most rapid increase is expected between 2010 and 2030 when the post World War II baby boom generation will be age 65 and over. In the next century, the only age groups in the U.S. expected to experience significant growth will be those past age 55. In this older age group, women will outnumber men by 2.6 to 1. By the year 2040 in the U.S., there will be 8 million to 13 million people 85 years of age or older; the estimate varies according to pessimistic to optimistic projections regarding disease prevention and treatment.

Unmarried women will be an increasing proportion of the elderly. Elderly women are more likely to be widowed (59%) than elderly men (22%).[30] Half of men 85 and older live with their wives, but only 10% of elderly women live with their husbands.[31] Because the unmarried tend to be more disadvantaged, there will be a need for more services for this segment of the elderly population. Older unmarried people are more vulnerable, demonstrating higher mortality rates and lower life satisfaction.

The Rectangularization of Life

The lifespan is the biologic limit to life, the maximal obtainable age by a member of a species. The general impression is that the human lifespan is increasing. Actually lifespan is fixed, and it is a biologic constant for each species.[20] In fact, differences in species' lifespans argue in favor of a species-specific genetic basis for longevity. If lifespan were not fixed, it would mean an unlimited increase of our elderly. But a correct analysis of survival reveals that death converges at the same maximal age; what has changed is life expectancy—the number of years of life expected from birth. Life expectancy cannot exceed the lifespan, but it can closely approximate it. Thus the number of old people will eventually hit a fixed limit, but the percentage of a typical life spent in the older years will increase.

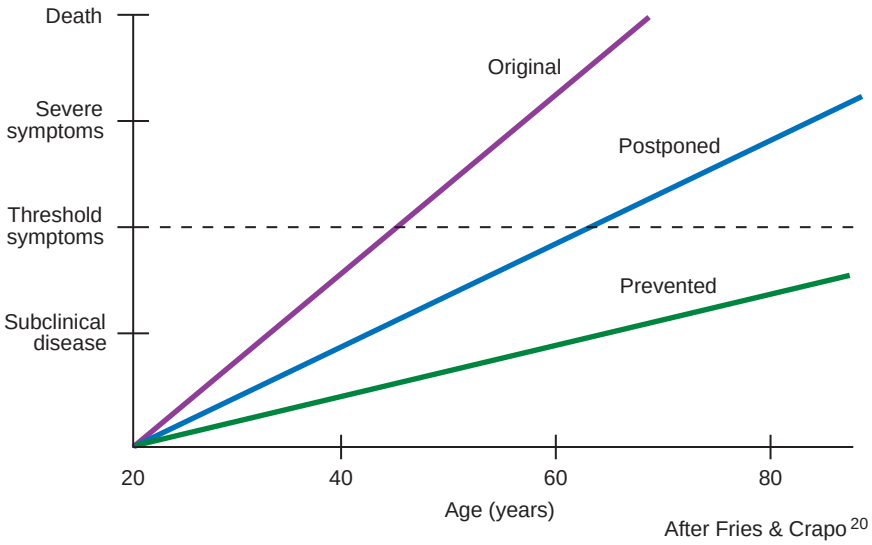
Our society has almost eliminated premature death. Diseases of the heart and the circulation, and cancers are now the leading causes of death. The reason for this is not an increase or an epidemic; it is a result of our success in virtually eliminating infectious diseases. Now the major determinant is chronic disease, affected by genetics, lifestyle, the environment, and aging itself. The major achievement left to be accomplished is in cardiovascular diseases. But even if cancer, diabetes, and all circulatory diseases were totally eliminated, life expectancy would not exceed 90 years.[18]

J.F. Fries described three eras in health and disease.[32] The first era existed until sometime in the early 1900s, and was characterized by acute infectious diseases. The second era, highlighted by cardiovascular diseases and cancer, is now beginning to fade into the third era, marked by problems of frailty (fading eyesight and hearing, impaired memory and cognitive function, decreased strength and reserve). Much of our medical approach is still based on the first era (find the disease and cure it), and now we have conditions that require a combination of medical, psychological, and social approaches. Our focus has been on age-dependent, fatal chronic diseases. The new challenge is with the nonfatal, age-dependent conditions, such as Alzheimer's disease, osteoarthritis, osteoporosis, obesity, and incontinence. It can be argued that health programs in the future should be evaluated by their impact on years free of disability, rather than on mortality.

The Concept of the Compression of Morbidity

Chronic illnesses are incremental in nature. The best health strategy is to change the slope, the rate at which illness develops, thus postponing the clinical illness, and if it is postponed long enough, effectively preventing it. There has been a profound change in public consciousness toward disease. Disease is increasingly seen as something not necessarily best treated by medication or surgery, but by prevention, or more accurately, by postponement.

Postponing illness was expressed by J.F. Fries as the ***compression of morbidity***.[20, 33] We would live relatively healthy lives and compress our illnesses into a short period of time just before death. Is this change really possible? A good affirmative example is the decrease in atherosclerosis in the U.S. Reasons include changes in the use of saturated fat, more effective detection and treatment of hypertension, increased exercise, and decreased smoking.



After Fries & Crapo[20]

Physician smokers have declined from a high of 79% to a small minority.[34] It is interesting, and amusing, to note that the greatest decrease has been among pulmonary surgeons, not surprising, while the least decrease has been among proctologists. From the mid 1970s to the early 1990s, smoking among physicians in the U.S. declined from 18.8% to 3.3%. Unfortunately, that still amounted to approximately 18,000 physicians who smoke. Approximately 35% of people in the U.S. who have not obtained a high school diploma are smokers, but only 12% of those with higher education are smoking, only 5.7% of those with

graduate degrees. Currently, approximately 23% of men and 18% of women are smokers.[17] Cigarette smoking among high school students peaked in 1997, then declined to the current level of 20%.[17] In addition, 14% of high school students smoke cigars and 8% use chewing tobacco. The use of chewing tobacco, pipe smoking, and cigars contributes significantly to morbidity and mortality. Tobacco, therefore, continues to be the single most preventable cause of premature illness and death in the U.S. It is important to note that smoking has a greater adverse effect on women compared with men.[35] Women who smoke only 1 to 4 cigarettes per day have a 2.5-fold increased risk of fatal coronary heart disease.[36]

Physicians and older patients may be skeptical that quitting smoking after decades of smoking could be beneficial, but the effects are at least partly reversible within 1 to 5 years after quitting. In the Nurses' Health Study, 61% of the excess risk of coronary heart disease mortality and 42% of stroke mortality was eliminated within 5 years after quitting smoking.[37] The improvement in respiratory disease mortality is slower, and a small increased risk of lung cancer mortality persists even after 30 years. However, by 20 years after cessation, all the excess risk of vascular mortality and death due to respiratory diseases other than lung cancer reached the level of a never smoker. Even older patients who already have coronary artery disease have improved survival if they quit smoking.[38] No matter how old you are, if you continue to smoke, you have an increased relative risk of death. But no matter how old you are, if you quit smoking, your risk of death decreases. Nevertheless, the risk of lung cancer remains elevated even in long-term ex-smokers.[39]

Since 1970, the death rate from coronary heart disease has declined approximately 50% in the U.S. Between 1973 and 1987 in the U.S., cardiovascular mortality declined in nearly every age group. In the combined age groups up to 54 years, cardiovascular mortality decreased 42%, and in people 55 to 84 years old, 33%.[35] Despite our progress, we must continue to exert preventive efforts on the risk factors associated with cardiovascular disease, especially obesity, hypertension, and lack of physical activity.

The effort to improve the quality of life has an important value to society; it will decrease the average number of years that people are disabled and a liability. Frailty and disability are now major health and social problems of society. Most significantly, this is a major financial challenge for health care systems and social programs. With evolution toward a rectangular society, the ratio of beneficiaries to taxpayers grows rapidly, jeopardizing the financial support for health and social programs. Compression of morbidity is at least one attractive solution to this problem.

Menopause as an Opportunity

Clinicians who interact with women at the time of the menopause have a wonderful opportunity and, therefore, a significant obligation. Medical intervention at this point of life offers women years of benefit from preventive health care. This represents an opportunity that should be seized.

It is logical to argue that health programs should be directed to the young. It makes sense to create good lifelong health behavior. While not underrating the importance of good health habits among the young, we would argue that the impact of teaching preventive care is more observable and more tangible at middle age. The prospects of limited mortality and the morbidity of chronic diseases are viewed with belief, understanding, and appreciation during these older years. The chance of illness is higher, but the impact of changes in lifestyle is greater.

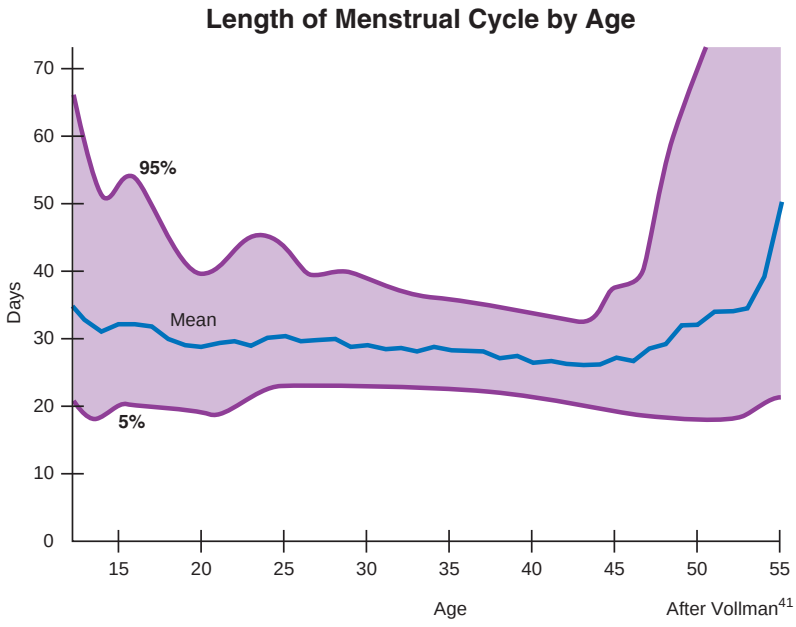
The Perimenopausal Transition

Definition of the Perimenopausal Transition

There is only one marker, menstrual irregularity, that can be used to objectively define and establish what is called the perimenopausal transition. This irregularity will be perceived by patients as skipped menstrual periods or longer durations (about 40 to 60 days) between periods.[40] There is no universal pattern; each woman will perceive a change that is her own individual characteristic alteration.

The ***menopause*** is that point in time when permanent cessation of menstruation occurs following the loss of ovarian activity. Menopause is derived from the Greek words *men* (month) and *pausis* (cessation). The years prior to menopause that encompass the change from normal ovulatory cycles to cessation of menses are known as the ***perimenopausal transitional*** years, marked by irregularity of menstrual cycles. ***Climacteric,*** an older, more general, and less precise term, indicates the period of time when a woman passes from the reproductive stage of life through the perimenopausal transition and the menopause to the postmenopausal years. Climacteric is from the Greek word for ladder.

Menstrual cycle length is determined by the rate and quality of follicular growth and development, and it is normal for the cycle to vary in individual women. Informative data come from two seminal longitudinal studies (with very similar results): the study of Vollman of more than 30,000 cycles recorded by 650 women and the study of Treloar of more that 25,000 woman-years in a little over 2,700 women.[41, 42] The observations of Vollman and Treloar documented a normal evolution in length and variation in menstrual cycles.



Menarche is followed by approximately 5–7 years of relatively long cycles at first, and then there is increasing regularity as cycles shorten to reach the usual reproductive age pattern. In the 40s, cycles begin to lengthen again. The highest incidence of anovulatory cycles is under age 20 and over age 40.[43,44] At age 25, over 40% of cycles are between 25 and 28 days in length; from 25 to 35, over 60% are between 25 and 28 days. The perfect 28-day cycle is indeed the most common mode, but it totaled only 12.4% of Vollman's cycles.

Overall, approximately 15% of reproductive-age cycles are 28 days in length. Only 0.5% of women experience a cycle less than 21 days long, and only 0.9% a cycle greater than 35 days.[45] Most women have cycles that last from 24 to 35 days, but at least 20% of women experience irregular cycles.[46]

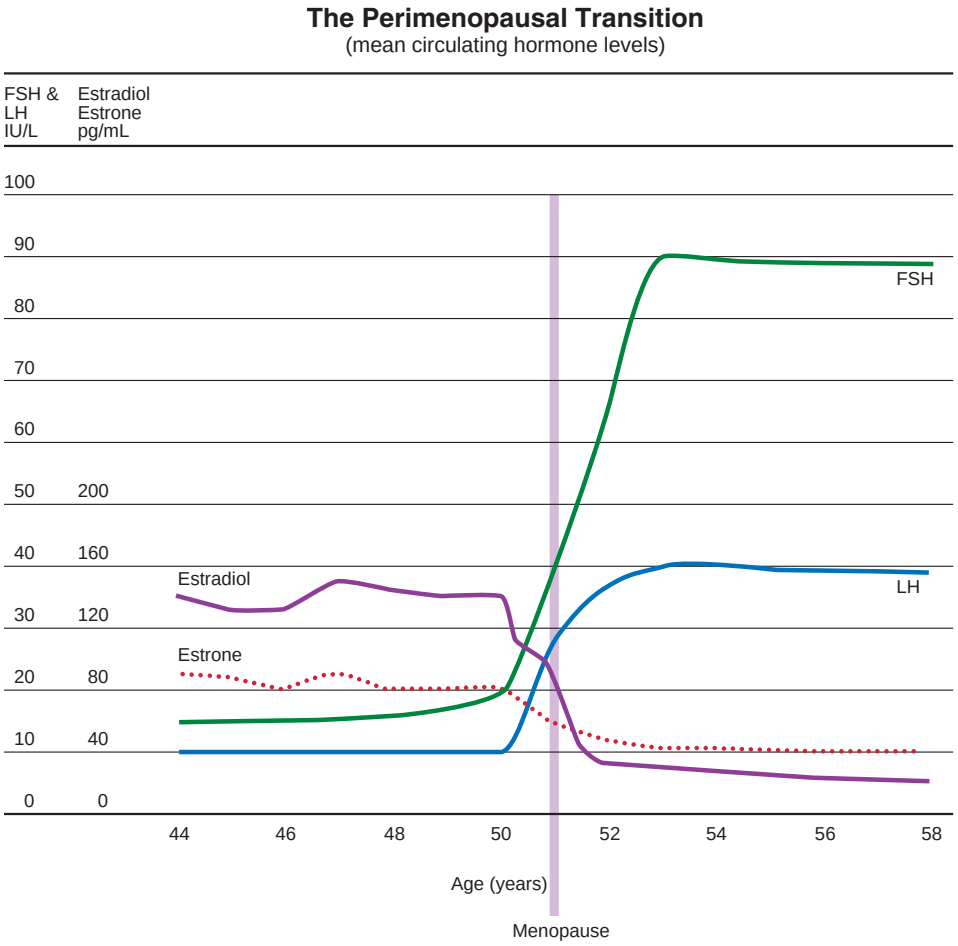
When women are in their 40s, anovulation becomes more prevalent, and prior to anovulation, menstrual cycle length increases, beginning 2 to 8 years before menopause.[42] Cycles greater than 40 days in length are prevalent in the year before menopause.[47] In an Australian longitudinal study, when cycle length exceeded 42 days, menopause predictably followed within 1 or 2 years.[48] This period of longer cycles uniformly precedes menopause no matter the age when menses cease, whether menopause is early or late.[49] The duration of the follicular phase is the major determinant of cycle length.[50, 51] This menstrual cycle change prior to menopause is marked by elevated follicle-stimulating hormone (FSH) levels and decreased levels of inhibin, but normal levels of luteinizing hormone (LH) and slightly elevated levels of estradiol.[52–58] ***Most importantly, even irregular cycles with long intervals (greater than 50–60 days) can be ovulatory, as many as 25%, meaning that late perimenopausal women can be at risk for pregnancy.***[59]

In the average woman, continuing follicular depletion and declining fertility begin at age 37–38, and menopause follows approximately 13 years later (average age 51). However, in epidemiologic studies approximately 10% of women in the general population become menopausal by the age of 45,[60, 61] probably because they were born with a smaller than normal ovarian follicular pool that is functionally depleted at an earlier age. Menopause occurs when the number of remaining follicles falls below a critical threshold, about 1,000, regardless of age.

Contrary to older belief (based on the report by Sherman et al., in 1976[50]), ***estradiol levels do not gradually wane in the years before the menopause, but remain in the normal range, although slightly elevated, until about 1 year before follicular growth and development cease.*** The Sherman et al. data were from a small cross-sectional study of one cycle collected from only 8 women, ages 46–56. More recent longitudinal studies of women as they pass through the perimenopausal transition reveal that estrogen levels do not begin a major decline until about a year before menopause.[56, 62, 63] Indeed, women experiencing the perimenopausal transition actually have higher overall estrogen levels, a response that is logically explained by an increased ovarian follicular reaction to the increase in FSH secretion during these years.[64] Variability in estrogen levels is characteristic of the perimenopausal transition, with greater variability observed in menstrual cycles that display greater irregularity.[65]

As noted, most women experience a 2- to 8-year period of time prior to menopause when anovulation becomes common.[42] During this period of time ovarian follicles continue their rate of loss until eventually the supply of follicles is finally depleted.[66,67] In a study of human ovaries, the loss that began when the total number of follicles reached approximately 25,000, usually at age 37–38, correlated with a subtle but real increase in FSH and decrease in inhibin.[68] These changes, including the increase in FSH, reflect the reduced *quantity* of aging follicles, and their reduced secretion of inhibin, the granulosa cell product that exerts an important negative feedback influence over FSH secretion by the pituitary gland. It is possible that both inhibin-A and inhibin-B may be involved, because luteal-phase levels of inhibin-A and follicular-phase levels of inhibin-B decrease with aging and antedate the rise in FSH.[69–71] A careful study in Australia, however, indicated that the increase in FSH was correlated only with a decrease in inhibin-B, and in response, estradiol concentrations increased slightly.[62]

Decreasing inhibin production could reflect either a shrinking number of follicles, or a reduced functional capacity of older follicles, or both.[72] ***The observation that preovulatory follicular fluid inhibin concentrations are similar in young and older cycling women suggests that the number of remaining follicles is the most important factor.***[73]



As FSH levels increase and the follicular phase becomes shorter, estradiol levels rise earlier, suggesting that higher FSH levels stimulate more rapid follicular development.[74] ***Careful studies indicated that the earlier acute rise in estradiol levels results from advanced follicular development at the beginning of the cycle and earlier selection of the dominant follicle.***[75, 76] Follicular phase and overall cycle length reach their nadir at approximately age 42. Over the subsequent 8–10 years preceding the menopause, average cycle length and variability steadily increase as ovulations become less regular and less frequent.[41] The age-related changes in the endocrine characteristics of the menstrual cycle that result from progressive follicular depletion correlate with a measurable decrease in ovarian volume and in the number of antral follicles observed by transvaginal ultrasonography during the early follicular phase.[77–83]

The inverse and tight relationship between FSH and inhibin indicates that inhibin is a sensitive marker of ovarian follicular competence and, in turn, that FSH measurement is a clinical assessment of inhibin.[53, 54] The decrease in inhibin secretion by the ovarian follicles begins early (around age 35), but accelerates after 40 years of age. This is reflected in the decrease in fecundity that occurs with aging (as discussed in Chapter 27). ***Furthermore, the ineffective ability to suppress gonadotropins with postmenopausal hormone therapy is a consequence of the loss of inhibin, and for this reason FSH cannot be used clinically to titer estrogen dosage in postmenopausal hormone therapy.***

The Michigan Bone Health and Metabolism Study is a longitudinal, assessment of the perimenopausal transition in a cohort of 629 women initiated in 1992–1993. The initial rise in FSH in these women was modest until 7 years prior to menopause, then accelerated with

an even greater increase in the 2 years before menopause, finally reaching a plateau about a year after menopause.[84] The major decrease in estradiol levels began about 2 years before menopause.[85] Declining levels of inhibin B and anti-müllerian hormone (AMH) reached a low to nondetectable point about 5 years before menopause.[86] Although the inhibin B and AMH results are in general agreement with other reports, the exactness of the timing is limited by the fact that the blood samples were obtained from only 50 women in the study. Nevertheless, the Michigan study confirms the validity of AMH as a marker for the ovarian reserve of follicles. Unlike inhibin B, AMH is not a participant in the feedback relationship between the ovary and the pituitary gonadotropins, rather AMH, a product of granulosa cells, reflects the number of follicles present in the ovaries awaiting FSH stimulation.[87] ***The variability in these measurements from individual to individual, however, precludes the practical use of these tests to predict with accuracy the future date of menopause.***

The perimenopausal years are a time period during which postmenopausal levels of FSH (greater than 20 IU/L) can be seen despite continued menstrual bleeding, while LH levels still remain in the normal range. Occasionally, corpus luteum formation and function occur, and the perimenopausal woman is not safely beyond the risk of an unplanned and unexpected pregnancy until elevated levels of both FSH (>20 IU/L) and LH (>30 IU/L) can be demonstrated.[55] However, even under these circumstances, fluctuations can occur, with a period of ovarian failure followed by resumption of ovarian function.[54] ***Because variability is the rule, it would be wise to recommend the use of contraception until the postmenopausal state is definitely established.*** According to the *Guinness Book of World Records*, a woman from Portland, Oregon, holds the modern record for the oldest spontaneous pregnancy, conceiving when 57 years and 120 days old. ***Several months of amenorrhea together with an FSH level of 40 IU/L or more are reliable signals that menopause is either near or already passed.***[88]

In the longitudinal Massachusetts Women's Health Study, women who reported the onset of menstrual irregularity were considered to be in the perimenopausal period of life.[89] The median age for the onset of this transition was 47.5 years. ***Only 10% of women ceased menstruating abruptly with no period of prolonged irregularity.*** The perimenopausal transition from reproductive to post-reproductive status was, for most women, approximately 4 years in duration. In the study by Treloar, the average age for entry into the perimenopausal transition was 45.1, and the age range that included 95% of the women was 39–51.[60] The mean duration of the perimenopausal transition was 5.0 years, with a range of 2 to 8 years.

The Perimenopausal Transition[42, 60, 89]
Average age of onset – 46
Age of onset for 95% of women – 39 to 51
Average duration – 5 years
Duration for 95% of women – 2 to 8 years

Preventive Health Screening of Healthy Perimenopausal Women

The most important contribution a clinician can provide to the perimenopausal woman is the education she needs and desires to make therapeutic choices. This early educational process will help to build a solid relationship with patients, a relationship they will want to

continue as they age. The following recommendations are derived from our own clinical experience:

- Provide guidance and education to facilitate a patient's decision making.
- Provide time and an appropriate location for sensitive and uninterrupted discussions.
- Use educational materials, especially handouts, but also explain them using your own words.
- Involve family members during counseling and educational visits.
- Be accessible. Consider designating a member of your staff as the menopause resource person. Encourage phone calls and emails.
- Be involved in community and hospital educational programs for the public.
- Use an effective, well-trained counselor for patients who need in-depth help in coping with life's trials and tribulations.

Preventive intervention during the perimenopausal years has three major goals. The overall objective is to prolong the period of maximal physical energy and optimal mental and social activity. A specific goal is to detect as early as possible any of the major chronic diseases, including hypertension, heart disease, diabetes mellitus, and cancer, as well as impairments of vision, hearing, and teeth. Finally, the clinician should help perimenopausal women to smoothly traverse the menopausal period of life. Preventive health care and management of the later reproductive years give clinicians an excellent opportunity to function as a woman's primary care provider.

SUMMARY—Preventive Health Screening of Healthy Postmenopausal Women

1. A complete medical history and physical examination should be performed every 5 years, at about age 40, 45, 50, and 55.

2. Annual visits should include a breast and pelvic examination (including a rectovaginal examination), recording of the body mass index (BMI), screening for sexually transmitted infections when appropriate, and a TSH assessment in the 40s and every 2 years beginning at age 60. Hypothyroidism increases with aging and is more common in women (Chapter 20).

3. Recording body height will detect any decrease associated with early osteoporosis. Bone mass should be measured in postmenopausal women who present with fractures, who have one or more risk factors for osteoporosis, or who are over age 65.

4. Annual screening mammography should begin at age 40 (discussed in Chapter 16).

5. At each visit, appropriate testing is scheduled for specific chronic conditions (including abnormal lipids), indicated immunizations are provided, and counseling covers changing nutritional needs, physical activities, injury prevention, occupational, sexual, marital, and parental problems, urinary function, and use of tobacco, alcohol, and drugs. Stool hemoccult testing should be performed annually after age 50.

6. Colonoscopy is recommended at ages 50 and 55, and if results are negative and there is no family history of colorectal cancer, colonoscopy need not be repeated.

The Age of Menopause

Designating the average age of menopause has been somewhat difficult. Based on cross-sectional studies, the median age was estimated to be somewhere between 50 and 52.[90] These studies relied on retrospective memories and the subjective vagaries of the individual being interviewed. Until recently, studies with longitudinal follow-up to observe women and record their experiences as they pass through menopause were hampered by relatively small numbers. The Massachusetts Women's Health Study provides us with data from 2,570 women.[89]

The median age for menopause in the Massachusetts Study was 51.3 years. Only current smoking could be identified as a cause of earlier menopause, a shift of approximately 1.5 years. Those factors that did not affect the age of menopause included the use of oral contraception, socioeconomic status, and marital status. Keep in mind that a median age of menopause means that only half the women have reached menopause at this age. In the classic longitudinal study by Treloar, the ***average*** age of menopause was 50.7, and the range that included 95% of the women was 44 to 56.[91] In a survey in the Netherlands, the average age of menopause was 50.2, and in an Italian longitudinal study, 50.9.[61, 92]

The Study of Women's Health Across the Nation (SWAN) is an ongoing, national study, recording the health of American women as they pass through the perimenopausal transition (http://www.edc.gsph.pitt.edu/swan/). The study began in 1994 in seven research centers and enrolled 3,302 participants with five racial/ethnic groups and a variety of backgrounds for an initial cross-sectional survey. In 1996, these women began a longitudinal, follow-up study with extensive data collection occurring annually.

In the SWAN study, the median age of menopause was 51.4, with an earlier onset associated with current smoking, lower education, and lower socio-economic status, whereas a later age was associated with parity and prior use of oral contraceptives.[93] In contrast, a Dutch study concluded that prior use of oral contraceptives was associated with an earlier (less than 1 year) menopause.[94] About 1% of women have been reported to experience menopause before the age of 40.[95] The SWAN study reported a similar percentage of 1.1%, with a slightly higher rate in black and Hispanic women and a lower rate of 0.5% in Chinese women and 0.1% in Japanese women.[96] Hispanic women experienced menopause about 6 months earlier compared with other ethnic groups, whereas Japanese women were about 3 months later.

Two large cohorts of European women reported average ages of menopause in various countries that centered around age 51, slightly higher in Northern Europe and slightly lower in Southern Europe.[97] Some countries, like India, report an average age of menopause as much as 5 years earlier.[98] In epidemiologic studies, approximately 10% of women in the general population become menopausal by the age of 45.[60, 61] Pedigree analysis has revealed that the genetic features of early menopause (age 40–45) and premature ovarian failure are similar and suggest a dominant pattern of inheritance through maternal or paternal relatives.[99, 100] There are two studies indicating that daughters of mothers with an early menopause (before age 46) also have an early menopause.[101–103]

There is sufficient evidence to believe that undernourished women and vegetarians experience an earlier menopause.[101, 104] Because of the contribution of body fat to estrogen production, thinner women experience a slightly earlier menopause.[105] Frequent consumption of alcohol is associated with a later menopause.[102] This is consistent with the reports that women who consume alcohol have higher blood and urinary levels of estrogen, and greater bone density.[106–110]

In multiple studies, there has been no correlation between age of menarche and age of menopause, with the exception of one Swedish study concluding that an earlier menarche

and earlier menopause go together.[61, 91, 101, 111, 112] In most studies, race, parity, and height have no influence on the age of menopause; however, three cross-sectional studies found later menopause to be associated with increasing parity.[61, 89, 93, 101, 105] Two studies found that irregular menses among women in their early 40s predicts an earlier menopause.[113, 114] A French survey detected no influence of heavy physical work on early menopause (before age 45).[115] An earlier menopause has been reported to be associated with living at high altitudes.[116, 117] And most intriguing, an earlier age of menopause has been reported in left-handed women compared with right-handed women.[118, 119] Finally, earlier menopause is associated with growth retardation in late gestation.[120]

It has been argued that premature ovarian failure can occur in women who have previously undergone abdominal hysterectomy or endometrial ablation, presumably because ovarian vascular flow has been compromised, but the only prospective study could find no elevations of FSH within the first 2 years after surgery.[121–123]

Multiple studies have consistently documented that an earlier menopause (an average of 1.5 years earlier) is a consequence of smoking. There is a dose-response relationship with the number of cigarettes smoked and the duration of smoking.[124, 125] Even former smokers show evidence of an impact.[97]

Unlike the decline in age of menarche that occurred with an improvement in health and living conditions, most historical investigation indicates that the age of menopause has changed little since early Greek times.[126, 127] Others (a minority) have disagreed, concluding that the age of menopause did undergo a change, starting with an average age of about 40 years in ancient times, and in Sweden an increase of about 1 year over the last 80 years.[112, 128] If there has been a change, however, history indicates it has been minimal. Even in ancient writings, an age of 50 is usually cited as the age of menopause.

Sexuality and Menopause

Sexuality is a lifelong behavior with evolving change and development. It begins with birth (maybe before) and ends with death. The notion that it ends with aging is inherently illogical. The need for closeness, caring, and companionship is lifelong. Old people today live longer, are healthier, have more education and leisure time, and have had their consciousness raised in regard to sexuality.

Younger people, especially physicians, underrate the extent of sexual interest in older people. In a random sample of women aged 50 to 82 in Madison, Wisconsin, nearly one-half of the women reported an ongoing sexual relationship.[129] In the Duke longitudinal study on aging, 70% of men in the 67 to 77 age group were sexually active, and 80% reported continuing sexual interest, while 50% of all older women were still interested in sex.[130] In the Postmenopausal Estrogen-Progestin Interventions (PEPI) trial, 60% of women 55–64 years old were sexually active.[131] In a national sample of American men and women, the prevalence of sexual behavior declined with aging; however, 26% of individuals age 75 to 85 years were still sexually active.[132] Therefore a significant number of postmenopausal women are sexually active, and only a relatively small percentage complain of sexual problems. The prevalence of self-reported sexual problems peaks in middle-aged women, sufficient to cause distress in about 22% of U.S. women, and about 12% of women aged 45 to 64.[133]

The decline in sexual activity with aging is influenced more by culture and attitudes than by nature and physiology (or hormones). The two most important influences on older sexual interaction are the strength of a relationship and the physical condition of each

partner.[131, 132, 134] The single most significant determinant of sexual activity for older women, therefore, is the unavailability of partners due to divorce and the fact that women are outliving men. Given the availability of a partner, the same general high or low rate of sexual activity can be maintained throughout life.[5, 135] Longitudinal studies indicate that the level of sexual activity is more stable over time than previously suggested.[136–138] Individuals who are sexually active earlier in life continue to be sexually active into old age. However, aging is associated with a decline in sexual function in many women, and this has been documented in the menopausal transition.[139, 140] A significant component of this decline can be attributed to menopausal symptoms associated with decreasing estrogen levels, a problem that is easily ameliorated by estrogen treatment.

There are two main sexual changes in the aging woman. There is a reduction in the rate of production and volume of vaginal lubricating fluid, and there is some loss of vaginal elasticity and thickness of the epithelium. Less vaginal atrophy is noted in sexually active women than in inactive women; presumably the activity maintains vaginal vasculature and circulation. The dyspareunia associated with postmenopausal urogenital atrophy includes a feeling of dryness and tightness, vaginal irritation and burning with coitus, and postcoital spotting and soreness. Of course, these changes are effectively prevented by estrogen treatment. Indeed, estrogen therapy has a positive impact on sexuality beyond its effects on vaginal tissue.[131] In an Australian study assessing changes in sexual functioning during the perimenopausal and menopausal transition, a correlation with a decline in sexuality was demonstrated with estradiol levels, but not with testosterone levels.[141] However, the prior level of sexual activity and the partner status and relationship were more important factors than hormone levels in determining midlife sexual function during the perimenopausal and menopausal transition.[142]

Illness and Sex

It is not uncommon to encounter women who have had surgery that affects sexuality. The list includes vulvectomy and surgery of the breast. Sexual problems are not limited, however, to surgical procedures and illnesses of the genitalia. Altered self-image can occur with diseases of any site; however, studies have not found hysterectomy to have a detrimental impact on sexuality.[131, 143]

Sexual counseling, to be effective, must be provided to couples both before and after surgery. It is not unexpected that the surgeon may not be fully capable of providing this counseling. A major contribution from an older woman's primary clinician is to arrange for competent and experienced sexual counseling. Unfortunately, most physicians operate on the principle that if no questions are raised there is no problem. The expert surgeon should be grateful for the help of experts in psychosexual therapy. Seek out the potential for post treatment sexual morbidity before the surgery. Assess the patient's abilities for coping and her sense of body image. Consider the quality of the patient's relationship, and be sensitive to the absence of a relationship. This entire effort may take some time. The normal state of presurgical anxiety, fear, and denial hampers good communication.

Antihypertensive agents are frequently responsible for male sexual dysfunction, but little information is available regarding female sexual function. However, remember that vaginal lubrication is the female counterpart to the male erection, and, therefore, vaginal dryness is a likely consequence. Adrenergic blocking agents are especially noted to affect libido and potency in men. Similarly, psychotropic drugs of all categories have been associated with inhibition of sexual function. Finally, one should always suspect alcoholism when patients complain of sexual dysfunction. Androgen treatment for decreased sexuality is discussed in Chapter 18.

Hormone Production After Menopause

Shortly after the menopause, one can safely say that there are no remaining ovarian follicles.[63, 144] Eventually there is a 10–20-fold increase in FSH and approximately a 3-fold increase in LH, reaching a maximal level 1–3 years after menopause, after which there is a gradual, but slight, decline in both gonadotropins.[145, 146] Elevated levels of both FSH and LH at this time in life are conclusive evidence of ovarian failure. FSH levels are higher than LH because LH is cleared from the blood so much faster (initial half-lives are about 20 min for LH and 3–4 h for FSH), and perhaps because there is no specific negative feedback peptide for LH like inhibin. The age-related decline in gonadotropin levels in the latter years of postmenopausal life is believed to reflect aging of the pituitary gonadotropin-secreting cells, specifically a decrease in the ability to respond to gonadotropin releasing-hormone (GnRH).

After menopause, the ovary secretes primarily androstenedione and testosterone, but the circulating level of androstenedione after menopause is about one-half that seen prior to menopause.[147] Most of this postmenopausal androstenedione is derived from the adrenal gland, with only a small amount secreted from the ovary, even though androstenedione is the principal steroid secreted by the postmenopausal ovary.[148, 149] Dehydroepiandrosterone (DHEA) and its sulfate (DHEAS), originating in the adrenal gland, decline markedly with aging; in the decade after menopause the circulating levels of DHEA are approximately 70% less and levels of DHEAS are about 74% less than the levels in young adult life.[150, 151]

Testosterone production decreases by approximately 25% after menopause, but the postmenopausal ovary in most, but not all women, secretes more testosterone than the premenopausal ovary, at least in the first years of the postmenopausal period. With the disappearance of follicles and estrogen, the elevated gonadotropins drive the remaining tissue in the ovary to a level of increased testosterone secretion. The ovarian cells of origin are uncertain; presumably the steroidogenic tissue is that which has accumulated from ovarian follicles undergoing atresia because stromal cells believed to be of mesenchymal origin lack steroidogenic capability.[152] Suppression of gonadotropins with gonadotropin-releasing hormone (GnRH) agonist or antagonist treatment of postmenopausal women results in a significant decrease in circulating levels of testosterone, indicating the gonadotropin-dependent postmenopausal ovarian origin.[153–155]

The total amount of testosterone produced after menopause, however, is decreased because the amount of the primary source, peripheral conversion of androstenedione, is reduced. The early postmenopausal circulating level of androstenedione decreases approximately 62% from young adult life.[150] The menopausal decline in the circulating levels of testosterone is not great, from no change in many women to as much as 15% in others.[56, 146, 150, 156, 157] In an excellent longitudinal Australian study from 5 years before menopause to 7 years after menopause, the circulating levels of testosterone did not change.[151] Indeed, because of a decrease in sex hormone-binding globulin, this Australian study calculated an increase in free androgens.

Later in the postmenopausal years, the circulating androgen levels are nearly all, if not all, derived from the adrenal gland. A careful study could detect no circulating androgens in postmenopausal women (averaging 12 years distant from menopause) with complete adrenal insufficiency, and no intraovarian testosterone or androstenedione.[158]

The circulating estradiol level after menopause is approximately 10–20 pg/mL, most of which is derived from peripheral conversion of estrone, which in turn is mainly derived from the peripheral conversion of androstenedione.[147, 160, 161] The circulating level of estrone

Blood Production Rates of Steroids[159]			
	Reproductive Age	*Postmenopausal*	*Oophorectomized*
Androstenedione	2–3 mg/day	0.5–1.5 mg/day	0.4–1.2 mg/day
Dehydroepiandrosterone	6–8	1.5–4.0	1.5–4.0
Dehydroepiandrosterone sulfate	8–16	4–9	4–9
Testosterone	0.2–0.25	0.05–0.18	0.02–0.12
Estrogen	0.350	0.045	0.045

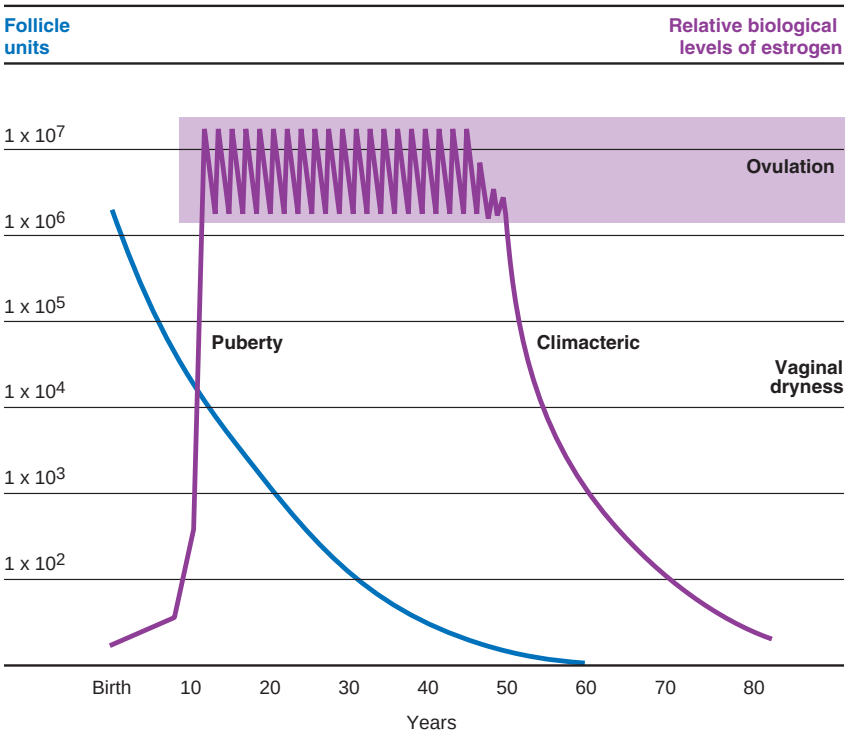
Changes in Circulating Hormone Levels at Menopause[56, 147, 160]		
	Premenopause	*Postmenopause*
Estradiol	40–400 pg/mL	10–20 pg/mL
Estrone	30–200 pg/mL	30–70 pg/mL
Testosterone	20–80 ng/dL	15–70 ng/dL
Androstenedione	60–300 ng/dL	30–150 ng/dL

in postmenopausal women is higher than that of estradiol, approximately 30–70 pg/mL. The average postmenopausal production rate of estrogen is approximately 45 μg/24 h, almost all, if not all, being estrogen derived from the peripheral conversion of androgens. The androgen/estrogen ratio changes drastically after menopause because of the more marked decline in estrogen, and an onset of mild hirsutism is common, reflecting this marked shift in the sex hormone ratio. With increasing postmenopausal age, a decrease can be measured in the circulating levels of dehydroepiandrosterone sulfate (DHEAS) and dehydroepiandrosterone (DHEA), whereas the circulating postmenopausal levels of androstenedione, testosterone, and estrogen remain relatively constant.[146, 147]

Estrogen production by the ovaries does not continue beyond the menopause; however, estrogen levels in postmenopausal women can be significant, principally due to the extraglandular conversion of androstenedione and testosterone to estrogen. The clinical impact of this estrogen varies from one postmenopausal woman to another, depending on the degree of extraglandular production, modified by a variety of factors.

The percent conversion of androstenedione to estrogen correlates with body weight. Increased production of estrogen from androstenedione with increasing body weight is probably due to the ability of fat to aromatize androgens. This fact and a decrease in the levels of sex hormone-binding globulin (which results in increased free estrogen concentrations) contribute to the well-known association between obesity and the development of endometrial cancer. Body weight, therefore, has a positive correlation with the circulating levels of estrone and estradiol.[147] Aromatization of androgens to estrogens is not limited to adipose tissue, however, because almost every tissue tested has this activity.

Eventually, the ovarian steroidogenic tissue is exhausted and, despite huge reactive increments in FSH and LH, no further steroidogenesis of importance results from gonadal activity. The postmenopausal ovary weighs less than 10 g, but it can be visualized by ultrasonography.[162] With increasing age, the adrenal contribution of precursors for estrogen production proves inadequate. In this final stage of estrogen availability, levels are insufficient to sustain secondary sex tissues.



In summary, the symptoms frequently seen and related to decreasing ovarian follicular competence and then estrogen loss in this protracted climacteric are:

1. Disturbances in menstrual pattern, including anovulation and reduced fertility, decreased flow or hypermenorrhea, irregular frequency of menses, and then, ultimately, amenorrhea.

2. Vasomotor instability (hot flushes and sweats).

3. Atrophic conditions: atrophy of vaginal epithelium; formation of urethral caruncles; dyspareunia and pruritus due to vulvar, introital, and vaginal atrophy; general skin atrophy; urinary difficulties such as urgency and abacterial urethritis and cystitis.

4. Health problems secondary to long-term deprivation of estrogen: the consequences of osteoporosis and cardiovascular disease.

A precise understanding of the symptom complex the individual patient may display is often difficult to achieve. Some patients experience severe multiple reactions that may be disabling. Others show no reactions, or minimal reactions that go unnoticed until careful medical evaluation.

It is helpful to classify the hormonal problems in three categories:

1. Those associated with relative estrogen excess such as dysfunctional uterine bleeding, endometrial hyperplasia, and endometrial cancer.

2. Those associated with estrogen deprivation such as flushes, atrophic vaginitis, urethritis, and osteoporosis.

3. Those associated with hormone therapy (Chapter 18).

Problems of Estrogen Excess

Exposure to Unopposed Estrogen

Throughout the perimenopausal period, there is a significant incidence of dysfunctional uterine bleeding. In the SWAN study, about 20% of cycles even early in the perimenopausal transition were anovulatory, associated with shorter intervals at the beginning of the transition and longer intervals later.[163] Irregular bleeding was most often due to anovulation, whereas heavy menstrual bleeding was associated with obesity and uterine abnormalities.

Although the greatest concern provoked by this symptom is endometrial neoplasia, the usual finding is non-neoplastic tissue displaying estrogen effects unopposed by progesterone. This results from anovulation in premenopausal women and from extragonadal endogenous estrogen production or estrogen administration in postmenopausal women. There are 4 mechanisms that could result in increased endogenous estrogen levels:

1. Increased precursor androgen (functional endocrine tumors, liver disease, stress).
2. Increased aromatization (obesity, hyperthyroidism, and liver disease).
3. Increased direct secretion of estrogen (ovarian tumors).
4. Decreased levels of SHBG (sex hormone-binding globulin) leading to increased levels of free estrogen.

In all women, whether premenopausal or postmenopausal, whether on or off hormone therapy, specific organic causes (neoplasia, complications of unexpected pregnancy, or bleeding from extrauterine sites) must be ruled out. In addition to careful history and physical examination, dysfunctional uterine bleeding requires endometrial evaluation. Transvaginal ultrasonographic measurement of endometrial thickness can be utilized in postmenopausal women to avoid unnecessary biopsies.[164] In perimenopausal and postmenopausal women with abnormal bleeding, endometrial biopsy is considered unnecessary when the endometrial thickness is less than 5 mm because the risk of endometrial hyperplasia or cancer is remote.[165–167] Substantial evidence is lacking to support the application of this criterion to premenopausal women. ***We believe that biopsy is unnecessary in perimenopausal women when the endometrial thickness is less than 5 mm, that biopsy is indicated when the clinical history suggests long-term unopposed estrogen exposure even when the endometrial thickness is "normal" (5–12 mm), and that biopsy should be performed when endometrial thickness is greater than 12 mm even when clinical suspicion of disease is low.***

If the uterus is normal on examination, for reasons of both accuracy and cost-effectiveness, the method of biopsy should be an office aspiration curettage, ***NOT*** the older, more costly and risky, in-hospital dilation and curettage (D&C). We recommend the use of a plastic endometrial suction device. It is easy to use, requires no cervical dilation, and is frequently painless. This device is as efficacious as older more painful techniques. Insertion should first be attempted without the use of a tenaculum. In many patients, this is feasible and avoids the sensation of the tenaculum grasping the cervix. Once the suction is applied, the endometrial cavity should be thoroughly curetted in all directions, just as one would with a sharp curette during a D&C. If the cannula fills up with tissue, a second and even a third cannula should be inserted until tissue is no longer obtained. Although most patients report no problems with cramps or pain, the application of suction in some patients stimulates cramping that usually passes within 5–10 minutes. Because cramping occurs in such a small minority of patients, it is not our practice to routinely give an inhibitor of prostaglandin synthesis. For repeat biopsies, in patients known to cramp, it is helpful to use such an agent at least 20 min before the procedure.

Less than 10% of postmenopausal women cannot be adequately evaluated by office biopsy. Most commonly, the reason is the inability to enter the uterine cavity. In such instances, dilation and curettage (D&C) are in order; however pretreatment with misoprostol or cervical laminaria may avoid a D&C. ***Furthermore, if the uterus is not normal on pelvic examination (enlarged and irregular), the office endometrial biopsy must yield to D and C with hysteroscopy in order to achieve accuracy of diagnosis.***

If the vulva, vagina, and cervix appear normal on inspection, perimenopausal bleeding can be assumed to be intrauterine in origin. Confirmation requires the absence of abnormal cytology on the Pap smear. The principal symptom of endometrial cancer is abnormal vaginal bleeding, but carcinoma will be encountered in patients with bleeding in less than 3% of postmenopausal endometrial biopsies.[168–170] Normal endometrium is found over half the time, polyps in approximately 3%, endometrial hyperplasia about 15% of the time, and atrophic endometrium in the rest of patients with postmenopausal bleeding. Postmenopausal bleeding should always be taken seriously. Approximately 10% of patients who have benign findings at the initial evaluation subsequently develop significant pathology within 2 years.[169] ***The persistence of abnormal bleeding demands repeated evaluation.***

Additional procedures include the following:

Colposcopy and cervical biopsy for abnormal cytology or obvious lesions**.**
Endocervical assessment by curettage for abnormal cytology (the endocervix must always be kept in mind as a source for abnormal cytology).
Hysterogram, hysteroscopy, or sonohysterography with endometrial biopsy if bleeding persists to determine the presence of endometrial polyps or submucosal fibroids, and to rule out the presence of endometrial cancer.[171]

Keep in mind that the pathologic reading, "tissue insufficient for diagnosis," when a patient is on estrogen-progestin treatment, often represents atrophic, decidualized endometrium that yields little to the exploring curet. If the clinician is confident in his or her technique, knowing that a full investigation of the intrauterine cavity has been accomplished, then *as long as the patient does not persist in bleeding*, this reading can be interpreted as comforting and benign, the absence of pathology.

In the absence of organic disease, appropriate management of uterine bleeding is dependent on the age of the woman and endometrial tissue findings. In the perimenopausal woman with dysfunctional uterine bleeding associated with proliferative or hyperplastic endometrium (uncomplicated by atypia or dysplastic constituents), periodic oral progestin therapy is mandatory, such as 5–10 mg medroxyprogesterone acetate or 200 mg micronized progesterone given daily for the first 14 days of each month. If hyperplasia is present, follow-up aspiration curettage after 3–4 months is required, and if progestin is ineffective and histologic regression is not observed, formal curettage is an essential preliminary to alternate therapeutic surgical choices. Progestins can mask abnormal tissue, and, therefore, follow-up biopsy is best scheduled 3 months after progestin treatment. Because hyperplasia *with atypia* carries with it a risk of cancer (even invasive), hysterectomy is the treatment of choice. Persistence or progression of abnormal endometrium was observed in 28.4% of women with complex hyperplasia and in 26.9% of women with atypical hyperplasia despite treatment with a progestational agent.[172] However, progestational response was better with higher dosage and longer duration treatment. If treatment with a progestational agent is elected, we recommend a minimum duration of 3 to 6 months, with 20 mg medroxyprogesterone or 40 mg megesterol acetate daily.

When monthly progestin therapy reverses simple hyperplastic changes (which it does in 95–98% of cases) and controls irregular bleeding, treatment should be continued until withdrawal bleeding ceases. This is a reliable sign (in effect, a bioassay) indicating the onset of estrogen deprivation and the need for the addition of estrogen. If vasomotor disturbances

begin before the cessation of menstrual bleeding, a combined estrogen-progestin program can be initiated as needed to control the flushes.

If contraception is required, the healthy, nonsmoking patient with normal blood pressure should seriously consider the use of estrogen-progestin contraception. The anovulatory woman cannot be guaranteed that spontaneous ovulation and pregnancy will not occur. ***The use of a low-dose estrogen-progestin contraceptive will at the same time provide contraception and prophylaxis against irregular, heavy anovulatory bleeding and the risk of endometrial hyperplasia and neoplasia.***

Clinicians have often utilized a traditional postmenopausal hormone regimen to treat a woman with the kind of irregular cycles usually experienced in the perimenopausal years. This addition of exogenous estrogen without a contraceptive dose of progestin when a woman is not amenorrheic or experiencing menopausal symptoms is inappropriate and even risky (exposing the endometrium to excessively high levels of estrogen). ***And most importantly, a postmenopausal hormonal regimen does not inhibit ovulation and provide contraception.***[173] The appropriate response is to regulate anovulatory cycles with monthly progestational treatment along with an appropriate contraceptive method or to utilize low-dose estrogen-progestin contraception. An oral contraceptive that contains 20 μg estrogen provides effective contraception, improves menstrual cycle regularity, diminishes bleeding, and relieves menopausal symptoms.[174] Treatment with the transdermal or vaginal method of estrogen-progestin contraception (Chapter 23) would also be appropriate.

A common clinical dilemma is when to change from estrogen-progestin contraception to postmenopausal hormone therapy. It is important to change because even with the lowest estrogen dose contraceptive available, the estrogen dose is 4-fold greater than the standard postmenopausal dose, and with increasing age, the dose-related risks with estrogen become significant. One approach to establish the onset of the postmenopausal years is to measure the FSH level, beginning at age 50, on an annual basis, being careful to obtain the blood sample on day 6 or 7 of the estrogen-progestin-free week in a standard 3-week regimen (when steroid levels have declined sufficiently to allow FSH to rise). Friday afternoon works well for patients who start new estrogen-progestin treatment on Sunday. When FSH is greater than 20 IU/L, it is time to change to a postmenopausal hormone program. Because of the variability in FSH levels experienced by women around the menopause, this method is not always accurate.[175, 176] Indeed, in some women, FSH will not rise until 2 weeks after the last steroid contraception exposure. A 2-week wait is not very practical and places the patient at risk for an unwanted pregnancy. The treatment-free week method is practical and works for most women. Women who are dependent on contraceptives to prevent pregnancy can be allowed to enter their mid fifties on low-dose estrogen-progestin contraception, and then empirically switched to a postmenopausal hormone regimen. ***The empirical approach is necessary with patients using the newer extended-day or continuous dosing regimens of estrogen-progestin contraception.***

Because of the favorable impact of locally released progestin on the endometrium, the levonorgestrel IUS (intrauterine system) is very effective for the treatment of menorrhagia, as effective as the administration of oral progestins (with less side effects), and compares favorably with endometrial resection or ablation. [177–181] In addition, this IUD can be used to treat endometrial hyperplasia.[182–187] Comparison studies of endometrial hyperplasia indicate that the levonorgestrel IUS is as effective, and probably better than standard treatment with an oral progestin.[183, 188, 189] The levonorgestrel IUS may be associated with a slight increase in the formation of ovarian cysts, but they are asymptomatic and resolve spontaneously.[190]

In postmenopausal women, one must view any adnexal mass as cancer until proven otherwise. Surgical intervention is usually necessary, and appropriate consultation must be obtained not only for the surgical procedure but also for suitable preoperative evaluation

and preparation. Nonpalpable, asymptomatic ovarian cysts are commonly detected by ultrasonography. Cysts that are less than 10 cm in diameter and without septations or solid components (unilocular) have a very low potential for malignant disease and can be managed with serial ultrasound surveillance (at 3 months, 6 months, 12 months, and then annually), especially if the serum CA 125 is normal.[191, 192] Surgery is recommended for symptomatic cases, if growth occurs, if internal echoes are obtained, if fluid develops in the pelvis, or if there is a family history of breast or ovarian cancer.

The Impact of Postmenopausal Estrogen Deprivation

The menopause should serve to remind patients and clinicians that this is a time for education. Certainly preventive health care education is important throughout life, but at the time of the menopause, a review of the major health issues can be especially rewarding. Besides the general issues of good health, attention is appropriately focused on cardiovascular disease and osteoporosis.

During the menopausal years, some women experience severe multiple symptoms, whereas others show no reactions or minimal reactions that can go unnoticed. The differences in menopausal reactions in symptoms across different cultures is poorly documented, and indeed, it is difficult to do so. Individual reporting is so conditioned by sociocultural factors that it is hard to determine what is due to biologic versus cultural variability.[193, 194] For example, there is no word to describe a hot flush in Japanese, Chinese, and Mayan.[195] Nevertheless, there is reason to believe that the nature and prevalence of menopausal symptoms are common to most women, and that variations among cultures and within cultures reflect not physiology, but differences in attitudes, societies, lifestyles, socioeconomic status, and individual perceptions.[196–201] Hormone levels during the perimenopausal years vary little among different ethnic groups; differences are mainly because of varying body sizes.[202]

Vasomotor Symptoms

The vasomotor flush is viewed as the hallmark of the female climacteric, experienced to some degree by most postmenopausal women. The term “hot flush” or “hot flash” is descriptive of a sudden onset of reddening of the skin over the head, neck, and chest, accompanied by an increase in heart rate and a feeling of intense body heat. The flush is sometimes concluded by profuse perspiration. The duration varies from a few seconds to several minutes and, rarely, for an hour. The frequency may be rare to recurrent every few minutes. Flushes are more frequent and severe at night (when a woman is often awakened from sleep) or during times of stress. In a cool environment, hot flushes are fewer, less intense, and shorter in duration compared with a warm environment.[203] Most importantly, hot flushing can affect a woman’s quality of life and interfere with work or recreational activities.

In the longitudinal follow-up of a large number of women, fully 10% of the women experienced hot flushes before menopause, while in other studies as many as 15–25% of premenopausal women reported hot flushes.[9, 89, 204, 205] The frequency has been reported to be even higher in premenopausal women diagnosed with premenstrual syndrome.[206] In the Massachusetts Women’s Health Study, the incidence of hot flushes increased from 10% during the premenopausal period to about 50% just after cessation of menses.[89] By approximately 4 years after menopause, the rate of hot flushes declined to 20%. In a community-based

Australian survey, 6% of premenopausal women, 26% of perimenopausal women, and 59% of postmenopausal women complained of hot flushing.[207] A large American cross-sectional survey reported that 57% of perimenopausal women and 49% of early postmenopausal women experienced significant hot flushing.[200] Another national survey in the U.S. reported hot flushing in 79% of perimenopausal women and 65% of postmenopausal women.[208]

In cross-sectional surveys, up to 40% of premenopausal women and 85% of menopausal women report some vasomotor complaints.[205] A longitudinal study in Gothenburg, Sweden, recorded a maximal prevalence of 60% at age 52–54, with a decline to 30% at age 60 and 9% at age 72.[209] In the SWAN study, 57% of perimenopausal women experienced hot flushing, and about 50% after menopause up to age 55.[210] There is no difference in the prevalence of vasomotor complaints in U.S. surveys of black and white women.[211, 212] Overweight women report more hot flushing, perhaps reflecting the effect of body fat causing a higher core body temperature.[200, 202, 213] Exact estimates on prevalence are hampered by inconsistencies and differences in methodologies, cultures, and definitions.[214] The prevalence in different societies is influenced by personal and social attitudes, individual psychological and physical health, familiarity with the portrayal of menopausal issues in the literature and media, ethnic variation, different diets, and dissimilar living conditions; however, accounting for cultural differences, the overall prevalence and experience are similar throughout the world.[215, 216]

Although the flush can occur in the premenopause, it is a major feature of postmenopause, peaking in the first year after the last menses, lasting in 50% of women for 4 to 5 years, but in some (as many as 25%) for longer than 5 years, and up to 15 years in 10%.[217] In an excellent Australian longitudinal cohort study, the average duration of vasomotor symptoms was 5.2 years (with a range of 2 to 10 years) in nonusers of hormone thrapy, and slightly longer, 5.5 years, in hormone users.[218]

The physiology of the hot flush is still not understood. Studies suggest that women with hot flushes have a more narrow zone of temperature regulation, and therefore, smaller changes in core body temperature produce compensatory responses, such as shivering or flushing.[219] MRI scanning of the brain during hot flushing indicates widely distributed cortical activation rather than a precise location.[220] Hot flushes are definitely brought about by a decline in estrogen; however, not all hot flushes are due to estrogen deficiency. Flushes and sweating can be secondary to diseases, including pheochromocytoma, carcinoid, leukemias, pancreatic tumors, and thyroid abnormalities.[221] Unfortunately, the hot flush is a relatively common psychosomatic symptom, and women often are unnecessarily treated with estrogen. ***When the clinical situation is not clear and obvious, estrogen deficiency as the cause of hot flushes should be documented by elevated levels of FSH.***

The correlation between the onset of flushes and estrogen reduction is clinically supported by the effectiveness of estrogen therapy and the absence of flushes in hypoestrogen states, such as gonadal dysgenesis. Only after estrogen is administered and withdrawn do hypogonadal women experience the hot flush. Although the clinical impression that premenopausal surgical castrates suffer more severe vasomotor reactions is widely held, this was not borne out in the only objective study ever performed.[222]

Although the hot flush is the most common problem of the postmenopause, it presents no inherent health hazard. The flush is accompanied by a discrete and reliable pattern of physiologic changes.[219, 223] The flush coincides with a surge of LH (not FSH) and is preceded by a subjective prodromal awareness that a flush is beginning. This aura is followed by measurable increased heat over the entire body surface. A flush is triggered by a small elevation in core body temperature. The body surface experiences an increase in temperature, accompanied by changes in skin conductance, and then the flush is followed by a fall in core temperature—all of which can be objectively measured. In short, the flush is not a

release of accumulated body heat but is a sudden inappropriate excitation of heat release mechanisms. Its relationship to the LH surge and temperature change within the brain is not understood. The observation that flushes occur after hypophysectomy indicates that the mechanism is not dependent on or due directly to LH release. In other words, the same brain event that causes flushes also stimulates gonadotropin-releasing hormone (GnRH) secretion and elevates LH. This is probably secondary to hypothalamic changes in neurotransmitters that increase neuronal and autonomic activity.[224]

Premenopausal women experiencing hot flushes should be screened for thyroid disease and other illnesses. A comprehensive review of all possible causes is available.[225] Clinicians should be sensitive to the possibility of an underlying emotional problem. Looking beyond the presenting symptoms into the patient's life is an important service to the patient and her family that eventually will be appreciated. This is far more difficult than simply prescribing estrogen, but confronting problems is the only way of reaching some resolution. Prescribing estrogen inappropriately (in the presence of normal levels of gonadotropins) only temporarily postpones, by a placebo response, dealing with the underlying issues.

A striking and consistent finding in most studies dealing with menopause and hormonal therapy is a marked placebo response (at least 51% in the first weeks of treatment)[226] in a variety of symptoms, including flushing. In an English randomized, placebo-controlled study of women being treated with estrogen implants and requesting repeat implants, there was no difference in outcome in terms of psychological and physical symptoms comparing the women who received an active implant to those receiving a placebo.[227]

A significant clinical problem encountered in our referral practice is the following scenario: a woman will occasionally undergo an apparently beneficial response to estrogen, only to have the response wear off in several months. This leads to a sequence of periodic visits to the clinician and ever-increasing doses of estrogen. When a patient reaches a point of requiring large doses of estrogen, a careful inquiry must be undertaken to search for a basic psychoneurotic or psychosocial problem. To help persuade a patient that her symptoms are not due to low levels of estrogen, we find it very helpful and convincing to measure the patient's blood level of estradiol and share the result with her.

The Hot Flush	
Premenopausal	10–25% of women
Perimenopausal	60%
Postmenopausal:	
No flushes	15–25%
Daily flushing	15–20%
Duration	1–2 years average
	5 or more years: 25%
Other Causes	Psychosomatic
	Stress
	Thyroid disease
	Subacute, chronic infections
	Pheochromocytoma
	Carcinoid
	Leukemia
	Cancer

Atrophic Changes

With extremely low estrogen production in the late postmenopausal age, or many years after castration, atrophy of vaginal mucosal surfaces takes place, accompanied by vaginitis, pruritus, dyspareunia, and stenosis. Genitourinary atrophy leads to a variety of symptoms that affect the ease and quality of living. Urethritis with dysuria, urgency incontinence, and urinary frequency are further results of mucosal thinning, in this instance, of the urethra and bladder. Recurrent urinary tract infections are effectively prevented by postmenopausal intravaginal estrogen treatment.[228] Vaginal relaxation with cystocele, rectocele, and uterine prolapse, and vulvar dystrophies are not a consequence of estrogen deprivation.

Deprived of estrogen, the vagina loses collagen, adipose tissue, and the ability to retain water. As the vaginal walls shrink, the rugae flatten and disappear. The surface epithelium loses its outer fibrous layer and thins to a few layers of cells, markedly reducing the ratio of superficial to basal cells. As a result, the vaginal surface is left friable, prone to bleeding with minimal trauma. While these changes are occurring, the blood vessels in the vaginal walls narrow, and secretions from sebaceous glands diminish. Over time the vagina itself contracts and loses flexibility, while the labia minora become paler and smaller. In addition, pH becomes more alkaline, making the vaginal environment less hospitable to lactobacilli and more susceptible to infection by urogenital and fecal pathogens. Infecting organisms can ascend into the urinary system to cause urethritis, urinary tract infections, and cystitis.

Dyspareunia, sometimes with postcoital bleeding, is the inevitable consequence of a severely atrophied vagina and scanty lubrication. Even for women who are not sexually active, atrophic vaginitis can cause itching, irritation, and burning. These symptoms often go unmentioned, and it is important to inspect for signs of vaginal atrophy even in the absence of complaints. ***Measuring pH is a simple way to determine estrogen's influence or absence. A pH greater than 4.5 is almost always observed with estrogen deficiency.***[229, 230]

Dyspareunia seldom brings older women to our offices. A basic reluctance to discuss sexual behavior still permeates our society, especially among older patients and physicians. Gentle questioning may lead to estrogen treatment of atrophy and enhancement of sexual enjoyment. Objective measurements have demonstrated that vaginal factors that influence the enjoyment of sexual intercourse can be maintained by appropriate doses of estrogen.[231] Both patient and clinician should be aware that a significant response can be expected by 1 month, but it takes a long time to fully restore the genitourinary tract (6–12 months), and clinicians and patients should not be discouraged by an apparent lack of immediate response. Raloxifene and tamoxifen have little impact on the vaginal epithelium, and vaginal dryness is worse with aromatase inhibitors. Sexual activity by itself supports the circulatory response of the vaginal tissues and enhances the therapeutic effects of estrogen. Therefore, sexually active older women have less atrophy of the vagina even without estrogen.

Although it is argued that genuine stress incontinence is not affected by treatment with estrogen, others contend that estrogen treatment improves or cures stress incontinence in over 50% of patients due to a direct effect on the urethral mucosa.[232–234] A meta-analysis concluded that improvement was reported only in nonrandomized studies.[235] Two randomized trials dedicated to this clinical problem failed to demonstrate a beneficial effect of estrogen treatment.[236, 237] Most cases of urinary incontinence in elderly women are a mixed problem with a significant component of urge incontinence that is believed to be improved by estrogen therapy. However, the Heart and Estrogen-progestin Replacement Study (HERS) randomized trial indicated a worsening of incontinence with hormone therapy for both urge and stress incontinence, and the Nurses' Health Study reported a small increase of incontinence in hormone users.[238, 239] ***There is no convincing support for a***

beneficial impact of estrogen treatment on incontinence. In the SWAN study, only 15% of incontinent women reported a worsening of urinary incontinence during the perimenopausal transition, largely because of weight gain.[240] The majority of incontinent women experienced either no change or an improvement. The SWAN study strongly documents that urinary incontinence is not a major symptom of menopause and the perimenopausal transition.[240, 241] ***Incontinence at midlife is not a consequence of hormonal changes, but largely the effect of excess body weight or diabetes mellitus.***

A decline in skin collagen content, elasticity, and skin thickness that occurs with aging can be considerably avoided by postmenopausal estrogen therapy.[242–246] The effect of estrogen on collagen is evident in both bone and skin; bone mass and collagen decline in parallel after menopause, and estrogen treatment reduces collagen turnover and improves collagen quality.[247, 248] One study demonstrated not only an increase in facial skin thickness, but an improvement in wrinkles with topical estrogen.[249] A randomized trial demonstrated improvements in skin elasticity, skin hydration, and skin thickness comparing hormone treatment with placebo.[250] More impressively, data from the U.S. First National Health and Nutrition Examination Survey indicated that estrogen use was associated with a lower prevalence of skin wrinkling and dry skin.[251] Smoking is a major risk factor for facial skin wrinkling, and hormone therapy cannot diminish this impact of smoking.[252] In a 1-year clinical trial, hormone therapy did not improve skin wrinkling already present.[253]

One of the features of aging in men and women is a steady reduction in muscular strength. Many factors affect this decline, including height, weight, and level of physical activity. Women currently using estrogen have been reported to demonstrate a lesser decline in muscular strength, although at least one study could detect no impact of estrogen.[254–259] This is an important issue because of the potential protective consequences against fractures, as well as a benefit due to the ability to maintain vigorous physical exercise.

Psychophysiologic Effects

The view that menopause has a deleterious effect on mental health is not supported in the psychiatric literature, or in surveys of the general population.[204, 205, 260, 261] The concept of a specific menopause-induced psychiatric disorder (involutional melancholia) has been abandoned. Indeed, depression is less common, not more common, among middle-aged women, and the menopause cannot be linked to psychological distress.[3–9, 262] The longitudinal study of premenopausal women indicates that hysterectomy with or without oophorectomy is not associated with a negative psychological impact among middle-aged women.[263] Longitudinal data from the Massachusetts Women's Health Study document that menopause is not associated with an increased risk of depression.[264] Although women are more likely to experience depression than men, this sex difference begins in early adolescence, not at menopause.[265]

The U.S. National Health Examination Follow-up Study includes both longitudinal and cross-sectional assessments of a nationally representative sample of women. This study has found no evidence linking either natural or surgical menopause to psychologic distress.[266] Indeed, the only longitudinal change was a slight decline in the prevalence of depression as women aged through the menopausal transition. Results in this study were the same in estrogen users and nonusers.

A negative view of mental health at the time of the menopause is not justified; many of the problems reported at the menopause are due to life events.[11, 12, 267, 268] Thus, there are problems encountered in the perimenopausal transition and the early postmenopause that are seen frequently, but their causal relation with estrogen is unlikely. These problems

include fatigue, nervousness, headaches, insomnia, depression, irritability, and palpitations. Indeed, at this stage of life both men and women express a multitude of complaints that do not reveal a gender difference that could be explained by a hormonal cause.[269, 270] Nevertheless, midlife women report complaints more often than men,[270] perhaps reflecting the generally negative perceptions and connotations our cultures and societies have attributed to the menopause.

Two longitudinal cohort studies assessed the new onset of depressive symptoms and disorders during the perimenopausal transition. The Penn Ovarian Aging Study followed over 8 years 436 women with no history of depression and correlated hormonal changes with the onset of depressed mood.[271] Fifty percent of the women developed an increase in measures of depression and 26% met the criteria for a clinical diagnosis of depressive disorder. Using the women as their own controls, the depression group was 2.5 times more likely to develop clinical depression comparing status during the perimenopausal transition to the premenopausal state. These symptoms during the perimenopausal transition were associated with greater variability (but no average differences) in estradiol levels, suggesting that fluctuations of estradiol can be an important destabilizing factor.

The Harvard Study of Moods and Cycles is a prospective cohort of women with and without histories of depression.[272] In the women who entered the perimenopausal transition, the risk of new depression was almost doubled compared with premenopausal women, from 9.5% to 16.6%, and this risk was linked to the presence of vasomotor symptoms. Importantly, a statistically significant increase in risk of new depressive symptoms was present only in women with a history of adverse life events (the events are not defined or specified in the report). Also of note, 83% of the women experienced no mood changes.

The SWAN study reported similar results. A first episode of depression in perimenopausal women was linked to poor physical health, anxiety disorders, stressful life events, and hot flushing.[273]

This area of concern has been very difficult to study. Inconsistent results can reflect variations in study designs, selection of subjects, methods used to measure mood, and the definition of menopausal status. ***However, the best reports provide reliable evidence of a vulnerable population of women.*** Depressive mood changes are influenced by other factors, including body weight, smoking, premenstrual syndrome (PMS, defined in Chapter 14), employment, and marital status. Premenopausal PMS is a strong predictor of depressive symptoms arising in the menopausal transition.

The most important questions are whether truly normal women experience an increase in depression during the menopausal transition, and are there subtle or even clinically apparent psychological problems that identify a susceptible subgroup? The cohort studies support the argument that there is a vulnerable group of perimenopausal women who are responsible for the increase of new depression observed during the perimenopausal transition. The data are consistent with the idea that fluctuations in hormone levels are related to mood symptoms, but it is impossible to know if this is a true cause and effect relationship.

In summary, most women (about 85%) experience the perimenopausal transition without mood difficulties. Some women are at greater risk of new onset depressive symptoms, and this is probably enhanced by hormonal variations and vasomotor symptoms. These vulnerable women are likely derived from a group of premenopausal women with underlying psychological problems (although "problem" may be too strong of a word). It is also possible that perimenopausal hormone changes create a state that makes an individual less able to deal with adverse events in life.

Attempts to study the effects of estrogen on these problems have been hampered by the subjectivity of the complaints (high placebo responses) and the "domino effect" of what

a reduction of hot flushes does to the frequency of the symptoms. Using a double-blind crossover prospective study format, Campbell and Whitehead concluded many years ago that many symptomatic "improvements" ascribed to estrogen therapy result from relief of hot flushes—a "domino" effect.[274] Studies that have controlled for menopausal symptoms conclude that mood is very affected by vasomotor symptoms and sleep disturbances, besides reflecting life problems.[139, 275]

A study of 2,001 Australian women aged 45–55 focused on the utilization of the health care system by women in the perimenopausal period of life.[14] Users of the health care system in this age group were frequent previous users of health care, less healthy, and had more psychosomatic symptoms and vasomotor reactions. These women were more likely to have had a significant previous adverse health history, including a past history of premenstrual complaints. This study emphasized that perimenopausal women who seek health care help are different from those who do not seek help, and they often embrace hormone therapy in the hope it will solve their problems. Similar findings have been reported in a cohort of British women.[276] It is this population that is seen most often, producing biased opinions among clinicians regarding the menopause. We must be careful not to generalize to the entire female population the behavior experienced by this relatively small group of women. Most importantly, perimenopausal women who present to clinicians often end up being treated with estrogen inappropriately and unnecessarily. Nevertheless, it is well established that a woman's quality of life is disrupted by vasomotor symptoms, and estrogen therapy provides impressive improvement.[277–279] Patients are grateful to be the recipients of this "domino" effect.

The Women's Health Initiative (discussed in Chapter 18) concluded that estrogen-progestin therapy had no beneficial impact on health-related quality of life.[280] However, only 12.7% of the participants had moderate to severe vasomotor symptoms at entry to the study, and the severity can be questioned because the participants were willing to take placebo medication. The overall baseline quality of life in this study was relatively high, and the study participants were older (the average number of years distant from menopause was 12+). This randomized, clinical trial did not study the appropriate population of women in order to assess the effect of hormone therapy on measures of quality of life.

The Women's International Study of Long Duration Oestrogen after the Menopause (WISDOM) trial was a randomized, controlled trial in the U.K., Australia, and New Zealand, of 3,721 women aged 50–69 treated with either combined 0.625 mg conjugated estrogens-2.5/5.0 mg medroxyprogesterone or placebo. [281] The original plan was to randomize 22,300 women to the study that would last 10 years. The study was canceled in October 2002 in reaction to the initial reports from the WHI. Unfortunately, the premature cancellation precludes the possibility of any long-term data from WISDOM. In 2,130 women who completed one year, there were statistically significant improvements in the treated women in the categories of vasomotor, sexual, and sleep symptoms. Treated women reported a reduction in aching joints and muscles, night sweats, insomnia, and vaginal dryness. The treated group reported more breast tenderness, but the percentages were notably low (16% in the treated group and 7% in the placebo group).

The WISDOM trial investigators argued that the small effects on quality of life reported by the WHI and HERS can be attributed to the insensitive measurement tools used in those clinical trials. The WISDOM trial used a survey tool specifically designed to assess postmenopausal physical and emotional wellbeing, plus a validated, generic questionnaire, the European quality of life instrument. Only the specific questionnaire detected significant changes; the European generic tool did not. This emphasizes the importance of using the appropriate study tool to investigate this area of postmenopausal health. Similar results with vasomotor symptoms, sleep, and joint complaints were actually reported by the WHI, but with a smaller difference between treated and placebo groups. The WHI survey had only one question devoted to sexuality.

The results of the WISDOM trial are not surprising; they reflect what all clinicians have observed in their own practices. The most important point to be made is this: the WISDOM, WHI, and HERS trials were all similar in that they enrolled postmenopausal women heavily tilted towards the oldest age group without symptoms. It is a simple and logical conclusion that hormone therapy in a younger, symptomatic group of postmenopausal women would produce greater quality of life benefits than that quantified in the clinical trials. All three clinical trials, therefore, underestimated the beneficial impact because of age and symptom status of their participants. However, in the WISDOM trial, even older postmenopausal women who were symptomatic benefited from hormone therapy. Age should not be the sole guiding factor in decision-making.

Emotional stability during the perimenopausal period can be disrupted by poor sleep patterns. Hot flushing does have an adverse impact on the quality of sleep.[282–284] Estrogen therapy improves the quality of sleep, decreasing the time to onset of sleep and increasing the rapid eye movement (REM) sleep time.[277, 285, 286] In the SWAN study, one-third of the women reported sleep problems, even without hot flushes or night sweats, and the prevalence of vasomotor symptoms was associated with an increased risk of sleep disturbances; hormone therapy improved sleep quality.[287, 288] Perhaps flushing may be insufficient to awaken a woman but sufficient to affect the quality of sleep, thereby diminishing the ability to handle the next day's problems and stresses. An improvement in sleeping with estrogen treatment can even be documented in postmenopausal women who are reportedly asymptomatic.[286]

Thus, the overall "quality of life" reported by women can be improved by better sleep and alleviation of hot flushing. However, it is still uncertain whether estrogen treatment has an additional direct pharmacologic antidepressant effect or whether the mood response is totally an indirect benefit of relief from physical symptoms and, consequently, improved sleep. Utilizing various assessment tools for measuring depression, improvements with estrogen treatment were recorded in oophorectomized women.[289, 290] In the large prospective cohort study of the Rancho Bernardo retirement community, no benefit could be detected in measures of depression in current users of postmenopausal estrogen compared with untreated women.[291] Indeed, treated women had higher depressive symptom scores, presumably reflecting treatment selection bias; symptomatic and depressed women seek hormone therapy. Others report that estrogen therapy has a more powerful impact on women's well-being beyond the relief of symptoms such as hot flushes.[277, 292, 293] In elderly depressed women, improvements in response to fluoxetine were enhanced by the addition of estrogen therapy.[294] In a 12-week, randomized, placebo-controlled trial of 55 perimenopausal women with clinically significant major depression, estradiol treatment with the 100 μg transdermal method significantly improved mood.[295] A similar American short-term study of 34 perimenopausal women with both major and minor depressions treated with 50 μg estradiol transdermally demonstrated improvements independently of an effect on vasomotor symptoms.[296] These small clinical trials argue that estrogen treatment is beneficial for the treatment of clinical depression. This conclusion is supported by the successful treatment of postpartum depression with estradiol treatment.[297, 298]

The most common cause of perimenopausal mood problems is already-existing depression,[10, 299] but there does exist a small population of women whose moods are sensitive to hormonal changes. In the American SWAN study, the prevalence of mood changes increased from the premenopause to the early perimenopause, from about 10% to about 16.5%.[299] There are three possible explanations: (1) the decline in estrogen at menopause affects neurotransmitters that regulate mood; (2) mood is adversely affected by vasomotor symptoms (the domino theory); (3) mood is affected by the vicissitudes of life that are commonly prevalent around menopause. Some would argue that these mood swings are in response to the hormonal fluctuations that occur during the perimenopausal years. These fluctuations do indeed occur,[64] but whether they cause any symptoms remains to be determined. It seems logical that individuals with mood problems can reflect all of these mechanisms.

Cognition and Alzheimer's Disease

Depending on the method of assessment, evidence for beneficial effects of estrogen on cognition can be found in the literature, especially in verbal memory.[300, 301] However, the effects in healthy women are not impressive, and perhaps of little clinical value. A short-term study failed to document an objective improvement in memory, although a slight improvement in mood was recorded.[302] Another short-term (3 months) randomized, double-blind study could detect no improvement in cognitive performance compared with placebo treatment.[303] The Melbourne Women's Midlife Health Project could not document an effect on verbal memory during the menopausal transition.[304] A longitudinal study in Chicago could not detect a cognitive decline through the menopause, as assessed by working memory and perceptual speed.[305] On the other hand, estrogen treatment of women immediately after bilateral oophorectomy was associated with improvement in certain, but not all, specific tests of memory, and healthy postmenopausal women taking estrogen scored higher on tests of immediate and delayed recall.[306–308] In a case-control study of women aged 55–93 years, estrogen users had better recall of proper names, but no improvement in word recall.[309] Women in the Baltimore Longitudinal Study of Aging who were using estrogen performed better in tests of visual learning and memory.[310, 311] In a New York City cohort of women, the use of estrogen was associated with better performance in tests of cognition, and better performance in verbal memory, but the cohort in the Study of Osteoporotic Fractures demonstrated no effect of estrogen use on the age-related decline in cognition.[312, 313] In Connecticut, a randomized, placebo-controlled trial demonstrated better reading ability and verbal memory in the estrogen-treated group of postmenopausal women.[314] Perhaps a lack of agreement is due to the variability in test vehicles and the specific aspects of memory function studied. Furthermore, there is impressive individual variability, and when differences have been observed they have not been large, and perhaps of little clinical importance. In addition, any beneficial effects may be attenuated by progestational agents.[301]

Another possibility for the variable effects of estrogen treatment on cognition is the variability among women in endogenous estrogen levels. Using sensitive assays for free, non-protein-bound estradiol and bioavailable (loosely bound) estradiol, cognitive decline occurred at a greater rate in women with low estradiol levels.[315] Studies of cognition may have to differentiate between low- and high-risk women according to endogenous, biologically active estradiol levels. Similarly, a beneficial effect on cognitive decline has been observed only in women negative for the gene associated with Alzheimer's disease, *APOE-ε4*, which encodes the ε4 allele of the glycoprotein known as apolipoprotein E, which has as one of its functions, the shuttling of lipids during neuronal repair.[316]

Up to three times as many women as men develop Alzheimer's disease. Estrogen is capable of protecting central nervous system function by means of multiple mechanisms. For example, estrogen protects against neuronal cytotoxicity induced by oxidation; estrogen reduces the serum concentration of amyloid P component (the glycoprotein found in Alzheimer's neurofibrillary tangles); and estrogen increases synapses and neuronal growth, especially dendritic spine density.[317–319] Estrogen protects against the cerebrovascular toxicity exerted by amyloid peptides, and promotes synaptic formation and neuronal growth and survival.[320–322] Progestational agents do not exert similar actions.

Case-control and cohort findings indicated that Alzheimer's disease and related dementia occurred less frequently (perhaps as much as 60% less) in estrogen users, and the effect was greater with increasing dose and duration of use.[323–325] In the Baltimore Longitudinal Study of Aging (a prospective cohort), the risk of Alzheimer's disease was 54% reduced; in a cohort in New York City, the risk was reduced 60%; and in the Italian Longitudinal Study of Aging, the risk was 72% reduced in estrogen users.[326–328] The findings are not uniformly positive; a case-control study with accurate information on clinical diagnoses and estrogen use from the U.K. General Practice Research Database could detect no impact of estrogen

treatment on the risk of developing Alzheimer's disease, but the number of estrogen users was very small.[329]

The short-term administration of unopposed estrogen to patients with Alzheimer's disease (secondary prevention) has been reported to improve cognitive performance, but mostly to have no effect.[330–334] The administration of combinations of estrogen and progestin has also failed to demonstrate a beneficial impact in Alzheimer's disease.[335] The presence of estrogen therapy has been reported to enhance the beneficial response to tacrine in women with Alzheimer's disease,[336] but overall, the evidence is consistent with a failure of estrogen to influence already-existing Alzheimer's disease or other forms of dementia.[337]

The data do support, however, a primary preventive effect. Most revealing is a prospective cohort study of women living in Cache County, Utah.[338] Hormone therapy provided about a 41% reduced risk of developing Alzheimer's with any use and an 83% reduction with 10 or more years of use. This cohort also demonstrated improved cognition in estrogen users.[339] Most importantly, if women had initiated hormone therapy within a period of time that encompassed 10 years before the development of clinical symptoms, there was no effect. The Utah study strongly suggested that hormone therapy must be used for a significant duration of time very early in the postmenopausal period in order to have an impact on the risk of Alzheimer's disease. As neurons become changed by the pathology of dementia, they lose their ability to respond favorably to estrogen.

The importance of timing is supported by findings from the large Women's Health Initiative (WHI) clinical trial. The oldest women in the WHI being treated with either estrogen alone or combined estrogen-progestin (treatment began at age 65 or older) had impaired cognition and an increased risk of dementia.[340–342] In a subset of these women, MRI scans demonstrated greater brain atrophy in the women receiving hormone therapy.[343] The mechanism for this adverse effect of hormonal therapy in old women may be a neurotoxic action because the WHI study with MRI scanning could not detect an increase in ischemic brain lesions.[344] Smaller MRI studies of younger women treated with hormone therapy found beneficial trophic changes in brain morphology, associated with improved cognition.[345–347]

The theme that emerges is that maintenance of health in target organs by estrogen requires normal tissue, a principle of timing that will also be discussed in regards to the heart in Chapter 18. Following the failure of secondary prevention trials to demonstrate a beneficial impact of hormone therapy on coronary disease in older women, it is increasingly argued that *healthy cardiovascular endothelium is needed to respond to estrogen*; that by the time, the endothelium is involved with excessive atherosclerosis, it is too late for estrogen to exert a beneficial effect. A similar argument is made for brain tissue, focusing on biochemical and signaling pathways that are progressively compromised with the neuronal involvement with disease.[348] *The requirement for normal tissue, at least in the heart and in the brain, would explain the beneficial effects in studies of primary prevention and the lack of effect in secondary prevention trials.*

Cardiovascular Disease

Diseases of the heart are the leading cause of death for women in the United States, followed by cerebrovascular disease and malignant neoplasms. In 2005, 1 in 6 female deaths was from coronary heart disease compared with 1 in 30 for breast cancer deaths.[349]

More female deaths in 2005 were caused by cardiovascular diseases than the combined total from cancer, chronic lower respiratory disease, Alzheimer's disease, accidents, and diabetes mellitus.

Most cardiovascular disease results from atherosclerosis in major vessels. The risk factors are the same for men and women: family history of cardiovascular disease, high blood pressure, smoking, diabetes mellitus, an abnormal cholesterol/lipoprotein profile, and obesity. However, when controlling for these risk factors, men prior to age 40 have a risk of developing coronary heart disease twice that of women. Even taking into consideration the changing lifestyle of women (e.g., employment outside the home), women still maintain their advantage in terms of risk for coronary heart disease. With increasing age, this advantage is gradually lost, and cardiovascular disease becomes the leading cause of death for both older women and older men.

Cardiovascular disease, especially atherosclerosis, is a consequence of multiple metabolic changes that interact with each other:

1. Adverse changes in the circulating lipid-lipoprotein profile.

2. Oxidation of low-density lipoprotein (LDL), producing a modified LDL that is chemotactic for circulating monocytes and inhibits macrophage motility (thus trapping macrophages in the intima), and that causes cell injury and death in the endothelium.

3. Endothelial injury and dysfunction affecting nitric oxide and prostacyclin production.

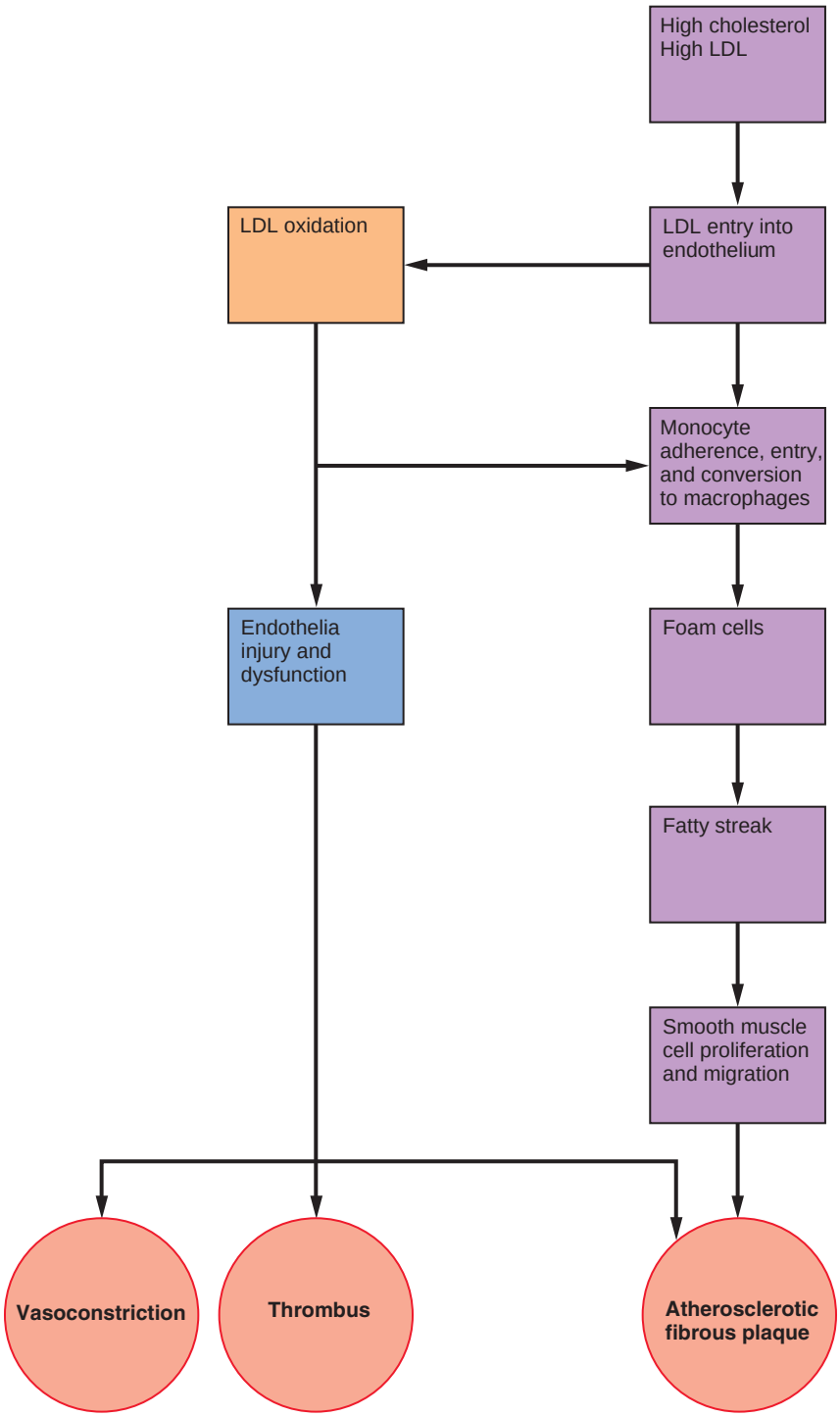
4. Macrophage migration and functions, influenced by growth factors and cytokines.

5. Proliferation and migration of smooth muscle cells, also influenced by growth factors and cytokines; these cells become the dominant cell type and the source of the connective tissue matrix in the atherosclerotic lesion, the fibrous plaque.

6. Vasoconstriction and thrombogenic events.

7. Remodeling of coronary arteries. An artery is able to respond to a developing atherosclerotic plaque by increasing its overall diameter in an attempt to maintain flow.[350] The mechanism of this adaptive remodeling is not known, but the extent of this process must affect the risk of occlusion and infarction.

There is an established sequence of events leading to atherosclerosis. The process starts with endothelial dysfunction that leads to the fatty streak in arterial vessels, the precursor to clinically significant lesions. The fatty streak lesion, therefore, antedates the fibrous plaque, developing under the endothelial surface and dominated by fat-laden macrophages (the foam cells). The damaged endothelium expresses cytokines, adhesion molecules, and other inflammatory agents that are involved in the formation of atherosclerotic plaques. The formation of a plaque is initiated by the aggregation and adherence of circulating monocytes (macrophages) to a site on the arterial endothelium, stimulating an inflammatory response. When the monocytes penetrate through the endothelium and enter the intima, they become loaded with lipids and converted to foam cells. Modification of LDL, especially oxidation, is crucial in this conversion of monocytes to foam cells. The adherence of monocytes to endothelium can be induced by elevated cholesterol and LDL-cholesterol in the circulation. Most of the cholesterol that accumulates in atherosclerotic plaques is derived from circulating LDL-cholesterol. As plaques become significant in size, they are prone to instability, rupturing and creating a prothrombotic state. Matrix metalloproteinase enzymes are secreted by inflammatory cells and smooth muscle cells. These enzymes digest the proteins



in the fibrous cap of an atherosclerotic plaque, making the plaque unstable and predisposed to rupture. ***Estrogen induces matrix metalloproteinase production or activity, which digests the fibrous cap of a plaque exposing the underlying thrombogenic collagen, and this is believed to be the mechanism involved in the adverse thrombotic effects of estrogen in the presence of established atherosclerosis.[351] In addition, 27-hydroxycholesterol, a cholesterol metabolite elevated in atherosclerotic lesions, competitively antagonizes estrogen receptor activity in cardiovascular epithelium.[352]***

During the reproductive years, women are "protected" from coronary heart disease. For this reason, women lag behind men in the incidence of coronary heart disease by 10 years, and for myocardial infarction and sudden death, women have a 20-year advantage. The reasons for this are complex, but a significant contribution to this protection can be assigned to the higher high-density lipoprotein (HDL) levels in younger women, an effect of estrogen and lower levels of testosterone. Throughout adulthood, the blood HDL-cholesterol level is about 10 mg/dL higher in women, and this difference continues through the postmenopausal years. Total and LDL-cholesterol levels are lower in premenopausal women than in men, although the levels gradually increase with aging and after menopause they rise rapidly.[353–357] After menopause the risk of coronary heart disease doubles for women as the atherogenic lipids at about age 60 reach levels greater than those in men. These changes can be favorably reduced by dietary modifications.[358, 359] Of course, these lipid changes at menopause (whether natural or surgical) can be reversed with estrogen treatment.[360]

Prospective studies have documented the strong association between total cholesterol and coronary heart disease in women, although coronary heart disease risk appears at higher total cholesterol levels for women than for men.[361, 362] Women with total cholesterol concentrations greater than 265 mg/dL have rates of coronary heart disease 3 times that of women with low levels. Even in elderly women, a high total cholesterol remains a significant predictor of heart disease but the strength of the association between the cholesterol level and cardiovascular disease decreases with aging, and by age 80 the cost and benefits may not justify cholesterol intervention.[363] This is the reason for ceasing lipoprotein screening after age 75 in patients with normal lipids. However, this decision should be individualized, taking into account the vigor and health of the patient.

The strongest predictor of coronary heart disease in women is a low HDL-cholesterol,[361,362,364] The average HDL-cholesterol in women is approximately 55–60 mg/dL. A decrease in HDL-cholesterol of 10 mg/dL increases coronary heart disease risk by 40–50%. In women (and men) who had normal total cholesterol and LDL-cholesterol levels, but low HDL-cholesterol levels, treatment with lovastatin reduced the risk of an acute major coronary event by approximately 37%.[365] High HDL-cholesterol levels are uncommon in women with coronary heart disease, but even women with high levels do develop coronary heart disease.[366] ***Because, the most powerful predictive value associated with HDL-cholesterol is the increased risk of coronary heart disease observed in individuals with low levels, it is appropriate to be concerned when HDL-cholesterol levels are less than 50 mg/dL.*** It should be emphasized that modest elevations in blood pressure markedly increase the risk associated with an elevated LDL-cholesterol or a low HDL-cholesterol.

Keep in mind that low HDL-cholesterol levels are a component of the metabolic syndrome related to insulin resistance. The metabolic syndrome is partly a result of heredity, but strongly influenced by obesity and physical inactivity. In the U.S., the overall estimated prevalence of the metabolic syndrome is 24%, higher in women (40% by age 60) and increasing with age.[367] The prevalence increases with increasing body weight, from about 5% in normal weight individuals to 60% in obese men and women, and the prevalence is highest in Mexican-Americans and lowest in blacks.[368]

The diagnosis of the metabolic syndrome in an individual requires that three abnormal findings are present out of the five following clinical characteristics:[369]

Hypertension—130/85 or higher.
Triglyceride levels—150 mg/dL or higher.
HDL-cholesterol levels—less than 50 mg/dL.
Abdominal obesity—greater than 35 inches waist circumference.
Fasting glucose—100 mg/dL or higher.

An increasing prevalence of metabolic syndrome during the perimenopausal and menopausal transition is correlated with increasing androgen dominance as estrogen secretion declines.[370, 371] Adiposity of the trunk is a risk factor for coronary heart disease in women and is associated with a relatively androgenic hormonal state, as well as hypertension, and disorders of lipid and carbohydrate metabolism.[372] Central fat distribution in women is positively correlated with increases in total cholesterol, triglycerides, and LDL-cholesterol and negatively correlated with HDL-cholesterol.[373] The atherogenic lipid profile associated with abdominal adiposity is at least partly mediated through an interplay with insulin and estrogen.[374] It is worth noting that there is a strong correlation between the magnitude of the worsening in cardiovascular risk factors (lipid and lipoprotein changes, blood pressure, and insulin levels) and the amount of weight gained during the menopausal transition.[375] Attention to weight gain during middle age is one of the most important components of good preventive health care. However, ***weight gain at menopause is not an effect of hormonal changes; it reflects diet, exercise, and aging.***[375]

Current recommendations regarding the optimal cholesterol/lipoprotein profile are more aggressive, urging more intensive treatment aimed at lowering LDL-cholesterol levels; in the presence of coronary heart disease, the goal is to lower LDL-cholesterol to less than 100 mg/dL.[376] Cholesterol-lowering drugs, specifically the statin family, have been repeatedly demonstrated in clinical trials to have a marked reduction in the risk of clinical cardiovascular events in both men and women.[377, 378]

Triglycerides are also an important risk factor for coronary heart disease in women, but are most commonly encountered in individuals with the metabolic syndrome.[376] If the triglyceride level is greater than 400 mg/dL and the HDL-cholesterol is less than 50 mg/dL, the risk of heart disease is substantially increased. Patients with an elevated triglyceride level and a positive family history for heart disease most likely have an autosomal-dominant disorder classified as familial combined hyperlipidemia. This disorder accounts for most myocardial infarctions in women less than 40 years old. Triglyceride levels of 150–200 mg/dL are considered borderline elevated. Triglyceride levels can be elevated because of obesity, smoking, and lack of exercise. Weight loss alone can return elevated triglyceride levels to normal.

Observational studies and clinical trials indicate that the major determinants of blood lipid levels are the same for both sexes. A diet high in saturated fatty acids and dietary cholesterol unfavorably increases blood lipids. Excess caloric intake and obesity decrease HDL-cholesterol and increase total cholesterol, LDL-cholesterol, and triglycerides. Smoking decreases HDL-cholesterol (and also produces lower estrogen levels and an earlier menopause). Genetic defects of receptor-mediated cholesterol uptake account for only a small percentage of hyperlipidemia in men and women. There is also evidence that men and women who had impaired fetal growth have increased levels of cholesterol and LDL-cholesterol in middle age.[379] The speculation is that impaired liver growth in utero produces a permanent adverse change in cholesterol and lipoprotein metabolism. Reduced fetal growth also leads in adulthood to insulin resistance and lower HDL-cholesterol levels, most severe in those who become obese.[380]

The Optimal Cholesterol/Lipoprotein Profile		
Total cholesterol	—	Less than 200 mg/dL
HDL-cholesterol	—	Greater than 50 mg/dL
LDL-cholesterol	—	Less than 100 mg/dL
Triglycerides	—	Less than 150 mg/dL

The Role of Estrogen Exposure

By a large margin, the leading cause of death among women continues to be coronary heart disease. Coronary atherosclerosis is a lifelong process that varies in its slope of development according to the presence or absence of risk factors. The landmark Pathobiological Determinants of Atherosclerosis in Youth (PDAY) study documented the presence of fatty streaks in adolescents and an increasing prevalence with increasing age.[381] The PDAY study further established that abnormal lipid profiles early in life are a major factor in determining the extent and age of onset of atherosclerosis.[382] It is important for clinicians caring for women to appreciate the importance of premenopausal atherosclerosis and to understand that appropriate medical interventions can reduce the risk of later clinical events. Because atherosclerosis begins early in life, it is logical to conclude that the postmenopausal risk of coronary clinical events is influenced by the degree of coronary artery atherosclerosis already present at the time of menopause.

Women with premature ovarian failure are at increased risk for cardiovascular disease.[383] In other words, there is an inverse relationship between the risk of cardiovascular disease and the age of menopause.[384] Endothelial function in women with premature ovarian failure is impaired, as measured by dilation of the brachial artery in response to blood flow, a response known to be mediated by estrogen-modulated endothelial nitric oxide.[385, 386] This association between endothelial dysfunction and hypoestrogenemia is reinforced by the observation that endothelial dysfunction in women with premature ovarian failure was improved by hormone therapy.[385]

An important contribution to the gender difference in cardiovascular disease prevalence and age of onset is the favorable effect of estrogen on important endothelial events. Vasodilatory and antithrombotic activities can be attributed to endothelial production of nitric oxide and prostacyclin, a process favorably influenced by estrogen. Hypercholesterolemia adversely affects this important endothelial process, and estrogen protects this important endothelial function in the presence of hypercholesterolemia.[387] Estrogen inhibits the oxidation of LDL, and also protects against the toxic effects of oxidized LDL on the endothelium. Women in the SWAN study who complained of hot flushing had more evidence of subclinical cardiovascular disease, such as aortic calcification, compared to women without hot flushes.[388]

A Chinese comparison study concluded that Chinese men and women with angiographically-determined coronary artery disease differ in the sex steroid environment presented to the heart by the circulation, compared with age-matched healthy individuals.[389] Straightforward reasoning has led investigators to connect the different prevalence of coronary artery disease in men and women to the obvious differences in circulating sex steroids determined by the testicles and ovaries. The newly appreciated importance of estrogen in the premenopausal years has added strength to this connection. For many years, it has been generally believed that higher estrogen exposure in women protects against coronary artery disease, and the difference in coronary artery disease prevalence between men and women diminishes after menopause because of the loss of estrogen. At the heart of the matter is the gonadal difference between men and women.

Acute coronary events in premenopausal women occur more frequently when estrogen levels are the lowest during the menstrual cycle.[390] In the national SWAN study, cardiovascular risk factors were more favorable in women with higher levels of estrogen and less favorable in women with longer menstrual cycles.[391] Even amenorrheic athletes in good physical condition have demonstrated endothelial dysfunction, a condition that responded favorably to estrogen-containing oral contraceptives.[392, 393]

In the WISE (Women's Ischemia Syndrome Evaluation) Study, a study of premenopausal women undergoing coronary angiography for suspected myocardial infarction, coronary artery disease was more prevalent in those women who had low estrogen levels because of hypothalamic suppression.[394] These findings are similar to the pioneering studies in monkeys that demonstrated acceleration of atherosclerosis in animals with low estrogen because of stress-induced hypothalamic suppression, an effect that could be prevented by oral contraceptive treatment.[395–397] Postmenopausal women studied with coronary angiography in the WISE study who had used oral contraceptives in the past had less coronary artery disease.[398] In addition, premenopausal women with coronary artery disease have lower circulating levels of estrogen compared with normal women.[399]

Depression is a recognized risk factor for heart disease, but its contribution to premenopausal atherosclerosis is just beginning to be appreciated. Premenopausal monkeys that exhibit depressive behavior (induced by their lower social rank in a colony of animals) develop a more adverse lipid profile and an increasing degree of atherosclerosis when fed atherogenic diets.[400] Premenopausal women with a history of recurrent depression and without known coronary disease are more likely to have coronary and aortic calcification, a marker for early atherosclerosis.[401] The SWAN study found more aortic calcification in black women with depressive symptoms, although an association between coronary calcification and depression in black or white women could not be detected.[402]

Thus hypoestrogenemia in the premenopausal years, whatever the cause, can increase the progression of atherosclerosis. This would include suppressed ovarian function associated with stress, depression, or athletic activity. The progressively deleterious effects of hypoestrogenemia include endothelial dysfunction, lower levels of HDL-cholesterol, an increase in central obesity, and possibly strengthening of depression. ***In monkeys and in women, lipid effects account for only 25–30% of the atheroprotective effects of estrogen.***[403, 404]

A vast literature has documented multiple mechanisms favorably influenced by estrogen that would inhibit the development of atherosclerosis.[405] Clinicians tend to view testosterone as an estrogen opponent, and this is supported, for example, by studies such as lipid responses to testosterone that move in the opposite directions to those of estrogen.[406] Both beneficial and detrimental vascular actions of testosterone have been documented with in vitro and animal studies. Clinical studies in women have in general supported an association between hyperandrogenism and an increase in risk for cardiovascular disease. The lipid and lipoprotein profile in androgenized women with polycystic ovaries (who are also exposed to relatively lower estrogen levels over time) is similar to the male pattern with higher levels of cholesterol, triglycerides, and LDL-cholesterol and lower levels of HDL-cholesterol, and this abnormal pattern is independent of body weight.[407–411] An adverse lipid and lipoprotein profile is a distinguishing feature of these patients even when body mass index, insulin, and age are controlled in case-control studies.[412] Subclinical atherosclerosis can be demonstrated by carotid ultrasonography to be prevalent in premenopausal women with a history of anovulation and polycystic ovaries.[413] In women undergoing coronary angiography, the prevalence of polycystic ovaries is increased, and women with polycystic ovaries have more extensive coronary atherosclerosis.[414] In the Nurses' Health Study, women with very irregular cycles compared to women with regular cycles had an adjusted increased risk of coronary heart disease.[415] ***Thus anovulatory women with polycystic ovaries develop risk factors for atherosclerosis and ultimately clinical disease comparable with that found in older, very overweight, postmenopausal women.***

Given the variability in circulating sex steroid levels between individuals and within individuals, we should not be surprised that random blood sampling does not always document meaningful differences in blood levels of gonadal steroids in men and women with coronary artery disease. Nevertheless, some cross-sectional studies have documented lower circulating testosterone levels in men with coronary artery disease and higher levels in men

with a reduced risk of metabolic syndrome.[416, 417] A prospective cohort study determined that lower testosterone levels were associated with an increased risk of developing metabolic syndrome in men.[418] Men with heart failure demonstrate improvements in symptoms and functional capacity when treated with testosterone.[419]

In healthy postmenopausal women, higher androgen levels are associated with increased metabolic markers for the risk of coronary artery disease, and increasing testosterone levels during the perimenopausal transition correlated with an increasing prevalence of metabolic syndrome in the SWAN study.[370, 420, 421] Higher testosterone levels, *but still in the normal range*, have also been found to correlate with a reduction in atherosclerosis progression in naturally postmenopausal women.[422, 423]

In summary, these studies suggest that higher testosterone levels within the normal physiologic range protect against atherosclerosis (perhaps by target tissue aromatization to estrogens), but that elevated androgen levels above the normal range, such as in anovulatory women with polycystic ovaries, increase the progression of atherosclerosis.

The important studies in women and monkeys reviewed above indicate that every woman has a trajectory of atherosclerosis, the slope of which determines the age of onset for clinical events.[424] The contribution of premenopausal atherosclerosis to the development of clinical events highlights the important role for clinicians in aggressively promoting preventive interventions that can favorably change the slope of development. An important risk factor is exposure to protective levels of estrogen at all stages of life. Conditions associated with hypoestrogenemia during the premenopausal years, therefore, require evaluation and treatment. There are many causes of hypoestrogenemia, and the treatments will vary according to the etiology. When indicated, appropriate hormone treatment can reduce the risk of cardiovascular disease later in life. In addition, appropriate interventions in insulin-resistant women with the metabolic abnormalities associated with polycystic ovaries can reduce the risks of both cardiovascular disease and diabetes mellitus.

A logical continuum of this reasoning is that hormone therapy in the early postmenopausal years likewise can provide primary prevention of clinical coronary disease. A meta-analysis of 23 randomized hormone therapy trials concluded that treatment reduced the risk of coronary heart disease events in younger women compared with older women (10 or more years since menopause or greater than 60 years of age).[425] This is a conclusion that is less firm than at first apparent, because most of these trials were not designed to measure an endpoint of cardiovascular disease. However, another meta-analysis by the same authors concluded that hormone therapy reduced overall mortality in women with an average age less than 60.[426] ***There is a growing story that adequate estrogen exposure prior to the onset of clinical events provides protection against cardiovascular disease.***

C-Reactive Protein

The development of atherosclerotic plaques involves the immune system (monocytes, cytokines, and cell adhesion molecules).[427] For this reason, studies indicate that C-reactive protein (CRP) is a marker of cardiovascular risk in men and women.[428–430] This risk is limited, however, to arterial disease; CRP levels are not linked to venous thrombosis or pulmonary embolus.[428] CRP predicts an increased risk of cardiovascular events even in individuals who have normal lipid levels, and, therefore, it is argued that both CRP and lipid profiles should be used for screening purposes.[430, 431]

CRP is a protein synthesized in the liver and atherosclerotic arteries, and was given its name because it reacts with the C-polysaccharide of *Streptococcus pneumoniae*.

Thus the circulating level of CRP increases in response to various inflammatory stimuli, but specifically bacterial infections and chronic inflammatory conditions such as systemic lupus erythematosus. Sensitive assays now detect small increases associated with low-grade inflammation in the vascular system.

Increased levels of CRP in patients with angina predict poor outcome, an increase in the relative risk of a coronary event. Prospective studies have documented an increased risk of cardiovascular events in patients without known cardiovascular disease who have high CRP levels, an association that is even greater in smokers.[432] Higher mean levels are found in both men and women who subsequently have myocardial infarctions. Stroke and peripheral vascular disease are also increased in men with higher CRP levels, but this has not been adequately assessed in women. In a meta-analysis of 14 prospective studies, individuals with CRP levels in the top third compared with individuals in the bottom third had a 2-fold increase in relative risk for coronary heart disease.[433] Thus, CRP levels have predictive value in both healthy individuals and individuals with cardiac disease. In addition, statin treatment lowers CRP levels,[434] and evidence indicates that statins and aspirin achieve greater benefits in individuals with high CRP levels.[428, 435]

In general, studies have indicated that oral estrogen treatment (with or without progestin) increases CRP levels and raloxifene does not. In a double-blind, randomized trial, postmenopausal hormone therapy and raloxifene equally lowered homocysteine levels, but estrogen-progestin treatment increased CRP levels whereas raloxifene had no effect.[436] These results were duplicated in a Dutch randomized study.[437] Tibolone increases CRP levels to the same degree as oral estrogen therapy.[438] A cross-sectional study found higher levels of CRP in healthy postmenopausal women using hormone therapy.[439] In the PEPI randomized trial, hormone therapy increased CRP levels, but the levels of E-selectin, another marker of inflammation, were reduced.[440]

An estrogen-induced increase in CRP levels may be due to estrogen's well-known effect to stimulate the hepatic synthesis of proteins, especially because of the first-pass phenomenon with oral administration. For this reason, transdermal estrogen treatment does not change CRP levels.[438, 441, 442] Studies with multiple inflammatory markers report that oral estrogen therapy increases only CRP, the only marker synthesized in the liver. In fact oral hormone therapy while increasing CRP, reduces the circulating levels of other inflammatory markers (E-selectin, intercellular adhesion molecule-1, vascular cell adhesion molecule-1, monocyte chemoattractant protein-1, and tumor necrosis factor-α) with inconsistent effects on interleukin-6.[443, 444] Most importantly, it is not certain that the decrease in CRP levels with statins and the increase with estrogen are instrumental in clinical outcomes or reflect other effects. Thus raising or lowering CRP levels will not necessarily increase and decrease the risk of clinical disease. A study from the Women's Health Initiative confirmed the correlation between baseline levels of CRP and an elevated risk of coronary heart disease, *but the increase in CRP induced by oral hormone therapy did not further increase the risk!*[445] ***The uncertainties and questions regarding the clinical meaning of CRP levels make it premature to conclude that changes in CRP levels with hormone therapy have a direct clinical consequence. This does not detract, however, from the use of baseline levels of CRP in an effort to assess the risk of cardiovascular disease in an individual patient.***

Homocysteine

Elevated homocysteine levels are associated with an increase in coronary heart disease.[430] Increasing folic acid intake and hormone therapy reduce the circulating levels of homocysteine. The value of homocysteine measurements in a screening program and the efficacy of therapy directed to lowering homocysteine levels remain to be determined.

Lipoprotein (a)

Lp(a) is composed of two parts, a lipoprotein particle similar to LDL and a glycoprotein that resembles a clotting protein. Lp(a) is an independent risk factor for coronary heart disease, and elevated levels are reduced by postmenopausal hormone therapy.[446] Unlike other markers, the levels of Lp(a) are unaffected by lifestyle, and no clinical use has been established for Lp(a) measurements.

Cardiovascular Disease: Concluding Thought

In the last 30 years, stroke mortality and mortality from coronary heart disease have declined substantially in the U.S. Improvements in medical and surgical care can account for some of this decline, but 60–70% of the improvement is due to preventive measures. Excellent data from epidemiologic studies and clinical trials demonstrate a decline in stroke and heart disease morbidity and mortality from smoking cessation, blood pressure reduction, and lowering of cholesterol.[447–449] It is now recognized that there is a strong and growing scientific basis for preventive medicine and health promotion efforts in clinical practice. Although, the most effective means to lower coronary heart disease in a population is through primary prevention, especially smoking cessation and body weight reduction, the important role for estrogen in maintenance of good cardiovascular health should not be ignored.

Osteoporosis

Bone is a very active organ. A continuous process, called bone remodeling, involves constant resorption (osteoclastic activity) and bone formation (osteoblastic activity). Both osteoblasts and osteoclasts are derived from bone marrow progenitors, osteoblasts from mesenchymal stem cells and osteoclasts from hematopoietic white cell lineage. Cytokines are involved in this development process, a process regulated by the sex steroids.

The amount of bone at any point of time reflects the balance of the osteoblastic and osteoclastic forces, influenced by a multitude of stimulating and inhibiting agents. Aging and a loss of estrogen both lead to excessive osteoclastic activity. A decrease in calcium intake and/or absorption lowers the serum level of ionized calcium. This stimulates parathyroid hormone (PTH) secretion to mobilize calcium from bone by direct stimulation of osteoclastic activity. Increased PTH also stimulates the production of vitamin D to increase intestinal calcium absorption. A deficiency in estrogen is associated with a greater responsiveness of bone to PTH. Thus, for any given level of PTH when estrogen is deficient there is more calcium removed from bone, raising serum calcium, which, in turn, lowers PTH and decreases vitamin D and intestinal absorption of calcium.

Osteoporosis, the most prevalent bone problem in the elderly, is decreased bone mass with a normal ratio of mineral to matrix, leading to an increase in fractures. Osteoporosis is a major global public health problem threatening more than 44 million individuals, and it is epidemic in the U.S., affecting 10 million Americans (four times more women than men).[450] The number of individuals with osteoporosis worldwide in 2010 is expected to total 52 million (35 million women). In addition, about 34 million people (80% women) in the U.S. with low bone mass (osteopenia) are at an increased risk for osteoporosis.

The increase in osteoporotic fractures in the developed world is partly due to an increase in the elderly population, but not totally. A comparison of bone densities in proximal femur bones in specimens from a period of over 200 years suggested that women lose more bone today, perhaps due to less physical activity and less parity.[451] Other contributing factors include a dietary decrease in calcium and an earlier and greater loss of bone because of the impact of smoking. Our Stone Age predecessors consumed a diet high in calcium, mostly from vegetable sources.[452] However, the impact of the tremendous increase in the elderly population throughout the world cannot be underrated. Because of this demographic change, the number of hip fractures occurring in the world each year will increase approximately 6-fold from 1990 to 2050, and the proportion occurring in Europe and North America will fall from 50% to 25% as the numbers of old people in developing countries increase.[453]

Pathophysiology of Osteoporosis

Osteoporosis is characterized by low bone mass and microarchitectural deterioration of bone tissue, leading to enhanced bone fragility and a consequent increase in the risk of fractures even with little or no trauma. The skeleton consists of two bone types. Cortical bone (the bone of the peripheral skeleton) is responsible for 80% of total bone, while trabecular bone (the bone of the axial skeleton—the spinal column, the pelvis, and the proximal femur) constitutes a honeycomb structure filled with red marrow and fat, providing greater surface area per unit volume.

The subsequent risk of fracture from osteoporosis in women will depend on bone mass at the time of menopause and the rate of bone loss following menopause.[454] Although the peak bone mass is influenced by heredity and endocrine factors, it is now recognized that there exists only a relatively narrow window of opportunity for acquiring bone mass. Almost all of the bone mass in the hip and the vertebral bodies is accumulated in young women by late adolescence (age 18), and the years immediately following menarche (11–14) are especially important.[455–457] Estrogen exposure during adolescence is vital. Individuals who experience a late menarche are characterized by bones with lesser density and a reduction in microstructural components, a bone quality that persists to menopause and is associated with an increased risk of fractures.[458] Women who experience amenorrhea during adolescence have an increased prevalence of osteoporosis.[459] Calcium supplementation in prepubertal and pubertal girls improves bone accrual, an important effect that could have long-lasting beneficial consequences.[460–463]

Although the onset of spinal bone loss begins in the 20s, the overall change is small until menopause.[456, 464] Bone density in the femur peaks in the mid to late 20s and begins to decrease around age 30. In general, trabecular bone resorption and formation occur 4 to 8 times as fast as cortical bone. Beyond age 30, trabecular resorption begins to exceed formation by about 0.7% per year. Bone loss accelerates after menopause as up to 5% of trabecular bone and 1–1.5% of total bone mass loss occurs per year in the first years after menopause. This accelerated loss continues for about 5 years, after which bone loss is considerably diminished but continues as the aging-related loss.[465] For the first 20 years following cessation of menses, postmenopause-related bone loss results in a 50% reduction in trabecular bone and a 30% reduction in cortical bone.[466, 467]

When estrogen levels decline, bone remodeling, resorption and formation of bone, increases. The site of remodeling is determined by a need to adapt to physical loads on a bone. Each remodeling locus is initiated by osteoclast excavation, a process that takes 3 to 5 months, followed by osteoblast refilling. Estrogen exerts a tonic suppression of remodeling and maintains a balance between osteoclastic and osteoblastic activity; in the absence

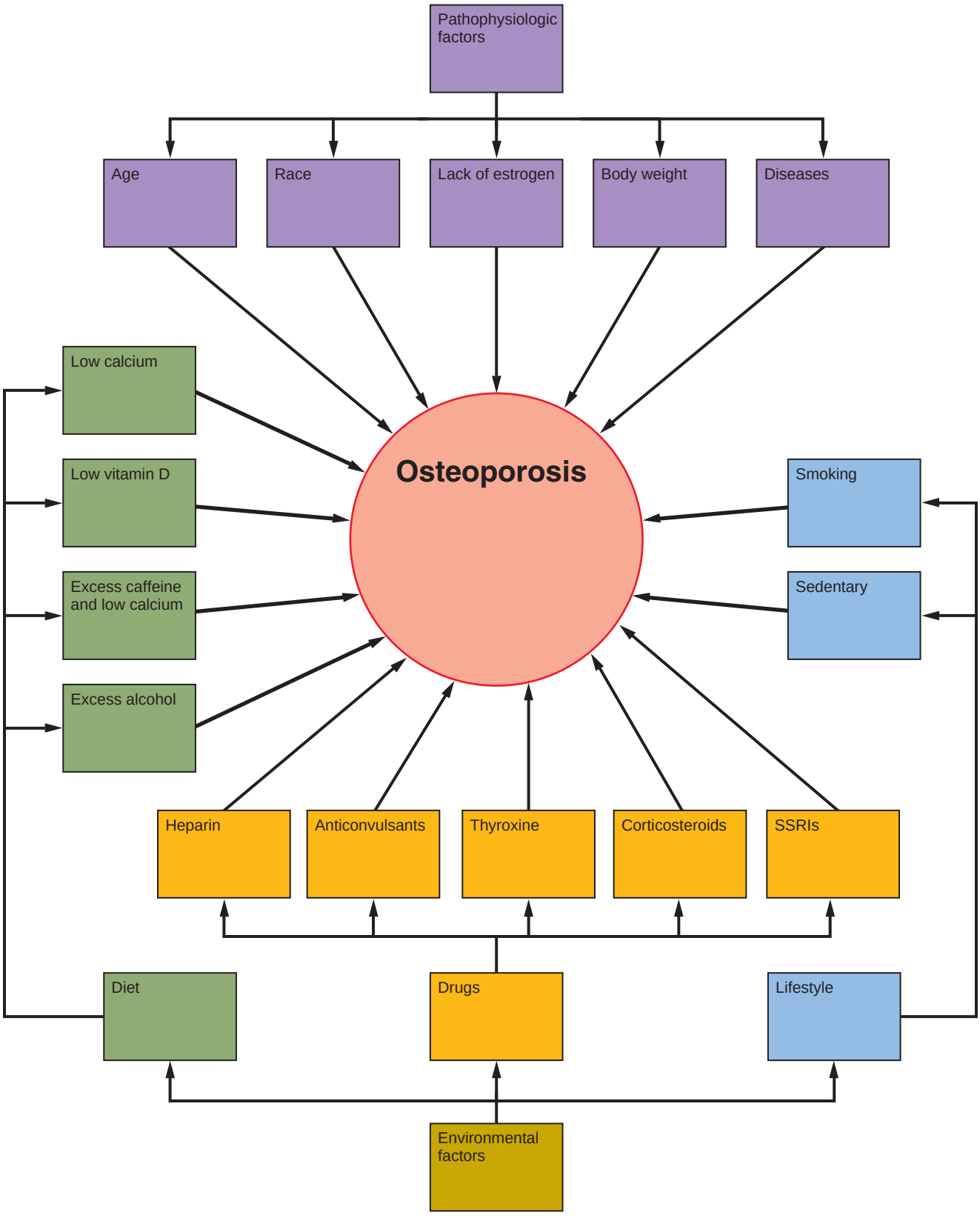
of estrogen, osteoclastic activity predominates, resulting in bone resorption. The precise mechanism of action for sex steroid protection of bones remains unknown; however a growing body of knowledge indicates complex interactions at the molecular level, with both a classic pathway involving genomic transcription by hormone receptors and a nongenomic pathway that inhibits apoptosis.[468, 469] Increased efficiency of calcium absorption, probably secondary to estrogen-induced enhancement of the availability of vitamin D, and a direct role for the estrogen receptors in the osteoblasts are likely important factors. Many estrogen-dependent growth factors and cytokines are involved in bone remodeling.[470, 471] Estrogen modulates the production of bone resorbing cytokines such as interleukin-1 and -6, bone stimulating factors such as insulin-like growth factor-I, colony-stimulating factor, osteoprotegerin, and proteins that are members of the transforming growth factor-β family.[472] Estrogen increases vitamin D receptors in osteoblasts, and this may be a method by which estrogen modulates vitamin D activity in bone.[473] There is little evidence that estrogen affects bone by altering the circulating calcitropic hormones.[474] Thus, the actions of estrogen are primarily direct effects on bone and important effects on vitamin D metabolism and renal and intestinal handling of calcium.

Estrogen is a critical hormone in both males and females. Males with mutations in the estrogen receptor-alpha or who have aromatase deficiencies grow slowly and have markedly reduced bone densities.[475, 476] Analysis of the decline in testosterone and estrogen circulating levels with aging indicates that the amount of bioavailable estrogen circulating in the blood is the most consistent predictor of bone density in men and women.[477] And most impressively, men with an aromatase deficiency, treated with estrogen, demonstrated that both androgens and estrogens are necessary in order for males to reach optimal bone mass.[478, 479]

Bone loss is slower in blacks. Bone mass, adjusted for body size, is greater in black women. It is believed that racial differences in bone density are established in childhood and early adolescence.[480, 481] More than half the difference, comparing black children with white children, is due to differences in body size and composition; a significant contribution is made by bone metabolism, a smaller portion by sex hormone levels, and only 2% of the difference is unexplained.[482] During adolescence, blacks absorb calcium more efficiently and bone turnover is greater with more bone formation.[483]

In general, bone mass is greater in black and obese women and less in white, thin, and sedentary women. Spinal bone density is similar among black and Asian women (lower in white women), and hip bone density is higher among blacks, whereas white and Asian women are lower.[484] Conventional wisdom has held that Asian women have even lower bone densities than white women. The SWAN study demonstrated that this apparent difference disappeared when the data were adjusted for body size.[484, 485] Both black and Asian women have higher bone density and lower fracture rates than white women. Despite the greater bone mass and lower fracture rates, the impact of osteoporotic fractures in black and Asian women is still considerable. Black and Asian women are subject to similar risk factors for osteoporosis (thinness, smoking, high alcohol consumption), and, most importantly, postmenopausal estrogen therapy is associated with protection.[486] Japanese women demonstrate the same amount and pattern of bone loss after menopause as white women.[487]

Although estrogen plays a principal role in regulating bone density, a genetic susceptibility for bone loss is important. A study of the premenopausal daughters of women with osteoporosis revealed a reduction in bone mass, suggesting a genetic influence as well as the sharing of a lifestyle that produces a relatively low peak bone mass.[488] Studies of twins and mother-daughter pairs indicate that up to 70% of the variation in bone density is determined by heredity.[489]


Pathophysiologic factors
Age
Race
Lack of estrogen
Body weight
Diseases
Low calcium
Low vitamin D
Excess caffeine and low calcium
Excess alcohol
Osteoporosis
Smoking
Sedentary
Heparin
Anticonvulsants
Thyroxine
Corticosteroids
SSRIs
Diet
Drugs
Lifestyle
Environmental factors

Variations in the gene that encodes the vitamin D receptor are prevalent in postmenopausal women with decreased bone densities.[490] The absence of vitamin D receptor gene alleleic polymorphisms was reported in older women who neither lose significant bone nor respond to calcium supplementation.[491] The inherited aspects of osteoporosis are likely to be influenced by multiple susceptibility genes. For example, low bone density is associated with specific alleles of *COLIA1*, one of the 2 genes that encode the 2 polypeptides of collagen.[492] Two single nucleotide polymorphisms, variants for key biological proteins (the osteoprotegerin and lipoprotein-receptor-related genes) are present in slightly over 20% of white people and are associated with an increased risk for osteoporosis.[493] The development of genetic markers to identify individuals at increased risk for osteoporosis is in its early stages, and will be one of the more complicated exercises in genomics before screening becomes reality.

The loss of bone in postmenopausal women is largely attributed to estrogen deficiency; 75% or more of the bone loss that occurs in women during the first 15 years after menopause is due to estrogen deficiency rather than to aging itself.[494, 495] Vertebral bone is especially vulnerable, beginning to decline as early as 20 years of age.[496] Vertebral bone mass is significantly decreased in perimenopausal and early postmenopausal women who have rising FSH and decreasing estrogen levels, whereas bone loss from the radius is not found until at least a year out from the menopause.[497] This early loss of axial skeleton bone suggests that the hypoestrogenic postmenopausal state is not the only cause of vertebral osteoporosis.[498] One obvious suspect is a decline in dietary intake of calcium and vitamin D in the premenopausal years; nevertheless, menopause and the loss of estrogen remain as the major contributors to bone loss. The risk of fracture, therefore, depends on 2 factors: the bone mass achieved at maturity and the subsequent rate of bone loss. A high rate of bone loss after menopause (the "fast loser") is highly predictive of an increased risk of fracture. The combination of a low bone mass and fast losing is additive, and thus, these individuals are at the highest risk of fracture. Fast losing probably reflects lower endogenous estrogen levels. The bone density, which is the threshold for vertebral fractures, is only slightly below the lower limit of normal for premenopausal women.[499]

Bone Loss During the Perimenopausal Transition

Should a clinician be concerned about bone loss and consider interventions during the perimenopausal years? Some studies concluded that calcium supplementation of perimenopausal women retards metacarpal and lumbar bone loss.[500, 501] Perimenopausal women in the SWAN study demonstrated decreasing bone densities that correlated with increasing FSH levels; however, accelerated bone loss did not occur until late in the perimenopause.[485] The amount of perimenopausal bone loss is small unless estrogen levels are below normal.[502–505] Healthy women (exercisers and non-exercisers) who are anovulatory or who have inadequate luteal phase function (and thus are exposed to less progesterone) do not have an increase in bone loss.[506, 507] ***Interventions and treatments to prevent future osteoporosis are not necessary in perimenopausal women who have adequate estrogen levels and who are eating normally.***

Signs and Symptoms of Osteoporosis

The osteoporotic disabilities sustained by castrate or postmenopausal women include back pain, decreased height and mobility, and fractures of the vertebral body, humerus, upper femur, distal forearm, and ribs. Osteoporosis is responsible for more than 2 million fractures per year in the U.S.[450] About one in two white women over age 50 will

experience an osteoporosis-related fracture. Back pain is a major clinical symptom of vertebral compression fractures. The pain with a fracture is acute, and then it decreases over 2–3 months, but lingers as chronic low back pain due to increasing lumbar lordosis. The pain subsides within 6 months unless multiple fractures produce a picture of constant pain. ***Many individuals will not know they have osteoporosis until they experience a fracture.***

Epidemiologic studies have revealed the following:[508, 509]

1. ***Spinal (vertebral) compression fracture.*** Symptomatic spinal osteoporosis, causing pain, loss of height, postural deformities (the kyphotic dowager's hump) with consequent pulmonary, gastrointestinal, and bladder dysfunction, is five times more common in white women than men. Approximately 50% of women over 65 years of age have spinal compression fractures; about two-thirds are clinically unrecognized. Each complete compression fracture causes the loss of approximately 1 cm in height. The average untreated postmenopausal white woman can expect to shrink 2.5 inches (6.4 cm). The most common sites for vertebral fractures are the 12th thoracic and the first three lumbar vertebrae. These physical changes also have a negative impact on body image and self esteem.

2. ***Colles' fracture.*** There is a 10-fold increase in distal forearm fractures in white women as they progress from age 35 to 60 years. A white woman has approximately a 15% lifetime risk of a forearm fracture. Colles' fractures are the most common fractures among white women until age 75, when hip fractures become more common.

3. ***Head of femur fracture.*** The incidence of hip fractures increases with age in white women, rising from 0.3/1,000 to 20/1,000 from 45 to 85 years. Eighty percent of all hip fractures are associated with osteoporosis. White women who are 50 years old have approximately a 14% lifetime risk of having a hip fracture; black 50-year-old women, 6%. This fracture carries an increased risk of morbidity and mortality. About 25% of patients over the age of 50 with hip fractures die due to the fracture or its complications (surgical, embolic, cardiopulmonary) within 1 year. The survivors are frequently severely disabled and may become permanent invalids. Hip fractures alone occur in about 300,000 individuals per year in the U.S. with a mortality of 40,000 annually and an associated cost of billions of dollars.[450]

4. ***Tooth loss.*** Oral alveolar bone loss (which can lead to loss of teeth) is strongly correlated with osteoporosis, and the salutary effect of estrogen on skeletal bone mass is also manifested on oral bone.[510, 511] Even in women without osteoporosis, there is a correlation between spinal bone density and number of teeth.[512] Tooth loss is also correlated with the use of cigarettes, a recognized contributor to bone loss. Postmenopausal women who use hormone therapy lose fewer teeth.[513–515]

Individuals at an increased risk for fracture can be identified by a careful history. The following risk factors are especially important for women:

- ***Aging: the risk of fracture doubles every 7–8 years after age 50.***
- ***Previous history of a fragility fracture.***
- ***Family history of fragility fracture in close relatives.***
- ***Smoking.***
- ***Being thin and small-framed.***
- ***Family history of osteoporosis.***
- ***Amenorrhea (hypoestrogenism).***
- ***Lifelong deficient calcium and vitamin D intake.***
- ***Use of bone-losing medications.***

- ***Sedentary lifestyle.***
- ***Excessive use of alcohol.***
- ***Rheumatoid arthritis.***

Postmenopausal women who have already experienced a vertebral fracture deserve aggressive intervention. The risk of subsequent, additional vertebral fractures is substantial; 20% of women experience another vertebral fracture within 1 year after the first fracture.[516] These women are also at increased risk for nonvertebral fractures.[517, 518] ***A complete evaluation for osteoporosis is an essential component of care for a patient presenting with a fracture. All osteoporotic fractures are associated with an increased risk of mortality that persists for 5 to 10 years after the fracture.[519]***

Because there are so many women with osteoporosis, a greater prevalence of depression in this population would amount to a clinical problem of considerable proportions. A cross-sectional subset in a large clinical trial reported a greater prevalence of depressive symptoms in women with fractures.[520] It is well recognized that fractures secondary to osteoporosis are accompanied by a reduction in psychological and physical well-being. As far as depression goes, it is difficult to know which came first, depression or fractures leading to subsequent depression. It has been reported that depressed people have a greater incidence of falls,[521] and thus it is not unreasonable to consider that depression comes first in some people. Furthermore, depressed people are sedentary and eat poorly, factors that favor bone loss. It is speculated that increased cortisol levels associated with depression might lead to bone loss, similar to that observed with the pharmacologic administration of glucocorticoids. On the other hand, a cohort study of American women, despite finding a link between depression and fractures, failed to detect an increase in depression associated with lower bone density measurements.[521] However, other studies *have* reported increases in depression associated with lower bone densities.[522–524]

Bone loss has been documented in an established rodent model for stress-induced depression, characterized by a decrease in osteoblastic bone formation that can be attenuated by an antidepressant drug.[525] In this experimental model, osteoblastic inhibition was mediated by stress-induced stimulation of the sympathetic nervous system. Although this response is associated with an increased secretion of adrenal glucocorticoids, the evidence also indicates a direct role for sympathetic fibers in bone.

We need to be aware that women who have experienced fractures may have depressive symptoms, and appropriate interventions can have a beneficial impact on quality of life. The important point is that depression and fractures are linked; one may precede the other and vice-versa in different patients.

Selective serotonin reuptake inhibitors (SSRIs) are the favored treatment for depression in older adults, a problem that affects about 10% of the older population. Several earlier studies had reported an increased risk of fractures with the daily use of SSRIs; however, these earlier studies were unable to control for the various factors that influence this risk, especially falls, depression, and bone density. An excellent prospective cohort study in Canada indicated that the increase in fractures persisted after controlling for these factors.[526]

There is a direct effect of SSRIs on bone. Components of the neural system are involved in bone metabolism, and serotonin receptors and serotonin transport have been identified in osteoblasts and osteocytes. The bone effects of parathyroid hormone and mechanical stimulation are modulated by the serotonin system. Mice with a mutation for the serotonin transporter develop less bone mass and strength.[527] Therefore daily SSRI use can impair bone formation, tilting the balance in favor of resorption and bone loss, and decreased bone densities have been reported in both male and female SSRI users (but not in users of tricyclic antidepressants).[528, 529]

Measuring Bone Density

There is a 50–100% increase in fracture risk for each standard deviation decline in bone mass (approximately 0.1 g/cm bone mass.[530] Measurement of lower bone mass in the hip is even more predictive; one standard deviation is associated with nearly a 3-fold increase in risk of fracture.[531] ***Although low bone density reliably predicts the risk of fracture, increases in bone density in response to treatment do not demonstrate a direct correlation with a reduction in fractures. Therefore, a few percentage point differences achieved by various treatments have little clinical meaning.***

The impressive correlation between fracture risk and low bone density has raised the question of whether it is of value to screen for osteoporosis. Keep in mind that because the rate of bone loss after menopause contributes equally to the risk of fracture as the total bone mass present at the time of the menopause, a normal bone density measurement at the time of menopause does not mean that the patient will not be at risk of fracture later in life. A relatively young woman with a low bone mass could be targeted for appropriate intervention; however, it is not cost-effective to attempt to screen all postmenopausal women with an expensive method, and attention is now returning to the methods of single photon, single-energy x-ray absorptiometry and ultrasonography because measurement of bone loss at the heel, the metacarpal, and the radius accurately assess future fracture risk.[532]

Bone density measurements are certainly useful when an individual woman requires the information in order to make an informed decision regarding hormone therapy. Indeed, a decision to use hormone therapy and better maintenance of a hormone program are correlated with patients' knowledge of their bone density measurements.[533, 534] Because smokers have lower estrogen levels on estrogen therapy, it is worthwhile to document the impact of treatment on bone density in order to consider whether dosage is adequate. Patients who have received long-term corticosteroid, thyroxine, anticonvulsant, or heparin treatment warrant bone mass assessment. Bone density measurements to monitor bone loss are performed every 2 years.

SUMMARY of Reasons to Measure Bone Mass

1. To help patients make decisions regarding hormone therapy.
2. To assess response to therapy in selected patients, e.g., smokers, and women with eating disorders.
3. To assess bone mass in patients being treated long-term with glucocorticoids, thyroid hormone, anticonvulsants, or heparin.
4. To confirm the diagnosis and assess the severity of osteoporosis to aid in treatment decisions, and to monitor efficacy of therapy.
5. To assess bone mass in postmenopausal women who present with fractures, who have one or more risk factors for osteoporosis, or who are over age 65.

Standard x-rays do not provide an early assessment of fracture risk; 30–40% of bone must be lost before radiographic changes become apparent. Photon absorptiometry measures the transmission of photons through bone. Single-photon absorptiometry uses a ^{125}I source of energy or, more recently, miniature x-ray tubes. These methods measure bone density in the radius and the calcaneus and are relatively inexpensive. These measurements

correlate with vertebral bone density and predict the risk of future fracture.[532] Dual-energy absorptiometry employs different photons from two energy sources. Dual-energy x-ray absorptiometry (DEXA) provides good precision for all sites of osteoporotic fractures, and the radiation dose is much less than for a standard chest x-ray. Whole-body scans by DEXA can measure total body calcium, lean body mass, and fat mass. Quantitative computed tomography (CT) for bone density measurements can be performed on most commercial computed tomography systems; however, radiation exposure is higher than with DEXA, and measurements of the femur are not available, although very accurate measurements of the spine are possible.

For high precision, the best information is provided by the DEXA technique (together with sophisticated software to give a precision of 1%), measuring the 3 sites of greatest interest, the radius, the hip, and the spine.[535] Better accuracy is gained by 3-site assessments because there can be differences among the sites. In other words, a normal value at one site does not preclude a low bone density at another site. For practical clinical use (and for screening), measurements are made at the lumbar spine, the hip and the femoral neck. Serial measurements are best made at least 2 years apart. Measurement of the bone density at the radius or calcaneus is more cost-effective, using single-source x-ray absorptiometry, and can be used for screening. It is anticipated that ultrasonography will prove to be a low-cost, effective method for bone mass assessment.[536] Ultrasonographic measurements of the calcaneus have been reported to be as accurate as femoral neck measurements by DEXA in predicting the risk of hip fractures.[537, 538] However, peripheral measurements are not accurate for monitoring response to treatment.

T Score — Standard deviations between patient and average peak young adult bone mass. The more negative, the greater the risk of fracture.

Z Score — Standard deviations between patient and average bone mass for same sex, age, and weight. A Z score lower than −2.0 (2.5% of normal population of same age) requires diagnostic evaluation for causes other than postmenopausal bone loss.

Definitions Based on Bone Mineral Density		
Normal	-	0 to –1 SD from the T-score reference standard (84% of the population)
Osteopenia	-	T-score –1 to –2.5 SD
Osteoporosis	-	T-score below –2.5 SD

The clinical relevance of a bone density measurement in a postmenopausal woman is estimated by using the T Score. For younger women, interpretation utilizes the Z Score.

Diagnostic Tests

Patients with osteoporosis should be screened for other conditions that lead to osteoporosis:

1. Serum parathyroid hormone, calcium, phosphorus, and alkaline phosphatase: for primary hyperparathyroidism.

2. Renal function tests: for secondary hyperparathyroidism with chronic renal failure.

3. Blood count and smear, sedimentation rate, protein electrophoresis: for multiple myeloma, leukemia, or lymphoma.

4. Thyroid function tests: for hyperthyroidism and excessive thyroid hormone treatment.

5. Careful history and, when indicated, appropriate laboratory studies to rule out long-term use of bone-losing medications, alcohol abuse, metastatic cancer, and chronic liver disease.

6. Serum 25-hydroxyvitamin D for vitamin D deficiency.

The presence of osteomalacia, to be suspected in all elderly patients with osteoporosis, can be detected by measuring the serum calcium, phosphorus, and alkaline phosphatase. These are all normal in patients with osteoporosis.

The effect of excess thyroid hormone on bone is not entirely clear. Although it is recognized that excess thyroxine treatment can cause bone loss, retrospective studies of thyroid function and bone mass have not produced uniform conclusions.[539] In the Study of Osteoporotic Fractures, women with a previous history of hyperthyroidism had an increased risk of subsequent hip fractures, and women taking thyroid hormone also had an increased risk of fracture (which did not reach statistical significance).[540] However, in a prospective assessment of TSH levels in the Study of Osteoporotic Fractures, no association could be detected between low TSH and bone loss.[541] Perhaps abnormal thyroid levels affect bone quality rather than quantity, and until accurate fracture data become available with this issue, it is prudent to maintain TSH levels within the normal range.

Should Osteopenia Be Treated?

Osteopenia as defined by bone density is very prevalent. Approximately 40% of American postmenopausal women have bone densities in the range of osteopenia. It is logical to expect a continuum of fracture risk extending from normal bone density to osteoporosis, and a longitudinal American study observed that osteopenia was associated with a 1.8-fold higher rate of fracture (compared with a 4-fold higher rate with osteoporosis).[542] Experts argue, however, that most women with osteopenia simply have lower than average bone mass (the osteopenic value represents the peak bone mass attained by that individual) and will not have progressive bone loss, and that in the absence of risk factors, treatment depends on the demonstration of continued bone loss with serial bone density measurements. Despite statistical adjustments, it is likely that the increased risk of fracture noted in the American cohort study was concentrated in individuals with risk factors. Certainly we should focus on lifestyle, diet, and calcium and vitamin D supplementation in women with osteopenia. ***We support preventive drug therapy for osteopenia in the presence of one or more risk factors for osteoporosis or documented progressive bone loss, or with a 10-year hip fracture probability of 3% or higher or a 10-year fracture probability of 20% or higher for any osteoporosis-related fracture according to FRAX.***

FRAX, the World Health Organization Fracture Risk Assessment Tool

The World Health Organization fracture risk assessment tool is known as FRAX, updated and calibrated according to risk in different geographic areas and available for use online

from the National Osteoporosis Foundation (http://www.nof.org/frax_update.htm) or directly from the FRAX website (http://www.shef.ac.uk/FRAX/).[543, 544] This algorithm predicts the 10-year major fracture risk and fracture probability, the likelihood of an older individual to experience over a period of 10 years a fracture because of low bone mass. The National Osteoporosis Foundation concluded that treatment is cost-effective if FRAX predicts a 10-year hip fracture probability of at least 3%, or an overall risk of any osteoporotic fracture greater than 20%.[545] The clinical application of FRAX produces no major changes in already existing guidelines.[546] An osteoporotic bone density and historical risk factors continue to be important factors indicating the need for treatment. FRAX is useful in identifying which individuals with osteopenia merit treatment; nevertheless, a careful evaluation of risk factors as reviewed in this chapter continues to be a reliable method to indicate fracture risk in an individual with osteopenia. For example, long-term use of corticosteroid or aromatase inhibitor treatment is an indication for bone loss prevention regardless of the bone density. In addition, FRAX utilizes bone density at the hip, but not at the spine. Treating a patient with low bone density in the spine but with normal hip bone requires careful consideration of the individual's history and risk factors. FRAX does not replace good medical judgment.

Biochemical Markers of Bone Turnover

There are many serum and urinary biochemical markers of bone turnover. Markers of bone formation include serum levels of osteocalcin, total and bone alkaline phosphatase, and procollagen peptide. Bone resorption is indicated by changes in urinary calcium, hydroxyproline, pyridinoline and deoxypyridinoline cross-links of collagen, telopeptides of collagen, and serum cross-linked telopeptides.

The urinary tests use a second morning voided urine specimen, and the immunoassays measure hydroxyproline, pyridinoline, deoxypyridinoline, the collagen cross-linked N-telopeptides, or the collagen cross-linked C-telopeptides.[547] These are structural proteins released by osteoclastic resorption of bone. A measurement of these markers provides a snapshot in time, the current state of bone metabolism. Because of individual variability, marker measurements correlate poorly with bone mineral density; however, changes in these markers quickly indicate responses to estrogen therapy.[548] Therefore, it has been proposed that efficacy of therapy (with either hormones or bisphosphonates) can be judged by comparing a baseline value with a single follow-up measurement after 1–3 months of treatment.[549] A lesser decrease in the urinary cross-linked peptides indicates a less than optimal bone response, identifying patients for further evaluation and specific added treatment. How reliable, practical, and cost-effective this approach is for routine clinical use is a reasonable question. ***Single, random measurements are not useful in predicting future bone loss, and bone turnover markers are insensitive indicators of response to raloxifene and calcitonin. Measurement of bone turnover markers is best confined to clinical trials, except in special situations as noted in this chapter.***

Hormone Treatment

Postmenopausal hormone therapy effectively reduces the number of all osteoporotic fractures, a conclusion now documented by randomized, placebo-controlled clinical trials.[550–552] The Women's Health Initiative (discussed in detail in Chapter 18) reported fracture outcomes based on an average of 5.6 years of follow-up, comparing placebo treatment with daily 0.625 mg conjugated estrogen and 2.5 mg medroxyprogesterone acetate (similar results were also reported in the estrogen-only arm of the clinical trial)[552, 553]:

WHI—Fracture Outcomes			
	Estrogen-progestin	*Placebo*	*Hazard Ratio*
Osteoporotic fractures	733 cases	896 cases	0.76 (0.69–0.83)
Spinal fractures	41	60	0.65 (0.46–0.92)
Hip fractures	52	73	0.67 (0.47–0.96)
Lower arm/wrist fractures	189	245	0.71 (0.59–0.85)

There are reasons to believe that the impact of hormone therapy is greater than the good results reported by the Women's Health Initiative (WHI). The Kaplan-Meier estimates of the impact of hormone therapy indicated that the reduction in fractures continued to increase over time, suggesting that a very powerful effect would be achieved with treatment of a long duration. In addition, spinal fractures included only clinically symptomatic fractures (known to represent about one-third of vertebral fractures); again the overall effect was probably greater if all vertebral fractures had been included. The beneficial effect was achieved even though the use of a bisphosphonate increased from 1% at baseline to about 6% in the estrogen-progestin group and 10% in the placebo group. The effect would have been even greater if bisphosphonates had not been prescribed by the participants' clinicians. The reduction in fractures was also underestimated because the WHI participants were not selected to obtain a group of women at high risk for fracture, but on the contrary, these women overall were at low risk for fractures. The report stated that hormone therapy decreased the risk of hip fracture by 60% among women with adequate calcium intake at baseline, but not in those with a lower intake. This underscores the importance of emphasizing to patients the consequences of inadequate calcium and vitamin D intake. The impact of hormone therapy was greater in leaner women as one would expect. ***Thus, the results of the WHI provide us with clinical trial evidence for a therapy that protects against fractures in a population not selected for having osteoporosis.*** Hormone therapy use in the U.S. declined by at least 50% following the initial publications from the WHI. This decrease was paralleled by a significant increase in the incidence of fractures among postmenopausal women![554]

Estrogen therapy stabilizes the process of osteoporosis or prevents it from occurring. Besides inhibiting osteoclastic resorption activity, estrogen increases intestinal calcium absorption, increases 1,25-dihydroxyvitamin D (the active form of vitamin D), increases renal conservation of calcium, and supports the survival of osteoblasts. With estrogen therapy one can expect a 50–60% decrease in fractures of the arm and hip,[555–558] and when estrogen is supplemented with calcium, an 80% reduction in vertebral compression fractures can be observed.[559] This reduction is seen primarily in patients who have taken estrogen for more than 5 years.[560, 561] ***Protection against fractures wanes with age, and long-term estrogen use is necessary to maximally reduce the risk of fracture after age 75.***

Because most osteoporotic fractures occur late in life, women and clinicians must understand that the short-term use of estrogen immediately after menopause cannot be expected to protect against fractures in the seventh and eighth decades of life. Some long-term protection is achieved with 7–10 years of estrogen therapy after menopause, but the impact is minimal after age 75.[562] In a prospective cohort study of women 65 years of age and older, in the women who had stopped using estrogen, and in those who were over 75 and had stopped using estrogen even if they had used estrogen for more than 10 years, there was no substantial effect on the risk for fractures.[563] The effective impact of estrogen requires initiation within 5 years of menopause and for current use to extend into the elderly years. The protective effect of estrogen rapidly dissipates after treatment is stopped because estrogen withdrawal is followed by rapid bone loss. In the 3- to 5-year period following loss of estrogen, whether after menopause or after cessation of estrogen therapy, there is an accelerated loss of bone.[564–568] The PEPI trial, however, found that this rate of bone loss was comparable to that in women

not on hormone therapy,[569] but in a Swedish case-control study, most of the beneficial effect of hormone therapy was lost 5 years after discontinuing treatment.[558] An American randomized trial concluded that elderly women who stopped taking hormones lost most of the bone density gained while on hormone therapy.[570] Two reports documented that less than 4 years of exposure to estrogen is followed by a period of time, perhaps 10 years, when bone density is better and fractures less prevalent than nonusers of hormone therapy.[571, 572] This protection is not as great as that obtained with long-term hormone use, and beneficial effects should not be expected after age 70 when fractures are most prevalent.

Maximal protection against osteoporotic fractures, therefore, requires lifelong therapy; even some long-term protection requires 10 or more years of treatment, and some protection against fractures is lost within 5 years of discontinuation.[573, 574] Standard doses of estrogen administered transdermally (50 μg) protect against fractures as well as standard oral doses.[558] A study of women randomized to treatment either with continuous transdermal delivery of estradiol 50 μg daily or oral estrogen demonstrated that both equally prevented postmenopausal bone loss.[575]

For many years, it was believed that estrogen therapy would either prevent or slow bone loss, but not produce a gain in bone density. Modern studies indicate that this is not the case. For example, in the PEPI trial, at the end of 3 years, the women receiving hormone treatment had experienced about a 5% gain in bone mineral density in the spine and 2% in the hip compared with approximately a 2% loss in the placebo group.[576] How long does this gain in bone continue? In a 10-year follow-up study of women receiving estrogen-progestin therapy, the spinal bone density steadily increased, reaching a level 13% over baseline after 10 years of treatment.[577] In another cohort study, hormone users did experience a loss of bone at the hip over a 10-year period of time; however the incidence of hip, wrist, and vertebral fractures was significantly reduced in the hormone users.[578]

The effect of starting treatment later in life is a controversial issue. The idea of postponing treatment to prevent osteoporosis until later in life has merit. Changes in bone density in the early postmenopausal years have no major effect on fractures later in life, except in individuals who already have low bone density. This amounts to only 5% of women in their early postmenopausal years, and most of this 5% will have risk factors such as smoking, fracture at a younger age, a thin body, or excessive alcohol consumption. The positive impact of hormone therapy on bone has been demonstrated to take place in women over age 65, and even over age 75.[560–563, 579–581] This is a strong argument in favor of treating very old women who have never been on estrogen. Estrogen use between the ages of 65 and 74 has been documented to protect against fractures.[557] However, a large cohort study could detect no significant impact on nonspinal fractures.[563] In a later analysis of this same cohort, some reduction in fracture risk was observed in women starting hormone therapy after age 60, but a statistically significant reduction was present only in those who started treatment before age 60 and remained on it.[582] Until we have better data, we should continue to promote early onset and prolonged use of hormone therapy if the objective is maximal protection against osteoporotic fractures.

Studies have demonstrated that a dose of 0.625 mg of conjugated estrogens is necessary to preserve bone density.[583] The conventional wisdom has stated that an estradiol blood level of 40–60 pg/mL is required to protect against bone loss.[584, 585] We now know that any amount of estrogen can have an impact, although it is very likely that some degree of protection is lost when doses are less than the equivalent of 0.625 mg conjugated estrogens or 1.0 mg estradiol.

The rate of bone loss and the incidence of hip and vertebral fracture is inversely related to the circulating estrogen levels in older women.[586, 587] Estradiol levels as low as 10 pg/mL have a beneficial impact on bone density and fracture rates compared with values below 5 pg/mL. ***Thus, any increment in estrogen, even within the usual postmenopausal range, will exert protective effects.*** This explains how a positive effect on bone was observed even

with the utilization of the vaginal ring that delivers a very small amount of estradiol with minimal systemic absorption.[588]

A lower dose of 0.3 mg daily of conjugated estrogens or 0.5 mg estradiol prevented loss of vertebral trabecular bone when combined with calcium supplementation (to achieve a total intake of 1,500 mg daily).[589–593] In a small study without calcium supplementation, the daily administration of 0.3 mg conjugated estrogens and 2.5 mg medroxyprogesterone acetate produced a slight increase in lumbar bone density with a lesser effect on the hip.[594] Even a dose of estradiol as low as 0.25 mg/day produced an increase in bone density.[595]

The Women's Health, Osteoporosis, Progestin, Estrogen (HOPE) trial involved 800 postmenopausal women in a dose-response evaluation of conjugated estrogens and medroxyprogesterone acetate. The lowest dose, 0.3 mg conjugated estrogens either unopposed or combined with 1.5 mg medroxyprogesterone acetate, produced a gain in bone density.[596] A major concern with lower doses is the possibility that there will be a significant percentage of nonresponders (discussed later). Nevertheless, a lower dose of estrogen may be more acceptable (fewer side effects) in elderly women. ***Patients electing to be treated with lower doses should have follow-up assessments for response with measurements of either bone density or urinary biochemical markers.***

While progestational agents are considered antiestrogenic, they have been reported to act independently, in a manner similar to estrogen, to reduce bone resorption.[597] When added to estrogen, progestins can lead to an apparent synergistic increase in bone formation associated with a positive balance of calcium.[598–601] On the other hand, good studies have failed to find a greater impact on bone, comparing estrogen alone to estrogen plus a progestin.[563] These different results are explained by the fact that the synergistic result of combining estrogen with a progestin is determined by the type of progestin, being limited to members of the 19-nortestosterone (norethindrone) family.[602] This could reflect an increase in free estrogen levels because of a reduction in sex hormone-binding globulin. Careful studies indicate that the addition of medroxyprogesterone provides an additional effect on bone only in women with established, significant osteoporosis.[601, 603, 604]

In the PEPI trial, at the end of 3 years, there was little difference comparing treatment with estrogen only to the groups receiving either sequential or continuous combinations of estrogen and medroxyprogesterone acetate or micronized progesterone.[576] Thus, the daily, continuous combination of estrogen-progestin is equally efficacious in maintaining bone density as the sequential regimens, although at least one study has indicated a slightly greater response in the lumbar spine with continuous treatment.[602, 605–607]

The addition of testosterone to an estrogen therapy program has been reported to provide no additional beneficial impact on bone or on relief from hot flushes.[608, 609] Others have demonstrated a greater increase in bone density with an estrogen-androgen combination compared with estrogen alone, although the blood estrogen levels achieved were higher than those associated with standard postmenopausal hormone therapy.[610] In another study, only a very pharmacologic dose of methyltestosterone added to the bone density achieved with estrogen alone.[611] ***Testosterone should not be prescribed with the expectation that androgen treatment will improve bone health.***

Follow-up Assessment of Hormone-Treated Women

Either a stable bone density or an increase indicates successful treatment; however, not all women will maintain or gain bone density on postmenopausal hormone therapy; in one study, 12% of treated women lost bone despite apparently good compliance.[612] In the PEPI 3-year clinical trial, where compliance rates were probably maximal, 4% of treated women

lost bone in the spine and 6% in the hip.[576] *It is worthwhile to measure the bone density in treated women when they are in their late 60s*. This problem of poor response is discussed near the end of this chapter.

Treatment with Estrogen Agonists-Antagonists

Raloxifene

Estrogen agonists-antagonists can have selective actions on specific target tissues. Raloxifene exerts no proliferative effect on the endometrium but produces favorable responses in bone and lipids.[613, 614] The changes in bone remodeling produced by raloxifene are consistent with an estrogen agonist effect.[615] As with estrogen therapy, discontinuation of raloxifene treatment is followed by a resumption of bone loss.[616]

The increase in bone density associated with raloxifene, 60 mg daily, is less than that seen with alendronate.[617–619] The MORE (Multiple Outcomes of Raloxifene Evaluation) study of raloxifene administration to osteoporotic women reported results from 8 years of follow-up.[620, 621] Women with low T-scores had approximately a 50% reduction in vertebral fractures with raloxifene treatment, and with previous vertebral fractures, approximately 35%. However, ***there has been no evidence of a reduction in hip or wrist fractures.*** Although the reduction in vertebral fractures is similar to that seen with alendronate and estrogen, why is there no decrease in hip fractures, despite a bone density response that is only slightly less than that associated with alendronate? There are at least two possible explanations. The study was of insufficient duration to demonstrate an impact in a population that was relatively young for hip fractures. Or a weaker impact on bone density produces a lesser effect in the hip, which has a combination of cortical and trabecular bone that is less responsive compared with the spine with its large content of trabecular bone.

A reduction in vertebral fractures equivalent to that with estrogen or bisphosphonates despite a lesser increase in bone density is strong evidence that fracture risk and reduction are not simply reflections of bone density. Thus a few percentage point differences comparing bone densities between two therapies do not translate into a difference in fracture protection. For this reason, fracture data are important, and we cannot conclude that small differences predict lesser or better performances in fracture protection. Another example is the greater gain in bone density associated with combination treatment with alendronate and estrogen; a greater gain in bone density does not necessarily mean a greater protection against fractures. ***In our view, raloxifene is an option for prevention of osteoporosis-related spinal fractures, especially for patients reluctant to use hormone therapy. We recommend, however, periodic evaluation of bone density in the hip, and if bone loss occurs, patient and clinician should consider another treatment option.***

Bazedoxifene

Bazedoxifene belongs to the estrogen agonist-antagonist family of drugs. It has favorable effects on bone and lipids, but does not affect the endometrium or the breast. Bazedoxifene in a dose of 20 mg daily decreased the risk of all clinical fractures in a randomized clinical trial, with a potency comparable to other anti-resorptive agents in postmenopausal women at high risk for fractures.[622, 623] In a subgroup of women at higher risk for fractures, bazedoxifene had a reduced risk of nonvertebral fractures (50% reduction with 20 mg), compared with *both* raloxifene and placebo. The only adverse event that differed with treatment was an increase in

venous thrombosis with treatment compared with placebo. The results of this trial indicate that the effect of bazedoxifene on bone should be comparable to that of estrogen and bisphosphonates.

The reduction of nonvertebral fractures with bazedoxifene compared with raloxifene should not be ignored. We have known for some time that even with 8 years of follow-up, raloxifene has no impact on the risk of hip fractures. This is likely because raloxifene is less potent, and thus the hip with a mixture of cortical and trabecular bone is more resistant to raloxifene's effects, compared with the spinal column that is composed of sensitive, trabecular bone. Bazedoxifene partnered with estrogen is called TSEC (tissue-selective estrogen complex). The idea is to gain the benefits of estrogen (bazedoxifene has little impact on hot flushes), protect the endometrium and possibly the breast, and enhance some actions of estrogen, such as a reduction in fractures. This approach to postmenopausal hormone therapy may eliminate the need for progestational agents.

Drugs in Development

Multiple drugs in the family of selective estrogen agonists/antagonists have potential applications for osteoporosis prevention and treatment. These include droloxifene, idoxifene, ospemifene, arzoxifene, lasofoxifene, and ormeloxifene, in addition to bazedoxifene. The drugs closest to clinical use are discussed in Chapter 18.

Calcium Supplementation

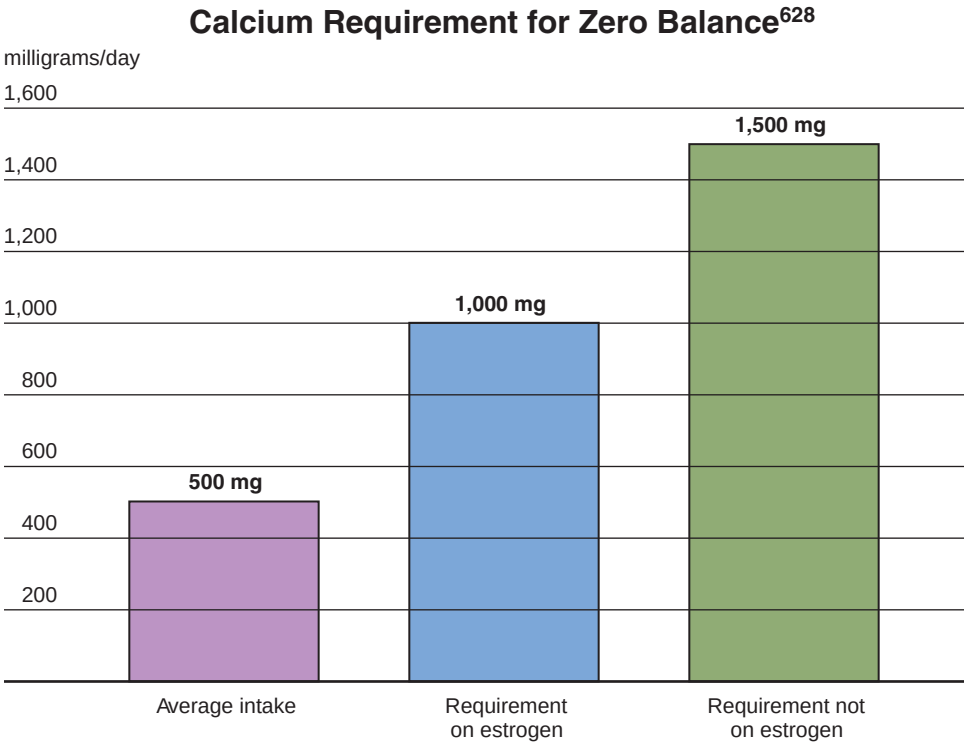
There has been considerable confusion over whether calcium supplementation by itself can offer protection against postmenopausal osteoporosis. This is partly due to the fact that calcium studies were performed in women who were in the very early postmenopausal years, in the midst of the rapid loss of calcium associated with estrogen deficiency, and this estrogen effect overwhelmed any responses to calcium. Studies that involved women beyond this early stage of the postmenopausal period definitely indicated a positive impact of calcium supplementation.[624–627]

Calcium absorption decreases with age because of a decrease in biologically active vitamin D and becomes significantly impaired after menopause. A positive calcium balance is mandatory to achieve adequate prevention against osteoporosis. Calcium supplementation (1,000 mg/day) reduces bone loss and decreases fractures, especially in individuals with low daily intakes.[628, 629] Estrogen acts to improve calcium absorption by increasing the levels of 1,25-dihydroxyvitamin D and makes it possible to utilize effective supplemental calcium in lower doses.

SUMMARY—Calcium Supplementation

1. In order to remain in zero calcium balance, women on estrogen therapy require a total of 1,000 mg elemental calcium per day.[628, 630, 631]
2. Because the average woman receives about 500 mg of calcium in her diet, the minimal daily supplement for women on estrogen equals an additional 500 mg.
3. Women not on estrogen require a daily supplement of at least 1,000 mg calcium to reach the recommended intake of 1,500 mg/day.

High-dose calcium supplementation can unmask asymptomatic hyperparathyroidism, causing abnormally high calcium blood levels and an increased risk for renal stones. Women receiving calcium supplementation in excess of 500 mg daily should have their blood levels of calcium and phosphorus measured yearly for the first 2 years. If normal, no further surveillance is necessary.



Even with the commonly used therapeutic doses of calcium, nearly 40% of postmenopausal women have inefficient absorption.[632] Estrogen improves calcium absorption and allows the utilization of supplemental calcium in effective doses without the side effects associated with higher doses (constipation and flatulence) that diminish compliance. We must emphasize that although calcium supplementation is important, it cannot provide the same degree of protection against osteoporosis as that achieved by hormonal therapy.[633, 634] Nevertheless, the beneficial impact of estrogen on bone is reduced in the absence of calcium supplementation.[590]

Improved calcium intake in adolescents results in significant increases in bone density and skeletal mass, providing protection against osteoporosis later in life.[635, 636] Calcium supplementation is far more important during adolescence than in the reproductive years when bone formation is minimal. Under age 25, during the years of bone accumulation, the daily calcium intake should be 1,500 mg.[628] This amount, 1,500 mg/day, is also recommended during pregnancy and lactation. Most calcium comes from dairy products; relying on other foods is not easy because it requires a high intake volume of other foods to provide the same amount of calcium in normal daily servings of dairy products. Foods with high amounts of oxalate and phytate, such as spinach, rhubarb, beans, peas, wheat bran, and beet greens, reduce calcium absorption.

Calcium Content of Foods

Yogurt (one cup)	415 mg
Yogurt with fruit (one cup)	345 mg
Juice fortified with calcium (one cup)	300 mg
Milk (one cup)	300 mg plus 100 IU Vitamin D
Ice cream (one cup)	175 mg
Cottage cheese (1 cup)	140 mg
Romano cheese (1 oz)	300 mg
Parmesan cheese (1 oz)	335 mg
Cheddar cheese (1 oz)	205 mg
Swiss cheese (1 oz)	270 mg
Mozzarella cheese (1 oz)	207 mg
Tofu (one cup)	150 mg
Broccoli, cooked (one cup)	140 mg
Beans, cooked (one cup)	80 mg

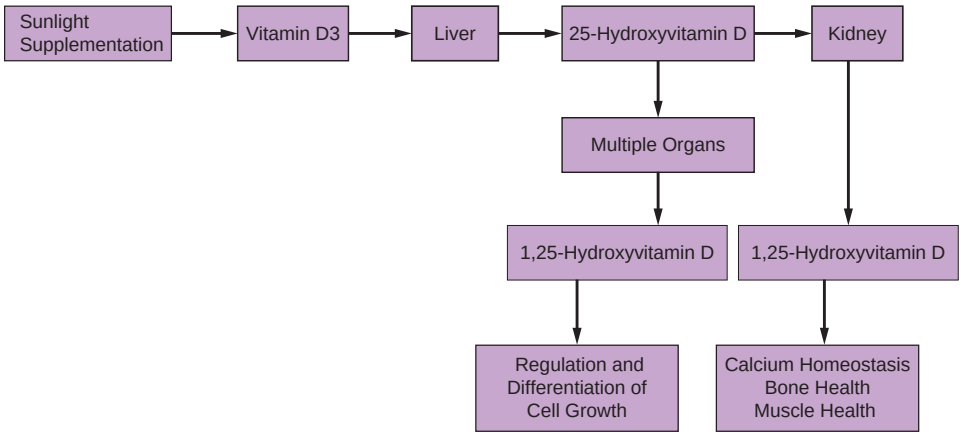
There are dozens of calcium supplements on the market, containing calcium carbonate, calcium lactate, calcium phosphate, or calcium gluconate. Calcium carbonate tablets are the cheapest and contain the most elemental calcium (40%). Calcium lactate tablets contain 13% calcium, calcium citrate 23%, and calcium gluconate only 9%. The calcium carbonate antacids are excellent, inexpensive sources. Be aware that aluminum-containing antacids such as Maalox, Mylanta, Gelusil, and Riopan can inhibit gastrointestinal absorption of calcium. Bone meal and dolomite as sources of calcium should be avoided because they are contaminated with lead. Calcium carbonate derived from "natural" sources (such as oyster shell), name brand formulations, and even "refined" products can also contain lead.[637] It is worth looking for specific products labeled as tested and containing no lead, having "purified" or "USP Verified Mark" on the label. Calcium citrate does not require gastric acid for absorption and is the best choice for older patients with reduced gastric acid production. Calcium supplementation is most efficient when single doses do not exceed 500 mg and when taken with a meal. Excess calcium supplementation (especially not with meals) is associated with a slight increase in risk for kidney stones.[638]

Vitamin D

Osteoporosis related to aging is due significantly to age-related changes in vitamin D and calcium metabolism.[639, 640] There is an age-related decrease in the ability of tissues to convert the major circulating form of vitamin D, 25-hydroxyvitamin D, to the active form of vitamin D (1,25-dihydroxycholecalciferol, better known as 1,25-dihydroxyvitamin D), and there is a decrease in the ability of the intestine to absorb dietary vitamin D. Exposure of the skin to ultraviolet rays in sunlight stimulates the formation of cholecalciferol (vitamin D3). One dose of sunlight that produces erythema is equivalent to 10,000 to 25,000 IU of vitamin D. Excessive exposure to sunlight cannot cause toxic levels of vitamin D because any excess in vitamin D3 is inactivated by ultraviolet radiation.

Vitamin D2 (ergocalciferol) in commercial supplements has been replaced with the more effective vitamin D3 (cholecalciferol), which is more than three times the potency of vitamin D2.[641] Cholecalciferol and ergocalciferol are converted in the liver to calcitriol

(25-hydroxyvitamin D), the major circulating metabolite that is active in stimulating the absorption of calcium and phosphate from the gastrointestinal tract. The circulating level of 25-hydroxyvitamin D and renal conversion to its active form, 1,25-dihydroxyvitamin D, are regulated by calcium, phosphorous, and parathyroid hormone.



There is now sufficient evidence to recommend that individuals should add 1,000–2,000 units of vitamin D to calcium supplementation. Older studies that failed to document a beneficial impact of vitamin D supplementation were handicapped by doses that were too low and inadequate calcium intake.[642] A pooled analysis of 68,500 patients in seven major vitamin D fracture trials concluded that vitamin D in appropriate doses coupled with calcium supplementation reduced fractures at all sites in men and women.[643]

Because adequate and active vitamin D depends on cutaneous generation mediated by sun exposure, women in the winter months can easily be relatively vitamin D deficient and lose bone.[644] In far northern and southern areas, the winter sunlight is inadequate to stimulate dermal activation. But even, in areas of the world where there is adequate sun exposure, inadequate dietary intake and an indoor lifestyle yield a significant number of women with abnormally low levels of circulating 25-hydroxyvitamin D.[645] In addition, clothing and sunscreens prevent the cutaneous production of vitamin D3. The National Health and Nutrition Examination Survey (NHANES) documented a rising prevalence of vitamin D insufficiency in the U.S. over the last two decades.[646] ***If uncertain regarding vitamin D supplementation, the serum level of 25-hydroxyvitamin D can be measured; a level below 30 ng/mL is abnormal.***[647]

The benefit of vitamin D supplementation is clear in older postmenopausal women, and the lack of side effects encourages us to recommend vitamin D supplementation as part of the overall program for osteoporosis prevention in younger postmenopausal women. But keep in mind that effective bone response to vitamin D requires adequate calcium intake.

One caution regarding vitamins: excessive intake of vitamin A, and specifically retinol, was reported to be associated with an increased rate of hip fractures in the women not using hormone therapy in the Nurses' Health Study; it would be better to avoid obtaining vitamin D by taking more than one multivitamin tablet daily.[648] It has been argued that excess retinol intake may explain the higher rate of fractures in Scandinavia.[649]

The Women's Health Initiative (WHI) conducted a randomized trial of calcium and vitamin D supplementation.[650] The 36,282 postmenopausal women were part of the WHI clinical trials involving postmenopausal hormone therapy or dietary modification. The average follow-up

was 7 years. Thirty-seven percent of the women were age 50 to 59, 45.5% were 60 to 59, and 17.5% were 70 to 79. The treated group was supplemented with 1,000 mg calcium and 400 IU vitamin D daily. Overall analysis indicated no significant reduction in fractures with calcium/vitamin D treatment. The women treated with calcium/vitamin D had a 17% greater risk of kidney stones. However, a closer look reveals some good news:

- Correcting for compliance by analyzing just those women who continued their medication revealed a statistically significant 29% reduction in risk for hip fractures. Only 59% of the treated women at the end of the trial were taking the intended dose.
- Women who were 60 years and older had a 21% significant reduction in hip fractures.
- The reduction in hip fractures was greatest (42%) in those women who combined calcium/vitamin D supplementation with postmenopausal hormone therapy.

The population at greatest risk for fractures (the oldest women in the WHI study) actually benefited, and the study confirmed something already known, that hormone therapy combined with calcium/vitamin D supplementation achieves the best results. Keep in mind that the women in this study were not at high risk for fractures. Indeed, whole body and spinal bone density increased in the placebo group. This is hard to explain; the average postmenopausal untreated woman loses spinal bone density. In this study, only hip bone density demonstrated a loss, and thus, it is not surprising that significant benefits were demonstrated only with hip fractures. The fact that most of these women were overweight probably contributed to the protection against bone loss in the spine. In a population of women losing bone density in both hip and spine, and in women with other risk factors for fractures, calcium/vitamin D supplementation should yield even better results than those reported by the WHI, including a reduction in spinal and arm fractures.

This same WHI study also assessed the impact of calcium/vitamin D on the risk of invasive colorectal cancer.[651] No difference was observed between the treated group and the placebo group, even when only women adherent to treatment were analyzed. However, it is recognized that the latency period for colorectal cancer is 10 to 20 years. The length of follow-up in this study may have been insufficient to detect an effect. Furthermore, colorectal cancer was not a primary outcome in the study design, and the study design was very complicated by the fact that the women were simultaneously enrolled in three overlapping trials (calcium/vitamin D, low-fat diet, and hormone therapy).

The impact of calcium/vitamin D supplementation on the risk of colorectal cancer remains unsettled. A possible reduction in colorectal cancer is still possible in those women who have low levels of calcium and vitamin D prior to treatment, as documented in the Nurses Health Study and as was the case in the WHI report.[651, 652] A 50% reduction in the risk of colorectal cancer has been observed comparing normal blood levels of 25-hydroxyvitamin D with the lowest levels.[653, 654]

What about the kidney stones? The women in the WHI trial were allowed to continue their own programs of supplementation. Thus many took calcium and multivitamins (which contain 400 IU vitamin D). The average daily calcium intake of the study population was 1,100 to 1,200 mg, 2-fold higher than the average American woman. The WHI does not provide data to answer this most important question: was the small increase in kidney stones observed in women who were taking excessive amounts of calcium and vitamin D?

The WHI has published multiple findings from the calcium/vitamin D randomized trial. The conclusions are uniformly negative, summarized as follows:

- The incidence of invasive breast cancer was similar in the treated and placebo groups. The risk of breast cancer was not associated with baseline 25-hydroxyvitamin D levels.[655]

- Calcium and vitamin D supplementation did not protect against a decline in physical functioning or performance.[656]
- The incidence of newly diagnosed diabetes was the same in the treated and placebo groups.[657]
- Myocardial infarction and stroke events were similar in the treated and placebo groups.[658]
- Calcium and vitamin D supplementation had no effect on blood pressure or the risk of developing hypertension.[659]

The Women's Health Initiative also reported the effects on overall mortality in their randomized trial of calcium and vitamin D supplementation.[660] After an average follow-up of 7 years, there was no impact on total mortality (744 deaths in treated women and 807 deaths in the placebo group). There was a nonsignificant reduction (about 10%) in risks for stroke and cancer mortality. Comparing the women older than 70 to those younger than age 70, there was a lower risk for mortality in the younger group, but this did not achieve statistical significance. When analysis was restricted to those participants who were adherent to therapy, the results for overall mortality and for the younger women were essentially unchanged. However, a nested case-control study assessed the results according to serum 25-hydroxyvitamin D levels at baseline; the results indicated that women with the lowest levels had a higher risk for death. The WHI investigators concluded that their results weakly support the hypothesis that calcium and vitamin D supplementation modestly reduces cancer and cardiovascular mortality. Overall, the results indicated no major benefit of calcium and vitamin D supplementation on the risk of death due to cancer or cardiovascular disease.

Smaller trials and observational studies of vitamin D supplementation have demonstrated reductions in total mortality, reductions in blood pressure and cholesterol levels, and reduced risks of stroke and coronary heart disease. Vitamin D insufficiency has been linked to a wide spectrum of problems, including an increase in autoimmune diseases such as rheumatoid arthritis, upper respiratory tract infections, diabetes mellitus, multiple sclerosis, cancer, and cardiovascular disease. Low levels of serum 25-hydroxyvitamin D have been reported to be associated with higher all-cause mortality, a greater prevalence of breast cancer, peripheral arterial disease, and higher cardiovascular mortality.[661–665] It is proposed that many tissues produce locally the active 1,25-dihydroxyvitamin D that exerts beneficial effects on cell regulation and differentiation before being inactivated within cells without reaching the circulation and interfering with calcium metabolism.[640]

What could explain the lack of strong agreement between the WHI results and the rest of the literature? The answer can be found in the characteristics of the women in the WHI and the doses of calcium and vitamin D used for supplementation. The women in the WHI were not at high risk for fractures, and most of the women were overweight. Only one-third of the participants had a low calcium intake at entry to the study, and 29% were already taking calcium supplementation. As previously noted, the average daily calcium intake of the study population was 1,100 to 1,200 mg, 2-fold higher than the average American woman.[650] Low serum levels of 25-hydroxyvitamin D were found in only 25% of the women in whom it was measured.

Currently, the recommended doses of vitamin D for supplementation far exceed the dose of 400 IU used in the WHI trial. Doses as high as 1,500 to 2,000 IU/day are being suggested as the levels required to achieve a beneficial impact. But even with the lower dose used in the WHI, there was a trend for a reduction in cardiovascular and cancer mortality, although the impact was nonsignificant in the WHI population.

The negative publicity associated with the WHI publications has obscured the importance of fracture risk in women. In the observational arm of the WHI, the risk of fracture was compared to the risks of cardiovascular events and breast cancer in a prospective cohort of

83,724 women, aged 70–79.[666] ***The recorded events indicated that the number of women who experienced a fracture in one year's time exceeded the combined number of women experiencing invasive breast cancer or cardiovascular disease (except in blacks in whom cardiovascular events dominated).*** And don't forget that in the WHI the women adherent to their calcium/vitamin D supplementation had a 29% reduction in risk for hip fractures, and the reduction in hip fractures (42%) was greatest in those women who combined calcium/vitamin D supplementation with hormone therapy.[650]

Because 60% of individuals who live in northern latitudes have low serum levels of 25-hydroxyvitamin D, the WHI results do not rule out the possibility of beneficial effects of adequate supplementation in a large segment of our population.[667] The WHI nested case-control study documented that the risk of hip fracture steadily increases in women as their serum levels of 25-hydroxyvitamin D decrease.[668]

Vitamin D is essential for adequate calcium absorption and maximal bone protection, but remember that supplemental vitamin D requires adequate calcium intake.[642, 643] Calcium absorption decreases and parathyroid hormone increases as the levels of 25-hydroxyvitamin D fall below 30 ng/mL. It is now recognized that vitamin D together with calcium supplementation in older men and women reduces the rate of fractures.[633, 669, 670] In addition, evidence indicates that adequate vitamin D may help prevent type 1 diabetes mellitus, hypertension, multiple sclerosis, and many cancers.[667] Normal vitamin D levels are important for muscle functioning; vitamin D supplementation alone and in combination with calcium lowers the risk of falling and improves lower extremity function in older men and women.[671, 672]

SUMMARY—Measurement of 25-Hydroxyvitamin D

1. Clinicians should be more aggressive in monitoring serum levels of 25-hydroxyvitamin D. A value less than 30 ng/mL is below normal; less than 20 ng/mL is a definitive indication of vitamin D deficiency. We believe that this measurement should be part of every older individual's annual medical assessment.

2. The dose of vitamin D3 supplementation can be easily titered according to the circulating level, but remember that it takes about 3 months to reach a new steady state after a change in dose.

3. Patients who lose bone despite adequate treatment for bone loss should have this serum measurement because inadequate calcium and vitamin D can be the reason for the loss.

Given these facts, how can we determine how much vitamin D to give to an individual patient? A collection of bone experts reached a consensus by considering the amount of vitamin D supplementation required to change blood levels of 25-hydroxyvitamin D and parathyroid hormone, correlating this information with the levels of vitamin D required to prevent bone loss.[673] ***To maintain the optimal serum level of 25-hydroxyvitamin D, greater than 30 ng/mL, it is now recommended that men and women of all ages, but especially 60 years of age and older, need a supplement of 1,000 to 2,000 IU vitamin D3 daily.*** To restore a low level of 25-hydroxyvitamin D to normal, 50,000 IU vitamin D3 (prescription required) is administered once weekly for 6 months. Toxic doses of vitamin D (enough to produce hypercalcemia) are far beyond these recommendations.

Bisphosphonates

Bisphosphonates are effective in preventing bone loss by enhancing osteoclast apoptosis and inhibiting bone resorption. The bisphosphonates bind to bone mineral where they remain for many years, making bone less susceptible to osteoclastic action. The first generation of bisphosphonates (etidronate) also inhibited bone mineralization, and therefore intermittent therapy was necessary. The second generation of bisphosphonates allows bone formation to occur while inhibiting bone resorption and makes it possible to use continuous therapy rather than intermittent therapy.

Diagnostic screening with the laboratory tests previously noted is important before the administration of bone-specific drugs in order to avoid inappropriate treatment in the presence of a secondary cause for osteoporosis, especially renal disease.

Oral bisphosphonates must be taken on an empty stomach with a full glass only of water, no other liquid, at least 30 min before any other food or liquid intake in order to achieve adequate absorption. A failure to remain upright for at least 30 to 60 min and until after the first food intake of the day after ingesting bisphosphonates can result in esophageal injuries, such as esophagitis, esophageal ulcers, and esophageal erosions with bleeding.[674] Upper gastrointestinal problems in clinical trials (with carefully instructed and monitored subjects) are similar comparing alendronate treatment with placebo, indicating that improper usage is the culprit.[675, 676] Risedronate (Actonel), 5 mg daily, is as effective as alendronate (Fosamax) for the prevention of bone loss, provides similar protection against fractures and may be better tolerated.[677–681] Other equally effective bisphosphonates include ibandronate (Boniva) and zoledronic acid (Reclast, Aclasta). Notably, zoledronic acid administered as a single 5-mg dose intravenously protects against bone loss for at least 2 years.[682, 683]

The mechanism for the gastrointestinal reactions is an interference with the normal healing process that repairs trauma associated with eating. With less frequent exposure, this healing is allowed to proceed unimpeded. Periodic administration with equivalent bone efficacy is possible because of the high affinity of bisphosphonates for bone. Thus weekly administration of alendronate and risedronate, each in a dose of 35 mg for prevention and 70 mg for treatment, reduces side effects and produces similar increases in bone density compared with a daily regimen.[684–688]

In women with osteoporosis, alendronate (10 mg daily) administration reduced the risk of all subsequent fractures by 30% and vertebral fractures by 50% in 3–4 years of treatment. [689–691] In normal postmenopausal women, alendronate increased bone density in both the spine and the hip, and the 5-mg dose (the preferred dose for preventive treatment) was more effective than 2.5 mg.[692, 693] In the data derived from follow-up of 4,432 women for an average of 4.2 years, a statistically significant reduced risk of fracture was demonstrated only in women with initial T scores of −2.5 or less, a 36% reduction in all fractures and a 50% reduction in vertebral fractures.[694]

Bisphosphonate treatment obviously benefits women who already have a low bone density or previous vertebral fractures. Risedronate reduces vertebral and hip fractures in women with osteoporosis with an impact similar to that of alendronate.[679, 695] After 5 years of treatment with risedronate, the risk of new vertebral fracture decreased by 59%, an improvement from the 49% reduction reported after 3 years.[696] Ibandronate reduced vertebral fractures by about 60% over 3 years; zoledronic acid reduced vertebral fractures by 70% and hip fractures by 41% over 3 years.[697, 698]

The EPIC (Early Postmenopausal Interventional Cohort) study concluded that over a 4-year period of time, alendronate and hormone therapy in the U.S. produce similar bone density results. The greater increase noted in Europe with hormone therapy probably reflects the use

of 19-nortestosterone progestins, which are known to have an additive effect on bone density when combined with estrogen. Combining bisphosphonate and hormone therapy produces an added gain in bone density. When women who were already taking hormone therapy also received alendronate (10 mg) for 1 year, the gain in bone density ranged from 0.9% in the femoral neck to 2.6% in the spine.[699] In women with osteopenia, combined therapy with alendronate 10 mg and 0.625 mg conjugated estrogens produced a 1% to 2% greater gain in bone density over a 2-year period of treatment; similar results were reported in a 1-year trial with risedronate.[700, 701] ***By no means is it certain that this difference will translate into a difference in the incidence of fractures later in life. Indeed, it is unlikely. Furthermore, there is a theoretical concern that oversuppression of resorption can ultimately yield more brittle bones.***

Compliance with alendronate has been overestimated by the clinical trials. It is well-recognized that participants in clinical trials are better motivated, better supported, and perform better. In the Kaiser Permanente Medical Care Program in California, about one-third of patients had acid-related complaints, and one in eight required treatment.[702] About 50% of the Kaiser patients did not comply with instructions, and about 50% discontinued therapy by 1 year.[702, 703] Analysis of general pharmaceutical claims databases in the U.S. revealed that only 43% refilled their first prescription, and after 2 years, only 20% were adherent to treatment.[704] Bone density measurements are recommended to assess compliance and to provide motivation for continuation. Programs with good patient support have reported long-term compliance with hormone therapy: 65% at 7.5 years in an Australian population and 61% at 7 years in the U.K.[705, 706]

Because of the apparent benefits associated with alendronate therapy, the clinical trials were halted after 4 years, although follow-up has indicated persistent gains in bone mineral density through 10 years.[707] There is evidence that a bisphosphonate already in the bone can recirculate when bone containing the bisphosphonate is remodeled.[708] Thus, perhaps long-term treatment is unnecessary; the optimal duration of treatment has not been established. Comparing bone loss after discontinuation of treatment, accelerated bone loss occurs after estrogen and raloxifene, but a residual effect on bone density is maintained for up to 7 years after alendronate is discontinued.[616, 709–711]

Recommended Dosages

	For Prevention	*For Treatment*
Alendronate	5 mg daily; 35 mg weekly	10 mg daily, 70 mg weekly
Risedronate	5 mg daily	5 mg daily
	35 mg weekly	35 mg weekly
	75 mg daily for 2 days each month	75 mg daily for 2 days each month
	150 mg monthly	150 mg monthly
Ibandronate	2.5 mg daily	2.5 mg daily
		150 mg monthly
		3 mg i.v. every 3 months
Zoledronic acid	5 mg i.v. every 2 years	5 mg i.v. annually

How Long Should Bisphosphonate Treatment Be Continued?

In an extension of the alendronate Fracture Intervention Trial (FIT), the participants were randomized after 5 years of treatment either to another 5 years of treatment or placebo.[712] The group that discontinued alendronate treatment (5 or 10 mg/day) experienced small losses of

bone mineral density over 5 years. The levels, however, remained above the pretreatment levels 10 years previously. There were no differences in nonvertebral fractures between the treatment and placebo groups, but there was a 2-fold higher rate of clinically recognizable vertebral fractures in the placebo group (5.3% vs. 2.4%). It was concluded that most women do not need long-term treatment, and that long-term treatment should be limited to high risk women (women with existing vertebral fractures or very low bone densities).

The rate of hip fracture among women who discontinued bisphosphonate therapy was compared with women who remained on treatment.[713] The study group consisted of 9,063 users who had been compliant for at least 2 years. The hip fracture rate about doubled in the women who discontinued treatment compared with those who did not. After discontinuation for a year or longer, the risk of hip fracture increased by 9 months in women with low compliance rates. In women with high compliance rates for 2 or 3 years, there were no significant differences in fracture risk after discontinuation for up to 1 year later. Why is there this lingering effect?

The unique tight binding of bisphosphonates to bone matrix causes this drug to remain in the body for decades. This is believed to be the explanation for why there is no rapid bone loss after discontinuing bisphosphonate treatment in contrast to the rapid loss that follows the termination of estrogen therapy. This is also the reason why concerns have been raised regarding long-term treatment because when bone remodeling releases bound bisphosphonate, it is free to be active again, and as a result the endogenous bisphosphonate is added to the administered bisphosphonate, raising dosage exposure. At this time, we don't know the lowest effective dose and the lowest effective duration of exposure. The potential risk that has been long recognized is that prolonged exposure to bisphosphonates or excessive dosage would oversuppress bone resorption, thus oversuppressing bone turnover and affecting the biomechanical strength of bone; indeed, allowing microcracks to accumulate. A unique fracture of the femor, a transverse fracture in an area marked by hypertrophy of the cortex, has been associated with long-term bisphosphonate use.[714]

Bisphosphonates and Jaw Osteonecrosis

In an analysis of information from the Surveillance, Epidemiology, and End Results (SEER) databank linked to Medicare claims, 16,703 cancer patients were identified who were treated with IV bisphosphonates (pamidronate and zoledronic acid) from 1995 to 2003.[715] When 14,349 treated patients were matched with 28,698 controls, the treated group had a 3-fold increased risk of jaw or facial bone surgery and a very large increased risk of osteomyelitis of the jaw. The estimated absolute risk equaled 5.48 events per 100 patients over 6 years. In addition, the risk increased with increasing cumulative dose.

The link between bisphosphonates and jaw osteonecrosis is accepted even though the studies contained small numbers of cases. It is acknowledged that this is a relatively rare complication. The mechanism is uncertain beyond the recognition that infection and blood flow changes are involved. It is postulated that a compromised healing ability of bone because of inhibition of bone turnover leads to sequestered osteomyelitis and necrosis. Studies have suggested that the risk of osteonecrosis of the jaw is greater in patients treated with zoledronic acid compared with pamidronate.[716–718] Awareness of this problem has led to increased attention to oral hygiene and the avoidance of tooth extractions in the high risk population of cancer patients receiving this treatment.

The risk of osteomyelitis and osteonecrosis of the jaw and face has been recognized for several years, but continues to be controversial. To be sure, most of the cases have been in cancer patients, usually treated with intravenous high doses, but this complication was

also reported in patients receiving oral bisphosphonate treatment for osteoporosis, with no history or evidence of malignancy.[718, 719] Experts in the field point out that many clinicians specializing in osteoporosis have never seen a case; the incidence with oral bisphosphonates is somewhere from 1 in 10,000 to 1 in 100,000. They further argue that a true cause and effect relationship would require appropriate controlled studies, and two are on-going. But clinical decision-making cannot be delayed until data are available, and it is by no means certain that the on-going trials will yield definitive results on such a rare event. Patients should be cautioned regarding this rare problem and urged to practice good dental care.

Bisphosphonates and Atrial Fibrillation

A meta-analysis presented at the meeting of the American College of Chest Physicians in October 2008 estimated that atrial fibrillation is observed in 2.5% to 3% of individuals treated with bisphosphonates, and that 1% to 2% experience hospitalization or death. A case-control study concluded atrial fibrillation could be attributed to bisphosphonates in 3% of previous or current users.[720] These positive reports were followed by a series of consistently negative studies. A larger Danish case-control study concluded that there was no link between atrial fibrillation and the use of bisphosphonates.[721] A Danish cohort study of fracture patients found no increase in the risks of stroke or myocardial infarction, and concluded that an increase in atrial fibrillation could be attributed to the use of bisphosphonates in individuals already at increased risk for cardiovascular events.[722] In two large American cohort databases of patients undergoing coronary angiography, there was no increase in atrial fibrillation or myocardial infarction associated with bisphosphonate treatment.[723] A retrospective cohort study from Taiwan compared bispohphonate use with the use of raloxifene and reported a similar rate of atrial fibrillation in both treatment groups.[724] An apparent increase in atrial fibrillation can be attributed to the fact that most bisphosphonate users are older, have more cardiovascular disease, and are at greater risk for atrial fibrillation.

Bisphosphonates and Esophageal Cancer

In a Letter to the Editor in the January 1, 2009, issue of the *New England Journal of Medicine,* the FDA reported 23 cases (8 fatal) of esophageal cancer in patients being treated with alendronate.[725] As of 2009, a total of 31 cases of esophageal cancer had been collected in Europe and Japan associated with alendronate, risedronate, ibandronate, and etidronate. Although this is a small number of cases with drugs that have been used for more than a decade by millions of people, the concern has added credibility because of the well-recognized side effect of esophageal injury with oral bisphosphonates. Nevertheless the FDA report was anecdotal in nature, with no comparison group. In response, American and European comparisons of treated patients to the U.S. expected incidence of esophageal cancer or to non-treated patients in national databases in the U.K. and Denmark failed to detect an increase in esophageal cancer in bisophosphonate-treated patients.[726, 727]

Bisphosphonates and Pain

The FDA has informed clinicians that patients taking bisphosphonates rarely experience severe bone, joint, or muscle pain that is relieved when treatment stops.

SUMMARY—Treatment with Bisphosphonates

1. An increased susceptibility to nonspinal fractures may occur relatively early when bisphosphonate treatment is combined with another antiresorptive treatment (such as estrogen), and this should be avoided because no additional benefit on fracture risk has been demonstrated with combined treatment.

2. Bisphosphonate treatment is best reserved for older postmenopausal women. It is not a drug of choice for the prevention of osteoporosis in relatively young postmenopausal women.

3. In all except very high-risk patients being treated with bisphosphonates, it would be wise to consider a time limit for duration of exposure. Bone density should be measured after 2 to 4 years of treatment, and if not in the osteoporosis range, treatment should be discontinued. Patients should be followed by monitoring of bone density, with a resumption of treatment in those who rapidly lose bone or in those who accumulate a loss of 5% to 10% in 1 year.

4. The onset of severe pain at any site is an indication to discontinue bisphosphonate treatment.

Treatment with Denosumab

RANKL (nuclear factor kB ligand) is secreted by osteoblasts and binds to its receptor, RANK, on the surface of osteoclasts, stimulating the osteoclasts to mature and to resorb bone. OPG is osteoprotegerin, a receptor produced by osteoblasts that binds with RANKL and prevents the activation of RANK, causing osteoclasts to undergo apoptosis. An agent that could act like OPG, therefore, would prevent bone loss. Denosumab is a human monoclonal antibody to RANKL that works like OPG, but rather than competing with RANK for its activator RANKL, it binds with great affinity to RANKL thus preventing activation of RANK. This is not the same mechanism as bisphosphonates and estrogens that bind to the surface of bone and interfere with osteoclast activity. Denosumab prevents both maturation and survival of osteoclasts, as well as active bone resorption by osteoclasts.

The impact of denosumab on bone density has been evaluated in Phase 2 clinical trials.[728] The participants were postmenopausal women up to age 80 who had osteoporosis based on low T-scores. For the first 2 years, the patients were randomized to one of 7 denosumab doses, alendronate, or placebo. After 2 years, patients in 5 of the treated groups were continued for 2 more years on 60 mg daily given subcutaneously only once every 6 months. One treatment arm was discontinued, and then after a year, treatment was resumed with 60 mg denosumab every 6 months. One treatment arm was discontinued after 2 years and followed. 262 women completed the study. Long-term treatment increased bone density in the spine and hip, whereas the placebo group continued to lose bone density. A reduction in markers for bone turnover was sustained. The group that discontinued treatment after 2 years lost bone density; after a year, retreatment with denosumab restored the bone to the level of the gain achieved in the first 2 years, levels similar to the groups who continued treatment for 4 years. No significant differences in the treatment and placebo groups were noted for adverse events. Similar results have been reported in a Phase 3 trial with 332 women in 21 centers in Canada and the U.S.[729] In a large, 3-year, worldwide multicenter trial involving 7,868 women with osteoporosis, the Fracture Reduction Evaluation of Denosumab in Osteoporosis Every

6 Months (FREEDOM) trial, treatment with 60 mg denosumab subcutaneously every 6 months reduced vertebral fractures by 68%, hip fractures 40%, and other nonvertebral fractures by 20%.[730] A separate phase 3 trial concluded that denosumab treatment increased bone mineral density at the hip and spine significantly more than alendronate.[731] There were no differences in major adverse events in these trials, including osteonecrosis or indications of suppressed immune function; however, eczema and cellulitis were significantly increased in the worldwide study.[730] Longer-term evaluation will be necessary for a full assessment of side effects that might occur with general use.

How do the results compare with bisphosphonates? The increase in bone density at the spine and hip is greater with denosumab, and total body bone density, as exemplified by the radius, increases with denosumab, but not with bisphosphonates. The requirement for subcutaneous administration might seem to be an obstacle, but on the other hand, a required visit to the office might increase continuation rates (although it is likely that a self-administration technique would be available). After discontinuation of denosumab, bone loss immediately begins. This quick reversibility contrasts with bisphosphonates, which bind tightly to bone remaining in bone for decades and maintaining a long-term effect; this can be an important advantage for denosumab.

There is a problem. The RANKL system operates in other tissues, significantly in the immune system. The number of subjects in the completed trials has been insufficient to achieve the statistical power to reassure us that infections and tumors won't arise in treated patients. Nevertheless, there are obvious advantages. Denosumab works fast; it's reversible; it needs to be administered only every 6 months. Denosumab (Prolia) was approved for the treatment of osteoporosis in high risk patients; however, it is likely to be effective for prevention of osteoporosis as well.

Calcitonin

Calcitonin regulates plasma calcium by inhibiting bone resorption and can be used in patients for whom hormone therapy is contraindicated. Studies with intranasal delivery of salmon calcitonin (200 IU daily) indicate it can increase bone density. Human calcitonin is available, but recombinant salmon calcitonin (Fortical and Miacalcin) is more potent. Calcitonin treatment should be combined with vitamin D and calcium supplementation. In a randomized 5-year trial, calcitonin treatment reduced spinal fractures with less of an effect (about 33%) compared with estrogen, alendronate, and raloxifene, and no reduction in hip fractures.[732] The lack of a dose response and a high drop-out rate raised questions regarding the potency and efficacy of calcitonin in preventing fractures.

Fluoride

The addition of fluoride, a potent stimulator of bone formation, can offer significant protection against osteoporosis. The clinical response depends on the formulation and dose. Slow-release sodium fluoride (25 mg b.i.d. given 12 of every 14 months) combined with calcium supplementation reduced the vertebral fracture rate with essentially no side effects.[733] Treatment is recommended for no longer than 4 years to avoid the toxic accumulation of fluoride in bone.[734] This treatment is reserved for patients with established postmenopausal osteoporosis; however, a commercial preparation has not been approved.

Strontium

Oral strontium ranelate (Protelos) is available in many countries. A dose of 2 gms strontium ranelate is dissolved in water and taken at bedtime. Clinical trials indicate that a reduction in spinal and nonvertebral fractures is achieved comparable to antiresorptive agents.[735, 736] The mechanism of action is unknown. The main side effect is nausea and diarrhea that usually resolves in a few months.

Tibolone

Tibolone is structurally related to the 19-nortestosterone progestins used clinically in oral contraceptives. It's chemistry, mechanisms of action, and clinical effects are discussed in detail in Chapter 18. Tibolone prevents bone loss in postmenopausal women as effectively as estrogen or estrogen-progestin therapy.[737–743] In a large, U.S. dose-response study with doses ranging from 0.3 to 2.5 mg daily, only the 1.25 and 2.5 mg doses produced progressive bone density increases in the femoral neck. Indeed, the impact on bone was essentially the same for the 2 highest doses, 1.25 and 2.5 mg. Although the 1.25 mg dose is acceptable for the prevention of bone loss, the 2.5 mg dose is more effective for the alleviation of hot flushes.[744]

The beneficial impact on bone can be attributed to tibolone's estrogenic metabolites acting through the estrogen receptor because it is blocked by an antiestrogen, but not by an antiandrogen or an antiprogestin.[745] Tibolone prevents the bone loss associated with GnRH agonist treatment (and the side effect of hot flushing).[746, 747] The bone mineral density data are similar to those associated with hormonal and alendronate therapy.

The LIFT study (Long-term Intervention on Fractures with Tibolone) was a randomized, placebo-controlled multicenter trial in 22 countries of tibolone, 1.25 mg, given daily over 3 years.[748] The 4,538 women who participated in the trial were age 60 to 85, all at high risk of fractures because of osteoporosis, and all treated with calcium and vitamin D supplementation. The study was stopped in February 2006 after a mean treatment of 34 months because of an increased risk of stroke. The risks of all events were assessed after 5 years of follow-up. The reduction of fractures was about four times as great in women who already had a vertebral fracture upon entry to the study compared with women who had not had a fracture at baseline. It is noteworthy that the number of falls in the treated group was 25% less. The increase in stroke was greater in the oldest women (over age 70).

Based on previous bone density studies, the results of the LIFT trial on fracture reduction were not unexpected. The magnitude of the effect is roughly comparable to those with estrogen, bispohosphonates, and raloxifene (with the important exception being a lack of effect of raloxifene on hip fractures). A reduction of breast cancer was comparable to that reported with tamoxifen and raloxifene, but this was not a primary endpoint of the study.

The reported risk of stroke is similar to that observed with estrogen. In the Women's Health Initiative, no increase in stroke was observed in women under age 60 who had an absence of stroke risk factors. It seems prudent to avoid the use of tibolone in elderly women and in women who are at risk for stroke (specifically those with hypertension, smoking, diabetes, or atrial fibrillation). The OPAL study (Osteoporosis Prevention and Arterial effects of tibo-Lone) was a 3-year, randomized, double-blind trial in six U.S. centers and five European centers, treating 866 postmenopausal women with either 2.5 mg tibolone daily, 0.625/2.5 mg daily of conjugated estrogens/medroxyprogesterone acetate, or placebo.[749] Unfortunately, the OPAL trial did not achieve its goal of providing robust data on cardiovascular effects, due to the older age of the women and the notably different results in American and European women. There continues to be good reason to believe that tibolone will have a neutral effect in terms of coronary heart disease.

In summary, tibolone treatment of postmenopausal women is as effective as estrogen therapy in relieving hot flushes, preventing bone loss, and increasing vaginal lubrication, but it stimulates libido to a greater degree than estrogen. There is less breast tenderness and mastalgia with tibolone. Endometrial safety has been reported to be comparable to that achieved with continuous combined estrogen-progestin regimens, and with a lower rate of breakthrough bleeding. The previously reported increased risks of breast cancer and endometrial cancer in observational studies very likely represent "preferential prescribing" of tibolone in Europe, discussed with references in Chapter 18. Women prescribed tibolone in Europe more often had chronic breast disease, a personal history of breast cancer, previous dysfunctional uterine bleeding, hypertension, and previous uterine operations. Most importantly, more women prescribed tibolone had a history of treatment with unopposed estrogen. Thus, clinicians were more likely to prescribe tibolone to women they believed were at higher risks for these two cancers, and this would yield higher rates in treated groups compared with control groups. The standard dose of tibolone for many years was 2.5 mg daily, but the new studies support the use of the lower dose, 1.25 mg, with no apparent loss of bone efficacy. Tibolone continues to be an appropriate choice for hormonal therapy, suitable for many postmenopausal women.

Teriparatide (Forteo)

Parathyroid hormone increases bone formation. Teriparatide (Forteo) is the recombinant human 1–34 amino acid fragment of parathyroid hormone. Another Canadian preparation encompasses the 1–84 sequence. This is the only treatment, besides fluoride, that directly stimulates osteoblasts to form new bone. Given to postmenopausal women with osteoporosis in a once-daily, subcutaneous dose of 20 μg, teriparatide produces a greater increase in bone density and possibly a greater reduction in fractures compared with estrogen or alendronate.[750, 751] ***Because of the expense and difficulty in self-administration, treatment with the parathyroid fragment is best directed to individuals with serious osteoporosis and at high risk for fractures.*** After a relatively short period of treatment, no more than 2 years, the gain in bone can be maintained with one of the anti-resorptive agents.[752, 753]

Alternative Therapies to Prevent Bone Loss

Phytoestrogens are effective in preventing bone loss in rats, but not in monkeys.[754–756] In women, some studies demonstrated at best a slight effect on spinal bone, but most found no benefit in either spine or hip.[757–760] Flaxseed supplementation had no effect on biomarkers of bone metabolism.[761] The difference between hip fracture incidence in Japanese and American women may be due to structural and/or genetic differences, not dietary intake of soy.[762]

Ipriflavone is a synthetic isoflavone; it is methylated dehyroxydaidzein, which is metabolized to daidzein. Studies with ipriflavone demonstrated prevention of bone loss over a year.[763–766] Overall the effect on bone was not as great as that observed with a standard dose of estrogen, or alendronate, perhaps not great enough to yield a benefit. A 4-year randomized trial in Europe assessed the effect of ipriflavone on bone density, urinary markers, and vertebral fractures in 474 women and could find no difference in the treated group compared with the placebo group.[767]

Equol is a bacterial metabolite, and the only hormonally active metabolite, of the soy phytoestrogen, daidzein. At least in vitro, equol stimulates gene transcription with both estrogen

receptors and with a greater potency than any other isoflavone.[768] Equol formation is totally dependent on intestinal microflora. The most important observation regarding equol is that most adults do not produce equol, even when challenged with high doses of soy.[769] This is a contrast to nonhuman primates and other animals; all that have been studied produce high levels of equol. Thus there are two human populations: equol producers and nonequol producers. The key question is whether equol producers receive greater clinical effects from phytoestrogens than non-equol producers. As noted, thus far the clinical effects of isoflavones on bone have not been impressive. In a 2-year randomized trial of postmenopausal women, isoflavone-rich soy milk increased spinal bone mass in the 45% of the subjects who were equol producers, with essentially no effect in non-equol producers.[769] Similar results were obtained in a Japanese trial testing the effect of isoflavone treatment in equol producers and non-producers.[770] Therefore, the population destined to receive a benefit from soy intake may be limited to equol producers. Studies need to be repeated measuring the responses in individuals who are identified as equol producers or nonequol producers. If the population destined to receive a benefit from soy intake is limited to equol producers, a convenient, inexpensive method must be developed to identify equol production. It may be possible to convert nonproducers to producers. A more straight-forward approach is to administer equol itself. Daidzein yields two forms in equol producers, the R-equol inactive isomer and S-equol, the active isomer that binds to estrogen receptor-β. S-equol has been synthesized and its administration is effective for the treatment of menopausal symptoms.[771] Another alternative is the S-equol supplement made by incubating equol-producing bacteria with soy isoflavones.[772, 773] It is absorbed readily with high bioavailability; low doses must be administered twice daily because of a relatively short half-life.

Statins

Statin treatment is, of course, a mainstay in programs designed to prevent cardiovascular disease. Prompted by animal experiments indicating that statins increase bone formation, case-control studies reported that statin treatment was associated with a reduced risk of fracture.[774, 775] However, analysis of a randomized trial of statin therapy and cardiovascular disease failed to detect an effect on fracture risk, and a cohort study reported that an apparent beneficial effect of statin on fracture risk was influenced by various confounding factors.[776, 777] In the prospective cohort arm of the Women's Health Initiative, bone densities and fracture rates were similar comparing statin users with nonusers.[778] Statin drugs should not replace proven effective medications to prevent or treat osteoporosis.

Thiazides

Older women are often treated with thiazides for hypertension. Thiazides reduce the urinary loss of calcium, induce a positive calcium balance, and treatment is associated with a higher bone density. It is useful to know that estrogen and thiazides are additive; a significantly higher bone density is achieved with combined use.[779]

Thiazolidinediones

Thiazolidinediones used to treat diabetes are associated with an increase in bone loss and fractures.[780, 781] Bone density measurements and careful assessment for fracture risk are important components of healthcare for these diabetic women, with emphasis on adequate calcium and vitamin D supplementation, and appropriate treatment with one of the antiresorptive agents where indicated.

Lifestyle Modifications

Lifestyle can have a beneficial effect on bone density. Physical activity (weight-bearing), as little as 30 min a day for 3 days a week, will increase the mineral content of bone in older women.[782] To be effective, exercise must exert a load on bone, especially the spine.[783] Ordinary walking will not suffice.[784] Even brisk walking achieves a significant increase in bone density only in the calcaneus, the site subjected to stress with walking.[785] However, brisk walking may slow the rate of bone loss in the hip.[786] In other words, weight lifting is better for the spine than is ordinary walking, although running probably helps hip bone mass. The activities that are beneficial are running, weight training, aerobics, stair climbing, and sports other than swimming. The effect of weight-bearing exercise on bone density is additive when combined with hormone therapy.[787] Although ordinary walking has little impact on bone density, it is still reasonable to expect walking to have an overall beneficial effect on the risk of fracture. Walking improves the cardiovascular status of patients and reduces body mass. These changes plus the exercise itself improve balance and decrease the risk of falling. For these reasons, walking, even after adjusting for bone density and body weight, is associated with a reduced risk of hip fracture.[540, 788]

The impact of exercise on bone is significantly less than that achieved by hormone therapy.[632] Women require the full combination of pharmacologic therapy, calcium and vitamin D supplementation, and exercise in order to minimize the risk of fractures. For each of these, the beneficial impact lasts only as long as the therapy is continued.

Cigarette smoking and excessive alcohol consumption are associated with an increased risk of osteoporosis. The magnitude of bone loss associated with cigarette smoking is consistent with a 40–45% increase in the risk of hip fracture.[789] Women who smoke also enter menopause earlier, and lose bone at a greater rate in the first years of the postmenopausal period.[790] ***The titration of estrogen dosage with circulating blood estradiol levels in smokers makes clinical sense, allowing the use of higher hormonal doses to maintain bone density.*** Monitoring of bone response with bone density measurements would further aid in achieving maximal effects of therapy.

A high coffee intake has been reported to be associated with an increased risk of osteoporosis.[791] However, this increase in risk is dependent on dietary calcium intake. In women who drank at least one glass of milk (300 mg calcium) per day throughout most of their lives, increasing caffeinated coffee intake was not associated with a lower bone density.[792] Repeatedly, we see the importance of teaching children and adolescents the merit of an adequate calcium intake; drinking nonfat milk throughout life is good for you. An adequate calcium intake compensates for "calcium robbers," such as caffeine and soft drinks. A British study concluded that an increase of only 300 mL of milk per day in adolescents increased bone density without an increase in weight or body fat.[636]

Remember that not all fractures are solely due to osteoporosis. Drug side effects, impaired vision, neurologic dysfunction, and muscular conditions all put patients at risk because more than 90% of fractures occur following a fall.[486] Interventions that reduce the odds of falling and enhance the ability to withstand the impact of a fall are important.[793] This includes patient education regarding hazards in the home, monitoring drug use, adequate nutrition, and a good exercise program. In addition, there is evidence that estrogen with or without added progestin improves muscle strength and balance.[254–259] On the other hand, some studies have not been able to document an increase in muscle strength or improvements in balance.[787, 794–797] Furthermore, the increase in muscle mass and strength in response to weight-bearing exercise was the same when hormone users were compared with non-users.[798] Nevertheless, a cohort study reported 71% fewer falls in early postmenopausal hormone users and 43% in late postmenopausal hormone users.[799]

Bone-Losing Medications

Clinicians should always remember that exposure to excessive thyroid and glucocorticoid hormones is associated with osteoporosis and an increased rate of fractures. The bone loss associated with glucocorticoid treatment is significantly prevented by hormone or bisphosphonate therapy.[800–803] ***Excessive thyroid effects can be avoided by annually monitoring treatment dosage with TSH levels.*** Specific preventive treatment should also be offered to patients using long-term anticonvulsants or heparin.

Considerable publicity has successfully raised clinical consciousness regarding the increased risk of fractures associated with the use of corticosteroids. We should similarly recognize the increased risk of fractures with the daily use of SSRIs. The Canadian Multicentre Osteoporosis Study Research Group reported the effect of daily SSRIs in a prospective cohort study in seven regional centers of 5,008 adults over the age of 50.[526] After adjusting for age, hip bone density, fractures at baseline, and estrogen use in women, daily SSRI use was associated with a 2-fold increased risk of fragility fractures. Daily use of SSRIs was also associated with about a 2-fold increased risk of falling, and these individuals had lower bone densities. Controlling for falls and lower bone density still left an increased risk of fractures in SSRI users that began after 1 to 1.5 years of use.

Does this side effect of SSRIs make sense? Is it the SSRI or the lifestyle associated with clinical depression? Serotonin does not cross the blood-brain barrier; in the brain serotonin activity is the result of synthesis, reuptake, and binding to a 5-hydroxytryptophan receptor, which is inhibited by SSRIs. Most of the circulating serotonin comes from synthesis in the duodenum by specialized neuroendocrine cells. Locally-released serotonin stimulates intestinal peristalsis, whereas the serotonin entering the circulation is taken up by platelets or serotonin can reach a target tissue like bone, linking bone formation to the gut. Mice with a mutation for the serotonin transporter protein develop less bone mass and strength.[527] Serotonin receptors and serotonin transport have been identified in osteoblasts and osteocytes. The bone effects of parathyroid hormone and mechanical stimulation are modulated by the serotonin system. Therefore daily SSRI use can impair bone formation, tilting the balance in favor of resporption and bone loss, and decreased bone densities have been reported in both male and female SSRI users (but not in users of tricyclic antidepressants).[528, 529]

It is not always easy to know which came first, depression or fractures leading to subsequent depression. It has been reported that depressed people and SSRI users have a greater incidence of falls,[521] and thus it is not unreasonable to consider that depression comes first in some people. However, orthostatic hypotension and syncope are more common in SSRI users, and this could also contribute to the greater prevalence of falls.

Depressed people are sedentary and eat poorly, factors that favor bone loss. Some have speculated that increased cortisol levels associated with depression might lead to bone loss, similar to that observed with the pharmacologic administration of corticosteroids. On the other hand, American studies, despite finding a link between depression and fractures, failed to detect an increase in depression associated with lower bone density measurements.[520, 521] However, other studies have reported increases in depression associated with lower bone densities.[522–524]

Obviously this is an unsettled picture, but we should be more aware of the possible increased risk of fractures with the daily use of SSRIs. Interventions that reduce the odds of falling and enhance the ability to withstand the impact of a fall are important. This includes patient education regarding hazards in the home, monitoring drug use, adequate nutrition, and a good exercise program. Aggressive monitoring of bone density is warranted; adequate calcium and vitamin D supplementation are necessary, and until more studies clarify this problem, it seems reasonable to consider treatment with one of the antiresorptive agents.

Proton pump-inhibiting drugs used to treat gastroesophageal reflux impair intestinal calcium absorption, resulting in secondary hyperparathyroidism and bone loss. Long-term use of omeprazole has been linked to an increase in hip fractures in case-control studies.[804, 805] These drugs should be added to the list of bone-losing medications that require careful monitoring of bone density and antiresorptive treatment when indicated. The list includes, of course, any treatment that produces a hypoestrogenic state, for example, an aromatase inhibitor.

SUMMARY—Bone-Losing Medications

1. Exposure to excessive thyroxine.
2. Glucocorticoid treatment.
3. SSRI therapy for depression.
4. Proton pump-inhibiting drugs.
5. Aromatase inhibitors.

Management of a Nonresponder to Hormone Therapy

There is a percentage of postmenopausal women on hormone therapy (from 5% to 15%, depending on compliance) who continue to lose bone and experience fractures.[576, 578, 612] In the PEPI 3-year clinical trial, where compliance rates were probably maximal, 4% of treated women lost bone in the spine and 6% in the hip.[576] Is this real, or does it represent a technical problem with bone density (regression to the mean)?[806, 807]

There are two reasons to question the prevalence of nonresponse to treatment. The first reason is the argument that the results of single bone density measurements display the phenomenon of regression to the mean (extreme results are in part due to random error). Analysis of the hip and spinal bone density data from the randomized clinical trials assessing the effects of alendronate and raloxifene in postmenopausal women with osteoporosis revealed that women who lost bone mineral density after 1 year of treatment were likely to gain bone during the second year.[806] The more extreme the measurement after 1 year of treatment, the more likely the next year's measurement indicated a reversal. Therefore, it has been recommended that bone treatments should not be discontinued when measurements after 1 year indicate loss of bone density because repeat results are usually closer to the mean.

The second good reason to question a relatively high percentage of nonresponse is found in a reanalysis of the PEPI data.[807] By focusing on those women who had replicate bone density measurements at each testing, true bone loss on hormonal treatment was rare. At the lumbar spine, only 1.5% of hormone users lost bone during the first year of treatment, and only 0.6% in the second and third years of the study. At the hip, these percentages were 2.3% for the first year and 0.4% in the second and third years, respectively.

These analyses concluded that nonresponse (or suboptimal response) is a real phenomenon, but the prevalence is less than previously suspected. However, there are two criticisms that challenge this conclusion. First, the data are derived from clinical trials where compliance

is much better than use of pharmacologic treatments in general populations. Second, and most importantly, clinical trial participants represent a very homogeneous group, the product of protocol inclusions and exclusions. This homogeneous group does not represent the general population where variations are more common. A good example is the wider variation in body weight in the general population. Therefore, we would expect to encounter in our practices more of the subgroup of individuals who metabolize and clear hormones at a greater rate. And more women in a general population will be responding less to standard hormone doses, as documented by bone density measurements. ***For this reason we advocate the use of bone density measurements to measure the response to treatment, and an increasing use of blood estrogen assays to assess the efficacy of hormone therapy. But keep in mind that because of the precision error with bone density measurements, a change of at least 3% to 5% is necessary to be clinically significant.***

Assessing Blood Estrogen Levels

It is worthwhile to measure the bone density in treated women when they are in their late 60s to detect poor responders. On the average, about 10–15% of women lose bone despite being prescribed hormone therapy. A Finnish 5-year clinical trial reported a prevalence of poor response based on bone density of 11% for spinal bone and 26% for the hip.[808] As expected, smoking and low body weight were common findings among the poor responders, but the most impressive characteristics were lower estradiol and higher FSH levels. It is only logical that there exists a group of women who metabolize and clear administered estrogens at a greater rate, and thus require a higher dose to sustain a protective effect on bone. Indeed, considerable variation in estradiol levels has been documented in individuals receiving both oral and transdermal hormone therapy.[809, 810] Marketing presentations by the pharmaceutical companies provide mean levels, suggesting stable and smooth maintenance of blood levels; however, the ranges, which are wide, are not revealed. An aim of individualizing hormone therapy is to determine the appropriate dose for the intended objective; in the case of bone, the target estradiol level should be 40–60 pg/mL, and a practical range for a blood sample derived during office hours from a patient taking her medication at night is 50–100 pg/mL. To minimize variation, we encourage clinicians to always use the same laboratory, after first establishing the values obtained using standard doses of estradiol in their own practices. The reliability of blood estrogen measurements is questionable when conjugated equine estrogens are the treatment being used, although some assays can provide reproducible results because of cross-reactions with the specific antibody in the assay.

As clinicians and the pharmaceutical industry promote lower doses of estrogen with the attractive notion that less is safer, a greater rate of poor response as measured by bone density can be expected. We are suggesting that once a poor responder is detected by measurement of bone density, titering of estrogen dose is indicated, utilizing the blood concentration of estradiol. We would further emphasize that appropriate detection and investigation will confirm that a relatively low blood estradiol level is the cause of most cases of poor bone response to hormone treatment. Measurement of vaginal pH from the lateral vaginal wall is very simple and inexpensive. It has been impressive in our experience and that of others how an acidic pH (less than 4.5) correlates with an appropriate dose of estrogen administration.[230, 811] This experience motivates us to suggest that this may be the best method to assess the adequacy of estrogen therapy. Assessing vaginal cytology is not useful. The vaginal mucosa is too sensitive to estrogen to allow dose-response titering.

Bone density measurements can also detect individuals who are responding poorly to bisphosphonate, raloxifene, or other treatments. Two considerations should come to mind. First, establish that proper intake of the medication is being practiced with good compliance. Second, rule out other causes of bone loss and make sure calcium and vitamin D supplementation

is adequate. Optimal vitamin D levels are necessary to maximize the response to any anti-resorption agent.[812] This emphasizes the need to titer the dose of vitamin D supplementation by measuring the circulating level of 25-hydroxyvitamin D. A value less than 30 ng/mL is below normal; less than 20 ng/mL is a definitive indication of vitamin D deficiency.

In a woman demonstrated to be losing bone despite hormone therapy, the following steps are recommended:

- **Check compliance and dose by measuring blood estrogen levels; adjust dose by titering with estradiol measurements. Consider the use of vaginal pH measurements.**

- **Rule out other causes of bone loss:**

  ***Hypogonadism:* e.g., eating disorders.**

  ***Drugs:* Heparin, anticonvulsants, glucocorticoids,**

  SSRIs, proton pump inhibitors, high intake of alcohol.

  ***Chronic Disease*: Renal and hepatic.**

  Endocrine Diseases: **Excess glucocorticoids.**
  Hyperthyroidism.
  Estrogen deficiency.
  Hyperparathyroidism.

  ***Nutritional*: Calcium, phosphate, vitamin D deficiencies.**

- **Follow with markers of bone turnover or bone density measurements.**

Conclusion

The menopause is a physiologic event that brings clinicians and patients together, providing the opportunity to enroll patients in health maintenance. The failure to respond appropriately (by either clinician or patient) easily leads to a loss of the patient from a practice, but equally, if not more, importantly, is the probability that the loss of a patient from a practice means that another woman has lost her involvement in a preventive health care program. Contrary to popular opinion, the menopause is not a signal of impending decline, but, rather, a wonderful phenomenon that can signal the start of something positive, a good health program.

All references are available online at: http://www.clinicalgynendoandinfertility.com

18 Postmenopausal Hormone Therapy

Postmenopausal hormone therapy had its beginning in the effort to alleviate specific symptoms associated with the decline in estrogen production at menopause. There is little question that women who suffer from hot flushes or atrophy of reproductive tract tissues can be relieved of their problems by the use of estrogens. In the 1990s, however, the focus of postmenopausal hormone therapy changed from short-term treatment to the preventive health care benefits associated with long-term treatment. It is almost certain that the long-term disabilities of osteoporosis can be largely prevented by therapy with estrogen and progestin. The long-term impact on cognition remains to be documented, but there is reason to believe that there will be a benefit in this area. However, long-term use was challenged by clinical trial data that were interpreted to indicate that hormone therapy did not protect against cardiovascular disease and that the risk of breast cancer was increased. The debate over these issues made decision-making by clinicians and patients very difficult. In this chapter, we offer our own interpretation as a guide for the clinical use of postmenopausal hormone therapy.

The evidence supporting many of the benefits with postmenopausal hormone therapy is also reviewed in Chapter 17, in which the effects of hormonal treatment are considered in

conjunction with the impact of the decrease in estrogen after menopause. In this chapter, we review the clinical aspects of postmenopausal hormone therapy, the impact of clinical trial results, and our methods of patient management.

History[1–4]

The existence of hormones was unknown 200 years ago. In the last half of the 19th century, a scattering of chemists and physiologists began to produce hormonally active extracts from glands, bile, and urine of animals. Adventuresome clinicians used these extracts to treat patients, for example supplying thyroid hormone to treat severely hypothyroid individuals, and the specialty of endocrinology was born. The word "endocrine" was adopted to designate the "glands of internal secretion," the multiple sources of hormones.

Charles Edouard Brown-Sequard, the son of a French woman and an American sea captain, was born on the island of Mauritius. Speaking fluent English and French, he practiced medicine and lectured in London and New York before settling in Paris. Brown-Sequard reported in 1889 that he was rejuvenated by the self-administration of extracts from dog testicles, most likely a placebo effect considering the scant amount of testosterone he could have extracted using his aqueous method, and he suggested that ovarian extracts would have the same revitalizing effect in women. Efforts to treat women around the end of the nineteenth century were largely unsuccessful, but in 1897, ovarian extract was reported to be effective for menopausal hot flushing.[5]

The first American attempt to treat menopausal symptoms is attributed to E.L. Sevringhaus and J. Evans of Madison, Wisconsin, who in 1929 administered a derivative from the amniotic fluid of cattle.[2, 6] In the 1930s, the ovarian hormones were isolated, and the "estrin" products and the synthetic estrogens, stilbestrol and ethinyl estradiol, were administered to menopausal women. Edgar Allen and Edward Doisy were the first to isolate the ovarian hormone, estrogen. Allen was born in Colorado, educated at Brown University, and served in France during World War I. In 1933, he became the chairman of the Department of Anatomy at Yale University. He died of a heart attack while on patrol off Long Island for the U.S. Coast Guard in February 1943. Doisy was born in Illinois and educated at the University of Illinois and Harvard. During World War I, he was assigned to the Rockefeller Institute in New York City and then to the Walter Reed Hospital in Washington. Doisy was the first chairman of biochemistry at the St. Louis University School of Medicine. He received the Nobel Prize in Medicine, along with Henrik Dam, in 1943 for his isolation and synthesis of vitamin K. Doisy died in 1986 at the age of 92.

In 1919, Allen and Doisy, both discharged from the army after World War I, joined the faculty at the Washington University School of Medicine in St. Louis. They became friends playing on a faculty baseball team and planned their first experiments while driving to work together. In 1922, Allen moved to the University of Missouri to be Professor of Anatomy, and Doisy went to St. Louis University, but they continued their collaboration. Doisy prepared ovarian extracts and mailed them to Allen for experiments. In 1923 and 1924, Allen and Doisy reported the isolation from pig ovaries and the administration to animals of "an ovarian hormone."

In 1926, Sir Alan S. Parkes and C.W. Bellerby coined the basic word "estrin" to designate the hormone or hormones that induce estrus in animals, the time when female mammals are fertile and receptive to males. Doisy and his students Veler and Thayer in St. Louis isolated a few milligrams of estrogen in crystalline form in 1929 from large amounts of urine from pregnant women. The terminology was extended to include the principal estrogens

in humans, estrone, estradiol, and estriol, in 1932 at the first meeting of the International Conference on the Standardization of Sex Hormones in London, although significant amounts of pure estradiol were not isolated until 1936. At this same meeting, the pioneering chemists were bemoaning the problem of scarcity that limited supplies to milligram amounts when a relatively unknown biochemist, A. Girard from France, offered twenty grams of crystalline estrogen derived by the use of a new reagent to treat mare's urine.[7]

In the 1920s, George W. Corner at the University of Rochester invited Willard Myron Allen, an organic chemist who was then a medical student, to join him in the study of the corpus luteum. Within 2 years, they had a pure extract, but it was not until 1934 that crystalline progesterone was isolated almost simultaneously in several countries. It took the corpora lutea of 50,000 pigs to yield a few milligrams. At the Second International Conference on Standardization of Sex Hormones in London, Corner and Allen proposed the name progestin. Others proposed luteosterone, and, at a cocktail party, the various biochemists agreed to call the chemical progesterone.[7]

Hormones were being administered to patients in the 1940s, but supplies were very limited. And with a scarce supply, hormones were incredibly expensive. Progesterone, for example, cost $200 per gram. "To secure barely enough androsterone to cover the head of a pin, Adolph Butenandt had had to start with nearly four thousand gallons of urine; to obtain less than one hundredth of an ounce of pure testosterone crystals, Ernst Laqueur had had to process nearly a ton of bulls' testicles. It took a full ton of cholesterol, from the spinal cords or brains of cattle or from the grease of sheep's wool, to yield just twenty pounds of the starting material from which progesterone ultimately could be obtained. Edward Doisy had had to process the ovaries of more than eighty thousand sows to get just twelve thousandths of a gram of estradiol."[8]

In the 1930s, the Ayerst Company was extracting estrogens from the urine of pregnant women. Limited by the problems of supply, low activity, and bad taste and odor, Gordon A. Grant, head of biochemistry for Ayerst, suggested in 1939 that they use urine from horses. The process produced sodium salts from the sulfate esters of the various estrogens, yielding a water-soluble conjugate. Premarin (conjugated estrogens) was approved in Canada in 1941 and in the U.S. in 1942 for the treatment of symptoms associated with menopause.[9] The tablets were and are still designated as variations of 1.25 mg, based on the equivalent amounts of Premarin and estrone (1.25 mg) that could produce the same effect in the Allen–Doisy bioassay (amount required to produce an increase in rat uterine weight). It was not until 1972 that the first quantitative analysis of Premarin was performed, based on gas chromatography. Modern studies indicate that there is a large number of steroids in Premarin, even androgens and progestins, but only the 10 estrogens are present in sufficient quantity to have clinical effects. Synthetic conjugated estrogens are available; one mixture, Cenestin, contains 9 estrogens and the other, Enjuvia, 10 estrogens.

Composition of Conjugated Estrogens (Premarin)	
Sodium estrone sulfate	**49.3%**
Sodium equilin sulfate	**22.4%**
Sodium 17α-dihydroequilin sulfate	**13.8%**
Sodium 17α-estradiol sulfate	**4.5%**
Sodium Δ8,9-dehydroestrone sulfate	**3.5%**
Sodium equilenin sulfate	**2.2%**
Sodium 17β-dihydroequilin sulfate	**1.7%**
Sodium 17α-dihydroequilenin sulfate	**1.2%**
Sodium 17β-estradiol sulfate	**0.9%**
Sodium 17β-dihydroequilenin sulfate	**0.5%**

Estrogen Formulations and Routes of Administration

Oral Administration

The relative potencies of commercially available estrogens are of great importance when prescribing estrogen, and the clinician should be familiar with the following potencies:

Relative Estrogen Potencies[10–15]

Estrogen	FSH Levels	Liver Proteins	Bone
Conjugated estrogens	1.0 mg	0.625 mg	0.625 mg
Micronized estradiol mg	1.0 mg	1.0 mg	1.0 mg
Estropipate (piperazine estrone sulfate) mg	1.0 mg	1.25 mg	1.25 mg
Ethinyl estradiol	5.0 µg	2–10 µg	5.0 µg
Estradiol valerate	–	–	1.0 mg
Esterified estrogens mg	–	–	0.625 mg
Transdermal estradiol	–	–	50 µg

The 17α-ethinyl group of ethinyl estradiol (by resisting metabolism) enhances hepatic effects, because no matter by which route it is administered, liver function is affected.[13] The same is true for conjugated equine estrogens. Contrary to the case with estradiol, the liver appears to preferentially extract ethinyl estradiol and conjugated equine estrogens no matter what the route of administration. Thus, the route of administration appears to influence the metabolic responses only in the case of specific estrogens, most notably estradiol.

A major factor in the potency differences among the various estrogens (estradiol, estrone, estriol) is the length of time that the estrogen binds to its receptor. The higher rate of dissociation with the weak estrogen (estriol) can be compensated for by continuous application to allow prolonged binding and activity. Estriol has only 20–30% affinity for the estrogen receptor compared with estradiol; therefore, it is rapidly cleared from a cell. However, if the effective concentration is kept equivalent to that of estradiol, it can produce a similar biologic response.[16] At least two studies have been unable to demonstrate prevention of bone loss with the administration of 2 mg estriol daily.[17, 18] In pregnancy, where the concentration of estriol is very great, it can be an important hormone not just a metabolite. Thus, higher estriol levels are not necessarily protective against potent estrogenic effects. Because estriol protects the rat against breast tumors induced by various chemical carcinogens,[19] it has been hypothesized that a higher estriol level protects against the more potent effects of estrone and estradiol. But, antagonism of estradiol occurs only within a vary narrow range of the ratio of estradiol to estriol, a range that is rarely encountered either physiologically or pharmacologically.[20] Below this range, estradiol is unimpeded, above this range estriol itself exerts estrogenic activity. The commercial preparation that contains estriol, estradiol, and estrone contains sufficient amounts of estrone and estradiol to produce standard clinical effects.

Esterified estrogens are synthetically prepared from plant precursors and are composed mostly of sodium estrone sulfate with a 6–15% component of sodium equilin sulfate. Estradiol valerate is rapidly hydrolyzed to estradiol; therefore, the pharmacology and effects are comparable at similar dosages.[21]

Transdermal Patch Administration

The patches first used for transdermal estrogen administration contained an alcohol reservoir; the estrogen was released through a semipermeable membrane attached to the skin with an adhesive. In the current generation of patches, the hormones are dissolved and distributed throughout the adhesive matrix. In a study of women who had previously discontinued patches because of skin irritation (contact dermatitis), skin reactions were less common with the newer matrix patches.[22] In addition, the matrix patches are better tolerated in tropical environments.[23] The patches are designated according to the amount of estradiol delivered per day: from 14 to 100 μg.

The concentration of estrogen in the hepatic portal system after oral administration is 4–5 times higher than that in the periphery.[24] Because of first-pass metabolism in the liver, oral estradiol results in a circulating estrone to estradiol ratio of approximately 3; with transdermal administration the ratio is 1. The first-pass effect may be important for lipoprotein effects. For example, short-term studies (6 weeks) could document increased catabolism of low-density lipoprotein (LDL-cholesterol) and increased production of apoprotein A-I with oral estrogen, but no effect with transdermal estrogen.[25, 26] A 2-year study in Los Angeles with a transdermal dose (100 μg) detected no significant change in HDL-cholesterol levels.[27] However, English data indicate that the transdermal administration of 50 μg estradiol twice a week is as effective as 0.625 mg oral conjugated estrogens, when combined with a progestin in sequential regimens, on bone density and lipids over a duration of 3 years.[28] Standard doses of estrogen administered transdermally (50 μg) protect against fractures as well as standard oral doses do.[29] As with oral estrogen, lower transdermal doses can produce effects on bone density and menopausal symptoms,[30] but a substantial number of women require higher doses.

The critical question is whether the first-pass effect of oral estrogen is *clinically* important. The different effects of oral and transdermal administration on metabolic parameters have been repeatedly compared over the years, but epidemiologic studies of clinical end points are not abundant, handicapped by the relatively small numbers of women using transdermal estrogen in most countries.

Clotting Factors. First-pass hepatic metabolism affects the synthesis of clotting proteins, markers of coagulation and fibrinolysis that can influence the risk of thrombosis and coronary heart disease events. Oral estrogen increases factor VII and prothrombin 1 and 2 fragment, whereas transdermal estrogen decreases factor VII.[31–34] Oral estrogen also increases circulating levels of matrix metalloproteinases, MMP-2 and MMP-9, enzymes that are associated with a tendency for clotting.[35] However, what is important is whether the different effects of oral and transdermal delivery on clotting factors translate into clinical differences and cardiovascular risk.

Activated Protein C (APC) Resistance and Risk of VTE. Resistance to APC is an important marker for venous thrombosis in individuals with inherited thrombogenic mutations and even in the absence of these mutations. Oral estrogen increases APC resistance, whereas transdermal estrogen has no significant effect on this marker.[36, 37] Based on this difference, one would predict that transdermal delivery of estrogen would be less likely than oral delivery of estrogen to be associated with venous thromboembolism (VTE).

A French case-control study (epidemiologic studies of the link between the transdermal route of administration and a relatively rare event are possible in France because of the popularity of the transdermal method) reported no increased risk of VTE in users of transdermal estrogen, as compared with a 4-fold increase in oral estrogen users.[38–40] Estrogen users who carried a factor V Leiden mutation or a prothrombin mutation had a 25-fold higher risk of VTE than did women who did not use estrogen and did not have either mutation. The women with a prothrombotic mutation who used transdermal estrogen had a VTE risk that was similar to that of women with a prothrombotic mutation who did not use estrogen. The French E3N

prospective cohort study also reported an increased risk of venous thromboembolism with current users of oral therapy, a hazard ratio of 1.7 (CI=1.1–2.8), a ratio that is similar to the usual 2-fold increase repeatedly documented in the literature, and no increase with transdermal estrogen.[41] Venous thrombosis is discussed in more detail later in this chapter.

Lipids and Hepatic Enzymes. Both oral and transdermal estrogen reduce total cholesterol, low-density lipoprotein cholesterol, and lipoprotein(a). Compared with transdermal estrogen, oral estrogen produces significantly greater elevations in high-density lipoprotein cholesterol and increases triglycerides, whereas transdermal estrogen decreases triglyceride levels.[31, 33, 42–44] Indeed, triglyceride levels markedly elevated in response to oral therapy return to normal when treatment is changed to transdermal administration.[45]

Inflammatory Markers. Women on oral estrogen have increased levels of C-reactive protein (CRP), whereas those taking transdermal estrogen do not.[31, 33, 42, 46–49] In fact oral hormone therapy while increasing CRP, as discussed in Chapter 17, reduces the circulating levels of other inflammatory markers (E-selectin, intercellular adhesion molecule-1, vascular cell adhesion molecule-1, monocyte chemoattractant protein-1, and tumor necrosis factor-α) with inconsistent effects on interleukin-6.[46, 47] Transdermal estrogen does not affect levels of these inflammatory markers. It is not certain that the decrease in CRP levels with statins and the increase with oral estrogen are instrumental in clinical outcomes or reflect other effects. Thus raising or lowering CRP levels will not necessarily increase and decrease the risk of clinical disease.

A longitudinal study of 346 postmenopausal women taking oral hormone therapy reported that elevated CRP was a strong predictor of future cardiac events, but only in those with increased IL-6 levels.[48] An increase in CRP alone was not associated with an excess of events. ***The difference in CRP levels between users of oral versus transdermal therapy, especially in younger postmenopausal women, is of little clinical significance.*** In fact, in the Estrogen Replacement on Progression of Coronary Atherosclerosis trial, estrogen-induced increases in CRP had no effect on disease progression, as measured by serial angiograms.[49] A study from the Women's Health Initiative confirmed the correlation between baseline levels of CRP and an elevated risk of coronary heart disease, *but the increase in CRP induced by oral hormone therapy did not further increase the risk!*[50]

Myocardial Infarction Risk. Both oral and transdermal administration of hormone therapy are associated with a decrease in myocardial infarction risk in observational studies.[51]

Metabolic Syndrome. In a 3-month randomized trial involving 50 obese women with metabolic syndrome, oral estradiol therapy worsened markers of the metabolic syndrome, including insulin resistance, suggesting a worsening of cardiovascular risk, whereas transdermal estradiol had minimal effects.[52]

Effects in Smokers. Limited evidence suggests that postmenopausal women who smoke may have a better cardiovascular response to transdermal estrogen than to oral estrogen, including greater reductions in total peripheral resistance, vascular sympathetic tone, and norepinephrine levels, and increased vascular responsivity.[53] Smokers receiving transdermal estradiol have decreased plasma viscosity and thromboxane B_2 levels.[54] These results raise the possibility, although the data are limited, that smokers may represent a group of women for whom transdermal estrogen would be an advantage.

Carbohydrate Metabolism. There is little difference between the oral and transdermal methods of delivery on carbohydrate metabolism. Both methods have a beneficial impact on central abdominal fat content, glucose levels and insulin resistance, associated with a reduced risk of developing adult-onset diabetes mellitus.[55–59]

Breast Cancer Risk. Oral conjugated equine estrogens/sequential medroxyprogesterone acetate decreased median levels of insulin-like growth factor-1 (IGF-1) by 26% and

increased median levels of sex hormone-binding globulin (SHBG) by 96% relative to baseline, whereas no change occurred with transdermal estradiol.[60] High IGF-1 and low SHBG levels are associated with increased breast cancer risk; however, it is difficult to make clinical conclusions based on these secondary markers. A German case-control study of 3,593 cases found no significantly increased risk of breast cancer with oral or transdermal hormone therapy.[61] Thus far, the epidemiologic data comparing oral and transdermal treatment are not sufficient to allow firm conclusions regarding breast cancer risk.

Colorectal Cancer Risk. In a case–control study, both oral and transdermal hormone therapy reduced the risk for developing colorectal cancer.[62] When transdermal therapy involved estrogen alone, the benefit was even greater.

Estradiol Levels in Users of Oral Versus Transdermal Estrogen. Studies comparing circulating estradiol levels in women receiving oral or transdermal estrogen reveal therapeutic estradiol levels predictive of a good bone response, but they also contain large standard deviations, indicating substantial variation among individuals.[63] Individual women metabolize estrogen differently, depending on the route of administration, liver function, skin absorption, body composition, body size, potential medication interactions, and the presence of binding proteins; all of which contribute to individual variations in serum estradiol levels.[64] In the future, measurement of serum estradiol levels may play a role in assessment of adequate treatment. This measurement will be especially useful for users of transdermal estrogen therapy, which produces more consistent estradiol levels than does orally administered therapy.

The only way to accurately compare clinical differences between oral and transdermal estrogen delivery is to establish that the two methods produce similar blood levels and that clinical differences reflect the first-pass effect through the liver. This is difficult to accomplish because the oral first-pass effect raises sex hormone-binding globulin (SHBG) levels such that total serum estradiol levels are greatly affected. A study of 18 women showed that oral estrogen increased SHBG by 67% to 171%, whereas transdermal estrogen did not alter SHGB levels.[65] Estrogen-induced changes in SHBG may be clinically significant because estrogen unbound to SHBG determines the estrogen effects of a given regimen. The only study that measured free estradiol levels, compensating for increases in SHBG, indicated at 12 weeks, that serum free estradiol levels in the oral group were similar to those in the transdermal group.[32] However, because these results were derived from only 18 women, the effect of oral and transdermal doses on free estradiol levels has not been reliably established. A potential advantage of transdermal treatment because it has no effect on SHGB levels is the absence of a reduction in free, unbound testosterone levels as is observed with oral therapy.[63] Thus, the transdermal method may be indicated in women with impaired sexual function.

SUMMARY

Based on the evidence to date, transdermal estrogen therapy is an option that should be offered to all women who wish to use hormone therapy. Sufficient evidence suggests that transdermal therapy is the method of choice for the following patients:

1. ***Women at high risk for VTE.***
2. ***Women with spontaneous or estrogen-induced hypertriglyceridemia.***
3. ***Obese women with metabolic syndrome.***

Transdermal estrogen therapy should be seriously considered in smokers, women with hypertension, and possibly for women with impaired sexuality.

Oral Versus Transdermal Administration. It is difficult to draw conclusions about clinical differences between oral and transdermal hormone delivery based on secondary markers. Epidemiologic studies on clinical events are needed. However, this is a challenge because of the relatively small number of women receiving transdermal estrogen. In addition, the studies must adjust for individual variability of dosing to ensure that circulating estrogen levels in the patients being studied are similar.

The Vaginal Administration of Estrogen—Very Low-Dose Method

Some patients do not gain full relief from the symptoms of vaginal atrophy with oral or transdermal administration of estrogen. Local vaginal administration makes sense for these patients. Vaginal treatment is especially helpful when a rapid response is desired. In addition, there are many women who desire the genitourinary effects of estrogen but either must or wish to avoid systemic therapy. Overall, there is no evidence that one method or preparation is superior to the others in achieving clinical response. ***Measurement of vaginal pH from the lateral vaginal wall is a simple and inexpensive way to assess adequate treatment of the vagina. It has been impressive in our experience and others how an acidic pH (<4.5) obtained from the lateral, outer third of the vagina correlates well with good estrogen effects.***[66–68]

Many clinicians believe that estrogen administered intravaginally is not absorbed, and systemic effects can be avoided. This is not the case. Estrogen in creams is absorbed very readily from a vagina with immature, atrophic mucosa.[69] Indeed, the initial absorption is rapid, and relatively high circulating levels of estrogen are easily reached. As the vaginal mucosa matures, absorption decreases.[70] This decline takes approximately 3–4 months, after which lesser but still significant absorption takes place.[71] Effective treatment of vaginal atrophy with minimal absorption can be achieved with the administration of 0.3 mg conjugated estrogens, 2–3 times per week.[72, 73] ***We believe that treatment with a vaginal cream longer than 6–12 months requires endometrial surveillance.***

The amount of estradiol delivered in low-dose tablet form or a ring is not sufficient to treat menopausal symptoms, but effectively improves local urogenital atrophy and reduces recurrent urinary tract infections. This has been accepted as a method to relieve atrophic vaginal symptoms in women with contraindications to estrogen treatment; however, systemic effects do occur.

Estring is a 55-mm diameter silicone ring that contains 2 mg estradiol, with a release rate of 7.5 μg/day for 90 days.[74] European studies have demonstrated that vaginal maturation can be achieved with this ring that can be left in place for 3 months, with a low-level of systemic absorption. [75, 76] The subjective symptoms associated with vaginal atrophy are rapidly relieved. No change in endometrial thickness was observed after 1 year of treatment.[77]

Vagifem is a tablet that contains 25 μg estradiol, and the initial dose of 1 tablet daily produces relief from atrophic symptoms within 2 weeks.[78] After the first 2 weeks, the maintenance dose is twice weekly, and endometrial thickness has been reported to not change from baseline; however, the study was only 6 months in duration.[79] One 2-month study found no evidence of endometrial stimulation; another reported 1 case of vaginal bleeding with endometrial proliferation.[80, 81] A smaller dose tablet, 10 μg, also improves vaginal atrophy, but it is not as effective as the larger dose.[82]

The systemic absorption of estrogen from the low-dose estradiol ring or tablet is very low, especially after the vagina achieves estrogen-induced maturation (about 3 months). Is this low level of absorption free of the risk of endometrial hyperplasia? The problem is that all stud-

ies have been too short (all 1 year or less, except one 2-year study) to determine long-term endometrial safety. Although systemic absorption occurs, the circulating estradiol levels with these low-dose methods remain in the normal postmenopausal range.[74, 79, 83–86] But, because the small increase in circulating estradiol levels causes distant target tissue responses (e.g., an increase in bone density or an improvement in the lipid profile[87, 88]), clinicians cannot assure patients that these methods are totally free of systemic activity. ***Although the change in blood levels is very slight, and for that reason not effective for the relief of vasomotor symptoms, we believe long-term treatment requires ultrasonographic monitoring of endometrial thickness with biopsy when indicated. This ultrasonographic approach is more preferable than complicating the treatment regimen with the addition of a progestational agent. We further suggest that each patient titrate her dose and schedule of treatment to balance an effective response with minimal dosing. For women who are breast cancer survivors and are considering this treatment, clinicians and patients must accept a small but real unknown risk.***

The Vaginal Administration of Estrogen—Standard Dose Method

A vaginal ring (FemRing, MenoRing) that releases estradiol acetate provides 50 or 100 µg estradiol per day over a 3-month time span.[89, 90] Estradiol acetate is a prohormone, which is rapidly hydrolyzed to estradiol that is reflected in blood estradiol levels similar to those achieved with oral and transdermal methods. The systemic levels achieved effectively suppress hot flushing, and a beneficial impact on bone is to be expected. Endometrial protection requires the addition of a progestin in the presence of a uterus.

Estradiol Implants

Estradiol pellets are available in doses of 25, 50, and 75 mg for subcutaneous administration twice yearly. The 25-mg pellet provides blood levels in the range of 40–60 pg/mL, levels that are comparable with those obtained with standard oral doses.[91, 92] However, the effect is cumulative, and after several years the blood levels are 2–3 times higher. Significant blood levels of estradiol will persist for up to 2 years after the last insertion. We believe that the estradiol pellets confer no advantages over the usual treatment regimens. We recommend that women receiving pellets be monitored with blood estradiol levels, and levels greater than 200 pg/mL (and preferably, 100 pg/mL) should be avoided by a greater interval between insertions.

Percutaneous Estrogen

Transdermal estradiol can also be administered by a gel, emulsion, or spray. The gel, available in various trade names (Divigel, Elestrin, Estrogel, Estreva Gel), is applied by a metered pump or from a foil packet once daily on an arm, anywhere from the wrist to the shoulder, or the thigh, without rubbing or massaging and alternating sides.[93, 94] The emulsion, Estrasorb, is packaged in foil pouches; usually two packets are applied daily, one to each thigh, and rubbed in thoroughly. Evamist is the transdermal spray, and the usual dose is one spray daily to the forearm (if more than one dose is required daily, each spray is on a separate site).[95] Simultaneous use of sunscreen on the site of administration should be avoided. If dosage is being monitored by blood estradiol levels, blood should be drawn from a site where transdermal estradiol has not been applied for several days. Although comparison studies have not been performed, it is reasonable to expect similar pharmacokinetics for all transdermal methods. As with pellets, we recommend that blood estradiol levels be monitored and maintained at a level below 100–200 pg/mL.

Monitoring Estrogen Dosage with Estradiol Blood Levels

Monitoring the estradiol blood level in postmenopausal women receiving hormone therapy is not as straightforward as it would seem. There are two primary difficulties. First, the clinical assays available differ considerably in their technique and quality (laboratory and antibody variations). Second, the various commercial products represent a diverse collection of estrogenic compounds, ranging from estradiol to unique equine estrogens. Although the body interconverts various estrogens into estrone and estradiol, is this process relatively consistent within and between individuals? A highly specific assay for estradiol will detect very low levels of estradiol in women receiving 0.625 mg conjugated equine estrogens; nevertheless, most clinical assays will report a level of 40–100 pg/mL in these women.

We find measurement of blood estradiol levels very useful in selected patients, such as the patient who requests ever-increasing doses of estrogen for the treatment of symptoms, which in the presence of very high blood levels of estradiol can be confidently diagnosed as psychosomatic. We further advocate titering of estrogen dosage with blood estradiol levels in women who fail to demonstrate a positive bone response on treatment, as discussed in Chapter 17. What each clinician must do is learn what blood level of estradiol as performed by the local laboratory is associated with the standard doses of hormone therapy (0.625 conjugated estrogens, 1 mg estradiol, 50 µg transdermal estradiol) and consistently use the same laboratory. In our laboratory this range is 40–100 pg/mL estradiol when the estrogen is taken the evening before the office visit (with transdermal administration blood sampling should be obtained the day before new patch placement); the range reflects individual variation including the variability from peak to nadir values. ***Remember that because FSH is regulated by a factor other than estrogen (i.e., inhibin), FSH levels cannot be used to monitor estrogen dosage.*** Postmenopausal hormone therapy will produce only a 10–20% decrease in FSH and LH, and there is great individual variability in the responses.[96]

Products containing ethinyl estradiol will not affect the measurement of circulating estradiol levels. Ethinyl estradiol circulates without being changed, and the antibodies in the immunoassays for estradiol will not recognize it. It is for this reason that women on oral contraceptives have very low measurements of estradiol. This problem for the postmenopausal use of ethinyl estradiol is not a major handicap because ethinyl estradiol is slowly metabolized, and blood levels are relatively stable with less variation from individual to individual compared with the other estrogen formulations.

Estrogen-Progestin Sequential and Continuous Regimens

Postmenopausal hormone therapy initially consisted only of sequential regimens that were logical reflections of the cyclic estrogen and progesterone patterns in a premenopausal menstrual cycle. Clinical trials established the doses and durations for progestin administration that would effectively protect the endometrium against unchecked proliferation.[97] Progestin withdrawal bleeding occurs in 80–90% of women on a sequential regimen,[98–100] and for this reason the continuous combined method of treatment evolved to improve patient continuance that was adversely affected by bleeding and other symptoms triggered

by the cyclic hormonal changes. The addition of a daily dose of a progestin to the daily administration of estrogen allowed the progestin dose to be smaller, provided effective protection against endometrial hyperplasia, and resulted in amenorrhea within 1 year of treatment in 80–90% of patients.[99, 101–103]

In the sequential regimen, estrogen is administered daily and progestins for 2 weeks of every month, using the *comparable* doses of the following progestins[100, 101, 104, 105]:

5 mg medroxyprogesterone acetate, or
0.7 mg norethindrone, or
1.0 mg norethindrone acetate, or
200 mg micronized progesterone.

In the daily continuous, combined regimen, progestins are combined with estrogen in the following *comparable* doses[102, 103, 106]:

1.5 or 2.5 mg medroxyprogesterone acetate, or
0.35 mg norethindrone, or
0.5 or 1.0 mg norethindrone acetate (0.1 mg dose is available), or
100 mg micronized progesterone or
2 mg drospirenone or
2 mg dienogest.

These hormonal regimens are combined with daily calcium supplementation (500 mg with a meal) and vitamin D (1,000–2,000 IU daily).

There has been a progressive decrease in dose used for postmenopausal hormone therapy. For many years, the standard dose of estrogen was 0.625 mg conjugated estrogens, 1–2 mg micronized estradiol, 1–2 mg estradiol valerate, or equivalent doses of other estrogens such as 5 µg ethinyl estradiol. Lower doses have been proven *on the average* to be as effective as these "standard" doses, providing clinicians and patients with more options. Conjugated estrogens in a dose of 0.3 or 0.45 mg effectively produce a gain in bone density when combined with 1.5 mg medroxyprogesterone acetate, and a dose of 0.5 mg micronized estradiol produces comparable effects.[107–110] The 0.45/1.5 mg and 0.3/1.5 mg conjugated estrogens/medroxyprogesterone acetate combinations improve vaginal atrophy, reduce hot flushing, and improve measures of sexual function in a pattern that is quantitatively and qualitatively similar to the 0.625/2.5 mg combination with less mastalgia.[111, 112] These lower-dose combinations are associated with less breakthrough bleeding and a higher rate of cumulative amenorrhea compared with older standard doses and retain the favorable changes in the lipid profile.[113, 114] At these lower doses of conjugated estrogens, the combination with progestin produces an additive effect; therefore, when these lower doses of estrogen are used without progestin, the effect on hot flushing will not be as great. In a dose-response study, the most efficacious dose of oral micronized estradiol was 1 mg/day.[115] The lower-dose combination of ethinyl estradiol and norethindrone acetate (2.5 µg/0.5 mg) is nearly as effective in treating hot flushes as the higher dose combination (5.0 µg/1.0 mg).[116]

Keep in mind our concern that with lower doses there will be more women who respond poorly, probably because of a greater rate of metabolism and clearance (discussed in Chapter 17).

Two metabolites of progesterone, allopregnanolone and pregnanolone, are believed to be responsible for progesterone's unique sedative effect. Treatment regimens with micronized progesterone should be taken at bedtime, and these estrogen-progesterone combinations are a good choice for women with sleep difficulties. A study in a sleep laboratory has demonstrated a significant improvement in sleep quantity and quality in women using a

sequential regimen of estrogen and micronized progesterone in contrast to no effect in the group using medroxyprogesterone acetate.[117]

Progestational Side Effects

Many women do not tolerate treatment with progestational hormones. Typical side effects include breast tenderness, bloating, and depression. These reactions are significant detrimental factors with continuance. However, appropriately designed, placebo-controlled studies fail to document adverse physical or psychological effects with short-term treatment utilizing medroxyprogesterone acetate, except for breast discomfort.[118–121] This suggests that progestin side effects other than mastalgia are related to duration of treatment or that only studies with large numbers of subjects will detect the small percentage of women who have problems (and both explanations are probably true).

Breast discomfort associated with postmenopausal hormone therapy can be attributed largely to progestins. In the PEPI randomized trial, an increase in mastalgia was observed *only* in 28.7% of the women receiving estrogen-progestin combinations, containing either medroxyprogesterone acetate or progesterone.[120] Comparison studies have not been performed to address whether this symptom is minimized by particular progestins. It has been our experience, that changing to a regimen containing norethindrone or norethindrone acetate has been beneficial (but this may reflect either a placebo response or diminishing severity with time).

Can the progestational agent be administered less frequently? We are secure in our position, supported by clinical data, that a daily combination program effectively prevents endometrial hyperplasia. A sequential regimen that incorporates progestin exposure for less than 14 days has over time an increased risk of endometrial hyperplasia.[122, 123] In a Finnish study, sequential regimens in standard schedules used for at least 5 years were associated with an increased risk of endometrial cancer.[124] Thus, sequential regimens with less than 14 days of progestin monthly or *even long-term use of recommended schedules* do not match the protection offered by the daily, continuous method of estrogen-progestin treatment.

Experience with extended cycle regimens is very limited. The administration of medroxyprogesterone acetate every 3 months was associated in 1 study with longer, heavier menses and unscheduled bleeding and a 1.5% incidence of hyperplasia at 1 year, whereas in another study, overall bleeding was less, but the incidence of hyperplasia was approximately 4%.[125, 126] In a Dutch study that was only 12 weeks in length, simple endometrial hyperplasia was encountered at the end of the unopposed estrogen phase.[127] In yet another study, there was no endometrial hyperplasia encountered by 143 women who completed 2 years of treatment; however, the progestin administered every 3 months was of high dosage, 20 mg medroxyprogesterone acetate daily for 14 days.[128] In Finland, the addition of progestin at 3-month intervals was associated with a striking increase in the risk of endometrial cancer when this regimen was used for many years.[124] Most impressively, the Scandinavian Long Cycle Study, a clinical trial scheduled to last 5 years, was canceled after 3 years because of a 12.5% incidence of endometrial pathology and 1 case of endometrial cancer.[129] Therefore, if a patient chooses an extended cycle regimen, endometrial monitoring is required. ***In our view, an annual endometrial biopsy is strongly recommended in estrogen users exposed only intermittently to progestin treatment. Any program that differs from the standard regimen is untested by clinical studies of sufficient length and patient numbers and, therefore, requires periodic surveillance of the endometrium. Even the long-term use of standard sequential regimens is subject to a small increase in the risk for endometrial cancer, and endometrial surveillance should be considered in women using this method.***

Progestins Available Worldwide		
		Estimated Comparable Oral Doses
Progesterone	Oral peanut oil tablet	200 mg
21-Carbon Derivatives:	Medroxyprogesterone acetate	5.0 mg
	Megestrol acetate	5.0 mg
	Cyproterone acetate	1.0 mg
	Dydrogesterone	10.0 mg
	Chlormadinone acetate	5–10.0 mg
	Medrogestone	10.0 mg
19-Nor Pregnanes:	Trimegestone	0.0625–0.50 mg
	Promegestone	0.5 mg
	Nomegestrol	5.0 mg
	Nomegestrol acetate	3.75–5.0 mg
	Demegestone	
	Nestorone (nonoral)	0.05–0.1 mg
19-Nortestosterone Family:		
Ethinylated:	Norethindrone	0.7–1.0 mg
	Norethindrone acetate	1.0 mg
	Levonorgestrel	0.075 mg
	Desogestrel	0.15 mg
	Norgestimate	0.09 mg
	Gestodene	0.20 mg
	Norethynodrel	
	Lynestrenol	
	Ethynodiol diacetate	
Nonethinylated:	Dienogest	2.0 mg
Derived from Spironolactone and Nonethinylated:	Drospirenone	2.0 mg

Some patients are very sensitive to medroxyprogesterone acetate. In our experience, these patients are often relieved of their symptoms by switching to norethindrone. In a sequential regimen, the dose of norethindrone is 0.7 mg (available in the progestin-only, minipill oral contraceptive; each pill contains 0.35 mg norethindrone). In the continuous, combined regimen, the dose of norethindrone is 0.35 mg daily. Commercial combination products are available containing estradiol and norethindrone acetate.

Progesterone can be administered in a vaginal gel that allows the delivery of very low doses that can effectively protect the endometrium with low systemic levels because of a first-pass effect on the uterus.[130] The administration of 90 mg every 2 days produces secretory changes in the endometrium.[131] An application of the 4% commercial preparation twice weekly protects the endometrium and is associated with amenorrhea in most patients. In a sequential regimen, the 4% preparation should be applied daily for at least 14 days each month. No long-term studies are available that document endometrial safety and metabolic effects.

The transdermal estrogen-progestin combinations incorporate norethindrone acetate in a daily dose of 0.140 or 0.250 mg; or levonorgestrel in daily doses of 0.007, 0.015, 0.030, and 0.040 mg/day; and in a sequential regimen, norethindrone acetate, 0.250 mg, or levonorgestrel, 0.010 mg.[132–134]

The Progestin Intrauterine Device

The contraceptive levonorgestrel-releasing intrauterine system (IUS) has been reconfigured in a smaller model (not yet available) that releases 10 µg of levonorgestrel per 24 hours; however, the larger, contraceptive levonorgestrel IUS (Mirena) can also be used in postmenopausal women.[135–139] The intrauterine presence of the progestin effectively protects the endometrium against hyperplasia and cancer.[140] The local site of action provides endometrial protection and escapes systemic progestin side effects; for example, estrogen's favorable lipid effects are not attenuated.[141] As with the oral continuous, combined regimens, there is irregular breakthrough bleeding in the first 6 months, and after 1 year, approximately 60–70% of the women are amenorrheic. The levonorgestrel system has the advantage of a 10-year duration of use. The frameless IUD has also been designed for postmenopausal use (FibroPlant-LNG), delivering 14 µg of levonorgestrel per 24 hours.[142, 143] These methods provide treatment options that minimize, if not totally eliminate, the systemic effects of progestins. See Chapter 25 for a complete discussion of advantages and problems.

Progestins for Hysterectomized Women

There are some special conditions that warrant the use of a combined estrogen-progestin regimen in hysterectomized women.

1. Because adenocarcinoma has been reported in patients with pelvic endometriosis who are treated with unopposed estrogen,[144–149] the combined estrogen-progestin program is strongly advised in patients with a past history of endometriosis. In addition, we have encountered a case of hydronephrosis secondary to ureteral obstruction caused by endometriosis (with atypia) in a woman on unopposed estrogen for years after hysterectomy and bilateral salpingo-oophorectomy for endometriosis.

2. Patients who have undergone procedures that have the potential to leave residual endometrium (e.g., a supracervical hysterectomy) should be treated with an estrogen-progestin combination. Responsive endometrium may be sequestered in patients who have undergone endometrial ablation,[150, 151] and combined estrogen-progestin treatment is recommended for these women.

3. It has been reported that patients who have had adenocarcinoma of the endometrium can take estrogen without fear of recurrence (discussed later in this chapter), but the combination of estrogen-progestin is recommended in view of the potential protective action of the progestational agent. Treatment can be initiated immediately postoperatively.

4. The combined estrogen-progestin approach makes sense for patients previously treated for endometrioid tumors of the ovary.[152]

Treatment with Androgens

The total amount of testosterone produced after menopause is decreased because the amount of the primary source, peripheral conversion of androstenedione, is reduced. The early postmenopausal circulating level of androstenedione decreases approximately 62% from young adult life.[153] Nevertheless, the menopausal decline in the circulating levels of testosterone is not great, from no change in many women to as much as 15% in others.[153–156] In an excellent longitudinal Australian study from 5 years before menopause to 7 years after menopause, the circulating levels of testosterone did not change.[157] Indeed, because of a decrease in sex hormone-binding globulin, this Australian study calculated an increase in free androgens. The total amount of testosterone produced per day, however, is slightly decreased because the primary source, the peripheral conversion of androstenedione, is reduced. Because of this decrease, some argue that androgen treatment is indicated in the postmenopausal period.

The potential benefits of androgen treatment include improvement in psychological well-being and an increase in sexually motivated behavior. ***Hypoactive sexual desire disorder is defined as a decrease in sexual activity sufficient to cause distress.*** Beneficial effects of androgen treatment have been reported with the administration of relatively large doses of androgen.[158] In a well-designed, placebo-controlled study, lower doses of androgen (but still very pharmacologic, 5 mg methyltestosterone) contributed little to actual sexual behavior, although an increase in sexual fantasies and masturbation could be documented.[159] The transdermal testosterone treatment of women improved sexual function compared with a placebo group only in the dose that raised circulating testosterone levels to about 100 ng/dL (the upper limit of normal for reproductive-age women is 80 ng/dL in most laboratories).[160]

Any benefit must be balanced by the unwanted effects, in particular, virilization (acne, alopecia, and hirsutism) and a negative impact on the cholesterol-lipoprotein profile. In a short-term study comparing a product with estrogen and a relatively low oral dose of testosterone (1.25 mg methyltestosterone) to estrogen alone, a negative impact on the lipid profile was apparent within 3 months.[161] Over a 2-year period, the administration of estrogen (1.25 mg) combined with 2.5 mg methyltestosterone produced a significant overall adverse impact on the cholesterol-lipoprotein profile.[162] In addition, 30% of the patients experienced acne, and 36% developed facial hirsutism. A lower dose of this combination (0.625 mg esterified estrogens and 1.25 mg methyltestosterone) also significantly lowered HDL-cholesterol.[163] The adverse impact on the lipid profile is less (and may even be avoided) by the parenteral administration of testosterone.[164] Of course, the clinical effects of these metabolic changes are not known.

It should be remembered that androgens do not protect the endometrium, and the addition of a progestin is still necessary. It is uncertain (and unstudied) how much aromatization, especially local aromatization in target tissues, of the administered testosterone increases the estrogen impact and whether this might increase the risk of endometrial and/or breast cancer. The addition of androgen does not reduce the amount of breakthrough bleeding women experience with a continuous combination regimen.[165] Adding testosterone to an estrogen therapy program has been reported to provide no additional beneficial impact on bone or on relief from hot flushes.[162, 166] On the other hand, others have demonstrated a greater increase in bone density with an estrogen-androgen combination compared with estrogen alone, although the blood levels achieved were higher than those associated with standard postmenopausal hormone therapy.[164] In another study, only a very pharmacologic dose of methyltestosterone added to the bone density achieved with estrogen alone.[167] A greater effect on bone associated with androgen treatment may be indirect, reflecting higher free estrogen levels because of a reduction in sex hormone-binding globulin and/or androgen-induced changes in muscle mass.

There is no doubt that pharmacologic amounts of androgen can increase libido, but these same doses produce unwanted effects.[168] In addition, patients on high doses of androgens often are somewhat addicted to this therapy. Small amounts of androgen supplementation can be provided in situations in which the patient and clinician are convinced that a depressed libido cannot be explained by psychosocial circumstances. In these cases, the lipid profile should be carefully monitored. Any positive clinical response may well be a placebo effect. ***Our preferred method is to use a testosterone product that can be titered by measuring the total testosterone blood level and maintaining the concentration in the range of 20–80 ng/dL.*** The products that are available in various parts of the world include testosterone undecanoate (used orally), sublingual micronized testosterone, intramuscular injections, subcutaneous implants, and transdermal preparations.[169] For example, the testosterone transdermal gel marketed for use in men (AndroGel), 5 gm/day, can be used at a starting dose of about 1 gm/day. Testosterone undecanoate produces very high testosterone levels with great variability and is not recommended.[170] A transdermal testosterone patch for women is available in many countries (Intrinsa, applied twice a week), but, in our view, monitoring with testosterone blood levels will still be important.

The initial clinical trials concluding that the 300 μg transdermal dose of testosterone was effective for low libido consisted of women with either surgical (1,172 women) or natural menopause (549 women) who were being also treated with estrogen.[171, 172] A 1-year, randomized, placebo-controlled clinical trial of 814 women with hypoactive sexual desire disorder and not on estrogen therapy from 65 centers in the U.S., Canada, Australia, the U.K., and Sweden assessed the impact of transdermal testosterone that delivered 150 or 300 μg/day.[173] The higher dose of testosterone increased sexuality (including desire, arousal, orgasm, and pleasure) by 1.4 episodes per month compared with placebo. This increase appeared as early as the second month of treatment. The lower dose did not differ from placebo. In the higher dose group, 30% reported unwanted androgenic effects (essentially an increase in facial hair). In addition 1 woman in the low-dose group and 3 women in the high-dose group developed clitoral enlargement (the enlargement resolved in the woman receiving the low dose, but not in the high-dose women).The frequency of acne, alopecia, and voice deepening was the same in all groups. It is certainly plausible that with longer exposure to the high dose, more and more women would develop androgenic side effects. There were four cases of breast cancer in the treatment groups and none in the placebo group; however, one was diagnosed after only 4 months of treatment and one had a bloody nipple discharge before the trial started.

In the Nurses' Health Study, the risk of invasive breast cancer associated with the use of combined estrogen and testosterone was nearly 2-fold increased.[174] This report from the Nurses' Health Study is complicated by the same problem in other breast cancer reports from this cohort: the hormone users (in this case, estrogen and testosterone) differ substantially from never users. This requires multiple statistical adjustments, a process that is further influenced by the number of cases involved. The analysis is limited by relatively small numbers; there were only 29 cases of breast cancer among the estrogen-testosterone users. Nevertheless, the results should raise caution regarding the postmenopausal use of androgens.

If testosterone affects breast tissue, does it do so directly or is it aromatized locally into estrogen? The majority of studies indicate that testosterone inhibits proliferation of breast cancer cell lines in vitro, as well as in vivo markers of breast epithelial proliferation in animals and women,[175, 176] suggesting that aromatization is of greater concern. Testosterone preparations such as implants and transdermal applications do carry the risk of target tissue aromatization, perhaps raising *local* estrogen levels to high levels in breast tissue. Perhaps an argument against this possibility was the failure to demonstrate any increase in breast density associated with transdermal testosterone treatment, even with the higher dose.[177] However, the mean age of the women in this study was 54.6, and an increase in breast density with estrogen-progestin therapy is largely observed in women over age 55.

In women younger than age 55, it is difficult to find any differences between hormone users and nonusers.[178]

In the transdermal testosterone trial with women *not* on estrogen treatment, there was a 10.6% incidence of vaginal bleeding in the women who had not undergone hysterectomy and were receiving the higher dose, compared with 2.6% in the placebo group and 2.7% in the low-dose group.[173] Was this due to aromatization of testosterone in the endometrium? There were no cases of endometrial hyperplasia or cancer in this trial, but again, a longer duration of exposure might have unwanted consequences. This issue cannot be resolved without long-term data. In addition, the long-term effects on the cardiovascular system are unknown.

Response in the clinical trials with transdermal testosterone did not correlate with testosterone levels at baseline, and higher levels during treatment did not predict androgenic side effects. This is not surprising because measurement of free and bioavailable testosterone is subject to considerable inaccuracy and variability. For this reason, testosterone levels cannot be used to diagnose the hypoactive sexual desire disorder.[179] The transdermal clinical trials have reported that all testosterone levels remained within the premenopausal ranges. However, the mean level of free testosterone was relatively high at 6.8 pg/mL, although within the reference range. According to the data in the supplemental appendix, available only online, the mean levels were at or above the upper end of the reference age. In addition, because of individual variability, there was a wide range of testosterone levels, with a significant number of values elevated above normal. For many women these are not physiological levels! Isn't the fact that 30% of the women receiving the high dose reported an increase in androgenic effects evidence of a pharmacologic effect? We don't know if it is possible to avoid unwanted consequences by careful monitoring of blood levels.

There is little doubt that the administration of pharmacologic amounts of testosterone can produce favorable effects on sexuality, but it remains doubtful that maintaining testosterone levels within the normal physiologic range can have a beneficial impact on health. Some women receiving pharmacologic amounts of testosterone develop very high circulating levels. The fundamental problem is that the long-term consequences of pharmacologic amounts of testosterone are unknown; however, long-term safety trials are underway.

If a clinician and a patient choose to use supplemental androgens, our advice is to select a treatment that can be monitored with measurements of total testosterone in serum. The choices include the testosterone transdermal patch, a testosterone skin gel (on the market for use in men), and testosterone compounded for individual use by a pharmacist. We are left with this question: Is a modest increase of 1 or 2 episodes per month sufficient to offset the unanswered question of long-term safety? Some women would say yes, but the clinician has an obligation to avoid excessive doses and to educate the patient regarding the unanswered questions.

Selective Estrogen Agonists/Antagonists (Selective Estrogen Receptor Modulators)

A greater understanding of the estrogen receptor mechanism (Chapter 2) allows us to understand how mixed estrogen agonists/antagonists can have selective actions on specific target tissues. New agents are being developed in an effort to isolate desired actions from unwanted side effects. Indeed, in time we can expect to see new products with progressively better agonist/antagonist profiles, yielding increasingly user-friendly drugs.

Raloxifene

Raloxifene exerts no proliferative effect on the endometrium but produces favorable responses in bone and lipids.[180–183] The MORE (Multiple Outcomes of Raloxifene Evaluation) study of raloxifene administration to osteoporotic women reported results from 8 years of follow-up.[184, 185] Women with low T-scores had approximately a 50% reduction in vertebral fractures with raloxifene treatment, and with previous vertebral fractures, approximately 35%. However, ***there has been no evidence of a reduction in hip or wrist fractures.*** The major side effect was about a 3-fold increase in venous thromboembolism. Raloxifene (and tamoxifen) share with estrogen an increased risk for venous thromboembolism.[186] The size of the risk is comparable for all three drugs, and nearly all the cases occur in the first 1 or 2 years of exposure. A small number of women experience hot flushing with raloxifene. Raloxifene treatment in the MORE trial had neither a positive nor a negative effect on cognition.[187]

Women who received raloxifene in the MORE trial had about an 80% reduction in the incidence of estrogen receptor-positive breast cancers. The CORE study, the Continuing Outcomes Relevant to Evista trial, was designed to measure the impact of four additional years of raloxifene (60 mg/day), beginning during the fourth year of the MORE trial.[188] Of the 7,705 participants initially randomized in the MORE trial, 3,510 women elected to continue raloxifene treatment (2,336 completed the CORE trial) and 1,703 continued on placebo (1,106 completed the trial). During the 4-year CORE study, raloxifene treatment was associated with a 66% (HR=0.34; CI=0.18–0.66) reduction of estrogen receptor-positive invasive breast cancer in the treated group. There was no difference in estrogen receptor-negative tumors. Over the entire 8-year period, the reduction in estrogen receptor-positive cancers reached 76%. In the 8-year period, there was no difference in the number of deaths in the two groups.

The Study of Tamoxifen and Raloxifene (STAR) trial enrolled 19,747 women at increased risk of breast cancer who were randomized to treatment with either raloxifene, 60 mg daily, or tamoxifen, 20 mg daily, in more than 500 centers in the U.S., Canada, and Puerto Rico.[189] After an average treatment period of almost 4 years, the numbers of invasive breast cancers were identical in the two groups of women. It was estimated that these results were equivalent to about a 50% reduction (based on the previous results in the tamoxifen prevention trial),[190, 191] but without a placebo arm, an accurate assessment was impossible. Thus, raloxifene appears to achieve the same reduction as tamoxifen in invasive breast cancers with a lesser increase in venous thrombosis, and perhaps no increase in cataracts and uterine cancer. Fractures, as well as strokes and heart attacks, were equally prevalent in the two groups. “Quality of life” was said to be the same for both drugs.

Tamoxifen has been demonstrated to reduce the incidence of both lobular carcinoma-in-situ and ductal carcinoma-in-situ.[190, 191] In the 7-year follow-up report of the tamoxifen for prevention study, the risk for breast cancer was 0.57 (CI=0.46–0.79), a 43% reduction, not the 50% cited in the results above, and the risk for in-situ disease was 0.63 (CI=0.45–0.89), a 37% reduction.[190] Not only did raloxifene not yield a reduction in in-situ cancers, the number with raloxifene in the STAR trial was greater. If raloxifene is truly preventing breast cancer, this should produce a reduction in in-situ disease. Perhaps with longer follow-up, a difference between the two treatment groups will no longer be apparent.

In a 2-year randomized trial in monkeys, raloxifene exerted no protection against coronary artery atherosclerosis despite changes in circulating lipids similar to those achieved in women.[192] The Raloxifene Use for the Heart (RUTH) study included more than 10,000 women from 26 countries, either at high risk for myocardial infarction or with known coronary heart disease.[193, 194] The participants were randomized to placebo or raloxifene, 60 mg daily, and followed for up to 7 years. There was no effect of raloxifene treatment on coronary heart disease events; however, there was a small increase in stroke mortality.

The results of the RUTH trial are not surprising. The known favorable impact of raloxifene on the cholesterol-lipid profile was not robust enough to prevent coronary events.

Because the Women's Health Initiative and the Nurses' Health Study reported a reduction in coronary events associated with estrogen therapy administered to young postmenopausal women, the RUTH trial performed a post hoc analysis of the impact of raloxifene according to age of the women at entry to the study as well as subgroups such as the use of medications, including statins.[194] Overall, raloxifene did not increase or decrease coronary events in either of the treated groups. The only category demonstrating a significant difference, a 40% reduction in coronary events, consisted of women less than 60 years of age. Despite the statistically significant reduction in coronary events in women under age 60, there was no relationship in any subgroup according to years since menopause, even in the group less than 10 years postmenopausal. The women who were less than 60 years of age were an average of 9.9 years since menopause, compared with 19.4 years for the overall study population. The finding of a beneficial effect of Raloxifene in the youngest postmenopausal women in the RUTH trial does not jibe with the failure to find a relationship with years since menopause. Out of the 10,101 women, there were only 360 to 460 women (14–18%) in each of the patient groups who were under age 60, *and only 134 women younger than age 60 experienced a coronary event*. Only one subgroup demonstrated a statistically significant different conclusion than the overall finding, *out of 51 analyzed subgroups*. A decision to use raloxifene should not be influenced by its effects on the cardiovascular system. This is a bone and breast decision.

In our view, raloxifene is an option for prevention of osteoporosis-related spinal fractures, especially for patients reluctant to use hormone therapy or in those wanting to combine some bone protection with a reduction in the risk of breast cancer. We recommend, however, periodic evaluation of bone density in the hip, and if bone loss occurs, patient and clinician should consider another treatment option.

Arzoxifene

Arzoxifene is an estrogen agonist-antagonist similar to raloxifene, originally studied for the treatment of breast cancer. Preclinical studies indicated that arzoxifene is an estrogen agonist in bone and on lipids, but an estrogen antagonist in endometrial and breast tissue. Arzoxifene, therefore, had the potential to be as effective as tamoxifen but be free of the risk of endometrial stimulation, and perhaps, venous thrombosis.

A Phase III clinical trial comparing arzoxifene and tamoxifen in the treatment for advanced local breast cancer or metastatic tumors was disappointing.[195] The trial was terminated when it became apparent that the results with arzoxifene were inferior to tamoxifen. Two other members of this drug family, droloxifene and idoxifene, have also failed to yield superior results to tamoxifen for the treatment of breast cancer. For this reason, attention was turned to another use for these agents.

A 2-year, randomized trial compared the bone density responses in 331 postmenopausal women treated with either arzoxifene, 20 mg/day, or placebo.[196] Bone density was slightly increased in the spine and the hip in the treated group compared with placebo. There was no evidence of endometrial stimulation in the treated group either on biopsied specimens or by measurement of endometrial thickness by transvaginal ultrasonography. Three patients in the placebo group and none in the treated group developed breast cancer. However, the initial results from a projected 5-year, phase III trial of 9,354 postmenopausal women indicated an increase in venous thromboembolic events, endometrial changes, and hot flushes, and no decrease in nonvertebral fractures in the treatment arm.[197] The overall impact was deemed insufficient to achieve a competitive edge, and development for osteoporosis prevention and treatment was terminated.

Ospemifene

Ospemifene (Ophena), another estrogen agonist-antagonist, is equally effective as raloxifene, in a dose of 90 mg/day, in suppressing bone turnover and avoiding endometrial and breast stimulation.[198] However, it is being primarily developed in a dose of 60 mg/day given orally for the treatment of vulvar and vaginal atrophy, sites where ospemifene exerts a significant estrogenic impact.[199] There is no effect on vasomotor symptoms. In preclinical studies, ospemifene suppressed the development of breast cancer in the drug-induced mouse model.

Drugs in Development

Several selective estrogen agonists/antagonists that were first tested for the treatment of breast cancer have potential for the prevention and treatment of osteoporosis. One possibility is droloxifene. Other members of this family with potential use are lasofoxifene and ormeloxifene. Preclinical studies have indicated the possibility of greater efficacy and potency in bone effects, but clinical use will depend on the outcomes of clinical trials.

The results from a 5-year, international, randomized trial with lasofoxifene given to postmenopausal women with osteoporosis indicated a reduction in both vertebral and nonvertebral fractures with a dose of 0.5 mg daily.[200] However, the reduction in nonvertebral fractures was small and consisted of forearm and wrist fractures; there was no significant reduction in hip fractures. There was a substantial decrease in the risk of estrogen receptor-positive breast cancer, a modest reduction in coronary heart disease and stroke, and no adverse effects on the endometrium. Treatment was associated with a 2-fold increase in venous thrombosis, hot flushing, and leg cramps. A clear-cut superiority over other treatment options was not apparent.

Keep two important points in mind as new data emerge in this slow and expensive process:

1. ***Comparison Phase III clinical trials are essential. Preclinical studies indicate potential, but only head-to-head comparisons tell us if a new drug is any better than what we already have. The comparison of these agonist-antagonist drugs with tamoxifen is a good example. Hoped-for superiority of the new drugs failed to emerge. In addition, the new drugs will have to perform better than the aromatase inhibitors. Comparison data are also required to determine whether one of the new drugs is superior in avoiding hot flushing and venous thrombosis.***

2. ***Fracture data for both the hip and spine are necessary. These drugs differ in potency as measured by bone density and biochemical markers of bone metabolism. Greater potency, therefore, gives some hope that one of these new drugs will overcome the serious drawback of raloxifene treatment, a lack of effect in preventing hip fractures.***

Bazedoxifene

Bazedoxifene belongs to the estrogen agonist-antagonist family of drugs. It has favorable effects on bone and lipids, but does not affect the endometrium or the breast.[201, 202] Bazedoxifene in a dose of 20 mg daily decreased the risk of all clinical fractures in a randomized

clinical trial, with a potency comparable to other anti-resorptive agents in postmenopausal women at high risk for fractures.[203, 204] In a subgroup of women at higher risk for fractures, bazedoxifene had a reduced risk of nonvertebral fractures (50% reduction with 20 mg), compared with *both* raloxifene and placebo. The only adverse event that differed with treatment was an increase in venous thrombosis with treatment compared with placebo; there was neither a beneficial nor an adverse effect on coronary heart disease or stroke. Treatment for 2 years had no effect on mammographic breast density.[205] The results of this trial indicate that the effect of bazedoxifene on bone should be comparable to that of estrogen and bisphosphonates.

The reduction of nonvertebral fractures with bazedoxifene compared with raloxifene should not be ignored. We have known for some time that even with 8 years of follow-up, raloxifene has no impact on the risk of hip fractures. This is likely because raloxifene is less potent, and thus the hip with a mixture of cortical and trabecular bone is more resistant to raloxifene's effects, compared with the spinal column that is composed of sensitive, trabecular bone. Bazedoxifene partnered with estrogen is called TSEC (tissue-selective estrogen complex). The idea is to gain the benefits of estrogen (bazedoxifene by itself has little impact on hot flushes), protect the endometrium and possibly the breast, and enhance some actions of estrogen, such as a reduction in fractures.[206] Bazedoxifene combined with conjugated estrogens effectively suppresses hot flushing, improves vaginal atrophy and lipids, prevents bone loss, and does not cause breast tenderness.[207–209] The combination of 20 mg bazedoxifene with conjugated estrogens prevents endometrial hyperplasia and has an extremely high rate of amenorrhea.[210, 211] This approach to postmenopausal hormone therapy may eliminate the need for progestational agents.

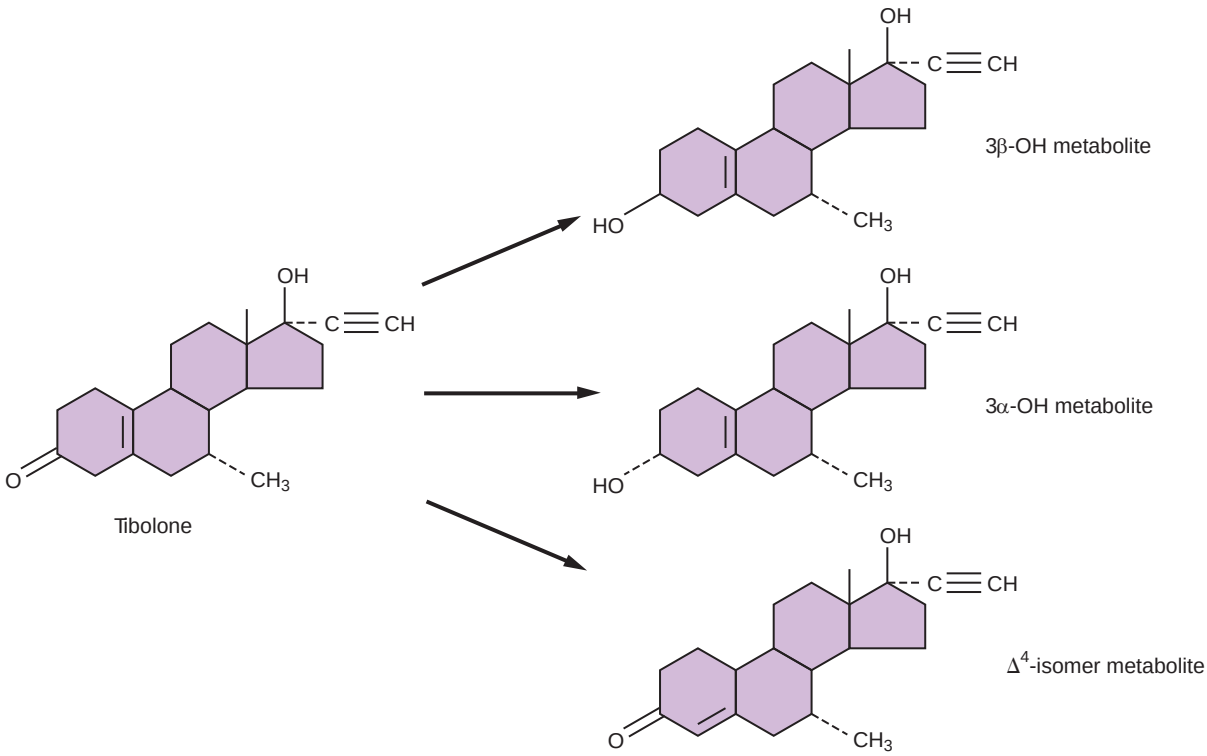
Tibolone

Tibolone is marketed for postmenopausal hormone therapy in many countries throughout the world, but not in the U.S. It was first introduced in the Netherlands, the home country of its manufacturer Organon (now Schering-Plough), which initiated research on this product in the 1960s. Although tibolone was specifically developed as a drug to treat osteoporosis, the clinical performance of tibolone led rapidly to its approval for the treatment of menopausal symptoms as well as prevention of osteoporosis. Various chemists and clinicians have tried to link tibolone with a popular acronym. In our view, this is unnecessary and further compounds the confusion created by multiple trade names (Livial, Liviel, Liviella, Livifem, Boltin, and Tibofem). Tibolone, the generic name, is a good name that is well established in history and deserves retention. Because of its unique metabolism, tibolone can exert different hormonal activities at different sites. This unique characteristic is precisely what makes the drug difficult to understand.

The Chemistry of Tibolone

Tibolone is structurally related to the 19-nortestosterone progestins that are used clinically in oral contraceptives; however, its activity depends on its metabolism. Tibolone {7-α,17-α-17-hydroxy-7-methyl-19-norpregn-5(10)-en-20-yn-3-one} is metabolized among human and nonhuman primates into three biologically active metabolites; the 3α-hydroxy (3α-OH) metabolite and the 3β-hydroxy (3β-OH) metabolite have estrogen agonist properties, whereas the Δ-4 ketoisomer has progestogenic and androgenic effects.[212, 213]

Although tibolone itself binds to the estrogen receptor, *in vivo* the activity of the 3-hydroxy metabolites is 100 times greater, with a greater affinity for the estrogen receptor-alpha than for estrogen receptor-beta.[213] The loss of the hydroxyl group at position 3 of the A ring eliminates estrogenic activity in the Δ-4 isomer. The Δ-4 isomer exerts its androgenic effects primarily in the liver and brain.



The conversion of tibolone into metabolites occurs chiefly in the liver and intestine. The metabolism of the parent compound is rapid and very near total, yielding mainly the 3α-OH and the 3β-OH metabolites in the circulation; the level of the 3α-OH metabolite is 3-fold higher compared with the 3β-OH metabolite.[214, 215] Tibolone and the Δ-4 isomer can be detected only at peak levels 2 h after ingestion, and, even then, the levels are very low, at the limit of detection. The half-life of the metabolites that predominate in the circulation (the 3α-OH and 3β-OH metabolites) is approximately 7 to 8 h, and a steady state is attained by day 5.[216] Eating does not affect the metabolism, and tibolone can be taken at any time of the day.[214] The pharmacokinetics of tibolone are not affected by impaired renal function. There is a very weak effect of the metabolites on cytochrome P450 enzymes, and no interference is to be expected with coadministered drugs.[215] After the initial metabolism of tibolone, the products are rapidly sulfated, and more than 75% of the metabolites circulate as the sulfates to be activated by tissue sulfatases.[215]

Tibolone is available in 2 daily doses, 1.25 and 2.5 mg. There is considerable variability (about 30–40%) within and between subjects, but the 1.25 and 2.5-mg doses produce the same bioequivalence as measured by maximum levels and areas under the curve for the 3α-OH and 3β-OH metabolites.[216] However, there are differences in clinical responses, which influence the choice of dose.

The metabolism of tibolone is not limited to the liver and intestine. Important effects are explained by specific local tissue metabolism. For example, the Δ-4 isomer is primarily produced within the endometrium, binds to the progesterone receptor, and protects the endometrium from the agonist effects of the two estrogenic metabolites.[217–220]

The Effect of Tibolone on Menopausal Symptoms

Menopausal symptoms provide the main motivation for women to use postmenopausal hormonal therapy. Tibolone must perform well in this category in order for it to be an attractive option for clinicians and patients. Clinical studies have established without question that tibolone exerts an estrogenic beneficial impact on hot flushing and vaginal dryness. Appropriate studies have documented that tibolone in a daily dose of 2.5 mg is as effective as standard postmenopausal hormone regimens in treating hot flushing.[221–227] The 1.25-mg dose takes longer to be effective, and this dose also has a higher incidence of persistent flushing.[228] In addition, tibolone effectively treats the side effect of hot flushing associated with gonadotropin-releasing hormone (GnRH) agonist therapy.[229] Fortunately, tibolone treatment provides an estrogenic effect on the vagina. Tibolone, 2.5 mg daily, relieves vaginal dryness and dyspareunia, and in most studies, tibolone is as effective as estrogen treatment.[223–227, 230–233]

A decided advantage for tibolone can be found in studies examining sexuality. In prospective, randomized trials comparing tibolone with estrogen or estrogen-progestin therapy, tibolone has been associated with a better sexual response.[224, 227, 234–237] An increase in libido has been reported in studies comparing tibolone with placebo,[233, 238] and the response has been greater than that with estrogen therapy, comparable to that associated with androgen treatment.[239] The overall effect has included an increase in sexual interest and sexual performance, specifically fantasies, arousal, and orgasm.

There are two possible mechanisms for tibolone's effect on sexuality: a direct androgenic effect of the Δ-4 isomer and/or an increase in the circulating level of free testosterone. Tibolone is associated with a profound change in the circulating levels of sex hormone-binding globulin, about a 50% decrease.[233, 240] This is undoubtedly due to the Δ-4 isomer and an androgenic effect on the liver. Tibolone treatment, therefore, produces a decrease in the concentration of total testosterone (bound and unbound) but a substantial increase in the amount of free, unbound testosterone. This hormonal profile is a striking contrast to that associated with estrogen therapy, which increases sex hormone-binding globulin and decreases both total and free testosterone levels. The androgen side effects of acne and hirsutism, however, have not been reported with tibolone treatment.

When women who had been on tibolone for 10 years were compared with a matched group, the treated women were less clumsy, less anxious in response to mild stress, and demonstrated better memory for facts, although there was no difference in memory for events and worse performance on sustained attention and planning.[241] Overall, tibolone exerts a positive effect on mood that is modest in impact.[240, 242] However, this is an area in which it is not easy to achieve consistent effects, a problem often due to the differences in measurement tools and definitions. The study of cognition is difficult because of the need to match treated and control groups for intelligence, age, occupation, education, and mental state (e.g., depression). Because of this difficulty, the literature reporting the effects of hormone therapy on cognition provides an inconsistent picture. This is further complicated by the sensitivity and appropriateness of the assessment tools that are used. This is an area that requires standardization and new approaches for research, not only for tibolone but for all pharmacologic treatments that affect the central nervous system.

The Effect of Tibolone on the Cardiovascular System

Consistent with reports of tibolone's effects on HDL-cholesterol in postmenopausal women, tibolone-treated monkeys have much lower HDL-cholesterol compared with control

monkeys.[243, 244] Although tibolone treatment resulted in lower circulating HDL-cholesterol, coronary artery atherosclerosis extent in monkeys was not significantly different from the control group. Similar results were observed in the carotid arteries.[244] That observation prompted the question of whether the HDL-cholesterol reductions noted among the animals treated with tibolone were associated with physiologically meaningful reductions in HDL-cholesterol function. HDL-cholesterol has a critical role in reverse cholesterol transport, the mechanism by which cell cholesterol (i.e., artery wall cholesterol) can be returned through the plasma to the liver for excretion.[245] Further, it has been found that cholesterol efflux capacity predicts the severity and extent of coronary artery disease in human patients.[246] Postmenopausal monkeys treated with tibolone had *no* reduction in cholesterol efflux.[247] This disassociation between reductions in circulating concentrations of HDL-cholesterol and the lack of change in HDL-cholesterol function suggests the likelihood that this may account in large part for the finding that coronary artery atherosclerosis was not increased or decreased in the monkey model.

Short-term clinical studies uniformly document that tibolone treatment, 2.5 mg/day, reduces HDL-cholesterol levels in women by about 20%; however, there is also a reduction in total cholesterol (about 10%) and triglycerides (about 20%) and a slight decrease or no change in LDL-cholesterol levels.[235, 248–254] In women, therefore, tibolone does not increase LDL-cholesterol, and the reduction in HDL-cholesterol is less that that recorded in monkeys. In addition, tibolone decreases LDL-cholesterol oxidation and produces a shift away from small dense LDL-cholesterol (which is more atherogenic); both changes would be beneficial.[254] The potential harmful effects associated with reductions in HDL-cholesterol are further balanced by tibolone-associated reductions in endothelin and lipoprotein(a), antiischemic effects detected in women with angina, and an improvement in insulin sensitivity.[253, 255–258] In longer term studies, HDL-cholesterol levels did not come back to baseline at the end of 2 years of treatment, but did return to baseline at the end of 3 years.[253, 259–261] And other studies have found that the decrease in HDL-cholesterol is statistically insignificant.[262, 263]

The recognition that reductions in HDL-cholesterol are potentially harmful is based on the important roles for HDL-cholesterol in the mediation of cholesterol movement from lipid-laden cells and inhibition of LDL-cholesterol oxidation. However, the experimental results in the monkey model indicate that reductions in HDL-cholesterol concentrations are not directly paralleled by reductions in important HDL-cholesterol functions. At least one reason for this lack of direct correlation is the complex nature of HDL-cholesterol lipoproteins, a heterogenous collection of particles that differ in their activities.[264] The overall change in HDL-cholesterol levels will not reflect specific changes in particles that can affect specific biologic activities. Similar to results in the monkey model, a randomized trial in women demonstrated that significant reductions in HDL-cholesterol levels (average 27%) caused by tibolone treatment, 2.5 mg/day, were due to a decrease in one subclass of HDL-cholesterol particles, and measures of HDL-cholesterol antiatherogenic functions (reverse cholesterol transport and inhibition of LDL-cholesterol oxidation) were not impaired.[258] The study was limited by the short, 12-week duration of treatment; however, the findings are consistent with those obtained in the 2-year monkey experiment. These human results were confirmed and strengthened by a study of 68 postmenopausal women randomized to daily treatment for 3 months with either 2.5 mg tibolone or placebo.[265] Changes in HDL-cholesterol were associated with an increase in hepatic lipase activity, an androgenic effect, again without impairing the ability of plasma to maintain cholesterol efflux.

Results in the monkey model are consistent with an overall neutral impact on the cardiovascular system.[244] A long-term (average of 7.5 years) follow-up of women treated with tibolone found no increase in carotid artery intimal media thickness and the number of atherosclerotic plaques, results that are consistent with the monkey model.[266] This neutral impact is further supported by failing to find any effect of tibolone on experimentally induced brachial artery dilation or on vascular resistance measured in the carotid and middle cerebral arteries.[262, 267] On the other hand, a method studying venous dilation in the hand found an improvement in

endothelium-dependent responses after tibolone treatment.[268] Myocardial infarction and heart failure have been reported to be associated with overactivity of the sympathetic component of the cardiac autonomic nervous system, and tibolone treatment decreases plasma levels of free fatty acids, an effect that results in an improved ratio of cardiac sympathetic tone to parasympathetic tone.[269] Another favorable effect connected to tibolone and its metabolites is a direct impact on endothelial cells that results in a beneficial decrease in endothelial-leukocyte adhesion molecules, another human finding similar to that in the monkey trial.[270]

The OPAL study (Osteoporosis Prevention and Arterial effects of tiboLone) was a 3-year, randomized, double-blind trial in six U.S. centers and five European centers, treating 866 postmenopausal women with either 2.5 mg tibolone daily, 0.625/2.5 mg daily of conjugated estrogens/medroxyprogesterone acetate, or placebo.[271] The arterial endpoint of the study was carotid intima-media thickness measured by ultrasonography every 6 months. Both the tibolone-treated group and the estrogen/progestin-treated group demonstrated an increase in intima media thickness over the time period of the study, at a rate significantly greater than the placebo group, leading to the conclusion that both tibolone and estrogen/progestin treatment increased atherosclerosis compared with the placebo group.

In the OPAL trial, European women differed from American women in multiple ways: higher lipids, higher blood pressure levels, more smokers. Hysterectomized women were excluded in the U.S., but not in Europe (28% of the study population). The overall mean results, indeed, indicated a difference comparing both treatment groups to placebo. But in the European women, atherosclerosis, measured by intima-media thickness, improved in the placebo group, making it easy to calculate a significant difference compared to the treated groups! In American women, there were no differences comparing the three treatment groups, all demonstrated progression of thickness. Thus the overall conclusion was inordinately influenced by the results in the European women. The investigators could not explain these differences. Unfortunately, the OPAL trial did not achieve its goal of providing robust data on cardiovascular effects, due to the older age of the women and the notably different results in American and European women. There continues to be good reason to believe that tibolone has a neutral effect on the cardiovascular system. In addition, tibolone does not adversely affect the blood pressure in women with established hypertension.[251]

A case-control study assessed the risk of venous thromboembolism in a very large population of postmenopausal women (23,505 cases and 231,562 controls) derived from the U.K. General Practice Research Database.[272] No increase in risk was observed with the current use of either tibolone or transdermal estrogen compared with a significant increase associated with the current use of oral estrogen.

The Effect of Tibolone on Diabetes

The administration of tibolone, 2.5 mg/day, to older women with type 2 diabetes mellitus produced no significant changes in the lipid profile.[273] Tibolone treatment is associated with an increase in insulin sensitivity in women with insulin resistance, although some have reported no effect in normal women.[250, 258, 274, 275] Therefore, tibolone is an attractive option for postmenopausal women with diabetes mellitus.

The Effect of Tibolone on the Uterus

Tibolone does not stimulate endometrial proliferation. This is because the predominant if not exclusive tibolone metabolite produced within the endometrium is the Δ-4 isomer,

which binds to the progesterone receptor and protects the endometrium from the agonist effects of the two estrogenic metabolites.[217–220] This protective effect has been documented in long-term (up to 8 years) human studies.[219, 220, 223, 231, 232, 276–278] Isolated cases of endometrial proliferation have been reported, for example, 4 of 150 women treated with 2.5 mg daily for 2 years.[279] In a 5-year follow-up, 47 of 434 women experienced bleeding, and of these 11 had endometrial polyps or fibroids, but there were two with simple hyperplasia and two with endometrial carcinoma in situ.[280] This underscores the standard clinical principle to investigate persistent vaginal bleeding in any postmenopausal woman. In the major U.S. clinical trial, three cases of endometrial cancer were observed, but, in each case, pre-existing carcinoma was later detected when the initial biopsy samples were more extensively examined.[253] Nevertheless, a second large 2-year trial was conducted, the THEBES Study, and no endometrial hyperplasia or cancer occurred in the tibolone-treated groups.[281]

The reported breakthrough bleeding rates with tibolone treatment have been comparable to treatment with combined, continuous estrogen-progestin therapy[224, 249, 276 226]; but well-designed comparison clinical trials indicate that the rate is less with tibolone.[225, 227, 233, 282, 283] In addition, amenorrhea is achieved more rapidly; 90% of tibolone-treated women are amenorrheic by 6 months.[225, 280, 284] Bleeding is less in older women and can be greater with the 2.5-mg dose compared with the 1.25-mg dose but the difference is too small to be detected in some studies.[228, 253, 285] Importantly, a lack of correlation has been observed between bleeding and endometrial thickness measured by ultrasonography.[285, 286] ***This again emphasizes the need to biopsy tibolone-treated women with persistent bleeding***.

Careful evaluations of women with fibroids who have been treated with tibolone have revealed no evidence of myoma growth, with up to 3 years of follow-up.[287–289] Furthermore, add-back treatment with tibolone effectively prevents flushing and bone loss and does not impair the fibroid response to therapy with GnRH analogs.[290]

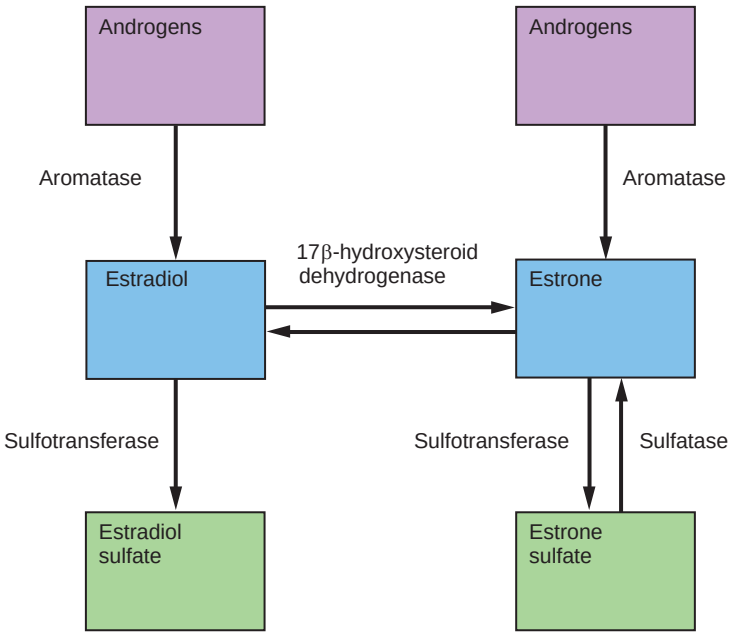
The Effect of Tibolone on the Breast

The breast is a complicated estrogen factory. Breast tissue, normal and abnormal, contains all the enzymes necessary for the formation of estrogens (sulfatase, aromatase, and 17β-hydroxysteroid dehydrogenase) and the conversion of estrogens into their sulfates (sulfotransferase). Estrone sulfate concentrations are high in the breast (higher than in plasma) and even higher in cancer tissue. This state is achieved in postmenopausal women with very low systemic levels of estrogen, indicating that a local mechanism is operative.

The major pathway of estrogen synthesis in human breast tumor cells is by conversion of estrone sulfate to estrone by estrone sulfatase, a pathway that is more important that the aromatase pathway.[291] Aromatase is an enzyme complex that produces the irreversible conversion of androgens to estrogens. The location of aromatase activity is predominantly in the stromal tissue of the breast. Comparing normal to tumor tissue, the levels of estrone sulfate and estradiol were higher in the tumor tissue.[292] Sulfatase activity is higher (130–200 times) than aromatase activity in all breast tissues examined, and the sulfatase and aromatase activity was higher in the tumor tissue than in normal tissue. Thus, estrogen concentrations in the breast are higher in women with breast cancer, and formation of estradiol from sulfated estrogen is the primary pathway. Most importantly, this increase in estrogen activity is independent of the estrogen receptor status of the tissue.

Tibolone and its metabolites inhibit estrone sulfatase and 17β-hydroxysteroid dehydrogenase in normal stromal cells and in hormone-dependent breast cancer cells (MCF-7 and T-47D).[293–296] This inhibits conversion of estrone sulfate to estradiol. In addition, tibolone

and its 3-hydroxymetabolites increase the conversion of estrone back to estrone sulfate by increasing the activity of sulfotransferase.[297] Tibolone and all three metabolites inhibit the conversion of estrone to estradiol by 17β-hydroxysteroid dehydrogenase.[294] Although these effects resemble progestin activity, tibolone is more potent. Tibolone increases aromatase activity in stromal cells but only at high concentrations that are beyond in vivo levels.[295] These tibolone-induced enzyme changes would lower the active estrogen concentrations in breast tissue.



In the rat and mouse breast cancer models (cancer induced by 7,12-dimethylbena{a} anthracene, DMBA), tibolone exerts protective effects to the same degree as tamoxifen.[298] However, tibolone is not antiestrogenic and does not inhibit aromatase. Therefore, the mechanism is explained by the enzyme effects summarized previously, inhibition of sulfatase and 17β-hydroxysteroid dehydrogenase and stimulation of sulfotransferase to increase the production of inactive sulfates.[295] In addition, tibolone increases cellular differentiation and stimulates apoptosis, at least with normal breast cells in vitro.[299] An increase in apoptosis is an action of the parent tibolone and its Δ-4 isomer. Thus, tibolone acts like progestins and weak androgens in breast cell line studies examining proliferation, differentiation, and apoptosis.

Tibolone and its metabolites do not display the same activity directed toward the sulfatase enzyme in all tissues. Strong inhibition of sulfatase is a major feature in breast cells, but tibolone and its metabolites inhibit sulfatase only moderately in the endometrium (contributing to an antiestrogenic action) and provide no inhibition in bone (allowing a greater estrogenic impact).[300]

Postmenopausal hormone therapy increases breast density on mammography in about 10–20% of estrogen users and about 20–35% of estrogen-progestin users, an effect that occurs within the first months of treatment. In contrast, tibolone does not increase breast density and causes far less mastalgia than that seen with estrogen treatment.[225, 237, 253, 283, 301–305] It is logical to conclude that these favorable responses are a consequence of the tibolone effects on the breast tissue enzymes involved in local estrogen production.

The Livial Intervention following Breast cancer: Efficacy, Recurrence, And Tolerability Endpoints (LIBERATE) trial was a multinational, placebo-controlled, randomized study of women with vasomotor symptoms who had had breast cancer surgically treated within

the previous 5 years.[306] The study was designed to demonstrate that tibolone was superior to placebo, but when the drug monitoring safety board notified the sponsor that there appeared to be an excess of breast cancers in the treated group, the trial was canceled July 31, 2007, 5 months before its scheduled end. The median duration of participation and treatment was about 3 years, with a wide range from a few weeks to almost 5 years. The participants used a variety of adjuvant treatments, mostly tamoxifen, 66.8%; 6.5% used aromatase inhibitors. The dose of tibolone was 2.5 mg daily. Final numbers for analysis were 1,556 women in the treated group and 1,542 in the placebo group. The women ranged in age from under 40 to 79, with a mean age of 52.7 years. 57.8% had positive lymph nodes and 70% had a tumor stage of IIA or higher. Estrogen receptor status was known in 2,808 women in whom the tumors were estrogen receptor positive in 77.8%. In the intent-to-treat analysis, the hazard ratio for recurrent breast cancer in the tibolone-treated women was 1.40 (CI=1.14–1.70). The absolute risk for tibolone was 51 cancers per 1,000 women per year, and 36 in the placebo group. The increase occurred only in women with estrogen receptor-positive tumors. There was no difference in mortality rates between the two groups during the 5-year study period. There were no differences in cardiovascular events or gynecologic cancers, and not surprisingly, vasomotor symptoms, quality of life measures, and bone density improved with tibolone treatment.

How do the LIBERATE results that indicate an estrogenic action of tibolone in breast cancer survivors jibe with the literature indicating that tibolone exerts a non-estrogenic effect on breast tissue? Indeed, it was realistic to expect tibolone to have a salutary effect on the breast. It is well documented that the breast responds to tibolone with less stimulation compared with estrogen, judged by changes in mammographic breast density and the characteristics of tissue obtained by fine-needle aspiration. In the LIFT clinical trial (discussed under bone) that had vertebral fractures as the primary endpoint and breast cancer as a secondary endpoint, the risk of breast cancer after 3 years was significantly 68% *reduced* with tibolone treatment, although the dose was lower, 1.25 mg daily.[307]

The previous literature documenting beneficial actions of tibolone on the breast reflected, however, the impact of tibolone on normal breast tissue, and tibolone's activity to lower local bioactive estrogen levels in target tissues might be lost in cancer cells. The contrary results in the LIFT trial could reflect its older population of women at high risk for fractures, a population that also differed by having lower body weights, no history of tamoxifen treatment, and lower risk factors for breast cancer.

Although the LIBERATE trial may apply to all breast cancer survivors, speaking strictly in a scientific sense, the results were derived mainly from tamoxifen users with 10-fold fewer users of aromatase inhibitors. The possibility that estrogen or tibolone would interfere with the beneficial effects of tamoxifen or aromatase inhibitors has always been one of the objections to treating breast cancer survivors with estrogenic hormones. In a subgroup analysis of the LIBERATE trial, the group of women who had used aromatase inhibitors had a greater risk of recurrent breast cancer compared with tamoxifen; however, the confidence interval was wide because of relatively small numbers. Possibly the estrogenic effect of tibolone would be more pronounced on an occult breast cancer in estrogen-depleted tissue compared with tissue where tamoxifen was bound to the estrogen receptor and prevented estrogenic stimulation. We don't know if the LIBERATE data are meaningful for future treatment regimens. Nevertheless, ***until there are new data, the use of tibolone in women with a history of breast cancer remains relatively contraindicated.***

The Effect of Tibolone on Bone

Tibolone prevents bone loss in postmenopausal women as effectively as estrogen or estrogen-progestin therapy.[235, 253, 308–312] The beneficial impact on bone can be attributed to the

estrogenic metabolites acting through the estrogen receptor because it is blocked by an antiestrogen but not by an antiandrogen or an antiprogestin.[313] In a large, U.S. dose-response study with doses ranging from 0.3 to 2.5 mg daily, only the 1.25 and 2.5-mg doses produced progressive bone density increases in the femoral neck.[253] Indeed, the impact on bone was essentially the same for the 2 highest doses, 1.25 and 2.5 mg. Although the 1.25-mg dose is acceptable for the prevention of bone loss, the 2.5-mg dose is more effective for the alleviation of hot flushes.[228] Tibolone prevents the bone loss associated with GnRH agonist treatment (and the side effect of hot flushing).[290, 314]

The LIFT study (Long-term Intervention on Fractures with Tibolone) was a randomized, placebo-controlled multicenter trial in 22 countries of tibolone, 1.25 mg, given daily over 3 years.[307] The 4,538 women who participated in the trial were age 60 to 85, all at high risk of fractures because of osteoporosis, and all treated with calcium and vitamin D supplementation. The study was stopped in February 2006 after a mean treatment of 34 months because of an increased risk of stroke. The risks of all events were assessed after 5 years of follow-up. Tibolone treatment reduced the number of vertebral fractures by about 45%, and nonvertebral fractures by about 25%. The reduction of fractures was about 4 times as great in women who already had a vertebral fracture upon entry to the study compared with women who had not had a fracture at baseline. It is noteworthy that the number of falls in the treated group was 25% less.

Based on previous bone density studies, the results of the LIFT trial on fracture reduction were not unexpected. The magnitude of the effect is roughly comparable to those with estrogen, bisphosphonates, and raloxifene (with the important exception being a lack of effect of raloxifene on hip fractures). The reduction of breast cancer was comparable to that reported with tamoxifen and raloxifene, but this was not a primary endpoint of the study. Although the difference was not statistically significant, there were four cases of endometrial cancer in the tibolone group and none in the placebo group.

The reported 2-fold increase in stroke in the LIFT trial was greater in the oldest women (over age 70), similar to that observed with estrogen. *It is best to avoid the use of tibolone in elderly women and in women who are at risk for stroke (specifically those with hypertension, smoking, diabetes, or atrial fibrillation).*

SUMMARY—Tibolone—Conclusion

Tibolone is an appropriate choice for hormonal therapy, suitable for many postmenopausal women. The standard dose of tibolone for many years was 2.5 mg daily, but the clinical trials support the use of the lower dose, 1.25 mg daily, with no major loss of efficacy. Because of its unique and varied metabolism, tibolone has different actions in different tissues, which provide an overall favorable risk-benefit profile. Tibolone treats menopausal symptoms, including hot flushes and vaginal dryness, as effectively as estrogen therapy and, most importantly, improves sexual response. Endometrial safety and prevention of bone loss are comparable to that achieved with continuous, combined estrogen-progestin regimens and with a lower rate of breakthrough bleeding.

The sum of the various biologic effects of tibolone and its metabolites on the cardiovascular system should neither increase nor decrease the risk of coronary artery disease. Thus far, there has been no indication for an increased risk of venous thromboembolism, but this is a potential side effect that requires further epidemiologic studies. Tibolone does not stimulate the proliferation of breast cells and affects enzyme activity in the breast to lower breast tissue concentrations of active estrogen. Tibolone does not increase breast density on mammography and does not increase the frequency of mastalgia. The reported increased risks of breast cancer and endometrial cancer in observational studies very likely represent "preferential prescribing" of tibolone in Europe.[315, 316]

Women prescribed tibolone in Europe more often had chronic breast disease, a personal history of breast cancer, previous dysfunctional uterine bleeding, hypertension, and previous uterine operations. Most importantly, more women prescribed tibolone had a history of previous treatment with unopposed estrogen. Thus, clinicians were more likely to prescribe tibolone to women they believed were at higher risks for these two cancers, and this would yield higher rates in treated groups compared with control groups.

Treatment Options for Hot Flushes

The treatment of choice for vasomotor symptoms is hormone therapy. However there exists a substantial number of women who either cannot or will not accept hormone therapy. The choices other than hormone therapy that have been available in the past offered only a modest benefit. Transdermal clonidine, applied with the 100-µg dose once weekly was a common choice, but the reduction in hot flushing was only slightly better than that obtained with placebo treatment.[317, 318] Clonidine, bromocriptine, and naloxone given orally are only partially effective for the relief of hot flushes and require high doses with a high rate of side effects such as drowsiness and dry mouth. Bellergal treatment (a combination of belladonna alkaloids, ergotamine tartrate, and phenobarbital) is slightly better than a placebo and a potent sedative in the short-term; however, one study documented a similar response with bellergal and placebo after 8 weeks.[319, 320] Veralipride, a dopamine antagonist that is active in the hypothalamus, is relatively effective in inhibiting flushing at a dose of 100 mg daily, but is associated with major side effects, including mastodynia and galactorrhea.[321–323] Medroxyprogesterone acetate (10–20 mg daily) and megestrol acetate (20 mg daily) are also effective (as effective as estrogen), but concerns regarding exogenous steroids, especially in patients who have had breast cancer, would apply to progestins as well.[324–326] Vitamin E, 800 IU daily, is barely more effective than placebo.[327] Dong quai, ginseng, black cohosh, isoflavones (including soy protein), yoga, and acupuncture all have little clinical difference compared with placebo treatment.[328–338]

In the last few years, selective serotonin reuptake inhibitors (SSRIs) have gained a reputation for significant efficacy in treating hot flushing. The drugs that have been studied include citalopram (Celexa), fluoxetine (Prozac), sertraline (Zoloft), paroxetine (Paxil), and serotonin and norepinephrine reuptake inhibitors, venlafaxine (Effexor) and desvenlafaxine succinate (Pristiq). In addition, an antiseizure medication, gabapentin (Neurontin), has been demonstrated to reduce vasomotor symptoms.

Drugs For Hot Flushing—Randomized Clinical Trial Results

Drug	*Dose*	*Reduction in Flushing*
Citalopram (Celexa) placebo	**20 mg/d**	**50%, same as placebo**
Fluoxetine (Prozac)	**20 mg/d**	**50%, same as placebo**
Sertraline (Zoloft)	**50,100 mg/d**	**40%, same as placebo**
Paroxetine (Paxil)	**12.5 mg/d**	**62%**
	25 mg/d	**65%**
Venlafaxine (Effexor)	**37.5 mg/d**	**37%**
	75 mg/d	**61%**
	150 mg/d	**61%**
Desvenlafaxine succinate (Pristiq)	**100 mg/d**	**64%**
Gabapentin (Neurontin)	**900 mg/d**	**50%**
Pregabalin (Lyrica)	**150 mg/d**	**65%**

In the study with paroxetine (the controlled release product), 61% of the treated group (a general population of postmenopausal women with only 12 individuals who were breast cancer survivors) at the end of the study had at least a 50% reduction in frequency and severity of flushing, an effect that was about 2.5 times better than placebo with the higher dose.[339] Venlafaxine was studied in breast cancer survivors; although the optimal dose was 75 mg, an appreciable response with 37.5 mg indicated that it would be worthwhile to begin treatment with the lower dose.[340, 341] The response was very rapid, within days, and therefore the dose can be increased in 1–2 weeks. The main side effects were mouth dryness, anorexia, nausea, and constipation. The efficacy of venlafaxine was demonstrated to be the same in women taking or not taking tamoxifen. Desvenlafaxine succinate is a metabolite of venlafaxine and equally effective as the parent compound.[342, 343] ***The effects of citalopram, fluoxetine and sertraline are no more effective than placebo, about a 50% reduction in short-term trials.***[344–348]

Gabapentin (Neurontin) is a γ-aminobutyric acid analogue that has been used for seizures since 1994. It is also effective for migraine headaches, tremors, and panic disorder. In a gabapentin clinical trial, 67% of the treated women experienced more than a 50% reduction in flushes at week 12, compared with 38% in the placebo group.[349] The most common side effects were somnolence (20%) and dizziness (13%). Peripheral edema occurs occasionally because of an induced decrease in serum protein. The potency of this agent appears to be more modest than the SSRIs in a dose of 900 mg/day.[350, 351] At higher doses, gabapentin was as effective as estrogen (about a 70% reduction in flushing); however, side effects are common at higher doses.[352]

Pregabalin (Lyrica), a more potent form of gabapentin, is an anticonvulsant drug that has been used in doses of 150–300 mg daily to treat anxiety and neuropathic pain, e.g., diabetic neurogenic pain, pain after shingles, and fibromyalgia. Side effects include dizziness, drowsiness, visual disturbances, tremor, weight gain, and a decrease in libido. In a phase III randomized trial, a dose of 75 mg b.i.d. (the recommended dose because of more side effects with higher doses), hot flushing was reduced by 65% after 6 weeks of treatment, an impact that was only 15% greater than placebo.[353] Although the data are limited to this short-term, small clinical trial, the effect of pregabalin appears to be comparable to gabapentin.

The SSRIs are the best choice after hormone therapy, although the reduction in hot flushing is considerably less than what can be achieved with estrogen treatment. It is worth trying to titer the dose down to its lowest effective level because of a bothersome incidence of decreased libido. In addition, clinical experience indicates that it is best to slowly titrate upward to the recommended dose and, likewise, to wean the patient slowly when discontinuing treatment. SSRIs are effective for flushing secondary to both tamoxifen and hypoestrogenemia, and the efficacy is similar in women with and without breast cancer.[354] ***An added advantage of the SSRIs is the fact that the clinical studies have also reported improvements in depression, anxiety, and sleep.***

Tamoxifen is converted to an active metabolite by enzymes that are inhibited by certain SSRIs. Paroxetine coadministration decreases plasma concentrations of the active tamoxifen metabolite.[355, 356] A lesser effect is associated with fluoxetine and sertraline. In a retrospective cohort study, only paroxetine use during tamoxifen therapy was associated with an increased risk of death due to breast cancer.[357] ***Paroxetine, fluoxetine, and sertraline are best avoided in women being treated with tamoxifen.***

Bioidentical Hormones

The adverse publicity following the publications from the Women's Health Initiative was a multibillion dollar bonanza for compounding pharmacies providing postmenopausal

hormones. Bioidentical hormones are now the focus of a political, financial, and legal conflict. Bruce Patsner, research professor in the Health Law & Policy Institute at the University of Houston Law Center, has written what is, in our view, a masterful analysis of the problem, with suggestions for its resolution.[358]

The History of the Conflict

The operations of a pharmacy are regulated in the individual states by state boards of pharmacy, in a system similar to the regulation of medical practice. The first federal law regulating drugs, the Federal Food, Drug and Cosmetic Act, was passed in 1938, at a time when most drugs were compounded according to a doctor's prescription. The American Pharmaceutical Association defines pharmacy compounding as the preparation of a prescription drug that is "individualized" to the needs of the patient. This changed after World War II with the development and growth of the pharmaceutical industry. The Kefauver-Harris Amendment in 1964 extended the role of the Food and Drug Administration (FDA) to include safety and efficacy.

In the 1990s, the FDA began to regard the drugs coming from compounding pharmacies as falling under the "new drug" regulations, and therefore, the FDA had jurisdiction over the marketing and promotion of those drugs. The pharmacy world was immediately challenged; there was no way that an individual pharmacy could carry out the kind of clinical studies required for the approval of new drugs. Thus, the pharmacists immediately realized that all compounded drugs would be illegal. At the same time, the FDA was a bit ambivalent, acknowledging that there were examples where the individual needs of a patient required the compounding of a drug, e.g., the creation of a liquid preparation when none was available. This was before compounding took to the Internet for marketing and promotion.

In 1992, the FDA issued its Compliance Policy Guide on Compounding, reserving a right for "selective enforcement," as a compromise between believing it was correct in assuming that compounded drugs represented new drugs and admitting that some patients required compounding. The pharmacy profession immediately rejected the idea that the FDA had any regulatory jurisdiction over pharmacies. The 1997 Food and Drug Modernization Act attempted to clarify the situation. An amendment was added to the existing laws stating that compounded drugs were not "new drugs," but at the same time the 1997 act prohibited the marketing of compounded drugs.

The pharmacy profession sued the FDA, arguing that a restriction on advertising and promotion of compounded drugs was unconstitutional, a restriction of free speech. The District Court ruled against the FDA in 1999. The FDA appealed and lost again in the 9th Circuit Court of Appeals, which further invalidated the entire 1997 act. In 2002, the U.S. Supreme Court upheld the Circuit Court decision.

The FDA issued a new Compliance Policy Guide in 2002, stating that selective enforcement would hinge on 3 major factors: (1). A potential adverse effect of a drug, (2). Whether drugs were compounded from non-FDA-approved components, and (3). Whether compounded drugs were similar to drugs already removed from the market for safety reasons. At this point, the FDA affirmed that it did not want to infringe on the traditional practice of compounding, the preparation of a drug according to a doctor's prescription to fit an individual patient's requirements.

Wyeth Pharmaceuticals filed a Citizen Petition with the FDA in October 2005 requesting that the FDA take action against several compounding pharmacies that were primarily

Internet-based. The Petition's major allegation was that these pharmacies were essentially manufacturing new drugs and should be subject to new drug regulations. On January 9, 2008, the FDA announced it would take action against seven pharmacies providing prescription bioidentical hormones, and issued warning letters that potentially could be followed by seizures of drugs and injunctions against production.

The FDA would like to regard compounded drugs as "new drugs," but the legal precedent has now been set by the courts: compounded drugs are not "new drugs." This was reaffirmed in a 2006 decision in the U.S. District Court for Texas. The FDA would further like to regard the giant compounding pharmacies, especially those operating over the Internet, as manufacturers, like pharmaceutical companies, but again, the court decisions have prevented the FDA from requiring compounding pharmacies to meet "new drug" standards. The "new drug" argument doesn't work.

"Bioidentical" and "natural" are often used in concert. Strictly defined, the hormones must be precisely the same as normal endogenous estrone, estradiol, and estriol, the three endogenous estrogens; progesterone, the progestational agent synthesized by the ovarian corpus luteum after ovulation, and testosterone and dehydroepiandrosterone, androgens made by the human body. To argue that products are not "artificial," begs the issue because even if the source is a steroid molecule derived from plants, chemical and manufacturing processing is still required. These terms obviously have marketing value, and the terms are used to imply greater safety, even greater efficacy. The situation is further compounded (pun intended) because it is likely most patients assume that the marketed bioidentical and natural hormones have been demonstrated by appropriate studies to be effective and safe. Of course this is not the case, although it seems like an obvious conclusion to view product A and product B to be the same if they are the same molecule. The problem is that compounding pharmacies are not required to compare the formulation with the performance of an approved product, nor is there any way for a patient to be assured that the dosage is correct (that the drug contains what it is supposed to contain).

The large compounding pharmacies meet the "individualization" requirement that usually comes from a clinician-patient interaction, by promoting salivary measurements of hormones, as interpreted by one of their employed clinicians to produce tailored hormone choices and doses. Assessment of this approach by researchers, as well as organizations such as the American College of Obstetricians and Gynecologists and the Endocrine Society, have concluded that the variations in salivary sex steroid levels from individual to individual and from specimen to specimen preclude clinical interpretation.[359–361] For most patients, laboratory testing is not necessary in hormone decision-making.

A Better Approach

The American Pharmacists Association and the National Association of Boards of Pharmacies define compounding as the steps required in order to provide a drug in response to a clinician's prescription according to an individual patient's needs, and the preparation of drugs in anticipation of a demand. Therefore, there are three people involved: the patient, the clinician, and the pharmacist. This is in contrast to the production of large amounts of a drug for a national market of unknown users. The American Pharmacists Association further says that ***if an FDA-approved product is commercially available that meets a patient's needs, it should be the drug provided.***

The key to the position of the pharmacists is the contention that there are circumstances, decided by the patient that makes the use of commercial products not a good choice. This seems reasonable, but it is also reasonable that this decision requires the involvement of the

clinician because ultimately a prescription is still required. The traditional view of compounding, therefore, is one of personal relationships with patient, clinician, and pharmacists. This becomes a totally different story when a large Internet pharmacy responds to thousands of prescriptions with no knowledge of the patients. What happened to the "individualization" aspect of compounding? Is it practicing medicine by the pharmacy to have a registered clinician employed by the pharmacy to interpret hormone levels and adjust doses?

Clinicians are appropriately frustrated by the claims made that bioidentical compounded drugs have greater efficacy and safety. Bruce Patsner argues that it should be accepted that bioidentical drugs from the big compounding pharmacies do not meet the definition of compounding supported by the pharmacist's own organization, the American Pharmacists Association—a personal relationship of patient, clinician, and pharmacist addressing an individual's needs.[358] The FDA can argue that the big operations are not legitimate compounding, but big commercial operations directed to unknown consumers.

Patsner also argues that the most vulnerable point is the false safety and efficacy claims. The contention should not be that the safety and efficacy claims are inaccurate, because the pharmacies can always compose their words to avoid legal assaults. The point of emphasis should go back to the pharmacist's published credo: if a commercial, approved product is available to meet the patient's needs, a compounded product is not indicated. Replacing a commercial product with a "natural," untested, unregulated product is not the same thing as prescribing a compounded product when no commercial product will meet the individual's needs.

Pastner summarizes his argument by saying that the large compounding pharmacies are not true compounders because they advertise and promote their products as replacements for commercially available, approved and tested drugs, and that the attempt at "individualization" uses an unsubstantiated method that marginalizes clinicians.[358]

The bioidentical hormones and the various commercial female hormone products are produced by pharmaceutical companies using similar methods that start from the same raw material, usually soy or yams. Some of the commercial products available consist of estradiol, testosterone, and progesterone, the exact same steroid drugs provided by compounding pharmacies. A major difference between the commercial products and products from compounding pharmacies is the important fact that commercial products are federally regulated and tested for purity and potency; compounding pharmacies have not been regulated in this fashion.

The compounded estrogen formulations that contain combinations of estrone, estriol, and estradiol contain sufficient estradiol to produce the same biologic effects associated with commercial preparations. There are no clinical studies documenting that these combinations confer better results or safety. Whether the presence of estriol reduces the risk of breast cancer has not been tested in appropriate clinical studies. Case-control data indicate that estriol used without a progestational agent increases the risk of endometrial cancer, thus its biologic behavior is similar to that of other estrogens.[362]

Custom-compounded formulations have not been proven to be safer or better, and should be regarded as having similar risks and benefits as commercial products. Given bioequivalent doses of various estrogens and progestational agents, one should expect the same biologic results. Claims for custom-compounded products have not been scientifically tested. For these reasons, unless well-designed studies document differences for a specific product, the same risks and benefits apply to equivalent doses of all products.

Salivary sex steroid levels vary widely between individuals, and from measurement to measurement within the same individual. Most importantly, the appropriate clinical studies to document correlation of salivary hormone levels with clinical state or responses have not been performed. Tailor-making a hormone therapy regimen according to

salivary testing is an appealing idea that has not been addressed scientifically, and given the variability in salivary hormone levels, it is unlikely that clinical studies would yield useful information.

Patients who wish to use products similar to endogenous hormones should be made aware of the content of available commercial formulations. Books and pharmacies promoting their own products should be viewed with caution; don't confuse marketing with science.

"Natural" (Alternative) Therapies

The business of selling alternative therapies is now a worldwide phenomenon. The promotion of many of these treatments relies on a network of alternative providers, authors, and compounding pharmacies. Why are herbs and botanicals not regulated? In the U.S., the Dietary Supplement Health and Education Act of 1994 deregulated the industry by classifying dietary supplements as neither foods nor drugs. Thus, manufacturers of dietary supplements are not required to demonstrate that they are safe or effective. In addition to a lack of regulation, there are many other problems associated with herbs and botanicals. The products vary in the amount and purity of active ingredients; indeed, products on the shelf often are adulterated and contaminated with drugs or metals.[363] And very importantly, there is enormous variation in the plants themselves because of genetic, harvest year, and processing differences and in individual metabolism of the products.

Phytoestrogens

"Phytoestrogens" is a descriptive term applied to nonsteroidal compounds that have estrogenic activity or are metabolized into compounds with estrogen activity. Phytoestrogens are classified into three groups: isoflavones, lignans, and coumestans.[364, 365] They are present in about 300 plants, especially legumes, and bind to the estrogen receptor. Soybeans, a rich source of phytoestrogens, contain isoflavones, the most common form of phytoestrogens, mainly genistein and daidzein, and a little glycitin.

PHYTOESTROGENS

1. **Isoflavones (Genistein, Daidzein, Glycitin)**
   soybeans, lentils, chickpeas (garbanzo beans)

2. **Lignans**
   flaxseed, cereals, vegetables, fruits

3. **Coumestans**
   sunflower seeds, bean sprouts

Isoflavones exist in plants bound as glycoside conjugates attached at the 3 position, called glycones. The carbohydrate component requires gut bacteria to remove the sugar moiety to produce active compounds, the aglycones. Individual variability in gastrointestinal microflora, as well as absorption, influences the bioavailability of isoflavones. Biochanin and formononetin are methylated precursors that are metabolized to genistein and daidzein. Red clover and lentils contain significant amounts of these precursors. The isoflavones are in the active, deconjugated forms in fermented soy foods like miso and tempeh. The concentration of isoflavones in tofu is highly variable.

The phytoestrogens are characterized by mixed estrogenic and antiestrogenic actions, depending on the target tissue and local estrogen concentration. Variations in activity may also be due to the fact that the soy phytoestrogens have a greater affinity for the estrogen receptor-β compared with estrogen receptor-α, although the affinity for the beta receptor is still only 35% that of estradiol.[366] Despite a low affinity for the alpha receptor, circulating levels many times that of steroidal estrogens produce the potential for biologic activity.

You can eat soybeans every day and never see a bean. Soybeans are defatted to produce soy flour. Soy flour is then prepared to remove the carbohydrates. 95% of soy flour is toasted and used as animal feed. Alcohol washing is used to get a taste-free product, but alcohol extraction removes the phytoestrogens.[367] SUPRO, known as "isolated soy protein," from Protein Technologies International, the major supplier for commercial products and research, is extracted by aqueous washing and retains the isoflavones.

The reason why most of the soybean crop is devoted to animal feed is because what is left after removing lipids is totally bland. The solution is to mix soybeans with other foods, e.g., beans and soups. Unfortunately they require standing in water for about 12 h and simmering for 2–3 h to be cooked. The average Japanese intake of isoflavones is about 50 mg/day.[368] The rest of Asia has an average consumption of about 25–45 mg/day, and Western consumption is less than 5 mg/day.[369, 370]

Alternative Therapies for Flushing

A belief that Asian women report fewer menopausal symptoms has been an underlying force in the promotion of isoflavones. However, this apparent difference in the prevalence of symptoms comparing Asia and the West may reflect cultural differences and not actual experience. An Italian study found a 45% reduction in flushing with 60 g of isolated soy protein daily (76 mg isoflavones), compared with a 30% reduction in the placebo group.[371] Two other studies, both with 50 mg/day of isoflavones, found a similar 15% reduction in the number of flushes compared with placebo.[372, 373] Another placebo-controlled short-term trial found a greater reduction in flushes with 70 mg isoflavones daily.[374] In a randomized, crossover study of a high dose of isoflavones, 150 mg/day, for flushes in breast cancer survivors, the treated group and the placebo group demonstrated equal effects.[331] The dose of 150 mg isoflavones per day was similar to three glasses of soy milk daily. Two Italian randomized trials found the same response to placebo and 72 or 80 mg/day isoflavones.[375, 376] An Australian study randomized women to 118 mg/day isoflavones or placebo and could detect no difference after 3 months in hot flushing, libido, vaginal dryness, or any of a long list of symptoms, and a Finnish randomized trial using 114 mg isoflavones found no effects on the vagina or on menopausal symptoms.[377, 378] In a randomized study in Iowa, no differences were found in hot flush frequency comparing isoflavone-rich soy protein to a whey protein control.[332] And finally, another randomized trial of breast cancer survivors found no difference comparing placebo with 90 mg isoflavones daily.[333]

Red Clover

Promensil is an extract of red clover *(Trifolium pratense)* containing formononetin, biochanin, daidzein, and genistein. Formononetin and biochanin are metabolized to daidzein and genistein, respectively. Red clover is a legume used to enrich nitrogen levels in soils. Promensil is produced by Novogen in Australia and marketed by Solvay in the U.S. A 500 mg tablet contains 200–230 mg of dried extract, which contains 40 mg of isoflavones. Two randomized, placebo-controlled studies of the effect of Promensil on hot flushes were

reported in 1999.[379, 380] Neither demonstrated a significant difference compared with the placebo group. In 1 of the reports, 4 times the recommended dose (4 tablets daily) also had no effect.[380] On the other hand, an appropriately designed Dutch study, 2 tablets daily, detected a significant reduction of flushing in a 12-week period of time.[381] A large placebo-controlled trial randomized 252 women with severe hot flushing to either Promensil (2 tablets daily) or another red clover extract Rimostil (2 tablets daily for an intake of 57 mg isoflavones).[382] The quantitative reduction in flushing (about 41% in 12 weeks) was identical in the Promensil, Rimostil, and placebo groups, although Promensil had a very slightly faster response. In another randomized clinical trial, the effect of red clover (a daily intake of 128 mg isoflavones) on vasomotor symptoms was no better than placebo treatment.[383] The best evidence indicates that the impact of red clover on vasomotor symptoms is the same as placebo treatment.

Why do these randomized, blinded, and placebo-controlled trials lack agreement? One reasonable explanation is that isoflavones have a mild impact on hot flushing, detectable only in women with frequent and severe flushing. A major clinical response should not be expected. Another possibility is the role of equol (see later discussion).

Other Alternative Treatments

One randomized, placebo-controlled trial examined the effect of dong quai on hot flushing.[328] No estrogenic effects could be detected on flushing, endometrium, or vagina. Ginseng has the same impact on menopausal symptoms as placebo treatment.[329] Similarly, vitamin E supplementation is ineffective for hot flushing.[327]

Evening Primrose

Evening primrose is often recommended for mastalgia, premenstrual syndrome, and menopausal symptoms. Evening primrose oil is extracted from the seed of the evening primrose; it provides linoleic and gamma-linoleic acids (precursors of prostaglandin E). Appropriately blinded and controlled studies have failed to find any differences comparing primrose oil with placebo.[384–386]

Black Cohosh

Black cohosh *(Cimicifuga racemosa)* is also called black snakeroot and bugbane. "Remifemin" is commercially available as an alcoholic extract of the root. A tablet contains 2 mg; the dose is 2 tablets b.i.d. or 40 drops of liquid extract b.i.d. Black cohosh has been heavily promoted, especially by German clinicians, for the treatment of menopausal hot flushes. Keep in mind that the study of hot flushing requires randomization to placebo treatment because placebo treatment is associated with an average 51% reduction in hot flush frequency.[387] Unfortunately, most of the early reports supporting the efficacy of black cohosh were case series or studies without placebo control groups or the studies did not directly and quantitatively measure hot flushing.

Black cohosh has been reported to contain formononetin, a methylated precursor that is metabolized to the two primary phytoestrogens, genistein and daidzein. More sophisticated analysis, however, using liquid chromatography methods, has failed to detect the presence of formononetin in various black cohosh preparations, nor in black cohosh roots and rhizomes.[388]

An older clinical trial was noteworthy and alone in finding a similar impact on hot flushing with black cohosh and placebo treatment.[330] Well-designed trials are confirming that early study and providing us with a uniform story. The Herbal Alternatives for Menopause (HALT) Study is centered in Seattle, Washington. This double-blind trial randomized 351 women to placebo or one of four treatment groups: (1) black cohosh 160 mg daily (note the higher dose); (2) a multibotanical treatment containing 50 mg black cohosh, alfalfa, chaste tree, dong quae, false unicorn, licorice, oats, pomegranate, and Siberian ginseng, 4 capsules daily; (3) the multibotanical plus counseling to increase dietary soy intake; (4) conjugated estrogens 0.625 mg with or without 2.5 mg medroxyprogesterone acetate.[389] After 1 year, no differences were observed in hot flushing comparing any of the three herbal treatment groups to placebo.[389] The herbal remedies also had no effect on sleep quality as reported after 3 months.[390]

A randomized trial in Chicago compared black cohosh, 128 mg, and red clover, 120 mg, to standard hormone therapy and placebo treatment.[383] Over a period of 1 year, only hormone therapy reduced vasomotor symptoms greater than placebo. In this same clinical trial, neither black cohosh nor red clover had an impact on measures of cognition.[391] A Mayo Clinic study reported the results of a double-blind, randomized, cross-over clinical trial to assess the efficacy of black cohosh for the treatment of menopausal hot flushes.[334] The dose was 20 mg b.i.d., the dose of the most commonly marketed black cohosh product in the U.S. The similarity of the studied product with Remifemin was confirmed by high performance liquid chromatography and proton nuclear magnetic resonance analysis. 132 patients were treated for two 4-week crossover periods. Black cohosh reduced hot flushing scores by 20% in the fourth treatment week compared with 27% in the placebo group; and frequency was reduced 17% on black cohosh and 26% on placebo. A randomized trial in Australia found no difference between placebo and a combination of black cohosh with Chinese herbs.[392]

Black cohosh is not estrogenic, and black cohosh has no effect on menopausal symptoms.

An expert committee of the U.S. Pharmacopoeia concluded that black cohosh may be associated with hepatotoxicity; however, a European review of cases with hepatoxicity emphasized the difficulty in establishing a cause-effect relationship.[393, 394] Hepatoxicity remains a concern, awaiting the accumulation of definitive data.

SUMMARY—Alternative Therapies for Hot Flushing—Conclusion

Thus far, all phytoestrogen products (this includes soy and red clover extracts) are proving to be no different than placebo for treating hot flushes. Estrogen products continue to be the most efficacious for this purpose. The serotonin uptake inhibitor class of antidepressants is next most effective. There has not been a head-to-head comparison study of estrogen and SSRIs, but it is reasonable to estimate that an appropriately chosen SSRI will reduce hot flushing by about 60% compared to 90% suppression with estrogen.

Ginkgo Biloba

Ginkgo biloba is an extract prepared from the leaves of the G. biloba tree. It contains flavonoids and unique terpene lactones. Ginkgo biloba is a multimillion dollar herb sold in the U.S. for the preservation of memory. In vitro studies suggested that ginkgo had antioxidant (from the flavonoids) and anti-amyloid (from the lactones) effects. Indeed, the biologic studies provided a rationale for the use of ginkgo to prevent dementia.

A randomized, double-blind, placebo-controlled trial comparing Ginkgo biloba with placebo for the prevention of dementia enrolled 3,069 elderly individuals (over age 75) in five academic centers in the U.S.[395, 396] The participants were randomized to b.i.d. doses of 120 mg ginkgo or placebo (45% female in the ginkgo group and 47% female in the placebo group). The ginkgo formulation and dosage were that used in many of the brands sold in the U.S. The dementia rate steadily increased in both groups over a 7-year period of follow-up, accumulating 277 cases (17.9%) in the treatment group and 246 cases (16.1%) in the placebo group. The rate of dementia did not differ between the two groups, nor did the rate of Alzheimer's disease. In addition, treatment with Ginkgo biloba did not produce less cognitive decline in either adults with normal cognition or with mild cognitive impairment. Other randomized trials have failed to demonstrate any beneficial effects on Alzheimer's, learning, memory, attention, verbal fluency, or concentration.[397, 398]

The American trial robustly demonstrated that Ginkgo biloba in the tested and commonly used dose did not delay the onset of dementia or cognitive decline.[395, 396] The concept of "delay" is important. A treatment that could delay the onset of dementia by 5 years would reduce the number of dementia cases by 50%. In fact, this clinical trial found a statistically significant increase in the risk for developing dementia with ginkgo treatment in the 25% of participants who had cardiovascular disease prior to enrollment. However, the authors appropriately urged caution in interpreting this subgroup analysis. ***A Cochrane review in 2007 of 35 clinical trials with 4,247 participants concluded that there was no convincing evidence that ginkgo treatment benefited individuals who already had dementia or cognitive impairment.***[399]

St. John's Wort

St. John's wort has been reported to be comparable to tricyclic antidepressants in treating mild to moderate depression, based on eight appropriate trials.[400] This is the conclusion of two meta-analyses.[401, 402] All studies were short-term, about 4–6 weeks in duration, and with small numbers. The treatment consisted of a 300-mg plant extract in tablet form, administered t.i.d. However, 2 large, American 8-week trials found no difference between treatment and placebo.[403, 404]

The U.S. Food and Drug Administration (FDA) issued an alert in February 2000 that St. John's wort may interact with drugs known to be metabolized by the cytochrome P450 pathway: theophylline, digoxin, immune suppressants, and oral contraceptives.[405] St. John's wort activates an orphan receptor that induces the expression of metabolic enzymes.[406] In clinically depressed individuals being treated with prescription antidepressants, manic reactions can result (the central serotonergic syndrome).

Phytoestrogens to Prevent Cardiovascular Disease

The cardiovascular story with phytoestrogens received a large boost in 1995, when a meta-analysis concluded that an intake of an average of 47 g soy protein/day lowered total cholesterol and LDL-cholesterol.[407] This was supported by studies in the monkey indicating that isoflavone increased HDL-cholesterol, enhanced vasodilation, and decreased atherosclerosis.[408]

Only intact soy protein has a beneficial effect on lipids. Separation of the protein component from dietary soy protein loses the effect. This effect depends on the inhibition of cholesterol absorption by the non-isoflavone protein.[409, 410] The mechanism involves up-regulation of the LDL-cholesterol receptor and catabolism of LDL-cholesterol, leading

to an increase in bile excretion. The soy peptide binds bile acids and prevents resorption. Alcohol extraction removes the isoflavones from soy protein and causes a loss of the beneficial effect on atherosclerosis in monkeys.[411] Thus, both the isoflavone portion and the protein component are required for a full cardiovascular effect. Non-alcohol-washed soy protein extract has been extensively studied in monkeys. This preparation lowers total cholesterol and LDL-cholesterol, and raises HDL-cholesterol,[412, 413] produces coronary artery vasodilation,[414] inhibits reduction in coronary flow after collagen induced platelet aggregation and serotonin release,[415] and inhibits atherosclerosis but not as robustly as estrogen.[408, 413]

In women, soy protein reduces total and LDL-cholesterol and does not affect triglycerides or HDL-cholesterol; ethanol-extracted soy protein has no effect.[416–419] The minimal dose is about 60 mg isoflavones daily, which is present in 25 g soy protein/day.[420] LDL-cholesterol must be above 130 mg/dL in order to have an effect. Studies of healthy men and women could detect no effect of phytoestrogens (25 to 80 mg isoflavones per day) on lipids or brachial vasodilation.[421–423] In a 12-week study of women with type 2 diabetes mellitus, dietary supplementation of 30 g soy protein (132 mg isoflavones) daily improved insulin resistance and glucose control in addition to lowering total cholesterol and LDL-cholesterol levels.[424] In addition, soy intake prevents LDL-cholesterol oxidation in hyperlipidemic men and women even when circulating LDL-cholesterol levels are unaffected.[425] Promensil, in a 10-week study, had no effect on lipids (it only contains isoflavones, no protein) but did improve arterial compliance.[426]

The U.S. FDA, in October 1999, authorized the use in food labeling of health claims related to the association between soy protein and reduced risk of coronary heart disease, "based on the totality of publicly available scientific evidence, soy protein included in a diet low in saturated fat and cholesterol may reduce the risk of CHD by lowering blood cholesterol levels."[427]

Remember that both protein and isoflavones are needed for a cardiovascular effect. Isoflavones by themselves have no effect on lipids.[373, 426, 428] Protein without isoflavones has no effect on vasodilation and atherosclerosis.[411] The FDA has stated that there is insufficient evidence to allow them to exclude alcohol-washed products from the health claim, but it makes sense that a combined protein-isoflavone product is best. Even in older women with moderate hypercholesterolemia, a high intake of soy phytoestrogens (purified isoflavones without protein) had no effect on the lipid profile.[429] And also remember, that there is no effect on the lipids in individuals who already have a normal profile. Even in individuals with high cholesterol levels, the beneficial impact of soy protein intake is modest and likely to have little clinical effect.

It will require appropriate clinical trials to determine how phytoestrogens compare in the cardiovascular system with estrogens and to determine the efficacy, safety, and correct dosage (studies thus far recommend a daily intake of 60 g soy protein). In addition, the intake of sufficient soy to produce a clinical response is not easy; intake is handicapped by gastrointestinal symptoms, a major alteration in diet or the use of an unpalatable supplement, and great variability in plant contents and products (due to processing). A dietary intake to match the isoflavone dose used in the studies on the lipid profile, for example, would require about 1 lb daily of tofu! In addition, individuals demonstrate great variability in absorption and metabolism. A user-friendly preparation must be developed that minimizes individual variability in response.

Phytoestrogens to Prevent Bone Loss

Phytoestrogens are effective in preventing bone loss in rats but not in monkeys.[430–432] In women, most studies have demonstrated at best a slight effect on spinal bone but no effect on hip bone.[30, 417, 433, 434] One 3-year, randomized trial demonstrated no effect on bone loss

in the spine and femur, with perhaps a modest bone-sparing at the femoral neck with 120 mg/day of isoflavones after adjustment for age and body fat.[435] A 1-year clinical trial could detect no impact of soy intake on bone mineral density in either equol producers or nonproducers.[436] Flaxseed supplementation had no effect on biomarkers of bone metabolism.[437] The difference between hip fracture incidence in Japanese and American women may be due to structural and/or genetic differences not dietary intake.[438]

Ipriflavone is a synthetic isoflavone; it is methylated dehyroxydaidzein, which is metabolized to daidzein. It was developed by Chiesi Pharmaceuticals in Italy. It is marketed in the U.S., and each tablet contains 150-mg ipriflavone combined with calcium (375 mg), vitamin D (187 IU), soy isoflavones (40 mg), and 3 mg boron. The Italian product is pure ipriflavone. The recommended dose is 600 mg/day, 2 tablets b.i.d. taken with meals. Studies with ipriflavone have demonstrated prevention of bone loss over a year.[439–442] Overall the effect on bone is not as great as that observed with standard doses of estrogen or bisphosphonates, perhaps not great enough to yield a benefit. A 4-year randomized trial in Europe assessed the effect of ipriflavone on bone density, urinary markers, and vertebral fractures in 474 women and could find no difference in the treated group compared with the placebo group.[443]

Phytoestrogens and Cognition

Phytoestrogens up-regulate cognition markers and improve memory in rats equally when compared with estrogen[444, 445] There is one human study that is disturbing. Men, in a National Institutes of Health study that began in 1965, reported their tofu consumption.[446] Cognition was tested in 1991–1993 when the men were age 71–93. Higher midlife tofu consumption (two or more servings per week) was associated with poor cognitive test performance, enlargement of ventricles, and low brain weight. Soy supplementation has been reported to improve measurements of memory and attention in postmenopausal women.[447, 448] On the other hand, randomized trials detected no effects of soy protein, red clover, or black cohosh on tests of memory, executive function, language, visual perception, cognition, or measures of quality of life.[391, 423, 449, 450]

Phytoestrogens and the Breast

In the parts of the world where soy intake is high, there is a lower incidence of breast, endometrial, and prostate cancers.[451] For example, a case-control study concluded that there was a 54% reduced risk of endometrial cancer, and another case-control study indicated a reduction in the risk of breast cancer in women with a high consumption of soy and other legumes.[452, 453] Daidzein and genistein urinary excretion are lower in Australian women who develop breast cancer.[454] High soy and tofu consumption and high urinary excretion of isoflavones have been reported to be associated with a lower risk of breast cancer in Singapore, China, Australia, and even in American women consuming a diet rich in isoflavones.[453, 455–459] These studies have supported the belief that high phytoestrogen intake protects against breast cancer. It is by no means certain, however, that there is a direct effect of soy intake.[460] Indeed, a 6-month study of the impact of administered soy protein on breast secretions in premenopausal and postmenopausal women revealed increased breast secretions with the appearance of hyperplastic epithelial cells.[461] Epithelial hyperplasia based on cytology in breast secretions was demonstrated in 7 of 24 (29.2%) of the subjects. Swedish and English cohort studies could not detect a relationship between dietary phytoestrogens and the risk of breast cancer.[462, 463]

Genistein increases epidermal growth factor in immature rat mammary tissue, and it has been hypothesized that earlier exposure to genistein promotes early cell differentiation leading to breast glands that are more resistant to the development of cancer.[464] On the other hand, using the chemically induced rat breast cancer model, no evidence of isoflavone inhibition on tumor development has been detected.[465] In the monkey, treated for 6 months, no proliferation was reported in either endometrium or mammary tissue.[466, 467]

One hypothesis speculates that phytoestrogens protect the breast by decreasing exposure to the more potent endogenous estrogens. The evidence does not support this idea. High-dose treatment (100 mg of daidzein plus 100 mg genistein) does lower estradiol and dehydroepiandrosterone sulfate levels in premenopausal women and increases cycle length.[468] However, these are extremely high doses. One study reported that treatment with Asian soy foods (approximately 32 mg isoflavones per day) was associated with a 9.3% significant decrease in luteal serum estradiol levels, but there were no other changes, including follicular-phase estradiol, progesterone levels, and sex hormone-binding globulin levels, or cycle length.[469] Interestingly, the reduction in luteal estradiol was observed only in Asian participants in whom urinary excretion of isoflavones was higher than nonAsians.[469] These same investigators reported that a high intake of the soy protein alone (with the isoflavones removed) reduced estradiol and progesterone levels throughout the cyle.[470] Other studies have found no effects on estradiol, FSH, LH, or sex hormone-binding globulin in premenopausal women,[471] and, most importantly, no effects on circulating hormones in postmenopausal women.[472, 473] The lack of an effect on gonadotropin and steroid levels is important, depriving the clinician of a method to assess dosage.

Catecholestrogens (2-hydroxy and 4-hydroxy estrogens) have long been proposed as a metabolite pathway that could be protective, or at least antiestrogenic. Hydroxylation in the 2 or 4 position produces inactive metabolites. In 1 study, 8 premenopausal women treated with a soy milk supplement increased their urinary excretion of 2-hydroxyestrone by an average of 47%.[474] Another study could detect no change in 2-hydroxyestrogens.[471] A study limited to Asian-American women also was unable to identify an impact of soy intake on overall estrogen metabolite excretion; however, an increase in catecholestrogens was observed with greater soy intake, balanced by a decrease in 16-hydroxylation.[475]

In response to soy, no significant increase in nipple aspirate levels of genistein and daidzein could be detected.[476] However an indication of estrogenic stimulation occurred, as measured by pS2 (a protein up-regulated by estrogen) levels, but there was no evidence of an effect on epithelial cell proliferation, estrogen and progesterone receptors, apoptosis, or mitosis. Thus, no antiestrogenic effect could be detected, and at best there was a very weak estrogen effect. In another study, 48 women with normal breasts received a 60 g soy supplement for 14 days, and in these women lobular epithelial proliferation and progesterone receptor expression increased, an indication of estrogen stimulation.[477] Some argue that the key to a beneficial impact on breast may be early exposure, and a sudden increase late in life of dietary phytoestrogens may be harmful. On the other hand, a Chinese cohort study of 5,042 breast cancer survivors documented a reduced risk of recurrence and death associated with increasing levels of soy intake, evident among women with either estrogen receptor-positive or receptor-negative disease and among either tamoxifen users or nonusers.[478] In an American cohort of 1,954 breast cancer survivors, a 60% greater decrease in breast cancer recurrence was observed in postmenopausal women using tamoxifen comparing the highest level of soy intake with the lowest.[479] ***Most of the evidence indicates that a high intake of phytoestrogens is associated with a reduced risk of breast cancer, including recurrence in breast cancer survivors. It is not known whether this effect is a marker for beneficial metabolic responses to phytoestrogens or whether there is a direct impact on breast tissue. The evidence also indicates that phytoestrogen consumption does not adversely interfere with tamoxifen's mechanism of action.***

Summary of Phytoestrogen Clinical Effects

Flushes	**Insignificant effect**
Coronary Heart Disease	**Weak impact**
Bone	**No effect**
Cognition	**Unknown**
Breast	**May be protective**
Endometrium	**No effect**
Vagina	**No effect**

There is agreement that phytoestrogens have no effects on the uterus or vagina.[328, 372, 373, 375, 379, 380, 466, 467, 472, 473, 480] ***A beneficial effect on vaginal dryness and dyspareunia cannot be expected; however, the lack of a proliferative stimulus on the endometrium is a wanted consequence of phytoestrogen supplementation.***

Currently the recommended intake expected to have some effect on coronary heart disease is 50–60 mg isoflavones per day, an amount that is in 25 g of soy protein aqueous extract. A beneficial impact on coronary heart disease in women with abnormal lipid profiles is to be expected, a consequence of a decrease in total and LDL-cholesterol and an increase in vascular reaction, but the actual clinical effect is unknown. Excess intake can cause gastrointestinal upset and flatulence, inhibition of enzymes necessary for the digestion of proteins, possibly obstruction of mineral uptake, and weight gain.

The Role of Equol

Equol is a bacterial metabolite and the only hormonally active metabolite of the soy phytoestrogen, daidzein. It is one of the estrogenic compounds in pregnant mare's urine, hence its name. At least in vitro, equol stimulates gene transcription with both estrogen receptors and with a greater potency than any other isoflavone.[481] Equol formation is totally dependent on intestinal microflora; therefore, strictly speaking it is not a phytoestrogen. To be accurate, equol is a nonsteroidal estrogen, a member of the isoflavone family, and exclusively a metabolic product of intestinal bacteria.

The most important observation regarding equol is that 30–50% of adults do not produce equol, even when challenged with high doses of soy.[482] This is in contrast to nonhuman primates and other animals; all that have been studied produce high levels of equol. Thus, there are two human populations: equol producers and equol nonproducers. The key question is whether equol producers receive greater clinical effects from phytoestrogens than nonequol producers. As noted, thus far the clinical effects of isoflavones on bone have not been impressive. In a 2-year randomized trial of postmenopausal women, isoflavone-rich soy milk increased spinal bone mass in the 45% of the subjects who were equol producers, with essentially no effect in equol nonproducers.[483] More profound beneficial effects on the lipid profile have been reported in equol-producing women.[482] Therefore, the population destined to receive a benefit from soy intake may be limited to equol producers. Studies need to be repeated measuring the responses in individuals who are identified as equol producers or equol nonproducers. If the population destined to receive a benefit from soy intake is limited to equol producers, a convenient, inexpensive method must be developed to identify equol production. It may be possible to convert nonproducers to producers.

An emerging approach is to administer equol itself. Daidzein yields two forms of equol, the R-equol inactive isomer and S-equol, the active isomer that binds to estrogen receptor-β. S-equol has been synthesized and its administration is effective for the treatment of menopausal symptoms.[484] Another alternative is the S-equol supplement made by incubating equol-producing bacteria with soy isoflavones.[485, 486] S-equol also blocks the activity of dihydrotestosterone, and thus, it has potential to treat androgenic effects such as acne, hirsutism, male pattern baldness, and prostate cancer.[487]

Estriol

Interest in estriol can be traced to Lemon's report in 1975 that estriol limited the growth of breast tumors in the chemical-induced rat tumor model.[19] However, it is usually overlooked that estradiol worked equally well in that model. Estriol treatment of postmenopausal women has no overall effect on lipids and no effect in the prevention of myocardial infarction.[488, 489] Estriol, without concomitant progestin treatment, does increase the risk of endometrial cancer with the long-term oral use of 1–2 mg/day.[362] At least two studies have been unable to demonstrate prevention of bone loss with the administration of 2 mg estriol daily.[17, 18] And one case-control study found no reduction in hip fractures with estriol compared with a lower risk with estradiol.[489] *There is no evidence indicating any beneficial effects unique to estriol.*

Transdermal Progesterone

Transdermal (or percutaneous) progesterone cream has been promoted to have multiple benefits. In order to achieve widespread effects, absorption must yield adequate blood levels. Two English randomized, blinded, placebo-controlled studies used 2–4 times the recommended dose and reported blood levels of about 1 ng/mL, supported by very low urinary pregnanediol levels.[490, 491] An American study achieved progesterone blood levels of 2–3 ng/mL with application twice daily.[492] An Italian 1-year study did not measure blood levels but could detect no effects on bone density, lipid profiles, or depression scores.[493]

These studies indicate very little systemic absorption of progesterone from the cream product (the levels do not reach normal luteal phase concentrations), and there is great variability.

An English randomized clinical trial using transdermal doses of 5, 20, 40, and 60 mg progesterone cream could detect no differences in measures of psychological, somatic, and vasomotor symptoms compared with placebo.[494] An Australian study of 16, 32, or 64 mg transdermal progesterone cream administered daily could detect no significant absorption and, most importantly, no endometrial response and no effect on flushes, lipids, bone, moods, or sexuality.[495, 496] Incidentally, this study found salivary progesterone levels to be so variable that they had no meaning. Progesterone cream can produce high salivary levels, without a significant change in serum or urinary levels (the mechanism is unknown).[497, 498] Red cell levels reflect serum levels and do not indicate preferential transport or sequestration.[498] ***Clinicians and patients should be aware that transdermal progesterone cream will not reduce hot flashes more than a placebo response, but most importantly, this treatment will not protect the endometrium against the risk of endometrial cancer associated with estrogen therapy.***

Wild yam creams are marketed as progesterone precursors or "balancing" formulas. Yam contains diosgenin, a plant steroid that can be converted to progesterone in a chemical laboratory but not in the human body. Predictably, a wild yam cream has no effects on a wide range of measurements in postmenopausal women.[499] Some do contain progesterone, added by the manufacturer. Creams with less than 0.016% progesterone can be sold over-the-counter. There is no evidence to indicate that these preparations produce systemic effects.

Better absorption is provided by progesterone gels, an alcoholic solution with hydroxypropyl methylcellulose and water.[500] With a 100 mg dose of a progesterone gel, serum progesterone levels are well into the luteal phase range, but clinical use awaits studies documenting the impact on endometrium.

Dehydroepiandrosterone (DHEA)

Adrenal androgen production decreases dramatically with aging. The mechanism is not known, but it is not due to the loss of estrogen at menopause nor can it be reversed with estrogen treatment.[501] The impressive decline (75–85%) in circulating levels of DHEA that occur with aging (greater in men than in women) has stimulated a search for a beneficial impact of DHEA supplementation.[502]

The only proven function of DHEA and its sulfate, DHEAS, is to provide a pool of prohormone for conversion to androgens and ultimately estrogens. By age 70 or 80, the circulating levels in men and women are about 10% of peak levels that occur between 20–30 years of age. DHEA supplementation does not produce improvements in menopausal symptoms, mood, libido, cognition, or memory, but it does increase testosterone and decrease HDL-cholesterol.[503] The acute administration of DHEA did produce a modest effect on sexual response in postmenopausal women, but the dose was enormous, 300 mg.[504]

Although low levels of DHEA and DHEAS have been reported to be associated with increased risk of cardiovascular disease in men, in women conflicting results are found in cross-sectional data. In a longitudinal study of 236 women, higher levels of DHEA and DHEAS in middle-aged women correlated with an *increased* risk of cardiovascular disease.[505]

DHEA supplementation, 50 mg/day, produced reproductive levels of DHEAS in elderly men, did not change levels of testosterone and dihydrotestosterone, and raised estradiol and estrone levels, although still within normal range.[506] In women, 25 or 50 mg/day increased testosterone levels, decreased sex hormone-binding globulin levels, and produced adverse effects on the lipid profile.[507, 508] Exogenously administered DHEA is converted to potent androgens and estrogens. Potential long-term effects include hirsutism, alopecia, voice changes, prostate and breast effects, and an increased risk of coronary heart disease. Supplementation with DHEA requires titering of dosage using the circulating level of total testosterone and keeping the concentration below 80 ng/dL. This is difficult because the U.S. Food and Drug Administration measured the DHEA content of 45 commercial products, and assayed values varied from 0 to 109.5%![509]

The daily intravaginal use of DHEA in low doses is reported to improve vaginal atrophy and sexuality, with little change in the serum levels of DHEA, estradiol, and testosterone.[510–513] Presumably, the DHEA is converted locally to estrogen and testosterone. Studies documenting the effects on target tissues, such as endometrium, bone, and liver, will be required to assess the long-term safety of this treatment.

There Is Only One Medicine

It is appropriate to inform a patient that when she uses preparations lacking in data regarding safety and efficacy, she is experimenting with her own body. Of course, every patient has the right to do so, but we have the obligation to provide this admonishment. An impressive number of patients will appreciate this advice and conclude that they would rather not be the subject of experimentation.

There is only one medicine. Anything claiming to treat or prevent health problems must withstand the rigor of scientific studies of efficacy and safety. Anything with the potential to affect health must be subject to this requirement. Those treatments that pass this testing will become part of our medical practice; those that fail will fall by the wayside. The simplicity and correctness of this argument are so overwhelming; this will be the future of alternative therapies.

Managing Bleeding During Postmenopausal Hormone Therapy

With sequential therapy, approximately 80–90% of women experience monthly withdrawal bleeding. With continuous, combined estrogen-progestin therapy, one can expect 40–60% of patients to experience breakthrough bleeding during the first 6 months of treatment; however, this percentage decreases to 10–20% after 1 year.[99, 100, 514] Although this percentage of amenorrhea with continuous, combined therapy is a gratifying accomplishment, the number of women who experience breakthrough bleeding is considerable, and it is a difficult management problem. Indeed, the single most aggravating and worrisome problem with daily, continuous therapy is this breakthrough bleeding.

Why call it breakthrough bleeding? The bleeding experienced by women on continuous, combined therapy is similar to that seen with oral contraceptives. It originates from an endometrium dominated by progestational influence; hence the endometrium is usually atrophic and yields little, if anything, to the exploring biopsy instrument. Breakthrough bleeding is due to a progestational effect on vascular strength and integrity, producing a fragility that is prone to breakdown and bleeding. It is helpful to explain to patients that this bleeding represents tissue breakdown as the endometrium adjusts to its new hormonal stimulation. From our experience with oral contraceptives, we have learned to be comfortable with this type of bleeding. We have learned, that for most patients, the incidence of breakthrough bleeding with oral contraceptives is greatest in the first few months of treatment and usually disappears in the majority of women. Indeed, this is the same pattern exhibited by postmenopausal women on continuous, combined therapy, and, therefore, the most effective management strategy is patient education and support.

There is no effective method supported by clinical studies, or a large experience, of drug alteration or substitution to manage this breakthrough bleeding. The breakthrough bleeding rate is only slightly better with a higher dose of progestin (5.0 mg medroxyprogesterone acetate) than with a lower dose (2.5 mg).[99, 514] Therefore, there is not a strong reason to use the higher dose, thus minimizing side effects. The best approach is to gain time, because most patients will cease bleeding. This means good educational preparation of the patient beforehand and frequent telephone or Internet contact to allay anxiety and encourage persistence. Estrogen-progestin combinations that contain a 19-nortestosterone progestin (e.g., norethindrone acetate) demonstrate the same pattern of bleeding, but fewer patients bleed in the first 6 months and the amenorrhea rate by 1 year is higher.[515–517]

OPTIONS FOR PERSISTENT BLEEDING

Sequential therapy
Vaginal hysterectomy
Endometrial ablation
The progestin IUD

There is a hard core of patients (10–20% at the end of 1 year) who continue to bleed. The closer a patient is to having been bleeding (either to her premenopausal state or to having been on a sequential method with withdrawal bleeding), the more likely that patient will experience breakthrough bleeding. Some clinicians, therefore, prefer to start patients near the menopause on the sequential method and convert to the continuous method some years later. We prefer to start with the continuous method because those women who achieve amenorrhea are highly appreciative. For the patients who persist in having breakthrough bleeding, it is better to return to the sequential program in order to have expected and orderly withdrawal bleeding instead of the irregularity of breakthrough bleeding.

Some patients may choose to undergo endometrial ablation to overcome the problem of breakthrough bleeding. But remember that concern still exists regarding the potential for isolated, residual endometrium to progress to carcinoma without recognition. Another option deserving of consideration is the progestin intrauterine device (IUD), discussed in detail in Chapter 25. The local release of progestin is effective in suppressing endometrial response and preventing bleeding, although there is a significant amount of breakthrough bleeding in the first year of use. The levonorgestrel-releasing IUS can be left in place for 10 years, a decided advantage.[136, 518] Finally, for some patients, vaginal hysterectomy will prove to be an acceptable alternative.

INDICATIONS FOR PRETREATMENT BIOPSY

Characteristics associated with a high risk of pathology
Previous unopposed estrogen therapy

INDICATIONS FOR ENDOMETRIAL BIOPSY DURING TREATMENT

Clinician anxiety
Patient anxiety
Treatment with unopposed estrogen
Endometrial thickness greater than 4 mm
Past history of unopposed estrogen therapy

It is not essential to routinely perform endometrial biopsies prior to treatment. Endometrial abnormalities in asymptomatic postmenopausal women are very rare.[514, 519, 520] A reasonable economic moderation would be to limit pretreatment biopsies (using a plastic endometrial suction device in the office) to patients at higher risk for endometrial changes: those women with conditions associated with chronic estrogen exposure (obesity, dysfunctional uterine bleeding, anovulation and infertility, hirsutism, high alcohol intake, hepatic disease, metabolic problems such as diabetes mellitus and hypothyroidism) and those women in whom irregular bleeding occurs while on estrogen-progestin therapy. In the absence of abnormal bleeding, a certain amount of trust in the protective effects of the progestin is justified, and routine, periodic biopsies are not necessary. ***However, women who elect to be treated with unopposed estrogen require endometrial surveillance at least once a year***.

It is appropriate to perform an endometrial aspiration biopsy when the patient's anxiety over the possibility of pathology requires this response. It is also appropriate to perform a biopsy when the clinician is concerned; with increasing experience with this method, it takes more and more to be concerned. If bleeding persists for 6 months, consider an office hysteroscopy; an impressive number of polyps and intrauterine fibroids will be discovered.

Abnormal endometrium is more frequently encountered in patients on combination estrogen-progestin when the patients have previously been treated for a period of time with unopposed estrogen. Breakthrough bleeding or unscheduled bleeding in these patients requires endometrial surveillance because an increased risk for endometrial cancer persists beyond the period of exposure to unopposed estrogen, and it is unknown

how effective the subsequent protective exposure to a progestin will be.[521–523] It is prudent to assess the endometrium in these patients prior to changing from unopposed to combined therapy. Clinicians should maintain a highly anxious state of mind with patients who have been treated previously with unopposed estrogen.

A combined estrogen-progestin program will not totally prevent endometrial cancer.[522] Vigilance on the part of the clinician, however, will detect endometrial cancer at an early stage, a stage that can be treated with excellent results.

It is common for women on a sequential regimen to begin bleeding while in the midst of progestin administration. The timing of withdrawal bleeding in women on a sequential estrogen-progestin program was suggested as a screening method for biopsy decision making. In women taking a variety of progestins for 12 days each month, bleeding on or before day 10 after the addition of the progestin was associated with proliferative endometrium. Bleeding beginning on day 11 or later was associated with secretory endometrium, presumably indicating less need for biopsy.[524] But does this correlate with the risk of hyperplasia and cancer? According to a study of 413 postmenopausal women, the day of bleeding did ***NOT*** predict endometrial safety.[525] Late regular withdrawal bleeding on a sequential program does not give 100% assurance that there is no hyperplasia and perhaps endometrial cancer. This uncertainty with the sequential program is another reason to turn to the daily, combined method where irregular bleeding and sonographic measurement of increased endometrial thickness provide good indications for endometrial biopsy.

If a patient has recurrent bleeding despite repeated medical therapy, submucous myomas or endometrial polyps must be suspected. Thorough curettage can miss such pathology, and further diagnostic study can be helpful. Either hysterosalpingography with slow instillation of dye and careful fluoroscopic examination or ultrasonography with instillation of saline into the uterine cavity or hysteroscopy may reveal a myoma or polyp. Hysteroscopy can also direct a more accurate biopsy of the endometrium.

Measurement of Endometrial Thickness by Transvaginal Ultrasonography

The thickness of the postmenopausal endometrium as measured by transvaginal ultrasonography in postmenopausal women correlates with the presence or absence of pathology. However, the severity of pathologic change does not correlate with the measured thickness.[526] Endometrial thickness (the two layers of the anterior and posterior walls in the longitudinal axis) under 5 mm is reassuring and allows conservative management.[527, 528] Endometrial thickness greater than 4 mm requires biopsy; it is estimated that 50–75% of bleeding patients on hormone therapy and evaluated by ultrasonography will require biopsy.[526, 529] An endometrial thickness less than 5 mm in women receiving hormone therapy, either a sequential regimen or a daily combination of estrogen-progestin, is reassuring.[528, 530, 531] It seems logical that endometrial thickness by ultrasonography in patients on a sequential regimen can be affected by day in the treatment cycle, and for that reason, ultrasonography assessment should be obtained toward the end of the progestin phase or at the beginning of the cycle.[532–534] An Italian study concluded that endometrial thickness measured soon after withdrawal bleeding in women on a sequential regimen was comparable to thickness in women on a continuous, combined program of estrogen-progestin treatment.[535] When a thick endometrium is associated with atrophic endometrium on biopsy, polyps are often present. Greater accuracy can be gained by the instillation of saline into the uterine cavity during ultrasonography.[536] Doppler velocimetry does not improve the accuracy of discriminating between normal and abnormal endometrium.[537] ***A clinician should not be satisfied with "normal" findings on ultrasonography if a patient has persistent bleeding. The pursuit of abnormal bleeding despite "normal"***

findings should reduce missed cases of pathology to nearly zero.[538] In this circumstance, hysteroscopy is recommended.

The Progestin Challenge Test

The administration of a progestational agent (e.g., 10 mg medroxyprogesterone acetate for 2 weeks) was developed by R. Don Gambrell Jr. as a means of detecting the presence of estrogen-dependent endometrium in postmenopausal women.[539] A withdrawal bleed would indicate that an endometrial response has occurred to the progestin, a response that requires previous endometrial stimulation by estrogen and indicates the need for endometrial assessment. In other words, the lack of a withdrawal bleed is reassuring for clinician and patient. Concern with this clinical maneuver has focused on whether there are false-negative and false-positive responses. Several studies are now available regarding the efficacy and validity of this method.[540–542] The published data indicate that most, but perhaps not all, women with endometrial proliferation, hyperplasia, and even cancer will respond with a withdrawal bleed after a progestin challenge, and ultrasonography measurement of endometrial thickness will be greater than 4 mm.[543, 544] In an experiment in monkeys, 1 of 14 animals treated with estrogen did not bleed and 5 of 13 placebo-treated animals did bleed.[545] The problem is that the studies thus far do not consist of very large numbers, and there is a lingering question whether a patient with abnormal endometrium will always bleed in response to progestin treatment and withdrawal.

Risks and Benefits of Estrogen-Progestin Therapy

Cardiovascular Disease—Evidence from Basic Science

A Favorable Impact on Lipids and Lipoproteins

The most important lipid effects of postmenopausal estrogen treatment are the reduction in LDL-cholesterol and the increase in HDL-cholesterol. Estrogen increases triglyceride levels and LDL-cholesterol catabolism as well as lipoprotein receptor numbers and activity, resulting in decreasing LDL-cholesterol levels.[546–548] Estrogen induces a change in LDL-cholesterol toward a smaller more dense particle, but it is in a form with a more rapid turnover in the circulation, allowing less time for oxidation and acquisition of cholesterol.[549, 550] The increase in HDL-cholesterol levels, particularly due to the HDL_2 subfraction, is to an important degree the consequence of the inhibition of hepatic lipase activity, which converts HDL_2 to HDL_3. Postmenopausal estrogen therapy with or without added progestin also produces a reduction in the circulating levels of lipoprotein(a).[551, 552]

The changes in circulating apoprotein levels mirror those of the lipoproteins: apolipoprotein B (the principal surface protein of LDL-cholesterol) levels diminish in response to estrogen, and apolipoprotein A-I (the principal apolipoprotein of HDL-cholesterol) increases. The HDL-cholesterol and triglyceride increases induced by estrogen treatment are attenuated if progestins are added in sufficient doses.[100, 102, 553–558] The concomitant administration of estrogen and an HMG-CoA reductase inhibitor (pravastatin) produced a more favorable change in the lipid profile in hypercholesterolemic women than either treatment alone.[559]

Direct Antiatherosclerotic Effects

Important studies in monkeys support the protective action of estrogen against atherosclerosis, emphasizing mechanisms independent of the cholesterol-lipoprotein profile. Oral administration of a combination of estrogen and a high dose of progestin to monkeys fed a high-cholesterol diet decreased the extent of coronary atherosclerosis despite a reduction in HDL-cholesterol levels.[560–562] In somewhat similar experiments, estrogen treatment markedly prevented arterial lesion development in rabbits, and this effect was not reduced by adding progestin to the treatment regimen.[563–566] These findings of a direct effect against atherosclerosis suggest that women with already favorable cholesterol profiles would benefit through this additional action. And, in considering the impact of progestational agents, lowering of HDL-cholesterol is not necessarily atherogenic if accompanied by an increased estrogen impact.

The monkey studies were extended to a postmenopausal model (ovariectomized monkeys). Compared with no hormone treatment, treatment with either estrogen alone or estrogen with progesterone in a sequential manner significantly reduced atherosclerosis, once again independently of the circulating lipid and lipoprotein profile.[567, 568] A direct inhibition of LDL-cholesterol accumulation and an increase in LDL-cholesterol metabolism in arterial vessels could be demonstrated in these monkeys being fed a highly atherogenic diet.[569] The daily administration of medroxyprogesterone acetate in this monkey model did not prevent the beneficial effect of conjugated estrogen on coronary artery atherosclerosis.[408]

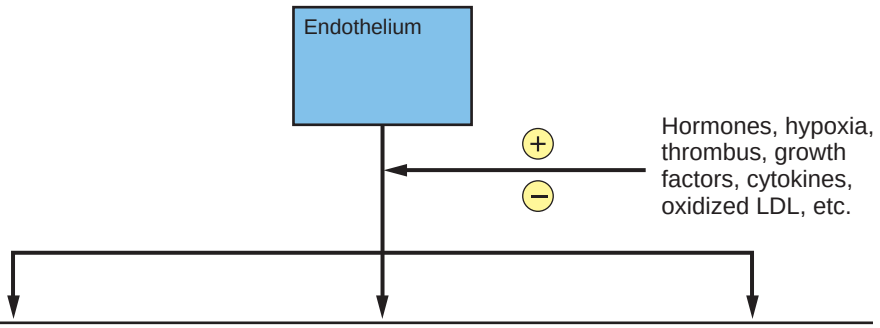
Estradiol fatty acid esters are present in low concentrations in the circulation, transported in lipoproteins. These esters are potent estrogens and protect against the oxidation of LDL-cholesterol; indeed, the antioxidant efficacy of estradiol may require esterification and incorporation into LDL-cholesterol.[570] Estradiol fatty acid ester concentrations are increased by oral estrogen but not by transdermal administration.[571]

Endothelium-Dependent Vasodilation and Antiplatelet Aggregation

Endothelium modulates the degree of contraction and function of the surrounding smooth muscle, primarily by the release of endothelium-derived relaxing and contracting factors. In hypertension and other cardiovascular diseases, the release of relaxing factors (which is probably one factor, nitric oxide) is blunted, and the release of contracting factors (the most important being endothelin-1) is augmented. The endothelins are a family of peptides that act in a paracrine fashion on smooth muscle cells. Endothelin-1 appears to be exclusively synthesized by endothelial cells. Endothelin-induced vasoconstriction is a consequence of a direct action on vascular smooth muscle cells, an action that is reversed by nitric oxide. Impaired release of nitric oxide, therefore, enhances endothelin action. Hypertension and atherosclerosis are believed to be influenced by the balance among these factors. Women have lower circulating levels of endothelin, and the levels are even lower during pregnancy and decrease in response to oral and transdermal estrogen treatment.[572, 573]

Nitric oxide (and estrogen) also inhibits the adhesion and aggregation of platelets in a synergistic manner with prostacyclin (also a potent vasodilator derived from the endothelium).[574, 575] Increased blood flow due to vasodilation and decreased peripheral resistance can be observed to occur rapidly following the administration of estrogen. This response can be produced by both transdermal and oral administration.[576–578] The synthesis and secretion of nitric oxide (the potent endothelial vasodilating product) can be directly stimulated by estrogen in in vitro experimental preparations of coronary arteries.[579] In both normal postmenopausal women and women with hypertension, hypercholesterolemia, diabetes mellitus, or coronary artery disease, the intraarterial infusion of physiologic amounts

of estradiol into the forearm potentiates endothelium-dependent vasodilation, and there is a dose-response effect.[580, 581] Similar brachial artery dilation has been reported with 0.3 and 0.625 mg conjugated estrogens.[582] Comparing brachial artery responses in women who are long-term hormone users (with or without progestin) with nonusers, improved endothelium-dependent vasodilation could be observed with standard doses.[583] In careful, randomized studies, the addition of norethindrone acetate or medroxyprogesterone acetate did not reduce the beneficial effect of estrogen on peripheral artery blood flow.[584, 585] However, not all studies agree; a Danish assessment of brachial artery responses demonstrated no difference between postmenopausal women on long-term combined estrogen-progestin therapy compared with postmenopausal women receiving no treatment.[586]



	Endothelin-1	Nitric oxide	Prostacyclin
Flow	Vasoconstriction	Vasodilation	Vasodilation
Platelet activity	Increases adhesion and aggregation	Inhibits adhesion	Inhibits aggregation
Vessel changes	Stimulates smooth muscle proliferation and migration	Inhibits smooth muscle proliferation and migration	Inhibits smooth muscle proliferation and migration

The synthesis of nitric oxide is involved in the regulation of vascular (and gastrointestinal) tone and in neuronal activity. A family of isozymes (nitric oxide synthases) catalyzes the oxidation of l-arginine to nitric oxide and citrulline. The action of nitric oxide synthase in the endothelium is calcium dependent, and its synthesis is mediated specifically by estrogen.[587] In animal experiments, the endothelial basal release of nitric oxide is greater in females, a gender difference that is mediated by estrogen.[579, 588] In women treated with postmenopausal estrogen and either cyproterone acetate or medroxyprogesterone acetate, circulating nitric oxide (as reflected in nitrite-nitrate levels) is increased, a consequence of estrogen-induced nitric oxide production in the endothelium.[589, 590] In contrast, long-term treatment with estradiol and norethindrone acetate was not associated with changes in nitric oxide, endothelin-1, prostacyclin, or thromboxane A_2.[591]

Acetylcholine induces vasoconstriction in coronary arteries; however, the direct administration of estradiol in physiologic doses into the coronary arteries of postmenopausal women with and without coronary heart disease converts acetylcholine-induced vasoconstriction into vasodilation with increased flow.[592] This favorable vasomotor response to acetylcholine can also be demonstrated in acute experiments with the transdermal administration of estradiol (achieving blood levels of 67–89 pg/mL).[593] This same estrogen-associated response is observed in women with coronary atherosclerosis comparing estrogen users to nonusers.[594] This is an endothelium-dependent response, mediated to a significant degree by an increase in nitric oxide.[595] The administration of standard doses of estrogen (with or without daily progestin) to women with coronary artery disease reduces the degree of ischemia and delays the onset of signs of myocardial ischemia on electrocardiograms and increases exercise tolerance.[596–600] This electrocardiographic response was not observed in women who presented with unstable angina.[601] In normal women, the standard oral 0.625 mg

dose of conjugated estrogens had no effect on hemodynamic responses to treadmill exercise.[602] In the monkey, the vasodilatory response to acetylcholine required a blood level of estradiol higher than 60 pg/mL.[603] Studies with transdermal estrogen treatment that indicate no effects on endothelial function need to be standardized according to blood levels of estradiol.

Endothelium-Independent Vasodilation

Estrogen causes relaxation in coronary arteries that are denuded of epithelium.[604] This response is not prevented by the presence of inhibitors of nitric oxide synthase or prostaglandin synthase. Thus this vasodilation is achieved through a mechanism independent of the vascular endothelium, perhaps acting on calcium-mediated events.[605] The vasodilation produced by sodium nitroprusside is endothelium-independent. In normal postmenopausal women and postmenopausal women with risk factors for atherosclerosis (hypertension, hypercholesterolemia, diabetes mellitus, coronary artery disease), the administration of physiologic levels of estradiol increased forearm vasodilation induced by sodium nitroprusside.[580] However, others have reported no effect of estrogen administration on endothelium-independent vasodilation.[578]

Actions on the Heart and Large Blood Vessels

Estrogen treatment increased left ventricular diastolic filling and stroke volume.[606–608,578] This effect is probably a direct inotropic action of estrogen that delays the age-related change in compliance that impairs cardiac relaxation.[609] In a 3-month study, medroxyprogesterone acetate (5 mg daily for 10 days each month) did not attenuate the increase in left ventricular output (systolic flow velocity) observed with estrogen treatment.[610] On the other hand, others have detected attenuation of estrogen's beneficial effects on compliance (stiffness) associated with combined estrogen-progestin treatment,[609, 611] And others have not been able to demonstrate an effect of short-term oral estrogen or long-term transdermal estrogen treatment on cardiac structure and function.[612, 613] The reasons for these differences are not apparent.

Improvement in Glucose Metabolism

An age-related decline in the basal metabolic rate is accentuated at menopause, associated with an increase in body fat, especially central (android) body fat.[614, 615] Insulin resistance and circulating insulin levels increase in women after menopause, and impaired glucose tolerance predicts an increased risk of coronary heart disease.[616, 617] Estrogen (with or without progestin) prevents the tendency to increase central body fat with aging.[618–621] This would inhibit the interaction among abdominal adiposity, hormones, insulin resistance, hyperinsulinemia, blood pressure, and an atherogenic lipid profile. Indeed, the Women's Health Initiative randomized, clinical trial documented improvements in fasting glucose and insulin levels in the estrogen-progestin treated group.[622] Hyperinsulinemia also has a direct atherogenic effect on blood vessels, perhaps secondary to insulin propeptides. In addition to its vasoconstrictive properties, endothelin-1 exerts a mitogenic effect and, therefore, contributes to the atherosclerotic process. Insulin directly stimulates the secretion of endothelin-1 in endothelial cells, and the circulating levels of endothelin-1 are correlated with insulin levels.[623]

Postmenopausal women being treated with oral estrogen have lower fasting insulin levels and a lesser insulin response to glucose.[555, 622, 624–627] In a 1-year randomized trial comparing

unopposed conjugated estrogens to the usual sequential and continuous regimens of conjugated estrogens and medroxyprogesterone acetate, no differences in the treatment groups were observed in the favorable decreases in fasting insulin levels.[555] Nonoral administration of estrogen has little effect on insulin metabolism, unless a dose is administered that is equivalent to 1.25 mg conjugated estrogens.[625, 628] Because a lower oral dose produces a beneficial impact, this suggests that the hepatic first-pass effect is important in this response, at least in normal women; reports with transdermal hormone therapy have indicated improvements in insulin resistance and hyperinsulinemia, but no effect in women with normal insulin sensitivity.[629, 630] In double-blind, cross-over, placebo-controlled studies of postmenopausal women with type 2, noninsulin-dependent diabetes mellitus, estrogen treatment improved all glucose metabolic parameters (including insulin resistance), the lipoprotein profile, and measurements of androgenicity.[631, 632]

The evidence strongly indicates that postmenopausal estrogen therapy improves glucose metabolism. Epidemiologic studies impressively document that this beneficial metabolic effect associated with estrogen lowers the incidence of adult-onset, type II diabetes mellitus. Three large cohort studies, the Nurses' Health Study, the Finnish Kuopio Osteoporosis Risk Factor and Prevention Study, and the French E3N study reported decreases in new-onset diabetes associated with estrogen therapy, 20% in American nurses who were ever-users, 18% in French ever users, and 69% in Finnish women who were current users.[633–635] In the French cohort, no effect of progestins was observed, and the reduction in the incidence of diabetes (32%) was greater with oral administration of estrogen compared with the transdermal method.[635] Clinical trial results are in agreement. In the HERS trial, the hormone-treated group developed diabetes at a rate that was 35% lower compared with the placebo group.[627] The Women's Health Initiative found a 21% significant reduction in estrogen-progestin users and a 12% reduction in estrogen-only users that did not achieve statistical significance.[57, 636]

Inhibition of Lipoprotein Oxidation

The oxidation of LDL-cholesterol particles is a step (perhaps the initial step) in the formation of atherosclerosis, and smoking is associated with a high level of lipoprotein oxidation. In animal experiments the administration of large amounts of antioxidants inhibits the formation of atherosclerosis and causes the regression of existing lesions. Estrogen is an antioxidant. Estradiol directly inhibits LDL-cholesterol oxidation in response to copper and decreases the overall formation of lipid oxides.[637, 638] Importantly, this antioxidant action of estradiol is associated with physiologic blood levels.[639] In addition, estrogen may regenerate circulating antioxidants (tocopherols and beta-carotene) and preserve these antioxidants within LDL-cholesterol particles. This antioxidant action of estrogen preserves endothelial-dependent vasodilator function by preventing the deleterious effect that oxidized LDL-cholesterol has on endothelial production of vasoactive agents.[640] In an assessment of peroxide formation by platelets, women treated with both estrogen and medroxyprogesterone acetate in a sequential regimen had greater antioxidant activity compared with the days on estrogen alone.[641] In a 1-year study, the presence of levonorgestrel did not attenuate the antioxidant activity of estradiol.[642]

A Favorable Impact on Fibrinolysis

Menopause is followed by increases in factor VII, fibrinogen, and plasminogen activator inhibitor-1 (PAI-1).[643, 644] These changes produce a relatively hypercoaguable state and are associated with an increased risk of cardiovascular events. Postmenopausal women treated with estrogen have lower fibrinogen and plasminogen levels. Reduced levels of

fibrinogen, factor VII, and PAI-1 have been observed in premenopausal women compared with postmenopausal women, and oral estrogen alone or combined with a progestin prevents the usual increase in these clotting factors associated with menopause.[645–649] This would be consistent with increased fibrinolytic activity, a possible cardioprotective mechanism probably mediated, at least partially, by nitric oxide and prostacyclin. Platelet aggregation is also reduced by postmenopausal estrogen treatment, and this response is slightly attenuated by medroxyprogesterone acetate.[574] In a randomized 1-year trial, the addition of medroxyprogesterone acetate, either sequentially or continuously, produced a more favorable change in coagulation factors compared with unopposed estrogen.[650]

The transdermal and oral routes of administration of estrogen (combined with medroxyprogesterone acetate) have puzzling differences in the reported effects on most hemostatic risk factors, such as factor VII, fibrinogen, PAI-1, and antithrombin III. In at least one study, however, antithrombin III levels were reduced by oral estrogen but not transdermal administration; however the values remained within the normal range.[651] In regards to PAI-1, studies with transdermal estrogen have provided conflicting data; for example, favorable changes in PAI-1 levels as well as no effect.[32, 652, 653] However, in a crossover study designed to compare 100 µg transdermal estradiol with 0.625 mg oral conjugated estrogens (both combined with 2.5 mg medroxyprogesterone acetate daily), only the oral estrogen had a favorable reduction in PAI-1 levels.[653] Appropriate doses of hormone therapy have been reported to not have an adverse impact on clotting factors.[646, 654, 655] One study found slightly increased clotting activation with transdermal administration of estradiol, but no change with oral conjugated estrogens.[656] Fibrinopeptide A is an indicator of thrombin generation, and, in 3-month studies, no significant alteration was produced by 0.625 mg conjugated estrogens in 1 and an increase in another.[657, 658] The clotting story is difficult to unravel. Perhaps one contributor to the uncertainty is a possible difference between short-term and long-term effects. [651, 654, 657]

Overall, estrogen treatment is associated with favorable changes consistent with an increase in fibrinolysis.[659, 660] How can there be favorable changes indicating an increase in fibrinolysis and at the same time an increased risk of venous thrombosis, and why in elderly women, especially those with clinically apparent coronary heart disease, does estrogen seem to have a prothrombotic effect? Decreases in antithrombin III and protein S associated with estrogen treatment, a hypercoagulable change, may have a greater impact on the venous system.[658] There also may be subtle variations of inherited susceptibilities that tilt the balance towards thrombosis; for example, concentrations of factors that favor arterial thrombosis have been reported (tissue factor pathway coagulation inhibitor and thrombin activatable fibrinolysis inhibitor) in women treated with estrogen.[661] Another possibility is that the fibrinolysis is a response to coagulation activity, and, therefore, not necessarily a beneficial response.

Estrogen has adverse effects on already established atherosclerosis. Matrix metalloproteinase enzymes are secreted by inflammatory cells and smooth muscle cells. These enzymes digest the proteins in the fibrous cap of an atherosclerotic plaque, making the plaque unstable and predisposed to rupture. Estrogen induces matrix metalloproteinase enzymes and decreases their specific inhibitors (TIMP); this is a mechanism involved in the prothrombotic effects of estrogen in the presence of established atherosclerosis. This effect of estrogen may be dose-related and might be avoided with transdermal administration.[662]

Inhibition of Intimal Thickening

Hypertension and atherosclerosis are associated with increased proliferation of vascular smooth muscle cells. This growth of smooth muscle cells is also characterized by migration into the intima. Arterial intimal thickening is an early indicator of atherosclerosis.

The proliferation and migration of human aortic smooth muscle cells in response to growth factors are inhibited by estradiol, and. importantly, this inhibition is not prevented by the presence of progestins.[663, 664] Nitric oxide, which is regulated by estrogen, also inhibits smooth muscle proliferation and migration.[665] Imaging studies have documented a reduction in intimal thickening in postmenopausal women who are estrogen users compared with nonusers, and this beneficial effect is not compromised by the addition of a progestational agent to the treatment regimen.[611, 666–668] Thus, postmenopausal hormonal therapy can bring about a reduction in atherosclerosis, and this effect is comparable with that produced by a lipid-lowering drug.[666, 669]

Protection of Endothelial Cells

Endothelial cells can respond to injury by initiating the clotting process. Animals studies indicate that estrogen accelerates healing and recovery of the endothelium in response to injury.[670] This is correlated with inhibition of intimal thickening and recovery of important functions such as nitric oxide production. In vitro studies of human endothelial cells demonstrate that estrogen can inhibit cytokine-induced apoptosis.[671] In the rat, medroxyprogesterone acetate blocked the estrogen-induced healing response after carotid artery injuries.[672]

Inhibition of Macrophage Foam Cell Formation

A feature of atherosclerotic plaque formation is monocytic infiltration into the arterial wall and the formation of macrophage foam cells. In a nonantioxidant activity, estrogen inhibits macrophage foam cell accumulation in atherosclerotic lesions.[673]

Reduction of ACE and Renin Levels

Although oral estrogen, but not transdermal estrogen, increases angiotensinogen levels, ACE (angiotensin-converting enzyme) and renin levels are decreased (with or without progestin) by both routes of administration.[674, 675] The angiotensin II receptor (the AT_1 receptor) is involved in vasoconstriction, aldosterone release, sodium and water retention, and growth and proliferation of myocardial and vascular cells. Estrogen induces down-regulation of the AT_1 receptor, and hypercholesterolemia is associated with AT_1 up-regulation and function.[676, 677]

Reduction of Adhesion Molecules

Adhesion molecules recruit leukocytes to the endothelium and play a role in attaching platelets to endothelium. Studies with multiple markers report that oral estrogen therapy increases only C-reactive protein (CRP), the only marker synthesized in the liver. In fact, oral hormone therapy although it increases CRP, reduces the circulating levels of other markers (E-selectin, P-selectin, intercellular adhesion molecule-1, vascular cell adhesion molecule-1, monocyte chemoattractant protein-1, and tumor necrosis factor-α) with inconsistent effects on interleukin-6.[46, 47, 678, 679] An increase in CRP levels may be due to estrogen's well-known effect to stimulate the hepatic synthesis of proteins, especially because of the first-pass phenomenon with oral administration. For this reason, transdermal estrogen treatment reduces adhesion markers but does not change CRP levels.[680–682]

Reduction of Homocysteine

Increased circulating levels of homocysteine are correlated with increased risks of atherosclerosis and thrombosis. Homocysteine levels increase after menopause and are associated with hypertension and degree of atherosclerosis.[683] Homocysteine levels are significantly lowered by estrogen or estrogen-progestin treatment, administered either orally or transdermally.[684, 685]

Cardiovascular Disease—Evidence from Observational Studies

A review of case-control studies in the literature finds overwhelming support for about a 50% reduced risk of coronary heart disease in estrogen users.[686–703] In three studies of women undergoing angiography, a comparison of coronary artery occlusion in users and nonusers of estrogen indicated a significant protective effect of postmenopausal estrogen.[695–697] Women using hormone therapy at the time of a myocardial infarction or with congestive heart failure have been reported to have an improved rate of survival.[704, 705] Little attention has been given to peripheral artery disease, but 1 case-control study did report a decrease in risk in users of hormone therapy.[706]

In a large number of cohort studies, most uniformly reported a reduction in coronary heart disease in estrogen users; only three produced conflicting data.[707–722] In the Nurses' Health Study with 20 years of follow-up, the age-adjusted relative risk of coronary disease in current users of hormone therapy was 39% reduced (RR=0.61;CI=0.52–0.71).[723] The benefit was observed with both the 0.625-mg and the 0.3-mg doses of conjugated estrogens. The beneficial impact was observed to diminish beginning 3 years after discontinuation. It was suggested that higher doses might be harmful because there was an apparent increase in the risk of coronary disease among women taking more than 0.625 mg conjugated estrogens per day. Current postmenopausal hormone users in the Nurses' Health Study have had a 37% reduced risk of mortality due largely to protection against coronary heart disease, an effect that was still present after adjusting for dietary factors, alcohol intake, vitamin or aspirin use, and exercise.[718]

Electron beam tomography (also called ultrafast computed tomography) can assess the presence of coronary artery disease by quantifying the amount of calcium in the coronary arteries, a measure that is known to correlate with the degree of disease and the risk of coronary events. Studies using this technique have demonstrated a lower prevalence of coronary artery calcium in women younger than age 60, a prevalence comparable to men (of any age) in women older than 60, and less calcium (and, therefore, less coronary artery disease) in women using postmenopausal hormone therapy compared with nonusers.[724, 725] In women with an average age of 59 who had used hormone therapy for an average of 9 years, coronary artery calcification was significantly reduced, with a greater effect observed with increasing duration of use.[726] This salutary effect of estrogen was confirmed in a substudy of the Women's Health Initiative estrogen-only arm.[727]

These observational studies have been criticized by arguing that estrogen treatment is a marker for variables (e.g., better diet and better health care) that place postmenopausal estrogen users in a low-risk group for cardiovascular disease (the "healthy user" effect). And indeed, women who choose to use hormone therapy have been reported to have a

better cardiovascular risk profile than nonusers.[728] This question was addressed by the Lipid Research Clinics study, the Leisure World Study, and the Nurses' Health Study.[711, 729, 730] These epidemiologists concluded that their evidence strongly indicated that in women receiving estrogen treatment who have the same risk factors for cardiovascular disease as those not receiving treatment the same beneficial effect of estrogen was present. This is especially the case in the Nurses' Health Study, in which the participants are of a relatively homogeneous socioeconomic group. A cohort follow-up study in southeastern New England documented similar levels of total cholesterol, HDL-cholesterol, body mass index, and blood pressure in estrogen users and nonusers, indicating that selection of significantly more healthy women for estrogen use could not fully explain the beneficial effect of estrogen on the risk of cardiovascular disease.[731] In a comparison of health variables among users and nonusers in south Australia, there was no evidence to support the presence of a "healthy user" effect.[732] In Chile, users and nonusers of hormone therapy had identical risk factors for cardiovascular disease.[733]

In contrast to the uniform results from observational studies of the association between postmenopausal hormone therapy and coronary heart disease, epidemiologic data over the last 30 years regarding estrogen use and stroke have not been consistent. The many studies have indicated either a small increase or no effect of postmenopausal hormone therapy on the risk of stroke or a reduction in risk associated with estrogen or estrogen-progestin use.[709, 711, 717, 723, 734–743] A prospective cohort study in Denmark recorded an increase in ischemic strokes, but *only* among hypertensive women, and a large cohort study from Sweden found no link between stroke and hormone therapy.[744, 745]

Within this confusing mixture of results on stroke, there was one consistent observation. The cohort studies (with a sufficient number of cases) that have assessed the impact of hormone use on the risk of death from stroke have all indicated a beneficial impact. For example, the National Health and Nutrition Examination Survey (NHANES) recruited a very large cohort of women in 1971–1975 for epidemiologic analysis. The follow-up longitudinal study of this cohort yielded a U.S. national sample of 1,910 white postmenopausal women. Postmenopausal hormone use in this cohort provided a 31% reduction in stroke incidence and a strongly significant 63% reduction in stroke mortality.[738] These relative risks were present even after adjusting for age, hypertension, diabetes, body weight, smoking, socioeconomic status, and previous cardiovascular disease. This study specifically addressed the criticism that one should expect less disease in estrogen users because they are healthier. After adjusting for physical activity as a marker of general health status, the risk estimates remained identical.

Hypertension is both a risk factor for cardiovascular mortality and a common problem in older people. Studies have either shown no effect or a small, but statistically significant, decrease in blood pressure due to estrogen treatment.[746–751] This has been the case in both normotensive and hypertensive women.[752–757] The addition of a progestin did not affect this response.[100, 758] Discontinuing hormone therapy in women with hypertension does not result in a decrease in blood pressure (an expected response if the treatment were raising blood pressure), and in some patients discontinuation is followed by an increase in blood pressure.[759] The acute administration of estrogen to women with hypertension is followed by decreases in blood pressure, pulse rate, and circulating levels of norepinephrine.[760] ***The very rare cases of increased blood pressure due to oral estrogen therapy truly represent idiosyncratic reactions. Blood pressure should be assessed every 6 months in hypertensive women being treated with postmenopausal hormones, and, if the blood pressure is labile, blood pressure should be measured every 3 months.***

Observational studies have also reported that hormone users have a decreased risk of developing venous leg ulcers or pressure ulcers.[761, 762]

Cardiovascular Disease—Evidence from Clinical Trials

The Women's Health Initiative

The Women's Health Initiative (WHI) was organized by the U.S. National Institutes of Health in 1992 to study the health of postmenopausal women and was scheduled to be completed in 2007.[763] From 1993 to 1998, the WHI enrolled 161,809 women aged 50 to 79 in 40 clinical centers. The major components of the WHI were: (1) two randomized trials of postmenopausal hormone therapy scheduled to conclude in 2005, (2) a dietary modification trial that randomized 48,000 women to either a sustained low-fat or a self-determined diet, (3) a calcium/vitamin D supplementation trial, and (4) an observational study. One of the randomized trials of postmenopausal hormone therapy, the combined estrogen-progestin arm (daily 0.625 mg conjugated estrogens and 2.5 mg medroxyprogesterone acetate), randomized 16,608 women to either treatment or placebo. The other hormone trial, an estrogen-only arm (daily 0.625 mg conjugated estrogens), randomized 10,739 hysterectomized women to treatment or placebo.

On May 31, 2002, the Data and Safety Monitoring Board (DSMB) made its periodic review of the data accumulated by the Women's Health Initiative. The DSMB made two recommendations that were announced on July 9, 2002: (1) to discontinue the trial arm administering daily estrogen-progestin and (2) to continue the trial arm with daily unopposed estrogen in hysterectomized women. The combined estrogen-progestin arm was discontinued after about 5 years of follow-up because of a statistically significant increase in invasive breast cancer and an increase in cardiovascular events.[764] The statistical parameters for benefit or harm were established in 1997 early in the study. When the increase in breast cancer exceeded the predetermined boundary, the DSMB was obligated to recommend discontinuation of this arm of the trial.

On March 2, 2004, the National Heart, Lung, and Blood Institute of the U.S. National Institutes of Health canceled the estrogen-only arm of the Women's Health Initiative. This arm of the WHI included 10,739 hysterectomized, postmenopausal women who had completed an average of 6.8 years of follow-up. The WHI Data and Safety Monitoring Board made their last periodic review of the study data in December 2003. The DSMB was not unanimous in its decision; some members wanted to stop the study and others wanted the study to continue after sending a letter to the participants describing the findings. Even though none of the findings had crossed the predefined boundaries, the NIH made the decision to stop the study on February 2, 2004. The decision was based on the following results:[765]

- An increased risk of stroke similar to that reported in the canceled estrogen-progestin arm of the WHI.
- No increase or decrease in coronary heart disease.
- A trend toward an increased risk of probable dementia and/or mild cognitive impairment.
- A reduction in hip fractures.
- No increase in breast cancer.

With the exception of breast cancer and coronary heart disease, the results in the estrogen-only arm were essentially identical to those in the estrogen-progestin arm of the study. But keep in mind that the populations in the two clinical, randomized trial arms of the WHI were not identical.[766] Considering risk factors for cardiovascular disease, the women in the estrogen-only arm were more obese, less active, and had more pre-existing cardiovascular

disease. The estrogen-only arm also differed in regard to risk factors for breast cancer: more early births and bilateral oophorectomy, and more and longer duration of previous hormone therapy. ***Therefore, these were two different trials with two different populations and treatments, making direct comparisons inappropriate.***

The published results of the WHI trial agree with more than 30 years of case-control and cohort data with the exception (as first presented by the WHI) of the cardiovascular results. The updated results on the risk of coronary heart disease (CHD) from the canceled estrogen-progestin arm of the WHI reflected central adjudication of the cardiac diagnoses in contrast to the initial report that relied on local diagnoses.[767] The final report covered an average of 5.6 years of follow-up, compared with 5.2 years in the initial report. Based on these data, there would be an increase of six cases of CHD per 10,000 women per year in the treated group.

Central adjudication disagreed with 10% of the diagnoses for myocardial infarction and 3% for death due to coronary heart disease. This small degree of disagreement changed the strength of the conclusions comparing the initial report[764] with the updated report. Indeed, the overall results by definition did not achieve statistical significance in the follow-up report, and only the first year results were statistically significant in the year-by-year analysis, a conclusion based on a difference of only 19 cases. In all of the WHI reports, the intent-to-treat analyses were adjusted for multiple outcomes, the Bonferroni adjustment. All adjusted results were not statistically significant. It is difficult to understand the clinical meaning of this manipulation, but most believe that this indicates a slightly lower mathematical conclusion than presented in the nonadjusted data. This, of course, would further weaken the power of the reported results.

Consider also the possibility of diagnostic bias. 40.5% of the estrogen-progestin group in the WHI (nearly 5,000 of the 8,500 in the treated group), in contrast to 6.8% of the placebo group, were unblinded because of vaginal bleeding. What was the impact on the clinicians' final management and diagnosis when told that the patient is in the WHI study and experiencing vaginal bleeding? This problem affects the data not only in regards to cardiovascular disease but also for breast cancer. Unblinding was not a problem in the estrogen-only arm of the WHI and no increase in coronary heart disease was recorded—was this because of an absence of diagnostic bias in the estrogen-only arm?

The characteristics of the participants are now well-known:

	Estrogen-Progestin Arm	*Estrogen-Only Arm*
Average age	**63.3 years**	**63.6 years**
Drop-out rate	**42%**	**53.8 %**
Drop-in rate (began hormone use)	**6.2%**	**5.7%**

The women in the estrogen-progestin arm were an average of slightly more than 12 years distant from menopause.[765] Most had been without hormone therapy for more than a decade. In the estrogen-only arm, the published results do not specify the number of years distant from menopause, but this duration may have been even greater, influenced by the ages of bilateral salpingo-oophorectomy. Women with significant menopausal symptoms were excluded from the study to avoid an exceedingly high drop-out rate in the placebo group. Women who had been on hormone therapy (about 25% of the participants in the estrogen-progestin arm and 35% in the estrogen-only arm) and then underwent a 3-month "washout" period and experienced menopausal symptoms were discouraged from participation (about 12.5% of the participants in the estrogen-progestin arm reported vasomotor symptoms on entry but were willing to be assigned to placebo, and, therefore, their symptoms were unlikely to have had a major disturbing effect). This exclusion means that

only a small number of women in the WHI were close to their age of menopause (about 16.5% of the participants in the estrogen-progestin arm were less than 5 years since their menopause). For example, there was only a total of 574 women who were age 50–54 in the estrogen-progestin arm.[768]

Over the time of the studies, the participants discontinued their medication at a consistently increasing rate, so that at termination about half were no longer adhering to treatment. The WHI investigators argued that the high drop-out rate could lead to an underestimation of adverse effects; however, this would not be the case if longer duration of treatment exerts a beneficial effect. For example, a case-control study in the U.K. found a significant reduction in the risk of myocardial infarction only with the use of hormone therapy for more than 5 years.[769] Indeed, a trend for an emerging protection against coronary heart disease was observed in both arms of the WHI with increasing duration of treatment. Analysis of subsamples in the WHI revealed that the treated group had greater reductions in total cholesterol, LDL-cholesterol, glucose, and insulin levels and greater increases in HDL-cholesterol and triglyceride levels. It is tempting to link these findings with the test for trend that revealed a favorable, decreasing relative risk of coronary heart disease over time, which was statistically significant. However, this analysis was hampered by decreasing numbers over time, and the conclusion was not a strong one.

In subgroup analyses, only the women in the estrogen-progestin arm who were 20 or more years distant from menopause had a statistically significant increased risk of coronary heart disease. Subtracting this group from the rest of the participants, coronary heart disease now was observed in an identical prevalence comparing the treated and placebo groups. It is not appropriate to conclude, based on the WHI, that hormone therapy increases the risk of coronary clinical events in all postmenopausal women; this conclusion can only be applied to a specific older group of women. Indeed as early as 2004, the data from the estrogen-only arm suggested that younger women experienced a reduced risk of coronary heart disease with estrogen treatment.[765]

A reanalysis of the coronary heart disease data in the canceled estrogen-progestin arm of the WHI, published 7 years after the initial report, contributed nothing new, confirming that a statistically significant increase in coronary heart events occurred only in the women 10 or more years distant from their menopause.[770]

One response to the publications from the Women's Health Initiative has been a scientific and clinical effort to assess and use lower doses of estrogen. Half of the standard dose of conjugated equine estrogens has been demonstrated to effectively treat menopausal symptoms and to prevent bone loss. It is reasonable to ask whether symptoms and bone are especially sensitive to the effects of estrogen, and whether lower doses of estrogen will beneficially impact other target tissues. The cardiovascular system is of obvious concern because it was already apparent that lower doses of estrogen do have a lesser effect on circulating lipids and lipoproteins.

Clarkson and his colleagues studied the effects of a lower-dose estrogen trial in a monkey model of coronary atherosclerosis.[771] The animals were fed an atherogenic diet for 10 months, calculated to induce atherosclerosis comparable to that observed in early postmenopausal women. After oophorectomy, the animals were randomized to treatment for 2 years with a placebo or a dose of conjugated equine estrogens equivalent to 0.3 mg/day in women. This dose had no effect on circulating lipid levels; nevertheless the treated animals had an average 52% reduction in coronary atherosclerosis. This degree of protection was similar to studies in this model using a dose of conjugated estrogens equivalent to 0.625 mg/day.

The reliability and value of results obtained in Clarkson's monkey model have stood the test of time. Repeatedly results from this model proved to be predictive of hormonal effects in women. Therefore, this experiment using a lower dose of estrogen is important,

providing information regarding the effect of a lower dose on coronary atherosclerosis. Consistent with some reports in women, the lower dose of estrogen had no effect on levels of LDL-cholesterol, HDL-cholesterol, or triglycerides. But the treatment markedly reduced the extent of coronary artery atherosclerosis, a further indication that estrogen-induced inhibition of atherosclerosis occurs to a large extent independently of changes in lipids and lipoproteins. In the dose-response clinical trial that led to FDA approval of lower doses of conjugated equine estrogens, the 0.3 mg dose still produced statistically significant beneficial changes in lipids and lipoproteins, and this dose prevented bone loss.[108, 114]

The WHI and Stroke

The WHI reported an overall increase in the estrogen-progestin arm of ischemic stroke, but no increase in fatal strokes.[764, 772] The increase in nonfatal ischemic stroke in the estrogen-only arm of the WHI was of similar magnitude.[765, 773] A randomized, double-blind, placebo-controlled secondary prevention trial (the WEST trial) of daily 1 mg estradiol therapy was conducted in postmenopausal women after a recent (within 90 days) ischemic stroke or transient ischemic attack (25% of the women).[774] After an average of 2.8 years of follow-up (range 16–50 months), there were no significant overall differences comparing the treatment and placebo groups in any of the assessed outcomes, including nonfatal stroke, fatal stroke, coronary death, nonfatal myocardial infarction, or transient ischemic attack. The WEST trial retrospectively analyzed the time course of cerebrovascular events and found a significantly increased risk of stroke only at 6 months based on 21 strokes in the estradiol group and 9 strokes in the placebo group.

A major limitation of the WEST study was the reduced compliance with treatment because of the problems associated with unopposed estrogen treatment. Over a 3-year period, 116 women in the estradiol group discontinued treatment (34%) compared with 79 in the placebo group (24%). ***Nevertheless, the clinical meaning is straightforward: patients should not be given estrogen treatment after a vascular event in the expectation that recurrent vascular events would be prevented by the initiation of estrogen treatment. However, this recommendation is specifically targeted to women with existing vascular disease.***

The Nurses' Health Study reported an update of its data on the use of hormone therapy and stroke, focusing on the timing of initiation of treatment and the effect of estrogen doses.[775] In the analyses adjusted for age, BMI, cholesterol levels, diabetes, hypertension, smoking, and family history of early coronary heart disease, the following relative risks were observed for ischemic stroke (there was no significant increase in hemorrhagic stroke):

Current use of estrogen alone-	RR=1.43 (CI=1.17–1.74)
Current use of estrogen-progestin-	RR=1.53 (CI=1.21–1.95)

Comparing initiation of hormone therapy near menopause with initiation 10 or more years after menopause, there was no major difference.

The Nurses' Health Study also reported an increasing risk of stroke with an increasing dose of estrogen:

0.3 mg estrogen	25 cases	RR=0.93 (CI=0.62–1.40)
0.625 mg	268 cases	RR=1.54 (CI=1.31–1.81)
1.25 mg	60 cases	RR=1.62 (CI=1.23–2.14)

It is not easy to derive a take-home message from the Nurses' Health Study report. The authors stated that their findings are "virtually identical to those of the WHI trials." However, in the last report from the WHI, when women with prior cardiovascular disease or those

older than 60 years were excluded, the risk of stroke in women less than 10 years since their menopause was not significantly increased.[776] Therefore, there is disagreement.

In women with risk factors for stroke, it is prudent to use low doses of estrogen and to vigorously address the risk factors, such as effective treatment of hypertension. Would the transdermal route of administration be safer? This is an important question that cannot be answered because of a lack of data, but because stroke risk is limited to ischemic events and it is possible that the transdermal route has a lower risk of thrombosis, it seems wise to promote this route of administration in older postmenopausal women and in women with risk factors for stroke.

The Secondary Prevention Randomized Clinical Trials

The Heart and Estrogen-progestin Replacement Study (HERS)

HERS was a randomized, double-blind, placebo-controlled clinical trial designed to determine whether daily treatment with 0.625 mg conjugated estrogens and 2.5 mg medroxyprogesterone acetate would reduce coronary heart disease events in women with pre-existing coronary disease.[777–779] 2,763 women (average age 66.7 years) were enrolled in 20 U.S. clinical centers and randomized to treatment and placebo beginning in February 1993 and ending in July 1998. Overall, there were 172 myocardial infarctions and coronary deaths in the hormone group and 176 in the placebo group—obviously no difference.

At baseline, the use of statins and aspirin was essentially equally prevalent in the treated and placebo groups (about 40% of the subjects used statins and 80% used aspirin). However, more women in the placebo group began treatment with statins, so that by the end of the follow-up period, the 69% versus 65% difference comparing placebo with treatment was statistically significant. The authors addressed this potential confounder by adjusting for the difference in statin use (as well as other confounders) and concluded that the adjusted analyses were essentially identical to the original analyses. However, no mention is made of the fact that the percentage use of statin use is impressively high. What if any beneficial effect of estrogen is lost because of the impact of statin therapy? Indeed, in a primary prevention trial, inhibition of atherosclerosis with estrogen treatment was observed only in women NOT receiving statins.[780] The HERS investigators compared coronary heart disease events in the hormone group with the placebo group in women not using statins or aspirin and found no difference, but this very important possible explanation for the lack of a beneficial effect of estrogen in HERS cannot be answered by the analysis of the HERS data because statin and aspirin treatment were not randomized, and the number of events in women not on statins or aspirin was very small.

Of the 2,763 postmenopausal women in the HERS trial, 2,321 (93%) agreed to be involved in additional follow-up evaluation, HERS-II. The original study[777] lasted 4.1 years, and the average extended follow-up equaled 2.7 years, for a mean total of 6.8 years. At the beginning of the follow-up period, the average age of the participants was 71 (67 at baseline and 74 at closure). The investigators could detect no significant differences in the rates of coronary events or secondary cardiovascular events comparing the treated group with the placebo group. There was no statistical trend for a beneficial effect of hormone therapy with longer duration of treatment. Because of the absence of a difference, the follow-up period, scheduled to last 4 years, was terminated early.

The additional follow-up period was unblinded; patients and physicians could choose to continue, discontinue, or initiate hormonal or other therapy. Hormone use in the original treated group in HERS declined from 81% after 1 year to 45% during the sixth year (and 11% were using preparations other than the original 0.625 mg conjugated estrogens and 2.5 mg medroxyprogesterone acetate). During the sixth year, 8% of the placebo group were now receiving hormone therapy. Raloxifene or tamoxifen had even been initiated, 3% in the hormone group and 4% in the placebo group. The investigators recognized this problem conceding that their power to detect an increasing benefit was eroded by the changing treatments; however, their analysis indicated an ability to detect at least an 18% reduction in cardiac risk.

The original HERS report indicated a 2–3-fold increase in deep vein thrombosis and pulmonary embolism in the hormone-treated group. In the follow-up period, there was no longer a statistically significant increase in deep vein thrombosis. There was no reduction in pulmonary embolism, but the number of events was too small to provide accurate assessment. The event rates for venous thrombosis were 5.9 per 1,000 women per year of treatment and 2.8 in the placebo group. Overall, there was a 48% increase in risk for biliary tract surgery in the treated group, 6 more cases per 1,000 women per year compared with placebo.

Intention-to-treat analysis compares all individuals in the treated group to all in the placebo group, regardless of individual compliance or completion of the study. Proponents argue that this is the best method of analysis for clinical trials because it reflects the full impact of randomization. Opponents contend that this method is intuitively wrong; how can the long-term benefit of a treatment be assessed if subjects receiving treatment for only a short period of time are included? HERS II performed an "as-treated" analysis, focusing on women with 80% or more compliance and found relative hazards (like relative risk) similar to those in the intent-to-treat analysis. However, the relative hazard in the as-treated analysis for primary coronary heart disease events in HERS II was reduced, although not statistically significant (RH=0.81; CI=0.52–1.32). Events were fewer in the as-treated analyses because only 37% of the events qualified. Adjustment for statin use was performed only in the intent-to-treat analysis ("only a trivial effect on the findings"). The HERS clinical trial results that reflect intent-to-treat analyses are compromised by a difficulty in detecting a long-term effect and the results that reflect the as-treated analysis are compromised by few events because of compliance and drop-out problems. Therefore, unanswered questions remain: whether reported increases in cardiovascular events early after the initiation of hormone therapy reflect a true risk of hormone therapy or the effect of reduced events in the placebo groups because of new onset treatment with statins and aspirin, or an effect limited to women with significant pre-existing coronary artery disease.

Other Secondary Prevention Trials

The results of a multicenter trial (the Estrogen Replacement and Atherosclerosis trial, ERA) examined the effect of postmenopausal hormone therapy on the progression of coronary atherosclerosis as assessed by angiography.[781] A total of 309 women were randomly assigned to receive either unopposed estrogen, 0.625 mg conjugated estrogens per day, a daily combination of estrogen and progestin, 0.625 mg conjugated estrogens and 2.5 mg medroxyprogesterone acetate or placebo. Over 3.5 years of treatment, angiography did not detect any differences in disease progression between any of the groups. The women in this study had documented heart disease on entry and were a relatively older group of women (mean age 65.8 years). Half had had a previous myocardial infarction. There were no reported increases in cardiac events in any of the three treatment groups.

The ERA trial joins the HERS trial in demonstrating no secondary preventive effect of postmenopausal hormone therapy on older women with significant coronary heart disease. ***Comparing the two trials, however, there is an important observation. The ERA trial***

contained an estrogen-only arm, and the absence of a difference between the estrogen-only arm and the estrogen-progestin arm argues against a negative effect due to the daily administration of medroxyprogesterone acetate.

Another secondary prevention, 3-year trial (Women's Estrogen-progestin Lipid-Lowering Hormone Atherosclerosis Regression Trial, WELL-HART) assessed whether unopposed estradiol or a sequential regimen of estradiol and medroxyprogesterone acetate could slow the progression of atherosclerosis.[782] The double-blind, placebo-controlled trial involved 226 postmenopausal women with an average age of 63.5 years (range 48–75), who already had at least one demonstrated coronary artery lesion. The results were based on follow-up angiograms in 59 women in the placebo group, 54 in the estradiol group, and 53 in the estradiol-medroxyprogesterone acetate group. A reduction of LDL-cholesterol to less than 130 mg/dL was achieved by dietary intervention, but coronary angiography to measure the change from baseline in the percent stenosis failed to demonstrate a difference among the groups receiving placebo; unopposed, daily 1 mg estradiol; or daily 1 mg estradiol and 12 days each month of 5 mg medroxyprogesterone acetate. ***There were also no differences in cardiovascular events during treatment. The results indicated that medroxyprogesterone acetate administered in a sequential regimen is not associated with an adverse cardiovascular effect.***

At least three other secondary prevention trials in women with coronary heart disease failed to demonstrate a beneficial impact of hormone therapy.[783–785] These various trials tested oral conjugated estrogens, oral estradiol valerate, and transdermal estradiol combined with either medroxyprogesterone acetate or norethindrone. ***The results of the secondary prevention trials in older women with established heart disease are uniformly consistent in finding no beneficial effects of hormone therapy, and the data indicate that different estrogens and different progestins behave similarly.***

The Timing Hypothesis

The timing hypothesis argues that estrogen can reduce the risk of coronary heart disease when administered to relatively young postmenopausal women before atherosclerosis has developed to the stage of unstable plaques (plaques with necrosis and inflammation).

The Women's Health Initiative investigators conducted a secondary analysis of the 2 canceled clinical trial arms.[776] The results in the estrogen–only arm, the combined estrogen–progestin arm, and with the participants combined were separated into age groups at randomization (50–59, 60–69, and 70–79) and according to years since menopause (<10, 10–19, and 20 or more). ***An increased risk for coronary heart disease was present only in the oldest women in the trials. There were no increases for CHD, stroke, or total mortality in women ages 50–59. In fact, only the increase in CHD events in women 20+ years since menopause reached statistical significance. There was no apparent increase in cardiovascular disease risk in treated women close to their menopause. Indeed, a statistically significant reduced risk was present for total mortality in women age 50–59.***

The data in this 2007 WHI report were not new. A careful reading of the initial adjudicated reports reveals that the risk of CHD was present only in the oldest women in the trials.[767, 786] In subgroup analyses, the only significant increase in stroke occurred in the estrogen–only arm in women ages 60–69. But as noted previously, when women with prior cardiovascular disease or those older than 60 years were excluded, the risk of stroke in women less than 10 years since their menopause was not significantly increased.

Will postmenopausal hormone therapy begun at or near the time of the menopause, and maintained for a relatively long duration of time, provide protection against coronary artery disease (primary prevention)? The design of the canceled arms of the WHI did not allow

an answer to this question. As previously noted, women with significant menopausal symptoms were excluded from the study to avoid an exceedingly high drop-out rate in the placebo group. The WHI addressed this problem by pointing out that the ratios of cardiovascular events in the treated and placebo arms were the same when assessed by decades of age, 50s, 60s, and 70s. However, this was not the critical analysis. By excluding women with menopausal symptoms, it is very likely that a small number of the participants were close to their age of menopause. Only 574 in both the treated and placebo groups were age 50–54 in the estrogen-progestin arm, and it is likely that even some of these women were relatively distant from their menopause. Even with the appropriate analysis according to years from menopause, the results are limited by very small numbers of women in their early postmenopausal years. Furthermore, the high drop-out rates in these clinical trials erode the statistical power for assessing year-by-year new statin or aspirin treatment.

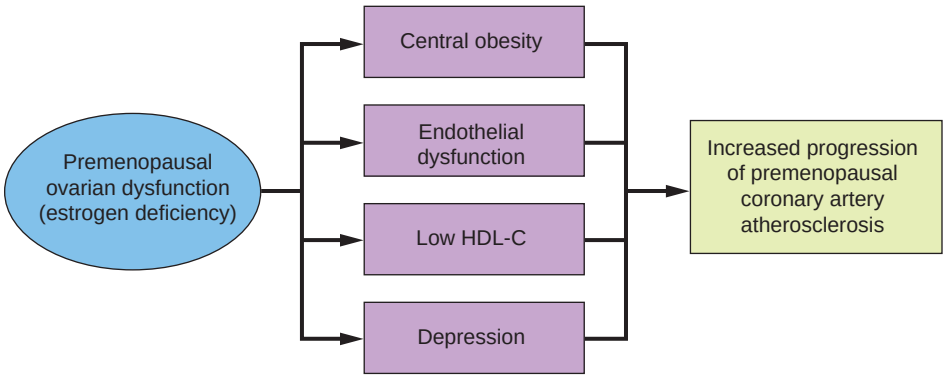
There is a small primary prevention trial of 199 healthy postmenopausal women randomized to a daily dose of 1 mg estradiol or placebo, and followed for 2 years measuring with ultrasonography the change in carotid artery intimal media thickness.[780] The women receiving estradiol had a small decrease in intimal media thickness in contrast to a marked increase in the placebo group. Interestingly, in those participants taking lipid-lowering medications, there was no difference comparing estrogen treatment with placebo (both groups had a small decrease in thickness), indicating that lipid-lowering drugs and estrogen had similar beneficial effects on atherosclerosis that were not additive. There have been two other similar ultrasonography studies. A 2-year Dutch study did not achieve a significant difference comparing treatment with placebo, but the results were handicapped by a very high drop-out rate and problems with follow-up.[787] A German study that found no effect of hormone therapy enrolled women who already had increased intimal media thickness, and the duration of the study was only 1 year.[788]

A meta–analysis of 23 randomized hormone therapy trials concluded that treatment reduced the risk of coronary heart disease events by 32% in younger women compared with older women (10 or more years since menopause or greater than 60 years of age).[789] This is a conclusion that is less firm than at first apparent, because most of these trials were not designed to measure an endpoint of cardiovascular disease. However, another meta–analysis by the same authors concluded that hormone therapy reduced overall mortality in women with an average age less than 60.[790] There is a growing story that adequate estrogen exposure prior to the onset of clinical events provides protection against cardiovascular disease.

The timing hypothesis originated in the hormone trials conducted in monkeys by the Wake-Forest group headed by Tom Clarkson.[791] This is randomized trial evidence, albeit in monkeys, and we should place the results at the head of the list of observations that support the timing hypothesis:

- Estrogen treatment initiated immediately after menopause in monkeys inhibited progression of coronary artery atherosclerosis by about 70%. When treatment was delayed by 2 years (equivalent to about 6 years in women), there was no effect.[791]
- Next in line, according to strength of evidence in our view, would be the WHI reports of reduced coronary artery calcium in estrogen-treated women and an increase in cardiac events only in the oldest women in the trials.[727, 776] The problem of low event rates in younger women in the WHI was addressed by lumping together, in 1 hazard ratio, myocardial infarction, coronary death, coronary revascularization, and confirmed angina—the risk in women aged 50 to 59 years for all of these events was significantly reduced (HR-0.66; CI=0.45–0.96).[786] In addition, the WHI measured coronary artery calcium in a substudy of the estrogen-only arm and found that women with bilateral oophorectomies who were not treated with estrogen had an increase in subclinical coronary heart disease.[792]
- Every woman has a trajectory of atherosclerosis development, the slope of which determines the age of onset for clinical events. Premenopausal women with lower

estrogen levels have higher cardiovascular risk factors and develop more and earlier coronary heart disease.[793] This includes suppressed ovarian function associated with stress, depression, or athletic activity. The importance of premenopausal estrogen is also supported by Clarkson's monkey studies. Premenopausal monkeys with normal ovarian function have less progression of coronary artery atherosclerosis as compared with monkeys with impaired ovarian function.[791] The Mayo Clinic Cohort Study of Oophorectomy and Aging included 1,274 premenopausal women with unilateral oophorectomy (followed for a median of 29.5 years) and 1,091 premenopausal women with bilateral oophorectomy (followed for a median of 25 years) who had surgery from 1950 through 1987, compared with 2,383 matched women from the same population of women who had not undergone oophorectomy.[794] Women with bilateral oophorectomy before age 45 had about a 5-fold increase in risk of mortality for neurological or mental diseases. The Mayo Clinic study also indicated that women with bilateral oophorectomy before age 45 experienced almost a 2-fold increase in mortality from cardiovascular disease, increased risks of parkinsonism, cognitive impairment and dementia, and an increase in depressive and anxiety symptoms later in life.[795–798]



- The Nurses' Health Study compared, after 24 years of follow-up, ovarian conservation (13,035 women) and bilateral oophorectomy (16,345 women) at the time of premenopausal hysterectomy.[799] It required bilateral oophorectomy in 220 women to achieve a reduction in breast and ovarian cancer in one case. However, total cancer mortality increased, most notably an increase in lung cancer, one case after each 190 surgeries. *An all-cause increase in mortality, coronary heart disease, and stroke was observed in those women who never used estrogen after surgery; this amounted to one additional death for every nine surgeries!*
- The results in observational studies strongly indicate that hormone treatment of young postmenopausal women reduces the risk of coronary heart disease. In the Nurses' Health Study, the reduction was approximately 50%.[723] The women in the Nurses' Health Study who were under age 60 when hormone treatment was initiated had a significant reduction in coronary heart disease risk compared with no effect in women over age 60.[800] A similar reduction was present in the observational arm of the WHI.[801] In fact, after adjustment for confounding influences in the WHI observational arm such as behavioral, dietary, physical activity, and cardiovascular risk factors, the relative risks for cardiovascular events were 30% to 38% lower than in the clinical trials. These data were derived from populations of women usually treated with postmenopausal hormone therapy, women close to their age of menopause.
- In a primary prevention trial using ultrasound measurement of carotid artery intima-media thickness, estradiol-treated women had slower progression of atherosclerosis.[780] These same investigators demonstrated no effect of estrogen treatment in older women who had angiographic evidence of coronary atherosclerosis.[782]

The message from multiple secondary prevention trials is clear: we should not prescribe estrogen to women with coronary heart disease in the expectation that treatment will reduce

subsequent cardiac events. The evidence is also convincing that progestational agents do not produce adverse cardiovascular effects. It remains very possible, indeed likely, that primary prevention of coronary heart disease can be achieved with estrogen administered at the right time of life. We await the results of two ongoing primary prevention trials, measuring carotid intima-media thickness with ultrasound, the KEEPS trial (www.kronosinstitute.org/keeps.html) and the ELITE trial (http://clinicaltrials.gov/show/NCT00114517).

A reasonable goal is to maintain a healthy level of estrogen during the premenopausal years and in the early postmenopausal period. Although the timing hypothesis has not been proven by randomized, clinical trials, the overall evidence is impressive, and in our view, sufficient to conclude that hormone therapy in the early postmenopausal years can provide primary prevention of clinical coronary disease. Clinical decisions reflect all of our knowledge (our education, the medical literature, and our experience), not just the data from randomized, clinical trials.

Coronary Heart Disease—Conclusion

The WHI concluded (and many individuals and organizations did as well) that hormone therapy is not a viable intervention for primary prevention of coronary heart disease. We cannot quarrel with the conclusion that postmenopausal hormone therapy does not reduce or slow the progression of established coronary heart disease. However, the WHI did not study the appropriate population in the appropriate time period to establish that hormone therapy does not exert a primary preventive effect on the risk of coronary heart disease.

The results of secondary prevention trials provide a reasonably solid basis not to recommend postmenopausal hormone therapy for women with existing atherosclerosis in the anticipation of preventing future cardiovascular events. The results also indicate that there is no need to avoid the use of medroxyprogesterone acetate, because there has been no difference observed comparing women treated only with estrogen to those treated with estrogen and progestin.

The cardiovascular results over the last few years support an emerging theme. The theme is: Healthy endothelium is needed to respond to estrogen. Experimental evidence in the monkey indicates that the beneficial effects of hormonal treatment are progressively diminished with increasing atherosclerosis.802 In postmenopausal women, the vasodilatory effects of estrogen dissipate with increasing age.803 By the time, the endothelium is involved with atherosclerosis, it is too late for estrogen to exert a beneficial effect. The clinical trial reports make an argument that the optimal approach to postmenopausal hormone therapy is to start treatment close to the menopause, avoiding a significant period of exposure to low estrogen levels prior to beginning therapy. And there continues to be good reason (a combination of biologic data and uniform agreement in a large number of observational studies) to believe that appropriately timed hormone therapy can have a beneficial role in the primary prevention of coronary heart disease.

Venous Thrombosis

Results from multiple studies indicate that postmenopausal hormone therapy increases the risk of venous thromboembolism (VTE) about 2-fold, mostly in the first year or two

of treatment, a conclusion supported by the results from the canceled estrogen-progestin arm of the WHI.[764, 804] The absolute risk in the WHI estrogen-progestin arm was 18 additional cases per 10,000 women per year. In the estrogen-only arm of the WHI, a smaller increase in deep vein thrombosis was observed, and an increase in pulmonary embolism was not statistically significant.[765, 805] There are some notable observations in the WHI data on venous thrombosis. Most of the cases occurred in the first 2 years of exposure, and the risk was greatest in the women over age 70 and in overweight women. The risk was higher in women who were susceptible to venous thrombosis, specifically those with the Leiden mutation. This raises a very important question: is it possible that the risk of venous thrombosis is very low in younger, normal postmenopausal women?

It should be emphasized that the risk of VTE appears to apply only to new hormone starters; women who have been on hormone therapy can be reassured that the evidence indicates that the increased risk of venous thrombosis is concentrated in the first 1—2 years of treatment. The actual risk is very low because of the low frequency of this event. If the relative risk is increased 2-fold, this would increase the incidence of venous thromboembolism by about 2 cases per 10,000 women per year of hormone use. Furthermore, venous thrombosis carries with it a very low risk of mortality, around 1%, and most of the fatal cases have followed venous thrombosis associated with trauma, surgery, or a major illness.

VTE is a risk that is reduced with the use of statins and low-dose aspirin,[806, 807] although it is not known whether statin and aspirin use would completely protect against the increased risk associated with hormone therapy.

Clinicians have argued that transdermal administration of estrogen is safer in regard to the risk of venous thromboembolism, considering the first-pass effect in the liver to be an important factor in the prothrombotic impact of oral estrogen. For example, it has been reported that the oral administration of estrogen compared with transdermal administration in male-to-female transsexuals is associated with a greater prothrombotic state and risk of venous thrombosis; however this effect could be at least partly attributed to major differences in the estrogen doses.[808] It would be better to have evidence uninfluenced by dosage, and for this purpose, we can consider responses of activated protein C resistance, recognized as a marker for venous thrombosis risk. In one randomized trial, oral estrogen treatment increased activated protein C resistance, but transdermal estrogen was no different than placebo.[37] Another randomized trial found that both routes of administration increased activated protein C resistance; however, the increase with oral estrogen was about 4 times greater compared with transdermal estrogen.[36] A French case-control study concluded that there was a 4-fold increase in venous thromboembolism with current use of oral estrogen but no increase in risk with transdermal estrogen.[38] In addition, this study reported that oral estrogen treatment adds to the risk of VTE associated with obesity, but transdermal estrogen does not.[809] The French study also reported (although limited by small numbers) that transdermal treatment, in contrast to oral estrogen, does not further increase the risk of VTE associated with Leiden or prothrombin mutations.[39]

But there are some problems with the French case-control study. There are wide confidence intervals for the significant odds ratio associated with oral estrogen treatment. Usually this reflects small numbers, but the number of cases and controls in this study should allow greater precision. It is possible that this imprecise conclusion is influenced by the fact that the cases and controls differed significantly in several characteristics that influence the risk of VTE, specifically greater body weight and a positive family history of VTE. We know that a 2-fold increased risk of VTE is uniformly reported, including the data from the Women's Health Initiative (WHI).[804, 805] It is very likely that the French estimate is higher compared with the usual 2-fold increased risk because of the confounding differences between their cases and controls. ***It is worth noting again that in the WHI, the cases of venous thrombosis were concentrated in the first years of exposure, in the oldest women in the study, and in the heaviest women in the study.***

The French case-control study found no increase in VTE risk associated with estrogen combined with progesterone or pregnane derivatives, and an increase with norpregnane derivatives. The pregnane group includes synthetic progestins familiar to us such as medroxyprogesterone acetate, chlormadinone, and cyproterone. The norpregnanes (progesterone without the 19-carbon) included two progestins, nomegestrol acetate and promegestone, that are not used in the U.S (nomegestrol acetate is the progestin in Uniplant, a single rod implant contraceptive, and combined with estradiol in an oral contraceptive). But can we make the conclusion that the norpregnanes are thrombogenic? The confidence interval in the norpregnane group was very wide, again apparently not due to small numbers, but this makes this conclusion shaky and suspect. Similar results were reported from the E3N French prospective cohort study, with an increased risk of venous thromboembolism of 1.7 (CI=1.1–2.8) associated with current users of oral therapy (a hazard ratio more in keeping with the usual 2-fold increase reported in the literature) and no increase with transdermal estrogen.[41] Again an increase was observed with norpregnane progestins, this time with tighter confidence intervals.

In the French studies, oral hormone users used almost exclusively estradiol in doses that averaged 1.5 mg/day. Transdermal users most commonly used an estradiol dose of 50 μg or less daily. To legitimately compare oral and transdermal methods, one would have to be sure the two groups had similar blood levels, to account for the wide variation in metabolism and clearance among individuals. It is possible that the difference between the oral and transdermal groups represent differences in estrogen doses. Nevertheless, the French conclusions are supported by a very large case-control study (23,505 cases of VTE) using the U.K. General Practice Research Database.[272] The use of transdermal estrogen alone or combined with a progestin was not associated with an increase in VTE, compared with about a 1.5-fold increase with oral estrogen alone and oral estrogen-progestin.

A greater safety with transdermal administration of estrogen in regards to VTE makes some sense because of the known lesser impact on clotting proteins when the first-pass liver effect is avoided. This is supported by transdermal's almost negligible effect on activated protein C resistance when compared with oral therapy.[36, 810] ***Therefore, the observational studies support the clinical choice of a transdermal method for women who are at higher risk for VTE. The data are too weak to make any statements with confidence regarding the effect of various progestins.***

Occasionally postmenopausal hormone therapy has to be considered in a patient who has experienced an episode of idiopathic venous thrombosis. A randomized trial assessing the risk of recurrent venous thromboembolism in women with previous episodes being treated with an oral hormone regimen was canceled after 1 year when eight women in the treated group developed recurrent venous thrombosis compared with one in the placebo group.[811] The eight women were studied retrospectively, and six of the eight had an inherited susceptibility for venous thrombosis. The numbers were not large, but the results are a strong argument for the recommendations detailed in Chapter 22 with oral contraceptives. There is no argument that inherited abnormalities in the clotting scheme increase the risk for venous thromboembolism. The evidence is mixed for arterial thrombosis, but most of the evidence fails to find an association between specific inherited defects (thrombophilia) and arterial thrombosis, although this is still unsettled.[812–815]

If a patient has a close family history (parent or sibling) or a previous episode of idiopathic thromboembolism, an evaluation to search for an underlying abnormality in the coagulation system is warranted before exposing the patient to exogenous hormone therapy. The following measurements are recommended, and abnormal results require consultation with a hematologist regarding prognosis and prophylactic treatment. The list of laboratory tests is long, and, because this is a dynamic and changing field, the best advice is to consult with a hematologist. When a diagnosis of a congenital condition is made, screening should be offered to other family members.

Hypercoaguable Conditions	*Thrombophilia Screening*
Antithrombin III deficiency	Antithrombin III
Protein C deficiency	Protein C
Protein S deficiency	Protein S
Factor V Leiden mutation	Activated protein C resistance ratio
Prothrombin gene mutation	Activated partial thromboplastin time
Antiphospholipid syndrome	Hexagonal activated partial thromboplastin time
	Anticardiolipin antibodies
	Lupus anticoagulant
	Fibrinogen
	Prothrombin G mutation (DNA test)
	Thrombin time
	Homocysteine level
	Complete blood count

Other risk factors for thromboembolism that should be considered by clinicians include an acquired predisposition, such as the presence of lupus anticoagulant or malignancy, and obesity, immobility or trauma. Varicose veins are not a risk factor unless they are very extensive, and, unlike arterial thrombosis, smoking either has no effect or at best is a weak risk factor for venous thromboembolism. Raloxifene (and tamoxifen) share with estrogen an increased risk for venous thromboembolism.[186] The size of the risk is comparable for all three drugs, about a 2-fold increase.

If a patient has a congenital predisposition for venous thrombosis or if she is considered to be at high risk for venous thromboembolism, the clinician and patient can consider the combination of hormone therapy (preferably transdermal administration) and chronic anticoagulation, in consultation with a hematologist, choosing from the options of statins, low-dose aspirin (75–162 mg), or low-dose warfarin.

There are no studies of venous thromboembolism following surgical procedures in postmenopausal hormone users. ***We recommend appropriate prophylactic anticoagulant treatment in hormone users anticipating immobility with hospitalization, especially if other risk factors (most notably, obesity) are present. Some patients may elect to discontinue hormone treatment 4 weeks prior to major surgery if extended immobility is to be expected, but this is an empiric, individual decision. Hormone therapy can be resumed when the patient is ambulatory.***

Breast Cancer and Postmenopausal Hormone Therapy

Women and clinicians are regularly reminded about the threat of breast cancer, in the media, by advertisements, and by the experience of a friend or family member fighting this disease. Breast cancer is a major focus in the health concerns and care for postmenopausal women because it has an increasing frequency with age. About 95% of all breast cancers and 97% of breast cancer deaths in the U.S. occur in women over age 40.[816] Thus, there are good reasons why this medical condition is a prominent factor in clinical decision-making regarding postmenopausal hormone therapy.

The long-term use of postmenopausal hormone therapy has been challenged by data that have been interpreted to indicate that the risk of breast cancer is increased in hormone users. The debate and publicity over this issue have made decision-making very difficult for both patients and clinicians. ***The most important unanswered question is whether postmenopausal hormone therapy initiates the growth of new breast cancers or whether the epidemiologic results reflect an impact on pre-existing tumors.***

Biological Plausibility

The most compelling reason to believe that long-term use of postmenopausal estrogen increases the risk of breast cancer is the inherent biologic plausibility. Factors known to increase a woman's exposure to estrogen are known to increase the risk of breast cancer; e.g., there is a small decrease in risk with late menarche and a moderate increase in risk with late natural menopause.[817] In premenopausal women who are overweight, the risk of breast cancer is lower compared with normal weight individuals, and in postmenopausal women, excess weight is associated with either an unchanged or slightly increased risk.[818–820] This is attributed to an increase in total and free estrogen levels in overweight postmenopausal women, in contrast to lower levels with increasing weight in premenopausal women. Postmenopausal obese women have later menopause, higher estrone production rates and higher free estradiol levels (lower sex hormone-binding globulin), and a slightly greater risk for breast cancer.[821] Greater bone density, believed to be a marker for estrogen exposure, is also associated with an increased risk of breast cancer. [822–824]

Studies seeking a correlation between circulating levels of sex hormones and breast cancer have yielded conflicting results. In the Rancho Bernardo cohort, no relationship between estrogen, androgen, and sex hormone-binding globulin levels and the incidence of breast cancer could be demonstrated.[825] Using serum collected earlier in life, no differences in endogenous hormones could be detected in 51 women who subsequently developed breast cancer; including the various estrogens, progesterone, androstenedione, and even sex hormone-binding globulin.[826] On the other hand, in a very large prospective study in Italy, estradiol, testosterone, and sex hormone-binding globulin levels were higher in postmenopausal women who subsequently developed breast cancer.[827] In a British report, women who subsequently developed breast cancer had higher levels of estradiol.[828] Two North American prospective studies also found higher levels of estrogen in women who subsequently developed breast cancer, and most impressively, an increasing risk of breast cancer correlated with increasing levels of free estradiol.[829, 830] In another study, women who developed breast cancer displayed higher levels of non-bound, free estradiol and lower levels of sex hormone-binding globulin (SHBG).[831] In the Nurses' Health Study, an association was reported between an increased risk of breast cancer and higher levels of estradiol, estrone, and dehydroepiandrosterone sulfate; whereas no association could be demonstrated with the percent free or bioavailable levels of estradiol, androstenedione, testosterone, or dehydroepiandrosterone.[832] ***The discrepancies among the various studies seeking a correlation between hormone blood levels and the risk of breast cancer reflect the fact that the differences are very small, and it is a struggle to achieve statistical significance.***

Overall, the biologic plausibility for an estrogen link with breast cancer is an impressive argument. This argument is further strengthened by the proven benefit of reducing the incidence of breast cancer with the antiestrogen, tamoxifen. However, as we shall see, establishing a clinically real cause-and-effect relationship with epidemiologic data requires more than the rationale of biologic plausibility.

- ***The risk of breast cancer is increased with increasing duration of lifelong exposure to estrogen.***
- ***Postmenopausal women who are overweight have a slightly increased risk of breast cancer.***

Results from Observational Studies

In the past decade, a large number of case-control and cohort studies indicated a slightly increased risk of breast cancer with postmenopausal hormone therapy. Overall, most of these studies indicated a greater risk associated with estrogen-progestin compared with unopposed estrogen treatment.[833–849]

A Reanalysis of the World's Literature

Prior to the publications from the Women's Health Initiative, the 1997 report of a reanalysis of the world's literature was the most referenced article on this subject. A team of epidemiologists invited all investigators who had previously studied the association of postmenopausal hormone use and the risk of breast cancer (51 studies) to submit their original data for a collaborative combined reanalysis, an undertaking more rigorous than a standard meta-analysis. This analysis reached the following conclusions[850]:

- Ever users of postmenopausal hormones had an overall increased relative risk of breast cancer of 1.14.
- Current users for 5 or more years had a relative risk of 1.35 (CI=1.21–1.49), and the risk increased with increasing duration of use.
- Current and recent users had evidence of having only localized disease (no metastatic disease) and ever users had less metastatic disease.
- There was no effect of a family history of breast cancer.

This was the first important indication that women with a positive family history of breast cancer do not have a further increase in risk with hormone therapy, a finding also reported in the Women's Health Initiative.[851] There were two other notable conclusions: all of the increase in risk was localized disease, and women who used hormone therapy and developed breast cancer had better survival rates than nonusers; these observations have been repeatedly confirmed.

The Million Women Study

The Million Women Study recruited 1,084,110 women between 1996 and 2001 from those invited by the U.K. National Health Service Breast Screening Programme to have screening mammography every 3 years (about half had ever used postmenopausal hormone therapy).[844] The study data were recorded from questionnaires returned prior to the initial mammography, and the women were followed to determine cancer incidence and death. The study is noteworthy for its large numbers and adjustments for the well-recognized factors associated with risk of breast cancer. No increase in risk of breast cancer was measured in past users of any hormone preparation, regardless of length of time since discontinuation, from less than 5 years to 10 or more years (with the exception of discontinuation in the year previous to diagnosis), and regardless of duration of use. Based on an average follow-up of 2.6 years (a very short exposure; indeed, the breast cancers were diagnosed on an average of 1.2 years after the study began), the relative risk for invasive breast cancer in current users of estrogen-only was 1.30 (CI=1.22–1.38), and for current users of estrogen-progestin, 2.00 (CI= 1.91–2.09).

There are many criticisms of the Million Women Study. For example, the study reported a lower risk of breast cancer for perimenopausal and postmenopausal women compared with premenopausal women despite the well-established fact that breast cancer risk increases with aging. There were many differences comparing users and nonusers, requiring multiple adjustments. Hormone use or nonuse was established at entry and not changed during follow-up despite multiple cross-overs in treatment among the women. Validation of the questionnaire data was claimed based on information obtained from only 570 women. Breast cancer mortality was assessed after an average of 4.1 years of follow-up, based on a total of 517 deaths; however, breast cancer was diagnosed very rapidly (an average of 1.2 years) and deaths occurred swiftly (within an average of 1.7 years). Current users and past users were compared with the never users, and, although the risk of mortality was increased, it did not reach statistical significance (1.22; CI=1.00–1.48). This finding was highlighted because it differed with a consistent story reported in the literature over a decade that hormone users had better survival rates. The Million Women Study calculated their risk of mortality by dividing deaths from breast cancer by total number of users or nonusers. When the data are recalculated appropriately by dividing deaths from breast cancer by total number of cases of breast cancer in the user and nonuser groups, the results agree with the literature—the risk of mortality was reduced about 27% in the hormone users!

The Danish Nurse Cohort

The Danish Nurse Cohort of 10,874 women was established in 1993.[852] The results reflected a variety of hormone products and regimens, and an increased hazard ratio for breast cancer was associated with hormone therapy (HR=2.42; CI=1.81–3.26), with a 2-fold increase in breast cancer mortality. Stated simply, hormone users had a greater mortality rate from breast cancer because they had more breast cancers. Even though the case fatality rate was better (more favorable prognosis), the increased incidence produced a net effect of an increase in mortality.

Danish case-control and cohort studies are noted for their ability to accurately obtain information from the national registries. Nevertheless, there were important limitations to this study. The most glaring problem was that not all causes of death could be verified and a death in a woman with a diagnosis of breast cancer was assumed to be a breast cancer death. The authors state that this problem would be balanced by similar numbers in the user and nonuser groups, but this is an assumption. In this population of older women, you would expect deaths from other causes to outnumber deaths from breast cancer. This single assumption by the epidemiologists could have skewed the results.

In addition, the case fatality and breast cancer mortality calculations were adjusted only for age. The authors stated that a relatively small number of deaths precluded multiple adjusting. But there are so many factors that influence the risk of breast cancer, including age of menarche, age at menopause, age at first full term pregnancy, parity and age at diagnosis, use of mammography, presence of benign breast disease (specifically with atypical hyperplasia), body size, and alcohol intake. How can we know that the results did not reflect an imbalance in some of these factors?

Hormone Therapy and Lobular Breast Cancer

A population-based, case-control study adds to the previous reports from this Seattle group of epidemiologists focusing on the relationship between postmenopausal hormone therapy and the histological subtype of breast cancer.[853] The main conclusions are an increased risk of tumors of the lobular type or with a lobular component, and an increase in the odds

ratio that appeared with 3 or more years of use with combined estrogen-progestin therapy. Similar results were reported in Swedish and German case-control studies.[847, 854]

It is a logical conclusion that the results reflect increasing proportions of hormonally sensitive tissue. Lobular tumors are characteristically estrogen receptor positive and more hormone sensitive. If hormone therapy is affecting pre-existing tumors, one would expect the hormonally sensitive lobular cancers to be detected earlier.

- ***Case-control and cohort studies have reported an increased risk of breast cancer associated with hormone therapy, greater with combined estrogen-progestin.***
- ***The reported increased risks of breast cancer associated with hormone therapy are due to estrogen receptor-positive tumors, perhaps mostly tumors with lobular tissue.***
- ***An increased risk of breast cancer is observed only in current users and is detected relatively rapidly.***

The Women's Health Initiative

The impression gained from the observational data concluding that exposure to estrogen-progestin was more harmful was reinforced when clinicians were confronted by the results in the 2 canceled arms of the Women's Health Initiative.

The second WHI report on breast cancer in the estrogen-progestin arm resulted in little change in the hazard ratios published in the initial report.[855] Invasive breast cancer was increased, 199 cases in the treated group and 150 in the placebo group (HR=1.24; CI=1.01–1.54). The absolute risk was four to six additional cancers per 10,000 women per year. The WHI performed subgroup analyses focusing on how prior hormone therapy use influenced the risk of breast cancer found in the estrogen-progestin trial arm.[856] Prior hormone users totaled 4,311 participants (26%), with 42% reporting less than 2 years of use (17% used hormones 5 to 10 years previously, and 26% more than 10 years before enrolling in the WHI study). Prior users had an increased hazard ratio compared to placebo (1.96; CI=1.17–3.27) in contrast to no increase among never users (1.02; CI=0.77–1.36). The WHI concluded that this difference could reflect an increasing risk with cumulative exposure to hormone therapy.

Many of the factors associated with the risk of breast cancer were slightly but significantly more prevalent in the group of prior hormone users in the WHI, such as younger age, more education, lower body mass, more physical activity, smoking, alcohol use, vasomotor symptoms, and lower bone density. The overall risk of breast cancer in the treated estrogen-progestin group was the same as previously reported by the WHI (1.24; CI=1.02–1.50).[855] However, after adjusting for the multiple factors recognized to influence the risk of breast cancer, *the hazard ratio was 1.20, and no longer statistically significant, CI=0.94–1.53).* The placebo group failed to demonstrate an age-related increase over the duration of the trial, and the treatment group and the placebo group differed in regard to breast cancer risk factors, forcing the investigators to perform multiple adjustments. Why did not the randomization of large numbers avoid this confounding problem? The strength of the WHI conclusion is limited by the fact that the participants in the treatment and placebo groups differed considerably.

The updated results in the canceled estrogen-only arm of the WHI failed to detect an increase in breast cancer risk associated with hormone therapy; indeed, the group adherent to treatment demonstrated a statistically significant reduction in invasive breast cancer (HR=0.67; CI=0.47–0.97).[857] This striking difference between the two canceled arms led

many to conclude that the results indicated an adverse effect of progestin exposure, or more bluntly, that there is no increase in risk with up to 5 years of estrogen-only use in contrast to an increased risk with estrogen-progestin.

It cannot be emphasized too strongly that it is not appropriate to compare the two canceled arms of the WHI and conclude that differences reflect the effects of progestin exposure. Even the WHI investigators cautioned clinicians to avoid comparing the two trial arms because the participants in the two arms differed considerably.[766] In regards to breast cancer risk, the women in the estrogen-only arm had a higher rate of previous exposure to hormones and for longer durations of prior use. It is possible that earlier and greater use of hormone therapy before participation in the study identified those individuals with pre-existing tumors who were then excluded from participation, accounting for the lower incidence of breast cancer in the treated group. ***The breast cancer results in the WHI do not allow us to answer the important question whether exposure to estrogen-progestin has a greater risk of breast cancer or whether pre-existing tumors respond differently to various hormone regimens, accounting for differences in epidemiologic reports.***

Long-Term Hormone Use

A cohort study assessed the risk of breast cancer in postmenopausal Finnish women using oral, transdermal, and vaginal estrogen-only or estrogen-progestin products containing either estradiol or estriol.[858, 859] The use of estradiol for 5 or more years and the use of estrogen-progestin for 3 or more years was associated with a statistically significant increased risk. The risk was similar comparing oral and transdermal therapy. An increase was observed in both localized and metastatic disease. A statistically significant increase was noted with carcinoma-in-situ.

Only 7% of Finnish postmenopausal women use hormone therapy for more than 5 years. This study reported an increase in breast cancer risk in this long-term user group of women. No increase in breast cancer risk was detected either in association with estriol given orally or with vaginal estrogen products; however, it is inappropriate to conclude that these formulations can be used without risk. To make this conclusion, users and nonusers of these formulations would have to be identical in terms of breast cancer risk factors, and to be comparable in terms of bioequivalent blood levels of estrogen. Only then could a valid comparison be made. This study could not adjust for these factors.

The use of postmenopausal hormone therapy in Finland can be accurately recorded because all treatments must be prescribed and then paid for by the National Social Insurance Institution. However the study is affected by an overwhelming problem: the results are questionable because of an inability to control for confounders. The risk was expressed as incidence ratios, calculated by dividing the observed number of cases by the numbers expected (based on the general statistics in Finland). Therefore, the study could not be controlled for confounders. Risk ratios less than 2.0 can easily be inaccurate because they can reflect confounding factors. It is well-demonstrated that hormone users differ when compared to nonusers in terms of recognized risk factors for breast cancer. The differences also include a greater prevalence of mammography among hormone users. A good example can be found in the report from the Nurses' Health Study that, like the cohort from Finland, indicated an increased risk of breast cancer with long-term users of estrogen-only.[860] The long-term users in the Nurses' Health Study had more bilateral salpingo-oophorectomies, more nulliparity, more benign breast disease, greater alcohol consumption, and they were thinner—all factors that make a comparison of users to nonusers very difficult.

In order to minimize the problem of confounders, the Finnish report argued that "there are no socioeconomic differences between postmenopausal hormone therapy users and the

general population in Finland," citing a previous report. This statement is not totally accurate. The previous report in 1999 was based on population surveys and measured only two things: length of education and rural vs. urban living.[861] A lack of educational differences was present in Finnish women under the age of 55, but older postmenopausal women had more years of education. In addition, there were regional differences at all ages, with the current use of hormone therapy being most common in the Helsinki area (especially among older women). Therefore the 1999 study does not imply a lack of differences in hormone users in Finland, in fact, just the opposite. Age information is not provided in the reports on breast cancer, but the longer-term hormone users were probably an older group of women, and according to the 1999 Finnish report, they do differ when compared to the general population of Finland. Remember, this cohort study is not comparing users with nonusers. It compares users to general population statistics. Therefore, we cannot know whether the results of this study reflect long-term use of estrogen, or whether the results reflect a greater prevalence of risk factors and mammography in the hormone-using group.

- ***The canceled estrogen-progestin arm of the WHI found a slight increase in risk of breast cancer that was significant in the fifth year of exposure.***
- ***The canceled estrogen-only arm of the WHI did not find an increase in breast cancer risk.***
- ***It is not appropriate to compare the two canceled arms of the WHI and make clinical conclusions because the participants in the two arms differed considerably.***
- ***There are no robust data linking long-term estrogen therapy with an increase in breast cancer risk.***

Hormone Therapy and Breast Cancer Outcome

Most of the studies that have examined the breast cancer mortality rates of women who had used postmenopausal hormone therapy at the time of diagnosis have documented improved survival rates.[718, 729, 737, 862–871] Even a study that indicated a reduction in mammographic sensitivity also reported smaller, more differentiated tumors among the users compared with the nonusers.[303] Evidence indicates that hormone users develop smaller, better-differentiated (lower grade) and lower stage tumors, evidence that is consistent with effects on pre-existing tumors and that surveillance/detection bias is not the only explanation for better survival.[843, 870–888] Lower grade tumors are present even when there is no difference in the prevalence of mammography comparing hormone users and nonusers or when the data are adjusted for the method of detection.[866, 868, 879, 889] This is not a totally uniform story in that at least 1 prospective study concluded that estrogen-progestin users had both lower and higher stage and grade tumors.[843] However, nearly all reports indicate that more tumors in hormone users are detected by screening mammography, and when assessing outcomes in all cancers detected by mammography, hormone users have more ductal in situ tumors, more node-negative cancers, smaller tumors, and less invasive disease and, thus, better survival rates.[871, 884]

In contrast, the WHI results in the estrogen-progestin arm indicated an earlier appearance of *worse* tumors than previously reported in case-control and cohort studies. The WHI argued that the results (both the invasive breast cancers and the mammography findings) are consistent with stimulation of growth of established breast cancers (supported by no statistical difference in in situ tumors) but at the same time a delay in diagnosis. This certainly challenges the idea that hormone users have better outcomes because of earlier detection. The WHI suggested that this disagreement could be because of a difference of mammography use in the observational studies. However, as pointed out, even studies that

examine tumor characteristics and outcome in users and nonusers who have equally used mammography, lower grade and lower stage disease with a better outcome is identified in the users. [866, 868, 884] In addition, a prospective cohort study found little impact of hormone use on mammography specificity.[890]

Differing with many reports in the literature, the WHI concluded that their results suggested that invasive breast cancers diagnosed in women who use hormone therapy may have a worse prognosis, basing this conclusion on the differences observed in tumor size and spread of disease. By now, it is well-recognized that the participants in the WHI represent an older postmenopausal population (average age 63 and an average of about 12 years since menopause in the estrogen-progestin arm). This older population is more likely to have pre-existing occult tumors that would become detectable quickly after hormonal stimulation. In addition, breast tissue in older postmenopausal women may respond differently to hormone stimulation than breast tissue in women close to their menopause. It is possible that the WHI results reflect this older population that might have occult tumors that are in fact larger and more prone to respond to hormonal stimulation than tumors in younger women.

There is another problem with the WHI data on tumor size and stage of disease. The WHI reported that estrogen-progestin users had slightly larger tumors (an average of 2 mm) and less localized disease. However, there are reasons, both apparent and not apparent, why the WHI data disagree with the great bulk of the literature. First, no nodes were examined in nearly 10% of the subjects who developed breast cancer; information was missing on node involvement in 4.0% of the treated group and 4.7% of the placebo group and on tumor size in 6.5% of the treated group and 6.0% of the placebo group. Because case numbers were not large, a change in a few cases could change the conclusions. Next, the WHI investigators assumed that tumors less than 1 cm in size with no node information should be classified as localized disease, and those greater than 1 cm in size were not staged. According to the U.S. SEER data, breast tumors 1 cm in diameter have a 20% incidence of positive nodes![891] Finally, there is a major problem that is hard to explain. The WHI conclusion of less localized disease in hormone users is based on, once again, a difference in the placebo group, a surprisingly low incidence of positive nodes. There is a linear relationship between tumor diameter and the percent of cases with positive nodes.[891] The average tumor size in the placebo group was 1.5 cm, and according to U.S. data, this should give about a 25% incidence of positive nodes, not the 15.8% that was reported.[891–894] Why is the placebo group different? Whatever the reason, it influences the conclusion.

The Women's Health Initiative investigators reported health outcomes at 3 years after the estrogen-progestin arm of the clinical trial was canceled.[895] This follow-up report included 15,730 participants who were followed from July 2002 to April 2005 (95% of randomized women). No increase in cardiovascular events, including venous thrombosis, in the women treated with estrogen-progestin was observed in the follow-up period. There were 79 cases of invasive breast cancer in the treated group in the follow-up period compared with 60 in the placebo group, a difference that gave a hazard ratio of 1.27, but it was not statistically significant. There was no difference in cases of colorectal cancer or fractures and a nonsignificant decrease in endometrial cancer in the treated group. There was a greater rate of mortality in the treatment group in the follow-up period, but this difference was small and not statistically significant.

A strong case can be made that the epidemiologic data on breast cancer reflect an impact of hormone therapy on pre-existing tumors. This WHI follow-up report concluded that the trend of increasing breast cancer during the trial period did not extend into the follow-up period. This is the pattern one would expect if hormone therapy is causing earlier detection of preexisting tumors only in current users. An adverse interpretation of the WHI follow-up report was based on a small and nonsignificant increase in mortality in the treated group, because of deaths from various cancers, but most prominently lung cancers (discussed later).

- ***Lower grade and stage of disease account for the better survival rates in hormone users who develop breast cancer compared with nonusers.***
- ***Hormone users who develop breast cancer usually have earlier detection of their tumors.***
- ***The survival and detection results are consistent with an impact of hormone therapy on pre-existing tumors.***
- ***The WHI results regarding grade and stage of disease disagree with uniformly contrary reports in the literature.***

The Impact of Hormone Therapy on Breast Density

An increase in mammographic breast density occurs at a greater rate in postmenopausal hormone users who develop mastalgia.[896] According to the PEPI and WHI clinical trials, about 25–30% of women receiving combinations of estrogen and progestin develop breast tenderness.[896, 897]

The subject of greatest clinical relevance, in our view, is the connection between breast tenderness and breast density, and ultimately the impact on the risk of breast cancer. Although the increase in new-onset breast tenderness is associated with a greater increase in mammographic breast density compared with women who do not develop mastalgia, it is worth noting, that one study asking whether hormone-induced changes in breast density increase the risk of breast cancer could find no evidence that this was so.[898]

A report from the WHI concluded that new-onset breast tenderness in the canceled estrogen-progestin arm of the WHI was associated with an increased risk of breast cancer (HR=1.48; CI=1.08–2.03) when compared to women who did not develop mastalgia.[897] The WHI report gave the impression that this linkage was of sufficient magnitude that it should be a major clinical concern, but there are reasons to place it in a different perspective.

The WHI compared the numbers for sensitivity, specificity, and positive predictive value with the Gail model, implying clinical worthiness by their similarity. However, the positive predictive value of new-onset mastalgia in the WHI was 2.7%, meaning that 97.3% of the women would be expected not to develop breast cancer. Furthermore, in the references cited by the WHI to buttress their use of the Gail model, one actually concluded that the Gail model is not particularly sensitive in identifying individuals at risk and the other was a study of false-positive mammogram screening.[899, 900] A positive predictive value of 2.7% is not clinically strong.

The WHI report appropriately acknowledges the increase in breast tenderness noted in some postmenopausal women treated with hormone therapy and the link between mastalgia and an increase in mammographic density. But then the WHI report discusses these subjects as if it is accepted that the hormone-induced increase in density is associated with an increased risk of breast cancer. Let's consider these issues.

Increased density impairs the detection of breast masses.[901] A failure to detect masses because of high density causes an increase in interval cancers (cancers that present between mammographic screenings, in other words, cancers diagnosed after a negative mammogram).[902] Difficulties in reading high-density mammograms also produce false-positive recalls (patients who are recalled for assessment and found not to have cancer). Being recalled for reassessment after an initial mammogram is a cause of significant psychological stress.[903] In addition, at least 25% of the overall cost of mammographic screening in one U.S. program was attributed to investigations of false-positive readings.[900]

These two problems, an increase in interval cancers (a decrease in mammographic sensitivity) and an increase in false-positive recalls (a decrease in mammographic specificity), are consistent with a possible decrease in the detection of cancer. Thus, the concern with dense breasts in postmenopausal women is a reduced quality of mammograms that would decrease the ability to detect early breast cancers.

Factors that are associated with greater breast density are nulliparity, older age at first birth, and current use of postmenopausal hormone therapy.[904] Mammographically dense breasts reflect a high proportion of stromal, ductal, and glandular tissue, associated with epithelial and stromal cell proliferation.[905] The likelihood of dense breasts in hormone nonusers decreases with advancing age and increasing body weight as glandular tissue is replaced by fat.[904] The link with nulliparity supports the contention that a full-term pregnancy early in life produces a change in structure in the breast that persists throughout life and is associated with resistance to proliferation.

Assessing the impact of breast density on the risk of breast cancer is complicated by two factors that produce heterogeneous data: (1) Results from programs with biennial screening are less favorable when compared with annual screening and the available data are derived from both, and (2) Results from recent years are better, reflecting improvements in technology. Nevertheless, high breast density (75% dense) on mammography in *hormone nonusers* is reported to be associated with a 4- to 6-fold increased risk of breast cancer.[906–909] ***Although hormone therapy increases breast density in some women, it is not certain that the short-term increase in density with hormone therapy changes an individual's risk of breast cancer.***

More current users of hormone therapy have dense breasts than nonusers.[904, 910–913] In women younger than age 55, it is difficult to find any differences between hormone users and nonusers.[178] The impact is essentially limited to women older than age 55. The effect of hormone therapy on breast density occurs rapidly; thus, duration of use has no effect.[178] In the PEPI 3-year randomized trial, almost all increases occurred within the first year, with an increase in breast density observed in 8% of estrogen users and 19–24% of estrogen-progestin users and only 2% in the placebo group.[914] The users of estrogen-progestin combined regimens had a greater risk of developing denser breasts compared with estrogen-alone treatment (7–13-fold greater in the PEPI trial with no differences observed comparing medroxyprogesterone acetate to micronized progesterone).[914] In the WHI, estrogen-progestin use increased density an average of 6.0% of users in the first year, with attenuation in the second year to 4.9%.[915] In careful studies, the daily, continuous, combined estrogen-progestin regimens have been reported to have a greater effect than sequential regimens, with an increase in density occurring within the first months of treatment and then maintained with no change.[302, 303, 916–918]

Therefore, hormone therapy increases breast density mainly in older postmenopausal women, more women respond to combined estrogen-progestin regimens (especially the daily, continuous programs), and the effect occurs within the first months of use and remains stable with no changes or some attenuation with increasing duration of use. However, this effect is only seen in at most about 25% of estrogen-progestin users, but usually around 10%, of hormone users; indeed, it should be emphasized that most women do not respond in this fashion.

Overall, studies have suggested that hormone users experience a decrease in mammographic sensitivity with a lesser impact on specificity (false-recall rates). However, the studies are based on small numbers of interval cancers, and it is uncertain how real or how large this effect is because of the difficulty in controlling for confounding factors (for example, age, age at menopause, and time since menopause). If the effectiveness of breast cancer screening is reduced by postmenopausal hormone therapy, one would expect an adverse impact on breast cancer mortality. Instead, a study that indicated a reduction in mammographic sensitivity also reported smaller, more differentiated (Grade I) tumors among the users compared with the nonusers,[303] and most of the studies that have exam-

ined the breast cancer mortality rates of women who had used postmenopausal hormone therapy, as reviewed above, have documented improved survival rates.

If breast density and breast cancer were a reflection only of hormone exposure, a strong preponderance of estrogen receptor-positive tumors would be expected in women with increased density. However, women with increased density demonstrated equal increases in risk for both estrogen receptor-positive and estrogen receptor-negative breast cancers, and 1 study found a preponderance of estrogen receptor-negative tumors in hormone users with dense breasts.[919, 920] This suggests that other factors besides hormone exposure are involved in the relationship between density and breast cancer. For example, there is an association between breast density and family history of breast cancer, indicating an underlying genetic basis for both density and breast cancer.[921] In a case-control study that assessed the relationship between hormone therapy and breast density, leaner women were more likely to increase their breast densities with hormone therapy, but there was no association between the response to hormones and family history, late age at first birth, or history of benign breast disease; the study concluded that recognized risk factors influenced the response to hormone therapy only to a minor degree, suggesting again that unknown genetic factors are involved.[922] In a study designed to correlate histologic findings in dense breast tissue, an increase in fibrous stroma and type 1 lobules was observed to be more prevalent in hormone users, but these changes were also present in non-hormone users, and overall there was no statistically significant difference between histologic features and breast density in women undergoing mastectomy for breast cancer.[923] If breast density in postmenopausal women were strictly related to the hormonal environment, a drastic reduction in the estrogen milieu of the breast should have a salutary impact on density. The MAP1 randomized trial evaluated the effect of letrozole treatment on breast density; no effect of aromatase inhibitor administration on breast density was observed despite one year of treatment.[924]

The increase in breast density associated with postmenopausal hormone therapy appears in some studies to be a transient, reversible change, a change not consistent with a *persistent* effect on cellular proliferation. After discontinuing hormone therapy, it was reported that breast density rapidly decreases so that former users do not display an increase compared to never users.[904, 913, 925, 926] In a retrospective analysis, regression of hormone-induced abnormalities was found to occur within 2 weeks of cessation of treatment.[926] Similar results were observed in a prospective study observing a reduction in density 3 weeks after stopping treatment.[927] However, in a large randomized trial of 1,704 women age 45 to 80, although suspension of hormone therapy for 1 or 2 months produced small but significant decreases in density, mammography recall rates of 10% to 12% were not affected.[928] And a nonrandomized before and after study (but of only 47 women) detected no significant changes in mammographic density after a 4-week cessation of hormone therapy.[929] Besides discontinuing hormone therapy, another approach is to consider lower doses of hormone therapy; there is some evidence that low-dose treatment has little effect on breast density.[930] The combination of the levonorgestrel IUD system and estrogen therapy in postmenopausal women does not increase breast density.[931]

- ***Some women develop an increase in breast density with the current use of postmenopausal hormone therapy, more often associated with continuous, combined use of estrogen-progestin regimens.***
- ***The older a postmenopausal patient is, the greater the risk of developing an increase in breast density with hormone therapy.***
- ***There is no strong evidence to indicate that new-onset tenderness or the increase in density with hormone therapy changes an individual's risk of breast cancer.***

- ***The evidence to support a recommendation that hormone therapy should be discontinued for several weeks prior to mammography in women who have dense breasts is mixed, and the only randomized study found no impact of suspension of treatment on a mammography recall rate of about 10%. Nevertheless, this is a reasonable recommendation.***
- ***In women who develop breast tenderness and/or an increase in density with postmenopausal hormone therapy, consider a dosage reduction of the administered hormones.***

The Effect of Hormone Therapy on Mammographic Screening

The impact of screening mammography has been established by multiple randomized trials: about a 22% reduction in breast cancer mortality in women 50 years and older, and 15% in women between ages 40 and 49. But at the same time it is recognized that it is difficult for mammography to detect noncalcified masses, especially in dense breasts. This sensitivity problem has been improved by digital mammography, but not eliminated.

Does the hormonal effect on breast density impair mammographic screening? In other words, is there an increase in interval cancers and false-positive recalls in postmenopausal hormone users? In a review of seven studies, there were relatively few interval cancers in the user groups, nevertheless six of the seven studies reported decreased mammographic sensitivity in hormone users with a small increase in interval cancers in users compared with nonusers.[932] Excluding women under age 50, the relative risk for an interval cancer was summarized as 1.7 (CI=1.2–2.4). American, Scottish, and Australian studies have indicated a 5–20% decrease in mammographic sensitivity in hormone users who have dense breasts.[933–936] A Finnish study concluded that women with the most dense breasts and using hormones had the highest relative risk of breast cancer, but this conclusion was based on only 4 cases of cancer in women with dense breasts.[937]

The risk of false-positive recall (mammographic specificity) was investigated in five studies.[933, 937–940] The rate of false-positive recall in nonusers ranged from 2.1% in the U.K. to as high as 14.7% in an American program; four of the five studies found a slight increase in the risk of false-positive recalls in hormone users. In a French study, mammographic sensitivity was reduced from 92% to 71% in users because of an incidence of interval cancers that was 3.5 times that of nonusers within the first year of the initial exam, and 1.7 times greater during the following 2 years.[938] Most of the hormone users were on combined estrogen-progestin regimens. The false-positive recall rate was only slightly higher, 3.3% in users and 2.8% in nonusers. A prospective study of screening mammograms from Massachusetts General Hospital concluded that recall rates were essentially the same comparing hormone users and nonusers, and that hormone therapy rarely causes a diagnostic dilemma.[939] However, in New Hampshire, increased breast density and use of hormone therapy independently increased the need for supplemental imaging.[940]

Treatment with either estrogen alone or estrogen-progestin in the WHI was associated with more abnormal mammograms compared with placebo.[857, 941] The difference was about 12% with estrogen-progestin and 8% with estrogen alone, but remember the characteristics of the participants differed in the two arms of the clinical trial. Nevertheless, the higher prevalence with estrogen-progestin is consistent with the other studies in the literature. Overall, the studies have suggested a decrease in mammographic sensitivity with a small impact on specificity (false-recall rates).

Adding Ultrasound to Mammography

The American College of Radiology conducted a prospective, multicenter, randomized trial in 21 centers in the U.S., with 2,725 women, designed to validate the performance of screening ultrasound in conjunction with mammography in women with dense breasts and at high risk for breast cancer.[942] Each patient underwent mammography and ultrasound in a randomized sequence. 40 cases of cancer were diagnosed, 12 on ultrasound alone, 12 on mammography alone, 8 suspicious with both techniques, and 8 with negative exams. Adding ultrasound yielded an additional 4.2 cancers per 1,000 high risk women. The false-positive rate for mammography alone was 4.4%, for ultrasound alone, 8.1%, and for combined mammography plus ultrasound 10.4%. ***Thus adding ultrasound to mammography screening in high-risk women with dense breasts improved the sensitivity of screening, but increased the rate of false-positive examinations. Breast cancer mortality was not an endpoint in this trial, but the fact that the cancers detected by ultrasound are usually asymptomatic, node-negative, and not detected by mammography should yield a reduction in mortality.***

Ultrasound screening can detect cancers not seen on mammography and its performance is not affected by dense breast tissue. Adding ultrasound to a screening program seems straightforward, but its impact on mortality reduction has not been measured in a large trial. In the single center studies of screening ultrasound that have been published, cancers have been found only by ultrasound, and most are small, early stage tumors. An Italian multicenter study reported that 29 cancers were found by ultrasound in 6,449 women with dense breasts and negative mammograms.[943] Nevertheless, a majority of facilities do not offer screening ultrasound because of a lack of qualified personnel and standardized protocols.

The problem with all screening methods is a substantial rate of false positives. In the American study, 91.4% of suspicious ultrasound findings were benign.[942] The positive predictive value for ultrasound was only 8.6%, but the value for mammography was 14.7%. Remember that ultrasound tends to find earlier tumors. The crucial question is how many false positives are worth the gain in additional cancer diagnoses. In the American study, the gain was an additional 29% (the number of cancers detected only by ultrasound). In women with elevated risks, this seems worthwhile. ***Women at high risk probably have a greater fear of diagnosing breast cancer late than of a false positive.***

Combining MRI with mammography yields a very high sensitivity, and this is now recommended for women at very high risk for breast cancer. MRI, of course, is the most sensitive technique, but it is very expensive, requires the intravenous injection of contrast, and isn't always tolerated by patients. Ultrasound has the advantage of being less expensive, easily tolerated, and widely available. Thus the combination of ultrasound and mammography seems best for women of intermediate risk. Ultrasound has a disadvantage of not detecting ductal carcinoma in situ, which is detected by mammography and MRI.

Mammography Screening—Conclusion

The final protocol for the best screening use of the three modalities, mammography, ultrasound, and MRI, will also require consideration of cost. The total cost is a complex summary of the technology, the time consumed, the increase in patient anxiety and discomfort, and the expense of additional testing because of false positives. ***Nevertheless, the evidence now seems sufficient to individualize decision-making and to recommend more than the single technique of mammography for high risk patients (defined as a combination of factors that produces a 3-fold increase in risk), especially in women with dense breasts. Thus far, over 90% of cancers detected only on***

ultrasound have been in women with dense breasts. Therefore, it seems advisable to add ultrasonography to mammography for hormone users who develop dense breasts and the density persists despite a short period without hormone therapy. In addition, digital screening mammography is superior to conventional film screening for women with dense breasts.[944]

- Hormone users who develop an increase in breast density have a small decrease in mammography sensitivity and an increase in false recalls.
- An adverse impact of this increase in breast density is not apparent in breast cancer mortality statistics in hormone users.
- Digital mammography is preferred for women with dense breasts.
- Ultrasonography should be added to mammography screening in hormone users who develop dense breasts with no regression after a short period without hormone treatment.

The Effect of Progestins in Combined Regimens

Recent studies support the conclusion that the relative risk of breast cancer is higher in users of combinations of estrogen and progestin. The effect is confined to estrogen receptor-positive, progesterone receptor-positive (ER+/PR+) tumors, mainly lobular cancers. ***What if this conclusion reflects early detection of better differentiated tumors, a consequence of a favorable response of pre-existing tumors to estrogen-progestin exposure?***

Almost every study that has reported an increase in breast cancer risk with postmenopausal hormone therapy has found the increase within a few years. Remember that although the doubling time of breast cancer is very variable, in general, a tumor doubles in size every 100 days. Thus, it takes a single malignant cell approximately 7 years to become detectable by mammography and 10 years to grow to a clinically detectable 1-cm mass.[945] The rapid finding of an increased risk within a few years suggests that the epidemiologic studies are detecting pre-existing tumors.

Older studies on the hormone receptor content in breast cancers diagnosed in hormone users were limited by small numbers. Furthermore, recent studies examining receptor status are likely to be more accurate in that receptor status assays have improved. In the Nurses' Health Study, the use of postmenopausal hormone therapy has been associated with a significant increase in ER+/PR+ breast cancer, but not receptor-negative disease.[946] This relationship, greatest in lean women, was stronger and observed sooner with the use of estrogen-progestin, a significant increase with 5 or fewer years of combined use and no increase with 5 or fewer years of the use of estrogen alone. In a cohort of women from the area of Lund, Sweden, an increased risk of breast cancer was reported only in users of continuous combined estrogen-progestin, and this increase was observed within 2 years of use.[947] Other epidemiologic studies have reported a similar greater risk in current users of continuous, combined estrogen-progestin, concentrated in estrogen receptor-positive disease.[839, 842, 845, 948, 949] Indeed, use of hormone therapy is the greatest predictor of estrogen receptor-positive disease.[950] In a retrospective study in the Northern California Kaiser program, only the current use of combined estrogen and progestin increased the odds of estrogen receptor-positive tumors.[951]

E3N is a French prospective cohort study initiated in 1990, which concluded that it would be preferable to use progesterone or dydrogesterone because estrogen used with these 2 progestins was not associated with an increase in the relative risk of invasive breast cancer.[952, 953] For any given progestin, the route of administration of estrogen (oral or transdermal) had no effect on relative risk. A statistically significant increase in relative risk was

associated with estrogen alone, RR=1.29 (CI=1.02–1.65) and with progestins other than progesterone or dydrogesterone, RR=1.69 (CI=1.50–1.91). The increased relative risks seemed to rapidly dissipate after discontinuation of treatment; although this analysis was initially limited by small numbers, it was confirmed in a later follow-up.[954]

The French study argued that their results indicate that the "natural" progestins are safer than "synthetic" progestins. But to accurately differentiate among various agents one would have to be certain that the doses administered represented bioequivalent doses in terms of target tissue impact, something that would be difficult to do. Let's focus on the rapidity at which cases of breast cancer were identified. The use of estrogen combined with "other progestins" had an increased relative risk even with less than 2 years of exposure, RR=1.37; CI=1.07–1.72. In their earlier report, an increased relative risk was even apparent with less than 1 year of exposure of estrogen combined with synthetic progestins.[952] ***The results in the French study could be due to earlier detection of pre-existing tumors, an effect accelerated by specific progestins with greater potency.***

A follow-up report from the French study indicated that the increased risk of breast cancer was evident only in women with recent use of hormone therapy and not in past users.[954] Furthermore, the risk with short-term use was confined to hormone use with synthetic progestins for 2 or fewer years in the 3-year period immediately following menopause. With longer duration of hormone use, the risk was apparent even in those who initiated treatment years after the menopause. The logical conclusion is that this striking finding with short-term use early in the postmenopausal years reflects an impact of hormone therapy on pre-existing tumors.

Molecular biology studies provide support for a favorable effect of estrogen-progestin exposure. In-vitro studies using microarray analysis have profiled the gene network both up- and down-regulated by estrogen.[955] Genes that are up-regulated by estrogen are down-regulated by estrogen-progestin treatment.[956] Comparing hormone users and nonusers, 276 genes were activated by hormone exposure (11 of the 13 women used estrogen-progestin, 2 estrogen alone). All patients in this cluster were free of recurrence 5 years after diagnosis. In a cohort of 131 women, those patients exhibiting the gene profile associated with estrogen-progestin exposure preferentially benefited from tamoxifen treatment.[956] This Swedish study found that estrogen-progestin use altered the gene expression profile only in estrogen receptor-positive cancers. Among the genes regulated, many were involved in either DNA repair or cell-cycle regulation. For example, the p63 gene, involved in tumor differentiation, was over-expressed in estrogen-progestin users. Previous reports have found this gene to be expressed in normal tissue, partially expressed in ductal hyperplasia, and not expressed in invasive cancers.[957]

It is well recognized that early pregnancy produces a mammary gland that is resistant to carcinogenesis. In rodents, this is accomplished by treatment with estrogen plus a progestin. The refractory phenotype that is produced is associated with progestin-induced changes in gene expression involved in cell proliferation.[958]

The progesterone receptor is induced by estrogens at the transcriptional level and decreased by progestins at both the transcriptional and translational levels (probably through receptor phosphorylation).[959] The progesterone receptor (in a fashion similar to the estrogen receptor) has two major forms, designated the A and B receptors.[960] The two forms are expressed by a single gene, a consequence of transcription from distinctly different promoters, in a complex system of transcription regulation.[961] Progestational agents can elicit a variety of responses determined by target tissue production and activity of the two receptor forms with dimerization as AA and BB (homodimers) or AB (heterodimer).

PR-A and PR-B are expressed in varying amounts in breast cancer and endometrial cancer cell lines. Studies indicate that the two receptors can be regulated independently;

e.g., the relative levels differ in endometrium during the menstrual cycle.[962] Tissue specificity with the progesterone receptor is influenced by which receptor and which dimer is active, and in addition, the transcriptional activities of PR-A and PR-B depend on target cell differences, especially in promoter context. In most cells, PR-B is the positive regulator of progesterone-responsive genes, and PR-A inhibits PR-B activity. Thus, repression of human estrogen receptor transcriptional activity (as well as glucocorticoid, mineralocorticoid, and androgen transcription) is dependent on the expression of PR-A.[963, 964] The PR-A and PR-B progesterone receptors have different molecular functions, affecting different genes, and, therefore, target tissue response to progesterone will be influenced by the differential expression of each receptor and the ratio of their concentrations, as well as the target tissue context of adaptor proteins.[965, 966]

PR-A positive breast cells exhibit more aggressive growth, and PR-A isoforms are dominant in the absence of progesterone. Even without its ligand, PR-A can exert gene regulation in estrogen receptor-positive breast cell lines.[967] In the absence of progesterone, PR-A up-regulates genes known to be associated with invasion and poor prognosis, including those genes that provide resistance to apoptosis. In the presence of progesterone, PR-B is a stronger regulator of gene transcription. The breasts of normal women express equal amounts of PR-A and PR-B.

PR-A excess and breast cancer are linked! Estrogen receptor-positive tumors with a higher rate of recurrence are rich in the PR-A isoform.[967] As breast cancers become less differentiated, metastatic tumors become dominated either by PR-A or PR-B. PR-A rich tumors with a low PR-A:PR-B ratio do poorly and respond less well to tamoxifen.[968] PR-A excess is also present in breast tissue from women with the *BRCA* mutations.[969]

Thus, progesterone receptors are not just passive markers of estrogen activity. ER+/PR+ tumors are well differentiated and have better outcomes. In the absence of progesterone, the unliganded PR-A can adversely influence the cell biology of estrogen receptor-positive tumors. Cells rich in PR-A are more likely to be invasive, poorly differentiated, and aggressive. In monkeys, the breast levels of PR-A were unchanged after 3 years of treatment with conjugated estrogens alone.[970] Treatment with conjugated estrogens and medroxyprogesterone acetate produced a decline in PR-A levels, producing a 10-fold beneficial change in the PR-A:PR-B ratio. ***It is possible that the exposure of an estrogen receptor-positive tumor to estrogen-progestin treatment can prevent an unfavorable PR-A:PR-B ratio, promoting the beneficial actions of PR-B. In addition, progestins may activate androgen receptors, a factor that has been demonstrated to inhibit growth and cause apoptosis in breast cancer cells.***[971, 972]

In a series of in vitro experiments, estradiol and its catechol metabolites induced neoplastic transformation in human breast epithelial cells.[973] This, of course, would be consistent with a genomic impact of estrogen that initiates breast cancer, an effect earlier than a promoting influence on already-established cancers. However, it is difficult to transfer in vitro effects on cell lines to the in vivo situation, and this is especially true with breast tissue that is bathed in a complicated and large collection of stimulating and inhibiting substances. Furthermore, the cells in the in vitro experiments were negative for both estrogen and progesterone receptors. These receptor-negative malignant cells may well be something different than the receptor-positive tumors associated with postmenopausal hormone therapy.

Favorable effects of progestin exposure are reflected in the studies of the association between endogenous hormone levels and the risk of breast cancer. A pooled analysis of nine prospective studies of postmenopausal women concluded that the risk of breast cancer increases with increasing concentrations of all endogenous estrogens and androgens, including estradiol, estrone, estrone sulfate, androstenedione, dehydroepiandrosterone, dehydroepiandrosterone sulfate, and testosterone.[974] The overall increase in breast cancer risk was about 2-fold comparing the lowest endogenous levels in postmenopausal women with the highest levels. Women who ultimately develop breast cancer do not have differ-

ent blood levels of progesterone.[826] Postmenopausal women who are overweight have an increased risk of breast cancer, and an analysis that adjusted for the increase in circulating estrogens associated with obesity concluded that the increasing risk with increasing body weight is the result of the increase in estrogens.[975]

In contrast to the endometrium, epithelial cell proliferation in the normal breast and in estrogen receptor-positive tumors reaches its peak during the progesterone-dominant luteal phase of the menstrual cycle.[976–979] This observation has been the driving force behind the argument that progesterone (progestins) is the major hormonal mitogen in the breast. However, studies do not support a major role for an adverse progestational influence. In animal models, it is estrogen that is the major inducer of proliferation and not progesterone. Indeed, evidence indicates that with increasing duration of exposure, progesterone can limit breast epithelial growth as it does with endometrial epithelium.[980–982] In vitro studies of normal breast epithelial cells reveal that progestins inhibit proliferation.[983] The story with human breast tissue specimens removed after the patients were treated with estradiol and progestin is more confusing, indicating on one hand that progestins inhibit in vivo estradiol-induced proliferation,[980, 982, 984] and on the other hand, markers of epithelial and stromal cell proliferation were higher in women being treated with estrogen-progestin.[176, 985] Nevertheless, progestins have been demonstrated to decrease anti-apoptotic protein expression,[986] and apoptosis in breast tissue is higher in the luteal phase than in the follicular phase.[987]

In the postmenopausal monkey model, greater breast cell proliferation was observed after 30 months of treatment with estrogen-progestin compared with estrogen alone.[988] In this same model, the administration of progesterone produced no differences in proliferation markers, but adding medroxyprogesterone acetate to estrogen increased breast proliferation by about 30% compared with placebo.[989, 990] However, in this 2-month study, the administration of progesterone inexplicably lowered the blood levels of estradiol and estrone by 30–50%, in contrast to no effect with medroxyprogesterone acetate. Thus the tissue results may reflect estrogen differences, not progestin differences. Nevertheless, the monkey experiments do not detect a beneficial effect of medroxyprogesterone acetate; a 2-year monkey study recorded lower levels of p53, a tumor suppressor gene, and lower levels of caspase-3, an enzyme involved in apoptosis.[991]

A prospective study of premenopausal women in Italy found that higher progesterone levels in the luteal phase were associated with a reduction in breast cancer risk.[992] However, a nested case-control analysis based on the Nurses' Health Study could find no influence of progesterone levels on breast cancer risk.[993] In this study, women with the highest levels of total and free estradiol in the early follicular phase had about a 2-fold increase in breast cancer risk, predominantly ER+/PR+ tumors, and a similar increase in risk was associated with higher levels of total and free testosterone and androstenedione. There was no association with luteal estradiol levels, and the authors speculated that this is because early follicular phase levels reflect non-ovarian target tissue levels, and also that progesterone down-regulation of estrogen receptors may occur in the breast during the luteal phase.

Although it is not a uniform story, it is possible that favorable breast tissue effects of progestins translate into better differentiation and earlier detection of pre-existing tumors. Impressive supporting data can be found in two important American studies. A retrospective cohort study in the Southern California Kaiser program found a reduction in breast cancer case mortality that was significant only among women with breast cancer who were users of estrogen-progestin, not with estrogen alone.[994] An increase in lower grade, lower stage, estrogen receptor-positive cancers was found only in current users of estrogen-progestin in a study remarkable for its size, 374,465 women screened in six U.S. mammography centers.[843] These data are consistent with a beneficial effect of estrogen-progestin treatment.

Breast cancer mortality was recorded in the Collaborative Breast Cancer Study Cohort, a prospective cohort of 12,269 postmenopausal women from Wisconsin, Massachusetts, and New Hampshire.[995] Women were followed for an average of 10.3 years after breast cancer diagnosis. After adjusting for BMI, smoking, and history of mammography screening, compared with nonusers, mortality from breast cancer was lower among current users of estrogen-progestin, and an even greater effect, a 40% reduction, with 5 or more years of use. These are striking data. The strength of the study is the large size of the cohort. Indeed, this is the strongest evidence thus far published that the use of estrogen-progestin is associated with the development of less aggressive breast cancers. Even in studies that adjusted for the prevalence of mammography screening, breast cancers in hormone users were smaller, had fewer positive axillary lymph nodes, and were of lower grade disease.

The following evidence supports a beneficial impact of progestins on pre-existing tumors:

- ***An increase in estrogen receptor-positive tumors is seen sooner with estrogen-progestin treatment, and greater risk is observed with continuous, daily estrogen-progestin use.***
- ***Genes up-regulated by estrogen are down-regulated by estrogen-progestin therapy.***
- ***Genes that are activated by estrogen-progestin are involved in DNA repair and cell cycle regulation.***
- ***Progestins decrease breast tissue levels of PR-A, causing a beneficial change in the PR-A:PR-B ratio that is associated with better differentiation and outcome.***
- ***A reduction in breast cancer case mortality has been reported with estrogen-progestin use, and not with estrogen alone.***

Prevalence of Breast Cancer and Hormone Therapy

Multiple reports have documented a U.S. decline in breast cancer incidence that paralleled the decrease in use of postmenopausal hormone therapy that followed the publications from the Women's Health Initiative.[996–999] A similar decline was documented in France, Scotland, Switzerland, and Australia.[1000–1003] The decline was partially influenced by a decrease in screening mammography in the U.S., but the correlation with hormone use exists even when the examined population includes only women screened with mammography. A similar pattern was evident in the WHI in the years following the cancellation of the estrogen-progestin arm as well as among the women in the observational arm, despite no change in the frequency of mammography in the WHI population.[1004]

This decrease in prevalence is consistent with the uniform findings in case-control and cohort studies of an increase in breast cancer risk only in current users, with a rapid reduction after cessation of treatment. An impact on existing tumors is supported by the other side of the coin, apparent in breast cancer statistics derived from the area around Geneva, Switzerland. Beginning in 1997, the peak breast cancer incidence in the Geneva area moved to a younger group of women (ages 60–64), with an increase occurring only in Stage I and Stage II disease with estrogen receptor-positive tumors in hormone users.[1005]

A 3-year follow-up from the WHI was the first of many anticipated follow-up reports from the canceled arms of the WHI.[895] The trend of increasing breast cancer during the trial period did not extend into the follow-up period. Thus it is very likely that the WHI results agree with the national data, and furthermore, the absence of a trend for increasing cases of breast cancer is the pattern one would expect if hormone therapy is causing earlier detection of pre-existing

tumors in current users. This agreement with national data was confirmed when the prevalence of breast cancer in the estrogen-progestin arm and the observational arm was reported by the WHI, indicating a marked decline after cancellation of the randomized trial.[1004]

The national decline in prevalence and the WHI results are both consistent with an impact of hormone therapy on pre-existing tumors. If hormone therapy is affecting pre-existing tumors, one would expect small, undetectable tumors to stop changing (at least temporarily) when women discontinue hormone therapy. This response would be consistent with the effects being reported, a decrease in estrogen receptor-positive tumors in younger postmenopausal women. The data most likely primarily reflect existing cancers just below the detection limit in 2002 that slowed or stopped their growing. Thus, a serious question is raised: What will the statistical data show in the coming years? Will some of the pre-existing tumors be overcome by body defenses and disappear? Will tumors that emerge later be of more advanced stage and grade disease with poorer outcomes?

- ***A decrease in breast cancer prevalence paralleled the decrease in hormone use following the publicity generated by the WHI publications.***
- ***The decrease in prevalence is consistent with the removal of hormonal effects on small undetectable tumors.***

Hormone Therapy and the BRCA Mutations

Women with either *BRCA1* or *BRCA2* germline mutations are advised to undergo bilateral prophylactic oophorectomy after childbearing because approximately 90% will develop breast or ovarian cancer. This surgery reduces the risk of ovarian cancer by about 90% and the risk of breast cancer by about 50%. These relatively young women must consider the postoperative consequences of surgical menopause in their decision-making. In a cohort of 462 women with *BRCA1/2* mutations from 13 medical centers in North America and Europe, the incidence of breast cancer was compared in 155 of the women who had undergone bilateral prophylactic oophorectomy with 307 women who did not have the operation.[1006] The women who had oophorectomy had a 60% reduction in the risk of developing breast cancer. Hormone therapy of any type did not alter the reduction in breast cancer experienced by the women undergoing oophorectomy. Thus, short-term use (several years) of hormone therapy did not have an adverse effect on the beneficial reduction in breast cancer risk following prophylactic oophorectomy. In a later follow-up of this group of women, 93 (60%) of the women who underwent oophorectomy used hormone therapy.[1007] The average length of follow-up was 2.6 years (more than 5 years in 16%) in the surgically treated group and 4.1 years (more than 5 years in 33%) in the non-oophorectomized group. There was no hint of a difference in breast cancer reduction comparing hormone users and nonusers. The findings were similar in 34 women who used a combination of estrogen and progestin, but the power of this finding was limited by the small number in this category.

A case-control study of 472 postmenopausal women with a *BRCA1* mutation found that women who used hormone therapy after prophylactic oophorectomy, either estrogen only or combined estrogen-progestin, not only did not have an increased risk of breast cancer, but hormone use was actually associated with a decreased risk.[1008] The findings were the same regardless of duration of use or current or past use. The conclusion is encouraging, but limited by the fact that 68% of the tumors in the study were estrogen receptor-negative, making the estrogen receptor-positive tumors (that are more likely to be influenced by hormone use) relatively small in number.

Women who are *BRCA* carriers face difficult decisions. The experience thus far indicates that hormone therapy can be used safely for several years. Continuing follow-up of these patients may extend this period of safety even longer.

- ***Prophylactic oophorectomy in women with BRCA mutations reduces the risk of breast cancer by about 50%.***
- ***Thus far, hormone use after prophylactic oophorectomy has not diminished the beneficial reduction in breast cancer risk.***

SUMMARY—Postmenopausal Hormone Therapy and Breast Cancer

1. The WHI agrees with case-control and cohort studies indicating that current use of hormone therapy has a slightly increased risk of breast cancer.
2. The increased risk is observed sooner with the use of combined estrogen-progestin regimens.
3. The increased risk with hormone therapy is confined to estrogen receptor-positive tumors, mainly lobular cancers.
4. It is still not known whether this finding is due to an effect of hormonal therapy on pre-existing tumors.
5. The epidemiologic data indicate that a positive family history of breast cancer or other risk factors should not be contraindications to the use of postmenopausal hormone therapy.
6. Women who develop breast cancer while using postmenopausal hormone therapy have a reduced risk of dying from breast cancer compared with never users. This is probably because of two factors: (1) increased surveillance and early detection; and (2) an effect on pre-existing tumors so that tumors appear at a less virulent and aggressive stage.

The most important unanswered question in regard to breast cancer is whether postmenopausal hormone therapy initiates the growth of new breast cancers or whether the epidemiologic results reflect an impact on pre-existing tumors. A summary of the wide range of evidence supports a favorable effect of hormone therapy on pre-existing tumors:

1. Epidemiologic studies find an increased risk within a few years of hormonal exposure.
2. Breast cancer associated with estrogen-progestin therapy is estrogen receptor-positive, lower-grade, lower-stage disease with better survival rates.
3. Epidemiologic studies find an increased risk only in current users; 5 years after discontinuation the risk returns to baseline.
4. A recent rapid decrease in breast cancer prevalence coincides with a decrease in the use of postmenopausal hormone therapy.

Postmenopausal hormone therapy may be associated with a small increase in the risk of breast cancer. Of course, even a small increase in risk for breast cancer is frightening for patients to contemplate. It is helpful to remind patients of the risk of lung cancer associated with smoking (a relative risk of 10–20), a risk magnitude that provides perspective on the possible risk associated with hormone therapy. It is also worth pointing out that the reported risk with hormone therapy is even smaller than that associated with recognized risk factors such as a positive family history, being overweight after menopause, and alcohol intake. In our view, because the literature is sufficiently strong, it is appropriate to share with patients an alternative explanation for the epidemiologic reports regarding breast cancer and postmenopausal hormone therapy. It is helpful to emphasize the possibility that the studies reflect an effect of hormone therapy on pre-existing tumors and that hormone users who develop breast cancer have a reduced risk of dying of breast cancer because their tumors are better differentiated, more localized, and smaller. Contrary to the prevailing belief, estrogen-progestin exposure may cause greater differentiation and earlier detection of pre-existing tumors, resulting in better outcomes.

Endometrial Neoplasia

There are two different types of endometrial cancer. The more uncommon form (perhaps 10% to 20%) develops rapidly, usually in older women, with a histologic pattern more characteristic of serous or clear cell carcinomas, in a background of atrophic endometrium. The more common form, endometrioid carcinoma, develops slowly from a precursor lesion in response to estrogen stimulation. This type is less aggressive, better differentiated, and responds to progestational treatment.

Estrogen normally promotes mitotic growth of the endometrium. Abnormal progression of growth through simple hyperplasia, complex hyperplasia, atypia, and early carcinoma has been associated with unopposed estrogen activity, administered either continuously or in cyclic fashion. Only 1 year of treatment with unopposed estrogen (0.625 mg conjugated estrogens or the equivalent) will produce a 20% incidence of hyperplasia, largely simple hyperplasia; in the 3-year PEPI trial, 30% of the women on unopposed estrogen developed adenomatous or atypical hyperplasia.[97, 100, 101] Some 10% of women with complex hyperplasia progress to frank cancer, and complex hyperplasia is observed to antedate adenocarcinoma in 25–30% of cases. If atypia is present, 20–25% of cases will progress to carcinoma within a year.[1009]

Approximately 40 case-control and cohort studies estimated that the risk of endometrial cancer in women on estrogen therapy (unopposed by a progestational agent) is increased by a factor of somewhere from 2 to 10 times the normal incidence of 1 per 1,000 postmenopausal women per year.[1010, 1011] The risk increases with the dose of estrogen and with the duration of exposure (reaching a 10-fold increase with 10–15 years of use, perhaps an incidence of 1 in 10 with very long-term use), and ***lingers for up to 10 years after estrogen is discontinued.***[1012–1014] The risk of cancer that has already spread beyond the uterus is increased 3-fold in women who have used estrogen a year or longer.[1012, 1015] Although most endometrial cancer associated with estrogen use is of low grade and stage, and associated with better survival (probably because of early detection), the overall risk of invasive cancer and death is increased. ***The risk of endometrial hyperplasia and cancer is not reduced by the administration of unopposed estrogen in a cyclic fashion (a period of time each month without treatment).***[1010, 1016]

A short-term study (2 years) indicated that estrogen-only treatment in one-half the usual standard dose of estrogen (in this case, 0.3 mg esterified estrogens) was not associated with an increased incidence of endometrial hyperplasia compared with a placebo group.[1017] In a similar 2-year study, endometrial stimulation with the transdermal delivery of a very low dose of estradiol, 14 μg/day, also did not differ compared with placebo.[1018] However, ***we have learned that long-term exposure to low levels of estrogen can induce abnormal endometrial growth (it just takes longer), and, in our view, lower-dose estrogen therapy requires either endometrial assessment annually or the addition of a progestin to the treatment regimen.*** This is supported by a case-control study from Washington that contained 18 cases and nine controls who had exclusively used only 0.3 mg/day of unopposed conjugated estrogens.[1019] The use of this half-dose estrogen was associated with an overall 5-fold increased risk of endometrial cancer, reaching a relative risk of 9.2 in current users for more than 8 years' duration. Although limited by small numbers, the conclusion is logical and consistent with our understanding of the importance of duration of exposure to any increased level of endometrial estrogen stimulation. In a randomized trial, endometrial hyperplasia was increased after 2 years of treatment with 0.3 mg conjugated estrogens without a progestin.[103]

The risk of endometrial excessive proliferation is reduced by the addition of a progestational agent to the treatment program.[97, 101] Although estrogen promotes the growth of endometrium, progestins inhibit that growth. This countereffect is accomplished by progestin reduction in cellular receptors for estrogen and by induction of target cell enzymes that convert estradiol to the excreted metabolite estrone sulfate. As a result, the number of estrogen receptor complexes that are retained in the endometrial nuclei are decreased in number, as is the overall intracellular availability of the powerful estradiol. In addition, progestational agents suppress estrogen-mediated transcription of oncogenes.

Reports of the clinical impact of adding progestin in sequence with estrogen include both the reversal of hyperplasia and a diminished incidence of endometrial cancer.[1020–1025] The protective action of progestational agents operates via a mechanism that requires time in order to reach its maximal effect. For that reason, the duration of exposure to the progestin each month is critical. Studies indicate that the *minimal* requirement is a monthly exposure of at least 10 days' duration.[521, 1026, 1027] About 2–3% of women per year develop endometrial hyperplasia when the progestin is administered for less than 10 days monthly. Although the older standard method incorporated the addition of a progestational agent for the last 10 days of estrogen exposure, most evidence has argued in favor of 12 or 14 days.

Important unanswered questions are the following: What is the actual incidence of endometrial cancer in very long-term users of postmenopausal hormone therapy, and are there differences among the various regimens and routes of administration? A case-control study from Seattle reported that the use of combined estrogen-progestin (essentially all sequential and oral) for 5 or more years was associated with an increased relative risk of endometrial cancer, even with 10–21 days of added progestin per month.[523] However, the increased risk was confined to those women who had been previously exposed to unopposed estrogen treatment; remember, after discontinuing unopposed estrogen treatment, the risk of endometrial cancer lingers for up to 10 years, even if a subsequent regimen includes a progestin. In the Swedish prospective cohort in Uppsala, a reduced risk of mortality due to endometrial cancer was observed in women receiving an estrogen-progestin combination; however, there were only two deaths, precluding statistical significance.[721] A case-control study from Los Angeles found no increased risk of endometrial cancer with the continuous, combined estrogen-progestin regimen or when at least 10 days of progestin were provided in a sequential regimen.[1027] ***Epidemiologic studies have suggested that continuous, combined estrogen-progestin regimens provide superior protection because long-term sequential regimens still carry a small increase in the risk of endometrial cancer.[124, 1028, 1029] In our view, an annual endometrial biopsy is strongly***

recommended in estrogen users exposed only intermittently to progestin treatment. Any program that differs from the standard regimen is untested by clinical studies of sufficient length and patient numbers and, therefore, requires periodic surveillance of the endometrium.

An attractive idea is that protection against endometrial cancer requires shedding of the endometrium. However, we know that at least one-third and up to one-half of the functioning endometrium is not lost during withdrawal bleeding, and it has not been established that endometrial shedding is essential to protect against cancer.[1030] It is just as logical to believe that prevention of growth with the development of atrophic endometrium is protective. Case-control studies have indicated that not only is the excess risk associated with unopposed estrogen prevented by continuous, combined estrogen-progestin regimens, but with increasing duration of use, the risk of endometrial cancer is lower than that in never users.[1028, 1031] In a small number of women who developed hyperplasia on a sequential regimen, conversion to continuous, combined treatment produced a return to normal endometrium, and in 345 women who completed 5 years of treatment with a continuous, combined regimen not a single case of hyperplasia was detected.[123]

The Women's Health Initiative reported a 21% decrease in endometrial cancers in the canceled estrogen-progestin arm after 5 years of this clinical trial, but this was not statistically significant.[1032] The WHI concluded that daily, continuous, combined estrogen-progestin treatment prevented the increase in endometrial cancer associated with unopposed estrogen. Adenocarcinoma of the endometrium (the cancer most likely to be affected by estrogen-progestin therapy) accounted for only eight cases in the treated group of 8,506 subjects and nine in the placebo group of 8,102 subjects, small numbers that make confident conclusions difficult. The literature on hormone therapy and the risk of endometrial cancer does not suggest that a beneficial reduction in risk with estrogen-progestin combined treatment should be expected within a time period of a few years, and if the estrogen-progestin arm of the WHI had not been canceled, we believe the reduced risk would eventually have achieved statistical significance.

The lowest daily dose of progestin that protects the endometrium has not been established. Currently, the sequential program with conjugated estrogens uses 5 or 10 mg medroxyprogesterone acetate and the combined daily method uses 1.5 or 2.5 mg. A 2-year study has indicated that 1.5 mg medroxyprogesterone acetate combined with 0.3 or 0.45 mg conjugated estrogens effectively prevents endometrial hyperplasia.[103] The dose of norethindrone that is comparable with 2.5 mg medroxyprogesterone acetate is 0.25 mg.[104]

Although the protective effect of progestin is considerable and predictable, it is unwise to expect all patients on estrogen-progestin therapy to never develop endometrial cancer. Appropriate monitoring of patients cannot be disregarded. Although routine assessments are not cost-effective, interventions directed by clinical responses are prudent and necessary. Greater surveillance is warranted in women on sequential estrogen-progestin regimens.

Ovarian Cancer

Prospective cohort studies concluded that the risk of fatal ovarian cancer is increased with long-term estrogen use.[1033–1039] And some case-control studies reported a small increase in risk in ever users that was higher with long duration of use.[1040] By no means is it certain if this association is real. A pooled analysis of 12 case-control studies could find no consistent evidence for an association between ovarian cancer and estrogen therapy.[1041] A meta-analysis concluded that there is a 14% increased risk of ovarian carcinoma among

ever users of hormone therapy and that there is a 27% increase in risk with more than 10 years of long-term use.[1042] However, the small increase in this meta-analysis of case-control studies was subject to multiple potential biases. Among the six studies included in the analysis of duration of use, only one reported a statistically significant increase in risk with 10 or more years of hormone therapy. Another meta-analysis concluded that there was no clear evidence of an increased risk of ovarian cancer with estrogen therapy and no effect of increasing duration of use.[1043]

Individual studies have been hampered by relatively small numbers, but the lack of a uniform and consistent association argues against a major impact of postmenopausal estrogen treatment on the risk of ovarian cancer. In a relatively large case-control study, no indication could be found for an association between postmenopausal hormone therapy and the risk of epithelial ovarian cancer, even with long-term treatment.[1044] Another case-control study reported a slightly increased risk, but it was not statistically significant.[1045] And another case-control study could find no increase in risk with ever use, past use, or long duration of use (and no differences comparing various estrogens and regimens).[1046] In a comparison of users and nonusers of hormone therapy in the state of Washington, the risk of epithelial ovarian cancer was increased among current and recent users of estrogen-only, but not in past users or in users of estrogen-progestin.[1047]

The canceled estrogen-progestin arm of the Women's Health Initiative reported an increase in ovarian cancer that was not statistically significant (Hazard ratio=1.58, CI=0.77–3.24), prompting this statement: "The possibility of an increased risk of ovarian cancer incidence and mortality remains worrisome and needs confirmation."[1032] The Kaplan-Meier curves suggested an increasing effect over time, but this, too, was not statistically significant. There were no differences reported in histologic type, stage, or grade (but the small numbers made it essentially impossible to assess subcategories).

All of the studies found it difficult to control for most of the factors that influence the risk of ovarian cancer. This is because there are multiple factors, and information regarding each factor is not readily available.

Factors That Decrease the Risk of Ovarian Cancer

Use of steroid hormone contraceptives.
Pregnancy and parity; a greater effect with a recent pregnancy and pregnancy at older age.[1048, 1049]
Breastfeeding.[1050]
Hysterectomy and tubal ligation.[1051]
NSAIDs.[1052]

Factors That Increase the Risk of Ovarian Cancer

Increasing BMI.[1053, 1054]
Infertility.[1055]
Caffeine intake.[1056]
Two or more eggs per week.[1057]
Family history of ovarian and breast cancer.[1038]

Mixed Reports on Decreased Risk

Alcohol intake.[1058]

Mixed Reports on Increased Risk

Cigarette smoking.[1059–1061]

Because of the many factors that influence the risk of ovarian cancer, case-control and cohort studies have found it difficult (in fact, impossible) to match cases and controls. Hormone users usually have used more oral contraceptives, have had fewer children, and are more educated and thinner. Adjustments have been made only for major factors, such as oral contraceptive use. The technique of meta-analysis is especially hampered by these confounding issues. The authors of the published meta-analyses[1041–1043] have inappropriately assumed that controlling for risk factors was uniformly accomplished in all studies.

A major problem has been the impact of endometrioid cancers, an ovarian cancer that logically can be expected to be influenced by estrogen therapy. In many of the studies, the overall results are swayed by the increase in endometrioid cancers, a cancer that could originate in hormonally-stimulated endometriosis.[1062] An accurate analysis requires a separate consideration of endometrioid cancers, but this is difficult because the small numbers do not allow effective sub-categorization.

An Australian case-control study reported a statistically significant increase only in the 18 cases with endometrioid cancer.[1063] A Swedish case-control study reported a small but significant increase in risk with unopposed estrogen and with sequential estrogen-progestin, but 49% of the cases were endometrioid cancers.[1040] In a cohort report from the Breast Cancer Detection Demonstration Project, only endometrioid cancer was significantly increased.[1034] In the WHI, there were two endometrioid cancers in the treated group and none in the placebo group.

It should also be noted that in one randomized trial, a case-control study, and two retrospective cohort analyses, no detrimental effect on prognosis after surgery for ovarian cancer could be detected in patients subsequently treated with hormones.[1064–1067]

Overall, there is an indication that ever users of hormone therapy, no matter what formulation, progestin, or treatment regimen, have a small increase in the risk of epithelial ovarian cancers. The data are consistent with a promotional effect on existing malignancies because the risk diminishes after discontinuation of treatment. It is not difficult to review the epidemiologic data and conclude that there is no uniform story, that there are studies with both positive and negative results, and that most of the studies struggled with limited power because of small numbers, and all of the studies are affected with confounding because of the difficulties in assessing and controlling for risk factors. The case-control and cohort studies irregularly controlled for level of education, parity, oral contraceptive use, BMI, tubal ligation, and family history of ovarian and breast cancer (not a single study controlled for all known risk factors!). It is a real possibility that there exists an increased risk for the hormonally sensitive ovarian cancer of the endometrioid type, and studies should carefully segregate this cancer for separate analysis. It is appropriate to emphasize the weak associations and the mixed story, but at the same time the seriousness of the specific relationship dictates that the association between postmenopausal hormone therapy and the risk of ovarian cancer remains an unresolved issue.

Colorectal Cancer

Most, but not all, cohort and case-control studies have reported a significantly reduced risk of colorectal cancer incidence and mortality in users of postmenopausal estrogen and estrogen-progestin.[1068–1076] The effect is greatest in current users and most studies have not indicated an increased effect with increasing duration of use; for example, the Nurses' Health Study (which found a 34% reduced risk in current users) could not demonstrate an added benefit with longer duration of current use.[1077] A reduction in fatal colon cancer has

been documented in current users.[1070, 1078] In addition, there appears to be a reduced risk of polyps, especially large polyps, among current and recent hormone users. A reduced risk of colorectal cancer has also been reported with a high intake of phytoestrogens.[463, 1079]

The canceled estrogen-progestin arm of the WHI reported a statistically significant 44% reduced risk of colorectal cancer achieved with only a few years of estrogen-progestin therapy.[622] In the estrogen-only arm of the WHI, there was no significant difference comparing the treatment and placebo groups.[1080] This result in the estrogen-progestin arm was not without concern, however, in that the treated group had more advanced disease. Indeed, the conclusion was largely because of a difference in localized disease, 10 cases in the treated group and 36 in the placebo group. The results suggest that already present cancers were influenced by hormone therapy to reach a more advanced stage, but that estrogen-progestin treatment reduced the risk of new colonic cancers. The estrogen-only arm of the WHI did not record a difference in colorectal cancer, but remember that this arm had two important problems: a very high drop-out rate and about 6,000 fewer participants. Furthermore, the WHI results must be confined to older postmenopausal women, an age group where carcinogenetic events are likely already to be underway.

One can only speculate regarding the mechanism of this benefit. The estrogen-induced change in the bile (a decrease in bile acids with an increase in cholesterol saturation) favors gallstone formation but may reduce promotion (by bile acids) of colonic cancer. Other possible mechanisms include a direct suppressive effect on mucosal cell growth and an effect on beneficial mucosal secretions. The colon contains only estrogen receptor-β, and the reduction in the risk of colonic cancer associated with postmenopausal estrogen therapy may reflect an antiproliferative activity of the beta estrogen receptor. ***This potential benefit deserves greater attention; colorectal cancer ranks third in women, both in incidence and mortality, and is more prevalent than cancers of the uterus or ovary.***[816]

Lung Cancer

The leading cause of cancer mortality in American men and women is lung cancer; 87% of the deaths occur in smokers and there are twice as many deaths in women as with breast cancer annually.[816] In a post-hoc analysis that combined data from zero to 4 years of follow-up with the treatment period in the canceled estrogen-progestin arm of the WHI, the incidence of non-small-cell lung cancer, the type that accounts for about 80% of lung cancer, was nonsignificantly increased, but the number of deaths and the number of poorly differentiated and metastatic tumors were increased in the treatment group.[1081] The cases were essentially limited to past and current smokers and to women over age 60. Although the WHI was not designed to assess lung cancer and chest imaging was not part of the study protocol, the results are provocative and concerning.

There is reason to believe lung cancer might be a target tissue for estrogen; at the same time, there is evidence to indicate that the impact is not detrimental, but protective. Estrogen receptors are present in normal and non-small-cell lung cells[1082]; however, case-control studies have indicated a decrease in risk for lung cancer, and specifically for non-small-cell tumors.[1083–1087] Two studies even reported a protective effect in hormone users against lung cancer especially in smokers.[1088, 1089] One study reported decreased survival in women with lung cancer who used hormone therapy,[1090] but others did not detect a decreased survival in lung cancer patients with a history of hormone therapy.[1091, 1092] The Nurses' Health Study found an increase in lung cancer mortality in women who underwent early bilateral oophorectomy and did *not* use estrogen.[799] Despite these encouraging findings, there still is concern because gene expression is stimulated in non-small-cell lung cancer cells by

estrogen, and proliferation of these cells is reduced by an estrogen antagonist.[1093, 1094] In the Rancho Bernardo cohort study, there was no significant association between hormone use and lung cancer; however, there was a suggestion that women over age 55 had a small non-significant increase in lung cancer in contrast to no increase in younger women using hormones (although not statistically significant, the results with stratification by age are similar to the WHI analysis).[1095] ***The overall data, including the WHI analysis, suggest that hormone therapy initiated in older women with a history of smoking may promote the growth of existing lung cancers. The WHI evidence in women under the age of 60 is reassuring and case-control and cohort data that reflect hormone use in a younger population than in the WHI indicate that estrogen is associated with some protection against lung cancer.***

Cervical Cancer

The association between postmenopausal hormone therapy and cancer of the uterine cervix has not been extensively studied. Evidence from one cohort study and one case-control study indicates that the postmenopausal use of estrogen does not increase the risk of cervical cancer.[1096, 1097] Indeed, these studies observed protection against cervical cancer in the estrogen users, but this may reflect detection bias (more examinations and Pap smears in estrogen users). Another case-control study suggested an increased risk in cervical adenocarcinomas, but there were only 13 cases.[1098] In a follow-up report of 120 women treated for Stage I and II cervical cancer, no adverse effects of hormone therapy on survival or recurrence were observed.[1099]

Malignant Melanoma

The possibility of a relationship between exogenous hormones and cutaneous malignant melanoma has been the subject of many observational studies. Accurate evaluation utilizing the Royal College of General Practitioners and Oxford Family Planning Association prospective cohorts and accounting for exposure to sunlight did not indicate a significant difference in the risk of melanoma comparing users of oral contraceptives to nonusers.[1100, 1101] Results with the use of postmenopausal estrogen therapy have not indicated a major impact. A slightly increased risk with long-term use of estrogen was noted in one case-control study (a conclusion based on 10–20 cases and not achieving statistical significance), whereas other case-control studies could find no association with postmenopausal estrogen treatment.[1102–1106] Others have reported slight increases in the risk of malignant melanoma associated with the use of exogenous estrogen, but all failed to reach statistical significance.[1096, 1107, 1108] In an analysis of cancer incidence in a Swedish cohort of women prescribed postmenopausal hormone therapy, no increase in malignant melanoma was observed.[863] On the adverse side, a Dutch case-control study reported a significant 42% increase in melanoma risk in postmenopausal hormone users, but it was based on only 33 cases and there was no consideration of sun exposure.[1109] ***No solid evidence indicates an increase in risk for cutaneous malignant melanoma with the use of hormone therapy.***

Metabolic Effects

Pancreatitis and severe hypertriglyceridemia can be precipitated by the administration of oral estrogen to women with elevated triglyceride levels.[1110, 1111] In women with triglyceride levels

between 250 and 500 mg/dL, estrogen should be provided with great caution, and a nonoral route of administration is preferred. ***The triglyceride response is rapid, and a repeat level should be obtained in 2–4 weeks. If increased, hormone therapy must be discontinued. A level greater than 500 mg/dL represents an absolute contraindication to estrogen treatment.*** Triglyceride levels in the normal range were not affected by progestins in the PEPI trial.[100] An exaggerated triglyceride response to estrogen might be attenuated by a progestin, especially a progestin of the 19-nortestosterone family, and, therefore, the daily, combination method of treatment could be considered for women with slightly elevated triglycerides. ***However, the treatment of choice is transdermal estrogen, a route of administration that does not affect triglyceride levels; indeed, triglyceride levels markedly elevated in response to oral therapy return to normal when treatment is changed to transdermal administration.***[42, 45]

Although physiologic and epidemiologic evidence indicates that estrogen use increases the risk of gallbladder disease, the overall impact is not great. The Nurses' Health Study indicated that oral estrogen therapy may carry a 1.5–2.0-fold increased risk of gallbladder disease.[1112] The risk of cholecystectomy appeared to increase with dose and duration of use and to persist for 5 or more years after stopping treatment. Other observational studies also reported increased risks of cholecystectomy in past and current users of estrogen.[779, 1113, 1114] At least two case-control studies concluded that estrogen use is not a risk factor for gallstone disease in postmenopausal women, although the statistical power was limited by small numbers.[1115, 1116] A cross-sectional study of gallstone disease could detect no association with postmenopausal hormone treatment.[1117] In the HERS clinical trial, the relative risk of gallbladder disease was 1.38; however, this did not achieve statistical significance.[777, 1118] The risk of gallbladder disease and gallbladder surgery was significantly increased in both the estrogen-progestin and estrogen-only arms of the WHI trial.[1119] This amounted to an increase of 20 to 30 cases per 10,000 per year in this older population of postmenopausal women. The routine, periodic use of blood chemistries is not cost-effective, and careful monitoring for the appearance of the symptoms and signs of biliary tract disease suffices. It is not certain that this potential problem is limited to oral therapy. Nonoral routes of estrogen administration have been reported to both increase and not increase biliary cholesterol saturation (a lithogenic response).[1120, 1121] In the Million Women Study in the U.K., the risk of gallbladder disease was lower with transdermal estrogen compared with oral estrogen, but the many problems with this observational cohort make it difficult to allow a confident statement.[1122]

Metabolic contraindications to estrogen therapy include chronically impaired liver function, acute vascular thrombosis (with or without emboli), and neuro-ophthalmologic vascular disease.

Weight Gain

The gain in weight that many middle-aged individuals experience is the result of lifestyle; specifically, the balance of dietary intake and exercise is tilted toward excessive caloric intake because of a decline in physical fitness and the age-related decrease in basal metabolic rate. Weight gain in women at menopause is not due to the hormonal changes associated with menopause.[1123–1125] Likewise, postmenopausal hormone therapy cannot be blamed for weight gain. The large Rancho Bernardo prospective cohort study and the randomized PEPI clinical trial documented that hormone therapy with or without progestin does not cause an increase in body weight.[1126, 1127] In fact, in the PEPI trial, the hormone-treated groups actually gained less weight than the placebo group. In the 2-year clinical HOPE trial assessing the efficacy of lower estrogen-progestin doses, hormonal treatment was associated with lesser increases in body weight and body fat compared with placebo.[1128]

After menopause, there is an increase in abdominal fat and total body fat that is associated with an increase in insulin resistance, a consequence of the decrease in estrogen levels.[1129]

Postmenopausal estrogen therapy maintains the premenopausal body habitus, preventing the increases in abdominal fat, insulin resistance, blood pressure, and diabetes mellitus associated with estrogen deficiency.[1130] Estrogen (with or without progestin) prevents the tendency to increase central body fat with aging.[618–621, 1131] This would inhibit the interaction among abdominal adiposity, hormones, insulin resistance, hyperinsulinemia, blood pressure, and an atherogenic lipid profile that results in the metabolic syndrome. An excellent randomized trial in Denmark documented less weight gain with hormone therapy because of a smaller increase in fat mass.[1132] In a substudy of the Women's Health Initiative, assessments of body composition by DEXA indicated that estrogen-progestin users had less fat and greater lean mass.[1133] This same salutary effect on central body fat has been observed with tibolone treatment.[1134] ***Rather than causing body weight gain, therefore, postmenopausal hormone therapy reduces the increase in insulin resistance and abdominal fat usually seen with aging, with a beneficial impact on the risks of hypertension, diabetes mellitus, and dyslipidemia.***

Presentations Requiring Clinical Judgment

Patients with Endometrial Cancer, Endometrioid Tumors, and Endometriosis

Gynecologic oncologists have reported that patients who have had Stage I and Stage II adenocarcinoma of the endometrium can take estrogen without fear of an increased risk of recurrence or a decrease in disease-free interval.[1135–1138] In a matched cohort of 249 women with Stage I, Stage II, and Stage III endometrial cancer with a long follow-up, there was no indication of an increase in recurrent disease with hormone therapy.[1139] Similar negative results were reported in a Turkish case-control study.[1140] The only randomized trial, organized by the Gynecologic Oncology Group, closed prematurely because of recruitment difficulties following the publicity associated with the Women's Health Initiative.[1141] Nevertheless, a total of 1,236 patients with Stage I or Stage II endometrial cancer were randomized to either estrogen-only or placebo, and although the participants constituted a low-risk group, the recurrence rate was low, 14 recurrences with five deaths in the treatment group and 12 recurrences and nine deaths in the placebo group. If a high-risk tumor is estrogen- and progesterone-receptor negative, it seems reasonable to allow immediate hormone therapy. Because the latent period with endometrial cancer is relatively short, a period of time (5 years) without evidence of recurrence would increase the likelihood of safety on an estrogen program. ***We recommend that hormone therapy be avoided in patients with high-risk tumors that are receptor-positive until 5 years have elapsed. The combination of estrogen-progestin is recommended in view of the potential protective action of the progestational agent. A similar approach makes sense for patients previously treated for endometrioid tumors of the ovary. In view of the fact that adenocarcinoma has been reported in patients with pelvic endometriosis and on unopposed estrogen, the combined estrogen-progestin program is also advised in patients with a past history of endometriosis.***[144–149]

Should a Woman Who Has Had Breast Cancer Use Postmenopausal Hormones?

The argument that postmenopausal hormone therapy should not be given to women who have had breast cancer is a reasonable one. It is based on the recognition of a large body of

evidence that indicates that breast cancer is often a hormone-responsive tumor. The overriding fear of many clinicians (and patients) is that metastatic cells are present (perhaps being controlled by various host defense factors) that will be susceptible to stimulation by exogenous hormones.[1142] However, many women who have had breast cancer are aware of the benefits of postmenopausal hormone treatment and are asking clinicians to help make this risk-benefit decision. In addition, some women suffer from such severe hot flushing and vaginal dryness that they are willing to consider hormonal treatment. Tibolone treatment of breast cancer survivors is relatively contraindicated as discussed earlier in this chapter.

The rate of recurrent breast cancer in hormone users has been reported in multiple case series with more than 1,000 breast cancer survivors.[1143–1159] It is reassuring that the recurrence rates in these reports are not different from the expected rate of breast cancer recurrence. In one series, 25 and then 77 women with breast cancer ranging from in situ to Stage III disease received estrogen-progestin therapy for 24 to 82 months; the recurrence rate was not greater than that expected.[1146, 1147] From this group of patients, 41 breast cancer survivors receiving hormone therapy had the same outcomes when compared with 82 women selected from a cancer registry and not taking hormones.[1147] In a report from Australia, 90 women with a history of breast cancer who were given a combination of estrogen and progestin had lower mortality and recurrence rates; however, the dose of progestin was very high (which in itself can be therapeutic) and treatment was not randomized.[1151] In a follow-up of 319 women treated with estrogen after treatment for localized breast cancer, only one patient developed recurrent disease.[1152] In a matched-controlled series of 277 breast cancer survivors, there was no difference in the estrogen-treated group and the control group for recurrent disease.[1159] A series with 114 women who received hormone treatment had a low rate of recurrence.[1148] These patients have had both positive and negative nodes and positive and negative estrogen receptor status. Although the results conform to an incidence of recurrent disease no greater than expected, the outcomes can reflect biases in clinician and patient decision-making that can only be overcome with a proper long-term, randomized clinical trial.

A case-control study of hormone therapy after breast cancer actually found a significant reduction in risk of recurrent disease, breast cancer mortality, and total mortality in hormone users.[1160] Again these are reassuring observational data that hormone therapy after breast cancer has no adverse impact on recurrence.

A U.S. trial at the University of Texas M.D. Anderson Cancer Center, provided estrogen to randomized women who had been treated for localized Stage I and Stage II breast cancer with either estrogen receptor-negative or status unknown tumors.[1161] After 5 years of follow-up, 56 women in the trial receiving estrogen were compared with 243 women with comparable disease, and there was no adverse effect of estrogen treatment on disease-free survival.[1162]

A multicenter, large case-control study of women younger than 55 years with breast cancer concluded that the use of oral contraceptives or postmenopausal hormone therapy either before or after diagnosis did not increase the risk of the first breast cancer or recurrent breast cancer.[1163] This negative finding was not changed by duration of use or age of oral contraceptive use, or by BMI, duration of use, or type of postmenopausal hormone use (estrogen-only or combined estrogen-progestin).

The HABITS Trial

"Hormonal replacement therapy After Breast cancer—Is it Safe?" (HABITS) began in multiple centers in Sweden in May 1997, to compare breast cancer survivors treated for at least 2 years with hormone therapy with treatment other than hormones. A similar trial was initiated in Stockholm. Because recruitment was slower than anticipated, the two trials

agreed in February 2002 to pool their patients and to use a joint safety and monitoring committee. In October 2003 the safety committee recommended that the trial be discontinued because there were 26 women in the treated group with new breast cancers compared with seven in the nontreated group. The HABITS trial was terminated in December 2003.[1164] Confronted with this outcome, the Stockholm investigators decided to cancel their trial as well even though the hazard ratio in the Stockholm patients was 0.82 (CI=0.35–1.9).

HABITS was a randomized but not placebo-controlled trial in which hormone therapy was compared to management without hormones in women with menopausal symptoms who had been previously treated for Stage I or Stage II breast cancer. Concomitant tamoxifen treatment was allowed in the HABITS patients but not aromatase inhibiters. Hormone therapy consisted of the variety of products and methods on the Swedish market, but not tibolone. Most of the treated women used products with the relatively high dose of 2 mg estradiol. After 4 more years of follow-up of 442 women, there were 39 cases of new breast cancer in women using hormone therapy compared with 17 in the non-treated group for a hazard ratio of 2.4 (CI=1.3–4.2).[1165]

The treated and non-treated groups of women in HABITS were very different in terms of characteristics and behaviors. More of the women in the treated group had hormone receptor-positive cancers (62.3%) compared with the non-treated group (54.5%). 11 women in the treated group never received hormones; 43 in the non-treated group did receive hormones. There was a very wide range of exposure times, ranging from 0 to 80 months. About one-third of the women who received hormones changed products during the study. The method of analysis of the HABITS data was intent-to-treat, and thus the impact of these differences cannot be ascertained.

Analysis of the new breast cancers in HABITS (either local recurrences or contralateral cancers) indicated statistically significant increases only in hormone receptor-positive cancers. However, when adjusted for use of hormone therapy before diagnosis of the original breast cancer, use of tamoxifen, and hormone receptor status, the hazard ratio was 2.2 with a CI of 1.0–5.1). By definition this is close, but not statistically significant.

The Stockholm trial reported in 2005, after a median follow-up of 4.1 years, 11 new breast cancers in the treated arm and 13 new breast cancers in the non-treated arm.[1166] Why the difference between the Stockholm trial and HABITS? The HABITS investigators suggest that their patients had more node-positive disease, and thus "probably" had more women with subclinical disease that would be stimulated by hormone therapy. Another possibility was more protection with higher tamoxifen use in the Stockhom trial, although the HABITS trial could detect no impact of tamoxifen. The HABITS investigators believe that another possible explanation was the greater use of norethindrone and norethindrone acetate in HABITS compared with the use of medroxyprogesterone acetate in Stockholm. All of these explanations are speculations; the difference between the two trials remains and calls into question the reliability and accuracy of the data.

The cancellation of HABITS and the Stockholm trials made it impossible for the English and Italian trials to continue recruitment, and they were also canceled. Thus we have no on-going clinical trials of estrogen or estrogen-progestin therapy in breast cancer survivors. In our view the data from the Swedish trials are confusing and not definitive.

Although intuitively it seems that the risk/benefit ratio would be more favorable in the presence of negative nodes, negative receptors, and small tumors, are negative estrogen and progesterone receptor assessments sufficient to conclude that the cancer is not sensitive to hormones? And if the patient is in the high cure category, does it make any difference what the receptor status is? The answers to these questions are not known. Receptor status is not absolute; it is always a relative measure. In the multiple case series discussed previously, the same results were observed in both estrogen receptor-positive and receptor-negative patients.

Patients and clinicians have to incorporate all of the previously mentioned considerations into this medical decision. But when all is said and done, patients have to take an unknown risk if they want the benefits of hormone treatment, and clinicians have to take an unknown medical-legal risk. Some patients will choose to take estrogen, judging the benefits to be worth the unknown risk. Until definitive data are available from clinical trials, clinicians should support patients in this decision. Other patients will prefer to avoid any unknown risks. These patients, too, deserve support in their decision.

Women with Diabetes Mellitus

Estrogen can improve the metabolic changes associated with diabetes. Indeed, in prospective studies of postmenopausal women with type 2, noninsulin-dependent diabetes mellitus, estrogen-only or estrogen-progestin therapy improved all glucose metabolic parameters, including insulin resistance, the lipoprotein profile, and measurements of androgenicity.[631, 632, 1032, 1167–1171] One study, however, could detect no impact with transdermal administration.[1169] These changes should reduce the risk of cardiovascular disease, and in a very large cohort of 24,420 women from the Northern California Kaiser Permamente Diabetes Registry, current use of hormone therapy reduced the risk of myocardial infarction, but in women with a recent heart attack, an increased risk of a recurrent myocardial infarction was observed in hormone users (again a difference between primary prevention and secondary prevention).[1172] Tibolone also has a beneficial impact in short-term studies on insulin resistance in normal women and in women with noninsulin-dependent diabetes mellitus.[250, 274] Raloxifene has no effects on glucose metabolism or insulin sensitivity in normal women but exerts a modest improvement in insulin resistance in women who are hyperinsulinemic.[1173, 1174]

Women with Liver Disease

Osteoporosis is a major consequence of chronic liver disease. Although other bone-preserving agents can be utilized, none provides the multisystem benefits associated with estrogen therapy. In an evaluation of liver chemistries in a group of patients with primary biliary cirrhosis, standard hormone therapy doses produced no adverse changes over a period of 1 year.[1175] Estrogen treatment, either oral or transdermal, has not been associated with worsening cholestasis in women with primary biliary cirrhosis.[1176, 1177] We recommend measurement of liver chemistries after 1 month of treatment and every 6 months, with continuing hormone therapy in the absence of deterioration.

A French cohort study concluded that hormone therapy protects against the progression of hepatic fibrosis from chronic hepatitis C.[1178] Most individuals with hepatitis C virus infection develop chronic disease, a major cause of worldwide morbidity and mortality from liver fibrosis. The time course is relatively slow, taking years to progress from infection to cirrhosis. Progression is increased by consumption of alcohol, excess body weight, diabetes, and the degree of fatty degeneration in the liver. The severity of liver fibrosis is greater in men, and progression in women accelerates around age 60. In vitro and animal experiments have documented a beneficial effect of estrogen on the development of fibrosis, an effect that is consistent with the data in the French study finding greater progression of fibrosis after menopause and amelioration with hormone therapy. Another French study, a retrospective survey, reported a greater rate of fibrosis progression with hepatitis C in postmenopausal and nulliparous women, and a lower rate in postmenopausal women treated with hormone therapy compared with nontreated women.[1179]

Liver fibrosis from hepatitis C infection is not the result of viral destruction of hepatic cells. Fibrosis is a response to the inflammatory activity incited by the virus. By now it is well known, that estrogen can suppress the secretion of proinflammatory cytokines. Because of the prevalence of hepatitis C infection, these French reports are very important. Many clinicians are reluctant to prescribe hormone therapy to women with a history of liver disease. However, as long as liver enzymes are normal, there is no reason to withhold treatment, and these French studies indicate that estrogen therapy is beneficial. Postmenopausal hormone therapy should be discussed when women present with a history of hepatis C infection.

The lesson learned from the above conditions is that estrogen alone and estrogen-progestin treatment including contraception is acceptable as long as liver enzyme function is normal. This caveat is also true for women who have undergone organ transplantation surgery.[1180]

Treatment in the Presence of Leiomyomas (Fibroids)

Uterine leiomyomas are monoclonal tumors that retain sensitivity to both estrogen and progestin (Chapter 4); therefore, it is appropriate to be concerned over whether leiomyomas will grow in response to postmenopausal hormone therapy. As assessed by vaginal ultrasonography, the number and size of uterine leiomyomas increased in women being treated with an intramuscular depot form of estrogen-progestin therapy.[1181] However, the hormonal dose in this study was relatively high, certainly higher than standard regimens. At the end of 1 year, women with small asymptomatic fibroids administered a daily combination of 0.625 mg conjugated estrogens and 2.5 mg medroxyprogesterone acetate had no sonographic evidence of growth in contrast to an increase in size observed with transdermal estradiol (50 μg) and 5 mg medroxyprogesterone acetate daily (a response that probably reflects the effect of a higher progestin dose).[1182] In follow-up studies with standard doses of estrogen-progestin or tibolone, ultrasonography detected no changes in uterine or myoma volumes.[1183–1185] Clinical experience indicates that fibroid tumors of the uterus almost always are not stimulated to grow by the usual postmenopausal doses of estrogen and progestin. Tibolone and raloxifene also do not stimulate myoma growth.[287, 1186, 1187] Nevertheless, pelvic examination surveillance is a wise course. For example, a vulvar leiomyoma with growth stimulated by estrogen-progestin treatment has been reported.[1188] A case-control study could find no statistically significant increase in the risk of uterine sarcomas associated with estrogen therapy.[1189]

Estrogen Therapy and Sleep Apnea

The low prevalence of sleep apnea in premenopausal women and the increased frequency after menopause suggest a hormonal link. In careful studies, however, postmenopausal hormone therapy had no significant adverse effect on sleep-disordered breathing in women with more than mild obstructive sleep apnea.[1190, 1191] And other sleep laboratory studies found that estrogen treatment reduced difficulties with sleep-disordered breathing.[1192–1194] The slight rise in basal body temperature induced by a progestational agent may be sufficient to disrupt the quality of sleep in some women, a problem that may be more noticeable with a sequential regimen and with nighttime administration.

Estrogen Therapy and Asthma

In some women, changes in asthma activity have been noted to correlate with phases of the menstrual cycle. The impact of postmenopausal hormone therapy on asthma activity has not been well investigated, but there is an indication that estrogen has an adverse effect. A worsening in spirometry assessment was detected in asthmatics after estrogen therapy; however, the difference was judged to be subclinical, and the patients did not report any changes in their perceptions of symptoms.[1195] A similar study could detect no adverse effects with either discontinuation or reinitiation of estrogen treatment.[1196] And another study concluded that estrogen therapy improved asthma and decreased the need for glucocorticosteroid treatment.[1197] In a prospective assessment of a cohort of women, the use of postmenopausal hormone therapy (estrogen with or without progestin) was associated with a 50% increase in the risk of developing adult-onset asthma, and the risk was greater with long-term use and with higher doses of estrogen.[1198] Similar results were reported in the French E3N cohort study, except an increased risk was observed only in estrogen alone users.[1199] In the Nurses' Health Study, newly diagnosed asthma was increased about 2-fold by hormone therapy.[1200] Because hormonal changes may precipitate asthmatic activity (e.g., catamenial asthma), attention should be directed to the symptomatic pattern and consideration given to the daily, continuous, combined regimen.

Asthma is another condition treated with glucocorticoids and associated with glucocorticoid-induced bone loss.[1201, 1202] Prevention with hormone therapy or bisphosphonate treatment warrants consideration.

Dementia and Cognition

The Women's Health Initiative Memory Study (WHIMS) was a randomized, clinical trial utilizing 4,532 of the participants in the WHI hormone studies. After the cancellation of the WHI estrogen-progestin arm, WHIMS analyzed the hormonal effects in its subset of participants, all age 65 and older.

The WHI concluded that estrogen-progestin and estrogen-only therapy increased the risk for dementia in women 65 years and older and did not prevent mild cognitive impairment.[1203–1206] However, the only statistically significant finding in the estrogen-progestin arm was increased dementia (vascular dementia, not Alzheimer's disease) in elderly women (22 cases in the treated group and 10 cases in the placebo group) who were 75 and older and who had been exposed to a relatively short-term of estrogen-progestin therapy. The estrogen-only arm of the WHI contained more obese women with pre-existing cardiovascular disease, and the trend for an increase in dementia likely reflected an effect in older women with established atherosclerosis. Will older women who have used hormone therapy for long durations early in their postmenopausal years be protected against dementia? The WHI report recognized that this hypothesis could not be tested in this clinical trial because of the older age of the study participants. A prospective study of a homogeneous population in Utah (thus minimizing, if not eliminating, the healthy user bias) concluded that a reduction in the risk of Alzheimer's required long-term treatment, initiated at least 10 years before symptoms of dementia appear.[1207] The favorable effects of hormone therapy on cognition and the risk of Alzheimer's disease appear to be limited to women who initiate treatment close to their menopause as discussed in Chapter 17.

Other Conditions

Close surveillance is indicated for some patients with seizure disorders and migraine headaches. Patients with migraine headaches often improve if a daily, continuous method of treatment is used, eliminating a cyclic change in hormone levels that can serve to trigger headaches. Conditions that do not represent contraindications include controlled hypertension, smoking, and varicose veins. The belief that estrogen is potentially harmful with each of these clinical situations is derived from old studies of high-dose oral contraceptives. Estrogen in appropriate doses is acceptable in the presence of these conditions.

No other cancers (in addition to those mentioned previously) are known to be adversely affected by hormone therapy. Postmenopausal hormone therapy can be administered to all patients with a history of cervical, ovarian, or vulvar malignancies.

Unusual anecdotal reports include the following:

1. Provocation of chorea by estrogen therapy in a woman with a history of Sydenham's chorea.[1208]
2. Exacerbation of pulmonary leiomyomatosis by estrogen therapy.[1209]
3. Psychiatric symptoms in response to estrogen in patients with acute intermittent porphyria.[1210]
4. Idiosyncratic ocular symptoms associated with estrogen.[1211]
5. Sudden deafness and tinnitus with commencement of hormone therapy.[1212]
6. Resolution of an infection with *Trichomonas vaginalis* after discontinuation of estrogen-progestin therapy.[1213]
7. Successful treatment of Sjögren's syndrome with tibolone.[1214]

Potential Benefits of Hormone Therapy

Estrogen Therapy and Rheumatic Diseases

No clear conclusion is apparent from the studies of estrogen's effect on rheumatic diseases, especially rheumatoid arthritis. Studies have indicated that exogenous estrogen, either oral contraceptives or postmenopausal therapy, protects against the onset of rheumatoid arthritis, whereas other studies find no effect.[1215–1218] These studies have been hampered by small numbers. In a randomized, placebo-controlled, clinical trial, maintenance of standard serum estradiol levels was associated with improvements in some measurements of disease activity in patients with rheumatoid arthritis.[1219] There has been no evidence that postmenopausal hormone therapy aggravates rheumatoid arthritis or causes a flare in disease activity.

In the Nurses' Health Study, the use of postmenopausal estrogen was associated with approximately a 2-fold increase in systemic lupus erythematosus, an observation based on 30 cases in past and current users of estrogen.[1220] In a follow-up of 60 postmenopausal

women with stable systemic lupus erythematosus, no adverse effects of hormone therapy could be detected.[1221] Patients with systemic lupus erythematosus develop early atherosclerosis and those treated with glucocorticoids are especially at greater risk for osteoporosis,[1222] but there is a concern that exogenous estrogen will increase flares and stimulate thrombosis because of the hypercoaguable state in patients with systemic lupus erythematosus. In 1- and 2-year randomized clinical trials, transdermal estradiol and oral estrogen-progestin regimens prevented bone loss with no increase in disease activity.[1223–1225] Importantly, no increase in arterial or venous thrombosis was observed with postmenopausal hormonal therapy in a longitudinal study of a large cohort of U.S. women with systemic lupus erythematosus.[1226] ***Postmenopausal hormone therapy can be considered in patients with stable or inactive disease, without renal involvement or high antiphospholipid antibodies.***

Bone loss associated with glucocorticoid therapy can be avoided with the usual postmenopausal hormone regimens.[1227, 1228] These patients are also excellent candidates for bisphosphonate treatment, another effective option that prevents glucocorticoid-induced bone loss.[1229, 1230] In addition, calcium and vitamin D supplementation are important to prevent bone loss associated with low-dose glucocorticoid treatment.[1231]

Estrogen Therapy and Osteoarthritis

Osteoarthritis is the most common form of arthritis in older people, and its prevalence increases rapidly in women after menopause. Osteoporosis protects against arthritis[1232]; therefore, the impact of estrogen therapy on osteoarthritis is a logical concern. Increasing severity of osteoarthritis of the knee has been reported to be associated with increasing bone density and the current use of postmenopausal hormone therapy in middle-aged women.[1233] However, estrogen treatment reduced osteoarthritis in a monkey model, and a cross-sectional study concluded that current users of estrogen had a reduced prevalence of osteoarthritis of the hip, and there was protection against the severity of osteoarthritis, with a greater effect with longer duration of use.[1234, 1235] Arthritic complaints are a major side effect of the low estrogen state induced in women with breast cancer treated with aromatase inhibitors, and osteoarthritis develops more frequently in women with the lowest levels of estrogen.[1236, 1237] Because there are no known treatments that modify the course of osteoarthritis, this potential benefit of postmenopausal hormone therapy deserves study by a randomized clinical trial.

Estrogen Therapy and the Oral Cavity

Oral complaints are common among postmenopausal women. The administration of estrogen provides significant relief from oral discomfort, burning, bad taste, and dryness[1238]. Estrogen therapy is also associated with a reduction in periodontal disease, including gingival inflammation and bleeding.[1239, 1240] These changes may reflect epithelial responses to estrogen by the oral mucosa, in a manner similar to that of the vaginal mucosa. Oral alveolar bone loss (which can lead to loss of teeth) is strongly correlated with osteoporosis, and the salutary effect of estrogen on skeletal bone mass is also manifested on oral bone.[1241, 1242] In the Leisure World Cohort, tooth loss and edentia were significantly reduced in estrogen users compared with nonusers (with a reduced need for dentures), and this beneficial effect was greater with increasing duration of estrogen use.[1243] Approximately a 25% reduced risk of tooth loss in current users of estrogen was observed in the Nurses' Health Study.[1244]

Professional singers have used hormone therapy to prevent what they view as unwanted voice changes associated with menopause.[1245] In prospective studies, objective voice analy-

ses have documented a more androgenic change in voice in the early postmenopausal years with a lesser change associated with estrogen treatment; although slightly attenuated by the addition of a progestin, the overall effect of estrogen treatment is to preserve voice quality.[1246–1248] Laryngeal cytology demonstrated epithelial maturation in women on estrogen treatment, and these women reported better voice quality and fewer voice changes compared with a control group.[1249]

Estrogen Therapy and Vision

There is some evidence that estrogen therapy improves visual acuity (or lessens the decrease occurring during the early postmenopausal years), perhaps due to a beneficial effect on lacrimal fluid.[1250, 1251] An increased prevalence of keratoconjunctivitis sicca (dry eyes) in menopausal and postmenopausal women, with symptoms of scratchiness, burning, and photophobia, is recognized by ophthalmologists.[1252] Although reports concluded that there was no effect or even a worsening of dry eyes with hormone therapy, a clinical trial indicated relief from dry eye symptoms with the use of topical estrogen eye drops.[1253–1257]

There is evidence that postmenopausal estrogen therapy has an effect that protects against lens opacities (cataracts).[1258–1262] Estrogen alone or estrogen-progestin treatment also lowers intraocular pressure in postmenopausal women with normal eyes or glaucoma.[1256, 1263–1265] In the Nurses' Health Study, current use of estrogen-progestin, but not estrogen alone, was associated with a reduced risk of glaucoma.[1266]

There is modest evidence that the risk of age-related macular degeneration is reduced in estrogen users, but some studies could detect no effect.[1267–1271]

Estrogen Therapy and Age-Related Hearing Loss

Demineralization of the cochlear capsule occurs with aging and with metabolic bone diseases, such as cochlear otosclerosis. This demineralization is associated with neural hearing loss. Postmenopausal women (age 60–85 years) who have lower than average femoral neck bone mass have an increased risk of having a hearing loss.[1272] This association between femoral neck bone mass and age-related hearing loss suggests that prevention of bone loss with estrogen therapy might also be exerted on the cochlear capsule. In addition, estrogen may have beneficial effects on cochlear blood flow and central nervous system auditory neurons. Hearing impairment in a Turkish family was associated with an inactivating mutation for estrogen receptor-β.[1273] Mouse knock-out studies indicate that estrogen receptor-β is important for the prevention of age-related hearing loss.[1274] Studies have documented better hearing levels in estrogen users, with an indication that the addition of a progestin attenuated the favorable effect of estrogen.[1275, 1276]

How Long Should Postmenopausal Hormone Therapy be Continued?

The answer to this question is relatively straightforward. A woman should continue her postmenopausal hormone regimen as long as she wants the benefits. Although some

estrogen effects will be long-lasting, the full impact is rapidly lost after discontinuation. For example, in the Nurses' Health Study, reduced risk of mortality (largely cardiovascular) was lost by the fifth year after discontinuing treatment.[718]

Should Hormone Therapy be Discontinued Abruptly or Gradually?

Menopausal symptoms after discontinuing hormonal therapy reoccur in a substantial number of women, disrupting quality of life in at least 25% of previously treated women, and in one Swedish study, 70%.[1277, 1278] It seems intuitively advantageous to encourage a tapering, gradual discontinuation program to minimize recurrence of menopausal symptoms. However, randomized trials comparing gradual discontinuation with abrupt cessation found no benefit with a tapering regimen.[1279–1282] Neither the recurrence rate nor the severity of menopausal symptoms differs comparing the two methods.

Should Very Old Women be Started on Hormone Therapy?

The positive impact of hormone therapy on bone has been demonstrated to take place even in women older than age 65.[1283, 1284] This is a strong argument in favor of treating very old women who have never been on estrogen who cannot take or afford the other alternatives for bone preservation. Estrogen treatment that is not begun until after age 60 can with long-term use achieve bone densities nearly but not totally comparable with those in women taking estrogen from menopause, and estrogen use between the ages of 65 and 74 has been documented to protect against hip fractures.[1285, 1286] Adding a pharmacologic regimen to an old woman's daily life is not a trivial consideration. This judgment requires the conclusion that a relatively youthful and vigorous elderly woman has something to gain from the treatment. Patients with osteoporosis and/or unfavorable lipoprotein profiles would certainly qualify, but an understanding of the results from the WHI must be attained with a thorough clinician-patient dialogue.

The primary and secondary prevention trials reviewed in this chapter have strongly indicated that estrogen administered to women with atherosclerosis is associated with an increased risk of arterial thrombosis. The mechanism, as reviewed in Chapter 17, is presumed to be the creation of a prothrombotic environment by the stimulation of metalloproteinase enzyme activity in unstable atherosclerotic plaques. This effect is enhanced by the production of 27-hydroxycholesterol in atherosclerotic sites, a cholesterol metabolite that competitively inhibits estrogen's beneficial actions within blood vessels. Statin treatment is known to stabilize atherosclerotic plaques rapidly, within 3 months. Although there are no studies to support this recommendation, it seems reasonable to consider initiating statin treatment for several months before starting estrogen therapy in older women.

Older women who have been deficient in estrogen for many years often experience side effects when standard doses of estrogen are initiated. Breast tenderness can be especially disturbing. Because of these side effects and the experience in the WHI with its older population, it is better to start older women with lower doses; e.g., the oral products at half the usual doses (0.3 mg conjugated estrogens or 0.5 mg estradiol) or a transdermal product that

delivers relatively low amounts of estrogen. An increase to standard doses is recommended if the response in bone density is not satisfactory.

Can the Diet Produce Variations in Systemic Estrogen Levels?

Oral estrogens have an extensive first-pass metabolism, both in the gastrointestinal tract and the liver. This metabolism consists chiefly of sulfation and hydroxylation. The cytochrome P450 system catalyzes the hydroxylation of estrogen, and antioxidants can inhibit this action. Flavanoids (e.g., naringenin and quercetin) are present in high concentrations in fruits and vegetables, and grapefruit juice inhibits estrogen metabolism, producing an increase in bioavailability that is consistent with an inhibition of hydroxylation.[1287, 1288] This raises the possibility that dietary interactions with food products could produce a clinical impact. There is great variability within individuals and between individuals in the pharmacokinetics of exogenously administered estrogen. It is possible that this variability partially reflects the dietary habits of individuals and not intrinsic metabolism. Because of this possibility, it seems prudent to recommend that patients take oral postmenopausal hormone therapy before they go to bed at night. This may minimize any effect of diet on blood levels of steroids.

An effect of alcohol ingestion by premenopausal women was not demonstrated on circulating levels of estrone, estradiol, dehydroepiandrosterone sulfate (DHEAS), or sex hormone-binding globulin in a cross-sectional study that depended on a questionnaire to assess alcohol intake.[1289] However, when alcohol is administered under experimental conditions, circulating estrogen concentrations are raised to high levels.[1290, 1291] And in a prospective cohort study of premenopausal women in Italy, higher estradiol levels were correlated with an increased alcohol intake over a 1-year period of time.[1292]

A Clinical Approach to Postmenopausal Hormone Therapy

We hope that we have convinced you that the menopause is a normal life event, not a disease, and that long-term postmenopausal hormone therapy is pharmacologic treatment that can provide preventive health care benefits. We have learned this from the women who revealed what they believe and what they experienced in the longitudinal studies of the last two decades. It only makes sense that trying to convince a woman she has a disease, when she does not believe it, will have a negative impact on the clinician–patient relationship. In addition, we believe our approach yields more willful and stronger decision-making that ultimately produces better continuation rates with treatment. Postmenopausal hormone therapy is an option that should be offered to most women as they consider their paths for successful aging, but the attitude and beliefs of the clinician have a major influence on the decisions made by patients. As beneficial as the impact of hormonal therapy may be, we must also emphasize the large improvement in health to be gained by lifestyle changes in smoking cessation, diet, regular exercise, and control of body weight.

It is the task of an epidemiologist to derive study conclusions based on study data. It is the obligation of a clinician to make a judgment whether the epidemiologist's conclusions have clinical meaning. For example, an epidemiologist can conclude that estrogen reduces coronary artery calcification and point out that a randomized clinical trial has not proved

that such a reduction lowers the risk of coronary heart disease. But it is appropriate for a clinician, knowing the correlation between coronary artery calcification and coronary heart disease, to conclude that estrogen reduction of coronary calcification will translate into less coronary heart disease. Medical judgments require more than absolute evidence from randomized trials, and medical judgments frequently do not have the luxury of postponing clinically meaningful decisions until data are conclusive.

Long-term postmenopausal hormone therapy is not precluded by the results reported by the WHI. There continues to be good reason to believe that there are benefits associated with treatment, including improvement of quality of life beyond the relief of hot flushes, maximal protection against osteoporotic fractures, a reduction in colorectal cancers and new onset diabetes mellitus, maintenance of skin turgor and elasticity, and the possibility of primary prevention of coronary heart disease and Alzheimer's disease. Of course, this should not detract or subtract from efforts to apply proven therapies (e.g., statins) and to support beneficial lifestyle modifications. Most importantly, there is an important message to be shared with colleagues and patients: contrary to the initial publicity, it is now apparent that the Women's Health Initiative results agree with 30 years of research.

The WHI Agrees With 30 Years of Research

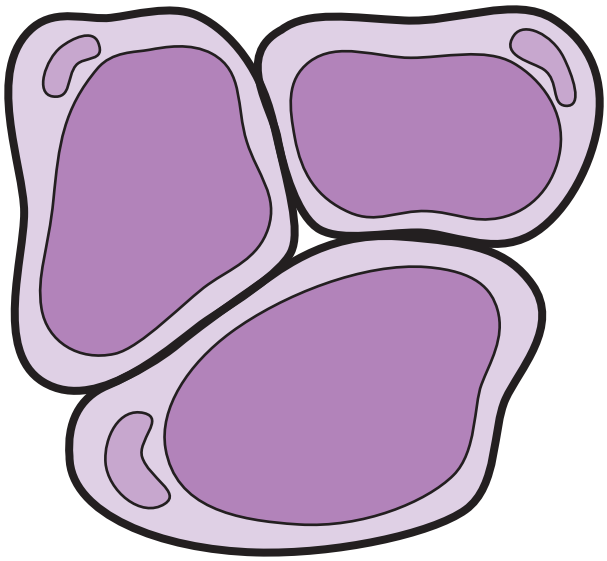
- Coronary heart disease: protection in young postmenopausal women.
- Stroke: no increase in early postmenopausal, healthy women.
- VTE: 2-fold increase in first years of use, concentrated in those at risk.
- Cancer: slightly increased risk of breast cancer or an effect on pre-existing tumors; reduction in colorectal cancer.
- Osteoporosis: reduction in fractures.
- Diabetes mellitus: reduction in new onset diabetes.

A theme has emerged from the epidemiologic confusion of the last decade. It takes healthy tissue to allow effective response to estrogen and maintenance of health. Experimental evidence in monkeys and women indicates that as endothelial cells become involved with atherosclerosis and neurons become affected with the pathologic process of Alzheimer's, beneficial responses to estrogen diminish.[802, 803, 1207] Maximal benefit, therefore, requires early onset of treatment, near the time of menopause.

The most effective and appropriate method to help in decision-making is to identify the specific goals and objectives of the individual patient—*let your patient be your guide*. Once an individual's objectives are identified, choices from multiple treatment options can be reviewed. Postmenopausal health and hormone therapy are subjects receiving enormous attention and research; therefore, decision-making should be at least an annual event, incorporating new knowledge as it appears. Approached in this fashion, the terms "short-term" and "long-term" and the imposition of time limits for therapy become meaningless. Clinician and patient together make an annual clinical judgment that is appropriately directed to accomplishing the individual patient's goals. ***The guiding principle is: the right dose for the appropriate duration according to an individual patient's needs.***

All references are available online at: http://www.clinicalgynendoandinfertility.com

19 Obesity



Because more than one-third of American adults are obese, the unrewarding fight against obesity is all too common, not only with our patients but also with ourselves. An astonishing 64.1% of U.S. women were overweight or obese (28.6% overweight and 35.5% obese) in 2008.[1, 2] Unfortunately, for over 100 years the incidence of obesity has been increasing in the U.S. and Europe, a reflection of an affluent society with an increasingly sedentary life combined with high caloric foods, but most of the increase has been recent, beginning in 1980.[3–6] About 59% of American adults do zero vigorous physical activity in their leisure time.[1] This change in lifestyle has produced a high prevalence of obesity with a similar trajectory in adults and children; almost 17% of school-aged children and adolescents in 2008 were obese and 32% overweight.[7, 8] The only good news is that the prevalence of obesity in the U.S. has not changed since 2003–2004; the upward trend has stopped.[2]

The lack of success in treating obesity is not due to an unawareness of the implications of obesity; there is a clear-cut, well-recognized relationship between mortality and weight.[1] The death rate from diabetes mellitus, for example, is approximately 4 times higher among obese diabetics than among those who control their weight. The death rate from appendicitis is double, presumably from anesthetic and surgical complications. Even the rate of accidents is higher, perhaps because fat people are awkward or because their view of the ground or floor is obstructed. The increase in mortality is not limited to obesity; all overweight individuals have an increase in the risk of death.[9] The Nurses' Health Study estimated that 23% of all deaths in nonsmoking middle-aged women are attributable to being overweight.[10]

The incidence of hypertension, heart disease, type 2 diabetes mellitus, metabolic syndrome, gout, gallbladder disease, obstructive sleep apnea, osteoarthritis, and all the most

prevalent cancers, including colorectal cancer, endometrial cancer, and postmenopausal breast cancer, is elevated in overweight people.[1, 11–13] Being overweight in adolescence is even a more powerful predictor of cardiovascular adverse health effects than being overweight as an adult.[14]

The increasing prevalence of obesity and its consequences now threaten to replace smoking as the primary cause of preventive mortality. A 40-year-old woman who is a nonsmoker can expect to lose 3.3 years of life if she is overweight (not obese, just overweight)![15] If obese, the loss is 7.1 years, and if smoking is added, the loss is 13.3 years. Unless there is a major change, it is estimated that the rise in U. S. life expectancy experienced over the last several decades will substantially slow down, offsetting the impact of declining smoking.[16]

When the personal and social problems encountered by obese people are also considered, it is no wonder that a clinician without a weight problem cannot comprehend why fat individuals remain overweight. The frequency with which a practitioner encounters the obese patient whose weight does not decrease despite a sworn adherence to a limited-calorie diet makes one question if there is something physiologically different about this patient. Is the problem due to lack of discipline and cheating on a diet, or does it also involve a pathophysiologic factor? Is the physiology of obese people unusual, or are they simply gluttons? ***Modern studies of obesity strongly indicate that this is a multifactorial problem, and that lack of willpower and laziness is not the simple answer.***

Definition of Obesity

Obesity is an excess storage of triglycerides in adipose cells. There is a difference between obesity and overweight.[17] Obesity is an excess of body fat. Overweight is a body weight, including muscle, bone, fat, and body water, in excess of some standard or ideal weight. The ideal weight for any adult is believed to correspond to his or her ideal weight from age 20 to 30. The following formulas give ideal weights in pounds:

Women: **100 + (4 × {height in inches minus 60})**
Men: **120 + (4 × {height in inches minus 60})**

At a weight close to ideal weight, individuals may be overweight, but not overfat. This is especially true of individuals engaged in regular exercise. An estimate of body fat, therefore, is more meaningful than a measurement of height and weight.

The most accurate method of determining body fat is to determine the density of the body by underwater measurement (hydrodensitometry). It certainly is not practical to measure density by submerging individuals in water in our offices; therefore, skinfold measurements with calipers have become popular as an index of body fat, or expensive imaging techniques can be utilized. These latter methods are not necessary for clinical practice. It is far simpler to utilize the body mass index nomogram, a method that corresponds closely to densitometry measurements.[18]

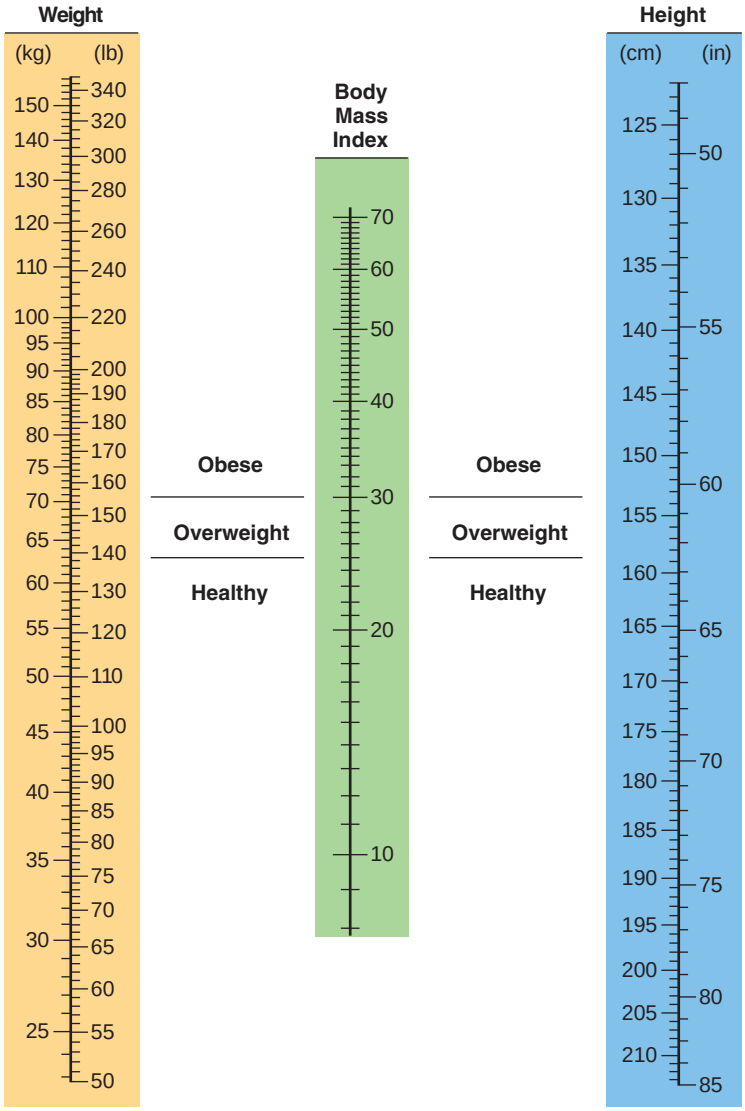
The body mass index (the Quetelet index) is the ratio of weight divided by the height squared (in metric units):

BMI = kilograms/meters^2

To use the nomogram for body mass index (BMI), read the central scale by aligning a straight edge between height and body weight. A body mass index of 25 or more warrants

treatment. Overweight is defined as a BMI of 25 or more (64.1% of American women in 2008). Obesity is defined as a BMI of 30 or more (35.5% of American women in 2008). A "good" BMI for most middle-aged people is in the range of 20 to 24.[19] Mortality is lowest in middle-aged women with a body mass index below 19.[10, 20] A BMI in the overweight or obese range predicts an increased risk for earlier death, no matter the age of the patient.

Overweight: **BMI = 25–29.9**
Obesity: **BMI = 30 and higher**



A person is obese when the amount of adipose tissue is sufficiently high (20% or more over ideal weight) to detrimentally alter biochemical and physiologic functions and to shorten life expectancy. Obesity is associated with four major risk factors for atherosclerosis: hypertension, diabetes, hypercholesterolemia, and hypertriglyceridemia. Overweight individuals have a higher prevalence of hypertension at every age, and the risk of developing hypertension is related to the amount of weight gain after age 25. The two in combination (hypertension and obesity) increase the risk of heart disease, cerebrovascular disease, and death. The Nurses' Health Study documented a continuing correlation between the body mass index and cardiovascular disease, diabetes, and cancer.[10, 21, 22] In other words, even a modest gain in adult weight, even in a range not considered to be overweight, increases the risk of cardiovascular and metabolic diseases. However, at any given BMI level, the

presence of an increase in abdominal fat, metabolic risk factors, or a strong family history of diabetes, hypertension, and heart disease increases the risk to good health.

It is well documented that women have a greater prevalence of obesity compared with men. One reason may be the fact that women have a lower metabolic rate than men, even when adjusted for differences in body composition and level of activity.[23] Another reason that more women gain weight with age is the postmenopausal loss of the increase in metabolic rate that is associated with the luteal phase of the menstrual cycle. The difference between men and women is even greater in older age.

Unfortunately, the basal metabolic rate decreases with age.[24, 25] After age 18, the resting metabolic rate declines about 2% per decade. The age-related decline in basal metabolic rate is not observed in women who continue to be involved in a regular endurance exercise program.[26] A 30-year-old individual will inevitably gain weight if there is no change in caloric intake or exercise level over the years. The middle-age spread is both a biologic and a psychosociologic phenomenon. It is, therefore, important for both our patients and ourselves to understand adipose tissue and the problem of obesity.

Physiology of Adipose Tissue

Adipose tissue serves three general functions:

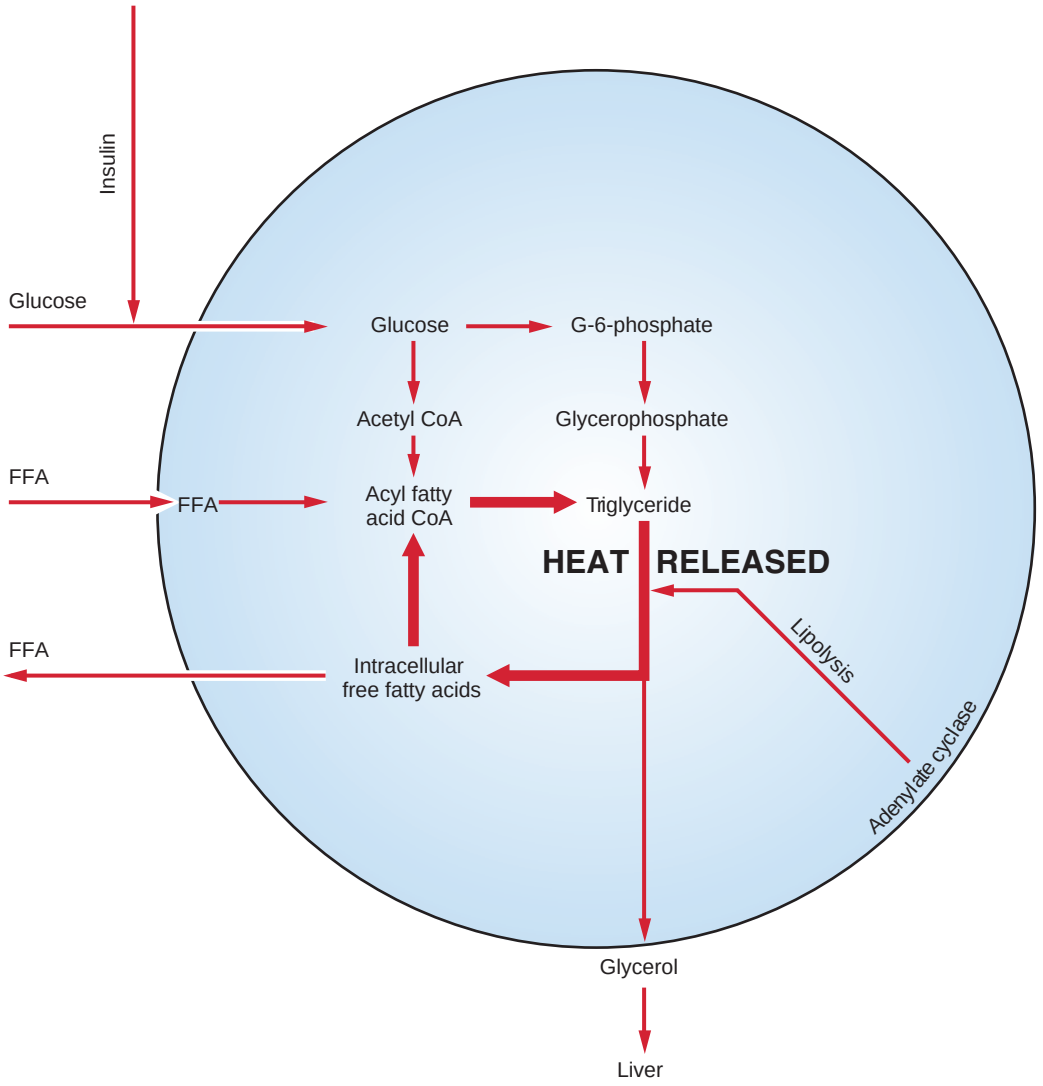
1. Adipose tissue is a storehouse of energy.
2. Fat serves as a cushion from trauma.
3. Adipose tissue plays a role in the regulation of body heat.

Each cell of adipose tissue can be regarded as a package of triglyceride, the most concentrated form of stored energy. There are 8 calories/g of triglyceride compared to 1 calorie/g of glycogen. The total store of tissue and fluid carbohydrate in adults (about 300 calories) is inadequate to meet between-meal demands. The storage of energy in fat tissue allows us to do other things besides eating. Our energy balance, therefore, is essentially equivalent to our fat balance. Thus, obesity is a consequence of the fat imbalance inherent in high-calorie diets.

The mechanism for mobilizing energy from fat involves various enzymes and neurohormonal agents. Following ingestion of fat and its breakdown by gastric and pancreatic lipases, absorption of long-chain triglycerides and free fatty acids takes place in the small bowel. Chylomicrons (microscopic particles of fat) transferred through lymph channels into the systemic venous circulation are normally removed by hepatic parenchymal cells where a new lipoprotein is released into the circulation. When this lipoprotein is exposed to adipose tissue, lipolysis takes place through the action of lipoprotein lipase, an enzyme derived from the fat cells themselves. The fatty acids that are released then enter the fat cells where they are reesterified with glycerophosphate into triglycerides. Because alcohol diverts fat from oxidation to storage, body weight is directly correlated with the level of alcohol consumption.[27]

Glucose serves three important functions:

1. Glucose supplies carbon atoms in the form of acetyl coenzyme A (acetyl CoA).
2. Glucose provides hydrogen for reductive steps.
3. Glucose is the main source of glycerophosphate.



The production and availability of glycerophosphate (required for reesterification of fatty acids and their storage as triglycerides) are considered rate limiting in lipogenesis, and this process depends on the presence of glucose.

After esterification, subsequent lipolysis results in the release of fatty acids and glycerol. In the cycle of lipolysis and reesterification, energy is freed as heat. A low variable level of lipolysis takes place continuously; its basic function is to provide body heat.

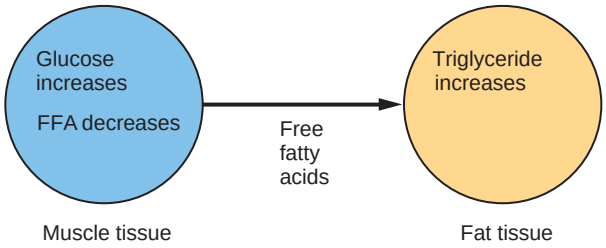
The chief metabolic products produced from fat are the circulating free fatty acids. Their availability is controlled by adipose tissue cells. When carbohydrate is in short supply, a flood of free fatty acids can be released. The free fatty acids in the peripheral circulation are almost wholly derived from endogenous triglycerides that undergo rapid hydrolysis to yield free fatty acid and glycerol. The glycerol is returned to the liver for resynthesis of glycogen.

Free fatty acid release from adipose tissue is stimulated by physical exercise, fasting, exposure to cold, nervous tension, and anxiety. The release of fatty acids by lipolysis varies from one anatomic site to another. Omental, mesenteric, and subcutaneous fat is more labile and easily mobilized than fat from other sources. Areas from which energy is not easily mobilized are retrobulbar and perirenal fat where the tissue serves a structural function.

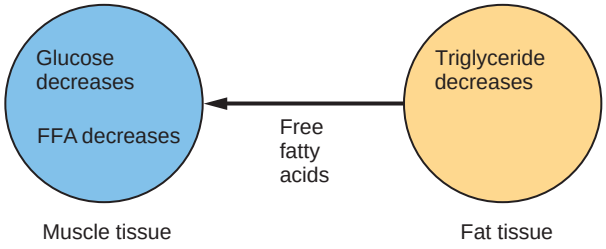
Adipose tissue lipase is sensitive to stimulation by both epinephrine and norepinephrine. Other hormones that activate lipase are adrenocorticotropic hormone (ACTH), thyroid-stimulating hormone (TSH), growth hormone, thyroxine (T4), 3,5,3'-triiodothyronine (T3), cortisol, glucagon, as well as vasopressin and human placental lactogen (hPL).

Lipase enzyme activity is inhibited by insulin, which appears to be alone as the major physiologic antagonist to the array of stimulating agents. When both glucose and insulin are abundant, transport of glucose into fat cells is high, and glycerophosphate production increases to esterify fatty acids.

Glucose abundant



Glucose reduced



The carbohydrate and fat composition of the fuel supply is constantly changing, depending on stresses and demands. Because the central nervous system and some other tissues can utilize only glucose for energy, a homeostatic mechanism for conserving carbohydrate is essential. When glucose is abundant and easily available, it is utilized in adipose tissue for producing glycerophosphate to immobilize fatty acids as triglycerides. The circulating level of free fatty acids in muscle,, therefore, is low, and glucose is used by all of the tissues.

When carbohydrate is scarce, the amount of glucose reaching the fat cells declines, and glycerophosphate production is reduced. The fat cell releases fatty acids, and their circulating levels rise to a point where glycolysis is inhibited. Thus, carbohydrate is spared in those tissues capable of using lipid substrates. If the rise of fatty acids is great enough, the liver is flooded with acetyl CoA. This is converted into ketone bodies, and clinical ketosis results.

In the simplest terms, when a person eats, glucose is available, insulin is secreted, and fat is stored. In starvation, the glucose level falls, insulin secretion decreases, and fat is mobilized.

If only single large meals are consumed, the body learns to convert carbohydrate to fat very quickly. Epidemiologic studies with schoolchildren demonstrated a positive correlation between fewer meals and a greater tendency toward obesity.[28] The person who does not eat all day and then stocks up at night is promoting an increase in fat.

Clinical Obesity

Leptin and the *Ob* Gene (the *LEP* Gene in Humans)

The hypothalamic location of the appetite center was established in 1940 by the demonstration that bilateral lesions of the ventromedial nucleus produce experimental obesity in rats. Such lesions lead to hyperphagia and decreased physical activity. Interestingly, this pattern is similar to that seen in human beings—the pressure to eat is reinforced by the desire to be physically inactive. The ventromedial nucleus was thought to represent an integrating center for appetite and hunger information. Destruction of the ventromedial nucleus was believed to result in a loss of satiety signals, leading to hyperphagia. Overeating and obesity, however, are not due to ventromedial nucleus damage but rather to destruction of the nearby ventral noradrenergic bundle.[29] Hypothalamic noradrenergic terminals are derived from long fibers ascending from hindbrain cell bodies. Lesions of the ventromedial nucleus produced by radiofrequency current fail to cause obesity. These lesions lead to overeating and obesity only when they extend beyond the ventromedial nucleus. Selective destruction of the ventral noradrenergic bundle results in hyperphagia. A sudden onset of hyperphagia can be due to a hypothalamic lesion. Possible causes include tumors, trauma, inflammatory processes, and aneurysms.

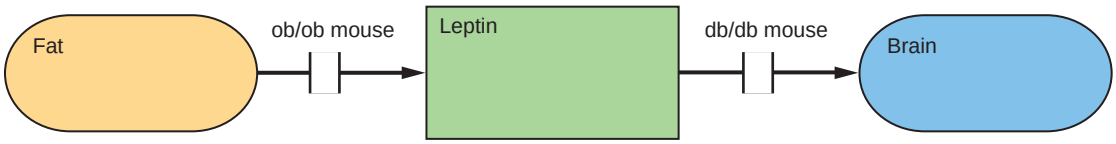
Signals arriving at these central nervous system (CNS) centers originate in peripheral tissues. Opiates, substance P, and cholecystokinin play a role in mediating taste, the gatekeeper for feeding, while peptides released from the stomach and intestine act as satiety signals.[30] In addition the CNS centers are regulated by locally released neuropeptides. Neuropeptides that inhibit appetite include corticotropin-releasing hormone (CRH), neurotensin, oxytocin, and cyclo(HisPro), a peptide derived by proteolysis of thyrotropin-releasing hormone.[31, 32] The control of food intake and energy expenditure is very complex, and no agent or system functions in isolation.

The word *leptin* is derived from the Greek word, "leptos," which means thin. ***Leptin is a 167-amino acid peptide secreted in adipose tissue, that circulates in the blood bound to a family of proteins, and acts on the central nervous system neurons that regulate eating behavior and energy balance.*** Rat studies in the 1950s suggested the existence of a hormone in adipose tissue that regulated body weight through an interaction with the hypothalamus.[33, 34] But it was not until 1994 that the *Ob* gene was identified, the gene responsible for obesity in the mouse.[35] In the human, this gene is known as the *LEP* gene.

Genetic Rodent Models of Obesity

	Single Gene Mutations	*Gene Product*	*Rodent Chromosome*	*Human Chromosome*
Mice:	ob/ob	Leptin	6	7
	db/db	Leptin receptor	4	1
	fat/fat	Carboxypeptidase E	8	11
	tub/tub	Phosphodiesterase	7	4
	Ay/Ay	Agouti protein	2	20
Rats:	fa/fa	Leptin receptor	5	1

There are four recessive gene mutations known in mice, and one dominant mutation (Ay/Ay).[36] Fat/fat mice are obese and remain insulin sensitive; the mutation decreases carboxypeptidase E, an enzyme that is involved in the conversion of prohormones to hormones; e.g., proinsulin to insulin. The biology of tub is yet unknown.



Ob/ob and db/db mice were described many years ago. The ob/ob mutation arose spontaneously in the Jackson Laboratory mouse colony in 1949. The ob/ob mouse is homozygous for a mutation of the *Ob* gene on chromosome 6, and the db/db mouse, discovered in 1966, is homozygous for a mutation of the *Db* gene on chromosome 4.[36, 37] These mice have been the subject of thousands of publications. The product of the *Ob* gene is leptin, and in the human, the *LEP* gene is located on chromosome 7q31.3. *Db* is the *diabetes* gene, and this is the locus of the mouse leptin receptor gene. Thus, the ob/ob mouse is obese because it does not produce leptin, and the db mouse is obese because it cannot respond to leptin; its leptin levels are very high (the mutation alters the leptin receptor).

The Leptin Receptor

The leptin receptor belongs to the cytokine receptor family.[37] There are multiple isoforms with at least three major forms, a short form and a long form, OB-R_S and OB-$R_{L,}$ plus a circulating protein that consists of the extracellular domain. The extracellular domain is very large with 816 amino acids. The intracellular domain of the short form contains 34 amino acids, and in the long form, about 303 amino acids. The short form has many variations, whereas the long form is the common signaling receptor. The only place that the long form is expressed in greater amounts than the short forms is in the hypothalamus, in the arcuate, ventromedial, paraventricular, and dorsomedial nuclei.[38, 39] High levels of the short-form leptin receptors in the choroid plexus indicate a transport role for the short form from blood into the cerebrospinal fluid to diffuse into the brain.[40]

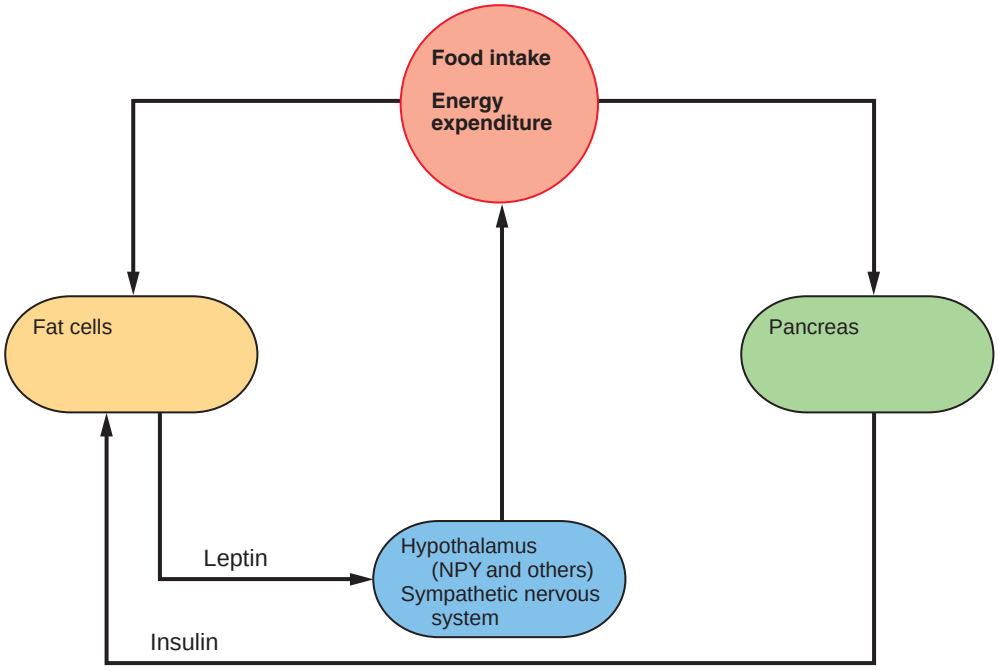
The class I cytokine receptor family (to which the long form belongs) acts by proteins that phosphorylate the receptor after binding and STAT proteins that are activated after phosphorylation, and then translocate to the nucleus and stimulate gene transcription. The long-form receptor works through STAT proteins, but gene knockouts specific for STAT proteins are not obese, indicating the presence of other signaling pathways activated by leptin.

The *Db* gene encodes the leptin receptor. The *Db* mutation converts the long form to the inactive short form, resulting in leptin resistance and obesity. The db/db mouse has a single G for T nucleotide substitution within the C terminal untranslated end of the short intracellular domain of the ob receptor. This results in a new splice site that creates an abnormal exon inserted into the messenger RNA (mRNA) that encodes the long intracellular domain of the ob receptor. As a result, the long-form mRNA in db/db mice encodes a protein with most of the intracellular domain truncated to be similar to the short form of the ob receptor. The fa/fa mutation is a glutamine for proline substitution in the extracellular domain, and in contrast to the db/db mouse, the fa/fa rat will respond to leptin, but only if it is delivered into the brain.[41]

The Physiologic Feedback Loop

Energy expenditure is composed of the basal metabolic rate, diet- and temperature-induced heat production, and the energy necessary for physical activity. Leptin induces weight loss in mice due to decreased appetite and food consumption and an increase in heat production and activity.[42] Leptin placed in the lateral ventricle of the rodent brain causes this weight loss, associated with a decrease in the hypothalamic peptide neuropeptide Y (NPY)

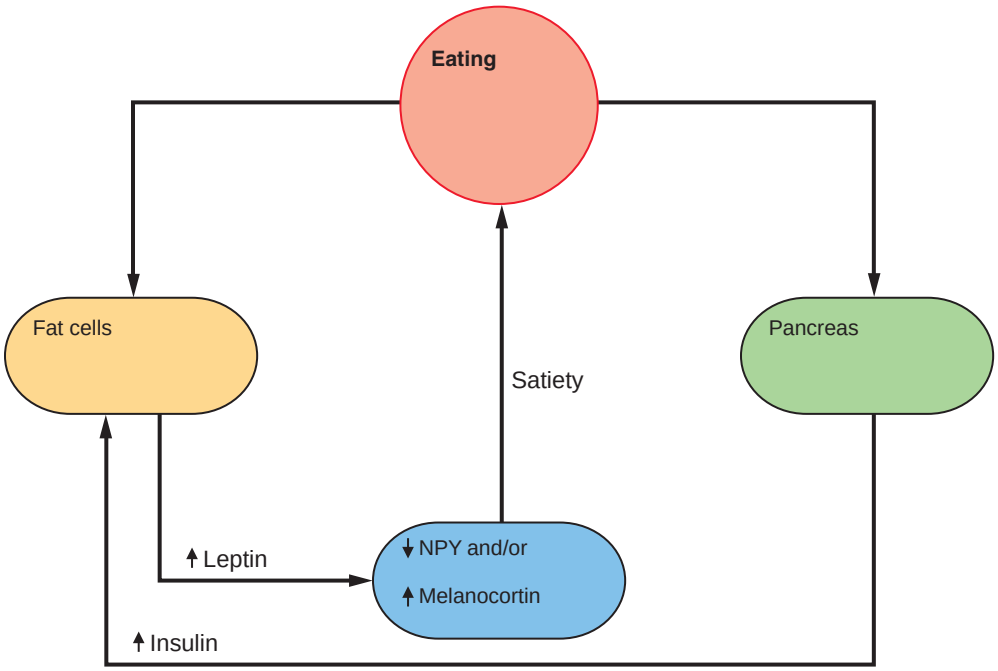
expression and secretion. NPY is a 36-amino acid polypeptide that is a potent stimulator of eating when injected directly into the rodent brain.[43] NPY has a tyrosine at both ends, hence the use of Y to stand for tyrosine. NPY stimulates food intake, decreases heat production, and increases insulin and cortisol secretion.



In rodents, insulin increases the expression of the *Ob* gene, but there is controversy in humans.[42] In humans, there is no surge of leptin after meals, and the acute administration of insulin does not stimulate an increase in leptin levels. However, an increase in *LEP* gene expression occurs in humans with chronic insulin stimulation, a situation that would be similar to overeating.[44] In addition, at least one study has found an acute increase in leptin levels following the induction of hyperinsulinemia (but only in women, not in men).[45] It is likely that insulin is at least one regulator of the *LEP* gene and its secretion of leptin. Thus, this is a physiologic feedback loop to maintain weight and energy.

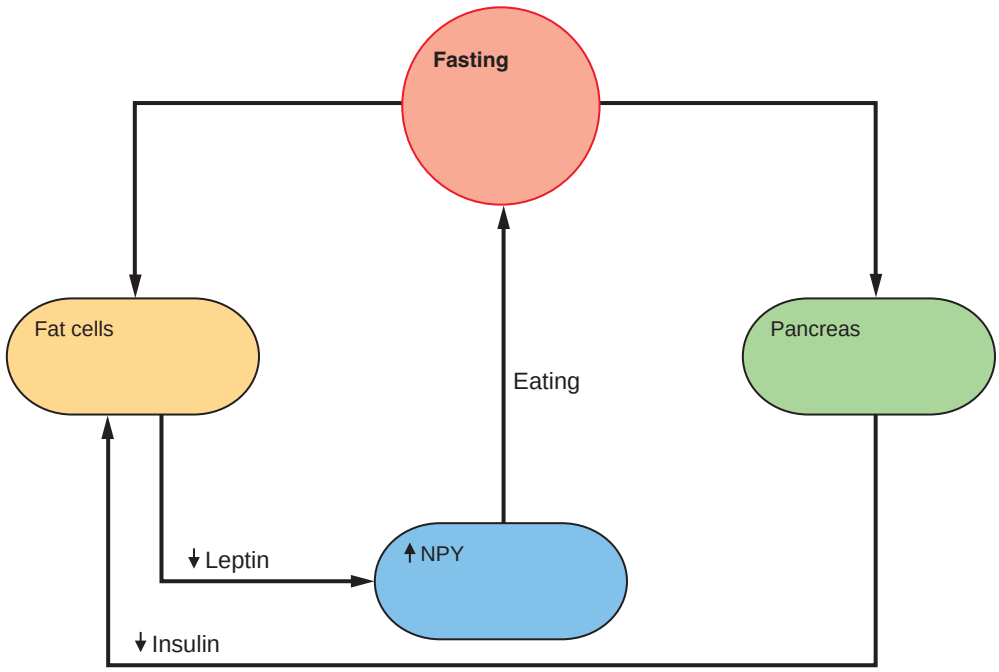
The Ay/Ay rodent becomes fat late in life. The Agouti gene encodes a protein produced by hair follicles. This protein binds to a melanocortin receptor in melanocytes in the skin, thus preventing melanocyte-stimulating hormone action. Agouti rats have high levels of this protein and have yellow instead of black fur. Melanocortins have been demonstrated in the hypothalamus, produced in the arcuate nucleus. Melanocortin binding to its brain receptor influences appetite.[46, 47] Knockout mice for the melanocortin receptor become obese, and these mice have high NPY expression. Ay/Ay rodents make an excess of Agouti protein, blocking melanocortin action. Thus, too little melanocortin could be another pathway for obesity. The importance of this pathway is evident in the report of mutations within the proopiomelanocortin (POMC) gene characterized by deficiencies in both melanocyte-stimulating hormone (MSH) and ACTH, yielding individuals with obesity, adrenal insufficiency, and red hair pigmentation.[48] These findings have culminated in the recognition that mutations in the melanocortin receptor gene account for the most common cause of familial human obesity that can be attributed to a single gene.[49]

Fasting and exercise decrease leptin secretion and increase NPY gene expression in the arcuate nucleus, followed by release of NPY by neuronal projections into the paraventricular nucleus. The neurons that respond to NPY originate in the arcuate nucleus and project into the paraventricular and dorsomedial nuclei. The arcuate nucleus lies outside the blood-brain barrier (there is no blood-brain barrier in the medial basal hypothalamus) and can be

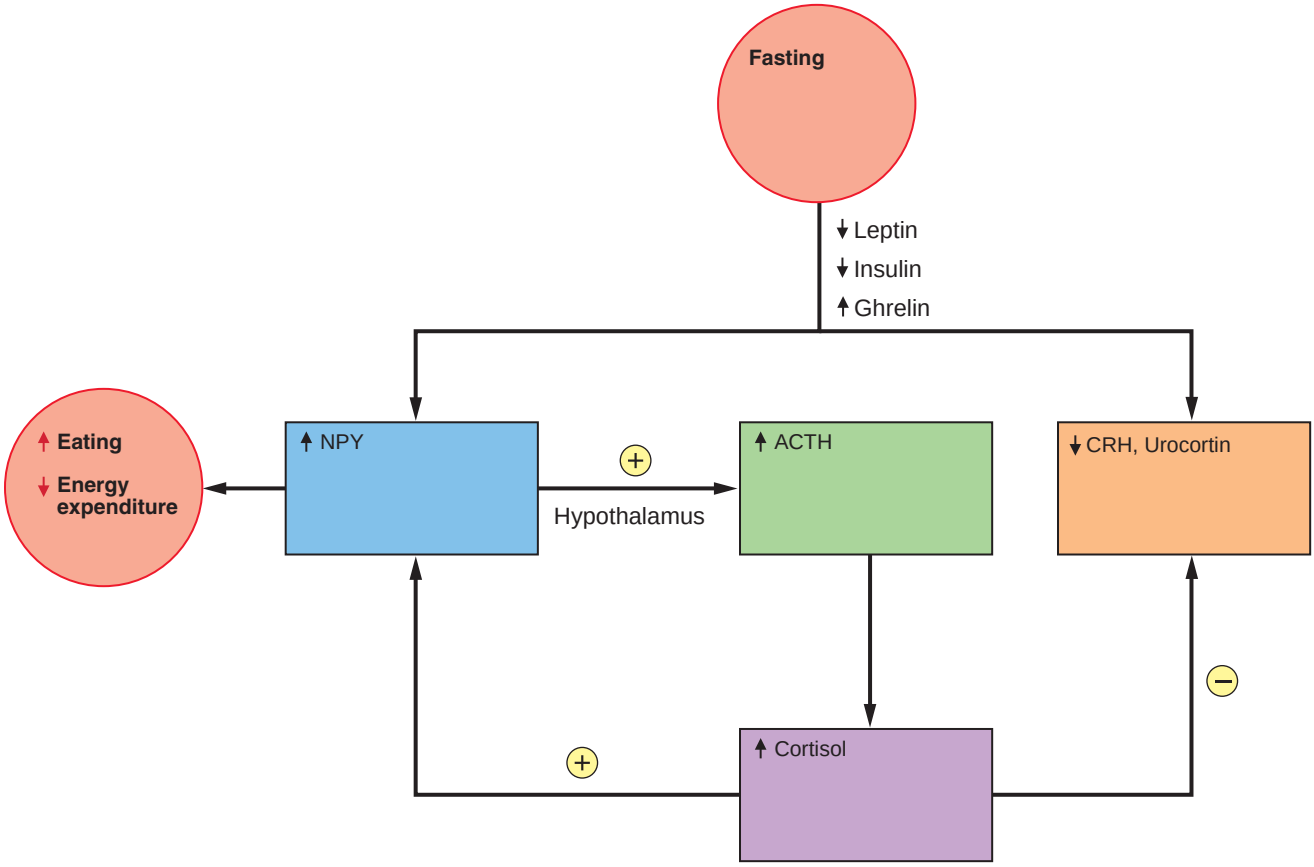


reached by leptin in the circulation. The NPY neurons stimulate feeding and inhibit heat production by inhibiting sympathetic nervous activity. In ob/ob mice with no leptin, NPY levels are high in the hypothalamus, and leptin treatment lowers NPY and restores everything to normal.[38]

In rodents, caloric restriction increases the expression of NPY in the arcuate nucleus and the release of NPY in the paraventricular nucleus. Leptin and insulin decrease during fasting, allowing an increase in the expression of the *NPY* gene. Because an increase in metabolism is unwanted during fasting, CRH decreases; however, there is an increase in cortisol that is a consequence of an NPY-induced hypothalamic signal, not yet identified. Fasting decreases leptin more than expected by the decrease in fat content, indicating the presence of other control mechanisms.[50]



In normal mice, leptin decreases NPY.[51] In most animal models, the NPY content in the hypothalamus is high with obesity, and brain infusions of NPY cause obesity and increased *Ob* gene expression.[52] Knockout mice deficient in NPY maintain normal weight and respond to leptin, indicating that NPY is part of a redundant system, not an absolute requirement.[53]



In the human, the *Ob* gene, known as the *LEP* gene, does not appear to be acutely regulated.[54] Feeding increases leptin levels, but slowly, correlating with insulin levels.[55, 56] Insulin provides negative feedback to the brain in a manner similar to leptin, including an effect on NPY.[43] The hyperphagia of diabetes mellitus reflects a deficiency in this insulin action. In fa/fa rats (obese because of a mutation in the leptin receptor), insulin does not affect neuropeptide Y expression, indicating that the inhibition of NPY exerted by insulin is mediated through the leptin-signaling system.[57]

CRH inhibits food intake and increases energy expenditure; thus, it is expected that weight loss would decrease CRH. In addition, leptin stimulates *CRH* gene expression (through NPY, POMC, and the melanocortin receptor pathway), and, therefore, lower leptin levels with fasting should lower CRH levels.[57] And indeed, CRH secretion is not increased after acute weight loss; however, it is well recognized that cortisol secretion is increased with stress and exercise. This may be due to a hypothalamic effect of NPY on ACTH secretion by the pituitary.[57] The specific food and energy response associated with CRH may also be mediated by urocortin, a CRH-related peptide.[58] Urocortin is very potent in reducing food intake, perhaps activating the energy system without activating the overall CRH mechanism.

In summary, circulating levels of leptin provide the hypothalamus with two important pieces of information: how much energy is available, stored in fat, and recognition of acute increases and decreases in energy intake leading to changes in appetite and energy expenditure.

Ghrelin

Ghrelin is a complex hormone, named for its ability to stimulate the release of growth hormone. Ghrelin participates in the regulation of food intake and energy metabolism, but also affects sleep and behavior, inhibits gonadotropin secretion, and is expressed in the ovaries and the placenta.[59, 60] It influences a wide spectrum of activities, including gastric motility and acid secretion, activity of the pancreas, and the secretion of growth hormone, prolactin, and ACTH.

Ghrelin is a 28-amino acid peptide discovered in 1999 that is secreted mainly in the upper portion of the stomach, but in other tissues as well, including the intestine, pituitary, hypothalamus, kidney, ovary, testis, and placenta.[61] The ghrelin circulating in the blood is predominantly from the stomach and intestine, and its target is the energy-regulating centers in the hypothalamus. Given to rodents, ghrelin acutely increases food intake and causes obesity. These actions of ghrelin are independent of its activity in stimulating growth hormone, prolactin, and ACTH secretion, an action that is probably mediated by different ghrelin receptors.[62] The targeted neurons are in the same network influenced by leptin (principally the NPY and the endogenous melanocortin receptor antagonist, agouti-related peptide, pathways), with ghrelin and leptin having opposing actions. Another stomach hormone, PYY, acts in a similar fashion as leptin, reducing food intake by modifying the NPY system; this leaves ghrelin as the only hormone known that stimulates food intake.[63]

The circulating level of ghrelin is lower in obese individuals, reduced with food intake and increased with fasting. Thus ghrelin levels are higher in individuals with anorexia, bulimia, or cachexia; and, in contrast to its appetite-stimulating action in healthy individuals and in patients with cancer cachexia, administration of ghrelin intravenously to young women with anorexia nervosa does not increase appetite.[64–66] These changes are opposite to those of leptin. Ghrelin is a signal to conserve energy by increasing appetite; leptin (and insulin) is a signal to expend energy.[64] The key regulator in these responses is the circulating glucose level, but in the case of ghrelin, the system is influenced by meals whereas leptin is modulated by fat mass. ***The connection between reproduction and the body's state of energy metabolism is now well established, and the complex leptin-ghrelin system is the means of communication. Specifically, ghrelin exerts a CNS inhibition to prevent reproduction in the presence of an energy deficit, with leptin and ghrelin operating as reciprocal regulators.***[67]

Adiponectin

Adiponectin is another polypeptide secreted by adipose tissue. It influences glucose regulation and fatty acid metabolism; adiponectin circulating levels are relatively high and are inversely correlated with percent body fat.[68–70] Overall adiponectin exerts a weight-losing effect (thus, levels are decreased in obese individuals), complementing the action of leptin in the brain. Adiponectin and its receptors are additional sites where genetic mutations may contribute to obesity. Genetic variants of the adiponectin gene have been described that are associated with the metabolic syndrome and diabetes mellitus.[71, 72]

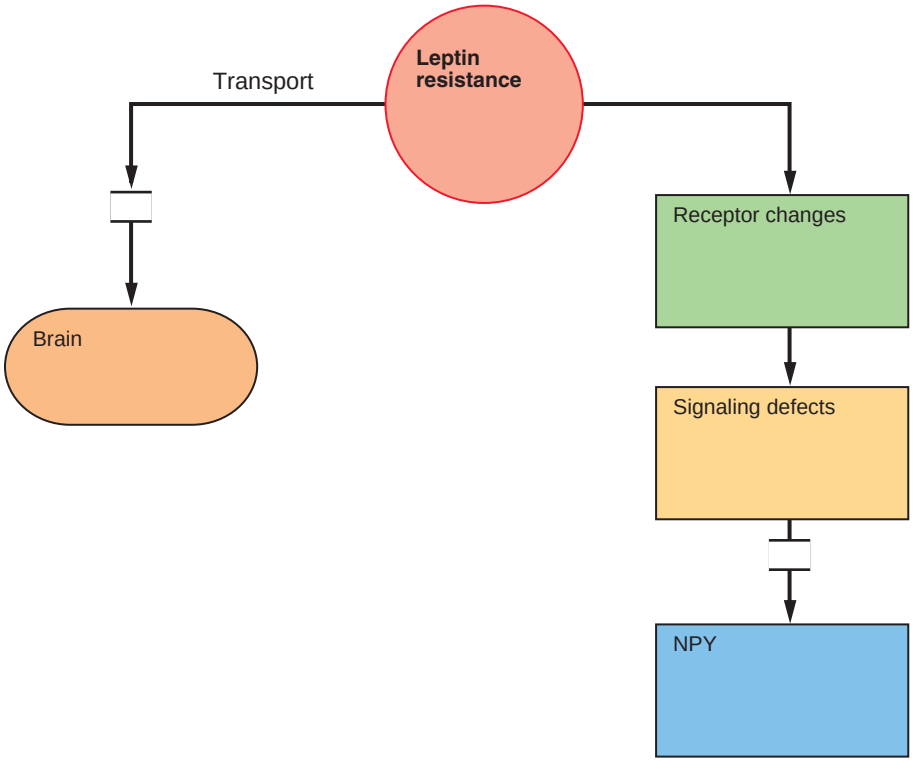
Leptin in Obese People

Most studies have indicated that nearly all obese individuals have elevated leptin levels, probably due to an increase in *LEP* gene expression and partly due to greater production because of larger fat cells.[73–75] The greater weight of these individuals represents the leptin level at which stability is achieved, when leptin resistance is overcome. In lean individuals,

leptin levels are similar in men and women, but in women, as weight increases, leptin increases 3 times more rapidly than in men.[45] Higher levels in women suggest a greater resistance to leptin, correlating with a greater prevalence of obesity.[76]

The incidence of obesity in black and Hispanic women is greater than that of white women.[1] Obese postmenopausal black women have 20% lower leptin levels than white women.[77] These lower leptin levels correlate with a lower resting metabolic rate in the black women. The lower levels may indicate a greater sensitivity to the leptin mechanism in black women.

In obese individuals, fat cells produce leptin normally and a prevalent genetic failure has not been identified. ***Thus it is hypothesized that obesity is due to resistance to leptin.***[78] This may be due to a transport problem of leptin into the brain, as indicated by the finding that leptin level differences between obese and lean individuals are greater in blood than in cerebrospinal fluid.[79, 80] At least in the mouse, this resistance is present only in the periphery in that leptin administered in the brain still works.[81]



In obese individuals, the majority of leptin is unbound and presumably active, consistent with the resistance hypothesis.[82] In lean individuals, the majority of circulating leptin is bound. Another important subject for research is the regulation of leptin binding to proteins in the circulation.

Why is Weight Loss so Hard to Maintain?

The average adult eats nearly 1 million calories per year.[36] When life was tougher, leptin served the purpose of meeting the threat of starvation. During periods of food availability, individuals with leptin defects could consume large amounts and store excess fat. Today, with no shortage of food, the individuals who survived during food shortages in the past

now become obese, and succumb to the complications of obesity. But the most vexing problem is that 90% to 95% of people who lose weight subsequently regain it.[83]

Weight loss in obese and lean people produces the same response, decreased leptin and insulin, and an increase in ghrelin. The obese person tends to regain weight because the lower leptin started from a different set point and is now lower than what the body views as required to be stable in terms of the amount of fat. The ghrelin response also makes it more difficult by stimulating appetite. When energy expenditure, food intake, and body weight are in balance, leptin and ghrelin are at a level consistent with a set point determined by this balance. With weight loss and loss of fat, leptin decreases and ghrelin increases, causing an increase in appetite and a decrease in energy. This is great when you lose weight with an illness, but not good for an overweight person trying to lose weight. Whenever a perturbation occurs, leptin and ghrelin levels change to restore the original status quo, making it difficult for overweight people to maintain weight loss. With both sustained weight loss and persistent weight gain, hormonal mechanisms as well as energy adjustments to maintain lower or higher body weights eventually lead to new set points. However, in the short-term, weight loss can be maintained only by a strict diet or an increase in physical activity to overcome the body's attempt to restore the original set point.

Circulating levels of leptin are correlated with the percent body fat.[73] In other words increased body fat increases the expression of the *LEP* gene in fat cells. The amount of leptin in the circulation, therefore, is a measure of the amount of adipose tissue in the body. For this reason, neither baseline levels nor initial changes in leptin predict whether weight loss can be maintained.[84]

A *reduction* of 10% in body weight is associated with a 53% reduction in serum leptin.[73] This would stimulate an effort to regain the weight. The key question is what happens if an individual can successfully maintain the lower weight. In a longitudinal study of obese people, when weight loss was maintained, leptin levels remained low.[85]

Therefore, whenever a perturbation occurs, leptin and ghrelin levels change to restore the original status quo. Thus, these responses work against attempts to lose weight. A change in energy intake causes a change in leptin levels.[57] The body then changes appetite and energy expenditure to conform to the new leptin level. A 10% increase in body weight is associated with a 300% increase in serum leptin.[55] The basic purpose is to conserve energy during periods of fasting and to avoid obesity during periods of excess. When caloric intake is reduced, the basal metabolic rate is reduced in a regulatory compensatory adaptation that makes maintenance of weight loss difficult.

Congenital Leptin Deficiency

The *LEP* gene (*Ob* gene in mice) has been sequenced from hundreds of obese individuals and mutations are rare.[78, 86]

The first definitive demonstration of a congenital leptin deficiency in humans was reported in two severely obese Pakistani children who were cousins.[87] Despite their mass of fat, their serum leptin levels were very low, and the molecular biology study of a fat biopsy revealed a homozygous deletion of a single guanine in the leptin gene. This mutation results in the introduction of 14 aberrant amino acids into the leptin peptide followed by a premature truncation. All four parents were heterozygotes. These children had normal birth weights, but immediately began gaining excessive weight with marked increases in appetites. In contrast to ob/ob mice, they did not have elevated cortisol levels; however they were hyperinsulinemic. Three obese members of a Turkish family were reported with an amino acid

substitution that impairs intracellular transport of leptin.[88] It is noteworthy that these individuals also displayed suppression of gonadal function. Congenital leptin deficiency is now recognized as a rare recessive inherited disorder, the result of mutations at various sites on the *LEP* gene and associated with hyperphagia and the onset of obesity at an early age.[89] Polymorphisms have been described in both the leptin gene and the leptin receptor gene, more prevalent in extremely obese patients, but a relatively low overall prevalence in obese individuals.[90]

With the complexity of the system for energy regulation and the large number of hormones and enzymes involved, we can expect to see multiple mutations identified. An obese individual was reported with a mutation in prohormone convertase (an enzyme that participates in the conversion of prohormones into hormones as does the carboxypeptidase E enzyme deficient in the fat/fat mouse).[91] Linkages to obesity have been described to regions near the leptin gene or the leptin receptor gene, perhaps indicating differences in regulatory elements of these genes.[92] Remember that melanocortin peptides, along with neuropeptide Y, are part of the leptin signaling pathway in the hypothalamus. Mutations in the melanocortin 4 receptor gene[93] are said to be the most frequent monogenetic cause of obesity, yet the 1% or 2% prevalence of this mutation in obese individuals is insufficient to warrant screening.[94] It is noteworthy that obese children with melanocortin 4 receptor gene mutations were able to lose weight with lifestyle interventions, but had greater difficulty in maintaining their weight loss.[95]

It is expected that only a small percentage (< 5%) of obese humans will have mutations in the leptin receptor or the LEP gene, but congenital leptin deficiency of the leptin receptor should be considered in children with hyperphagia and severe obesity.

Leptin and Reproduction

Several observations support a role for leptin in reproductive physiology.

1. Leptin administration accelerates the onset of puberty in rodents.[96]

2. Leptin levels increase at puberty in boys.[97] Indeed, nocturnal levels of leptin increase in monkeys before the nocturnal increase in pulsatile secretion of LH.[98]

3. Low leptin levels are present in athletes and in patients with anorexia and delayed puberty.[99]

4. The ob/ob mouse undergoes normal sexual development, but remains prepubertal and never ovulates; fertility is restored with leptin administration.[96] A similar presentation and response were reported in a child with congenital leptin deficiency.[100]

Leptin levels are greater in females than in males, and in premenopausal women compared with postmenopausal women.[101] In girls, leptin levels are higher and decrease with increasing Tanner stages of puberty.[102] Thus with puberty, there is increasing sensitivity to leptin. Or in another way to look at this relationship, decreasing leptin during puberty may allow greater food intake for growth by lowering the satiety signal.

The effect of leptin on reproduction can be viewed as an additional role in maintaining responses to stress. Weight loss is known to be associated with an increased adrenal response and a decrease in thyroid function; these endocrine changes, along with suppression of the estrous cycle, occur in fasted mice and are reversed by treatment with leptin.[103]

The puzzle is why CRH is elevated in stress amenorrhea (especially that associated with weight loss) in contrast to the decrease with fasting in normal and obese individuals. One possibility is that the decrease in leptin and increase in NPY associated with stress-related weight loss is the expected response, but it is inadequate to suppress the stress-induced increase in CRH. The blunted patterns in amenorrheic athletes support this. The increase in CRH and resulting hypercortisolism further increase metabolism and weight loss.

Athletes with cyclic menses demonstrate a normal diurnal rhythm in leptin levels. However, amenorrheic athletes do not have a diurnal pattern.[104] Both cycling athletes and amenorrheic athletes have low leptin levels (3-fold reduction) that correlate with reduced body fat, but the levels are further lowered by hypoinsulinemia and hypercortisolemia. In addition, amenorrheic athletes have a blunted leptin response to the increase in insulin following meals.

In postmenopausal women, leptin levels decrease with endurance training, and hormone therapy has no effect.[105] This indicates that the gender difference (higher levels in women) is due to a difference in fat content, not a hormonal difference. Because of the tight connection between insulin and leptin levels in mice, and the now well-recognized prevalence of hyperinsulinemia in women with polycystic ovaries, it makes sense to examine leptin levels in these women.

Studies, controlling for weight, detected no differences in leptin levels comparing women with and without polycystic ovaries.[106–109] In women with polycystic ovaries, the relationship between leptin and body weight is maintained. Thus, in contrast to the rodent model, hyperinsulinemia and insulin resistance do not affect leptin levels in these women. However, a role for leptin in the changes associated with polycystic ovaries should not yet be discounted. There may be subtle differences that have biologic consequences. At least one study demonstrated a correlation between leptin levels and 24-h insulin levels in women with polycystic ovaries.[106] Furthermore, a drug that lowers insulin resistance, troglitazone, inhibits transcription of the *LEP* gene.[110]

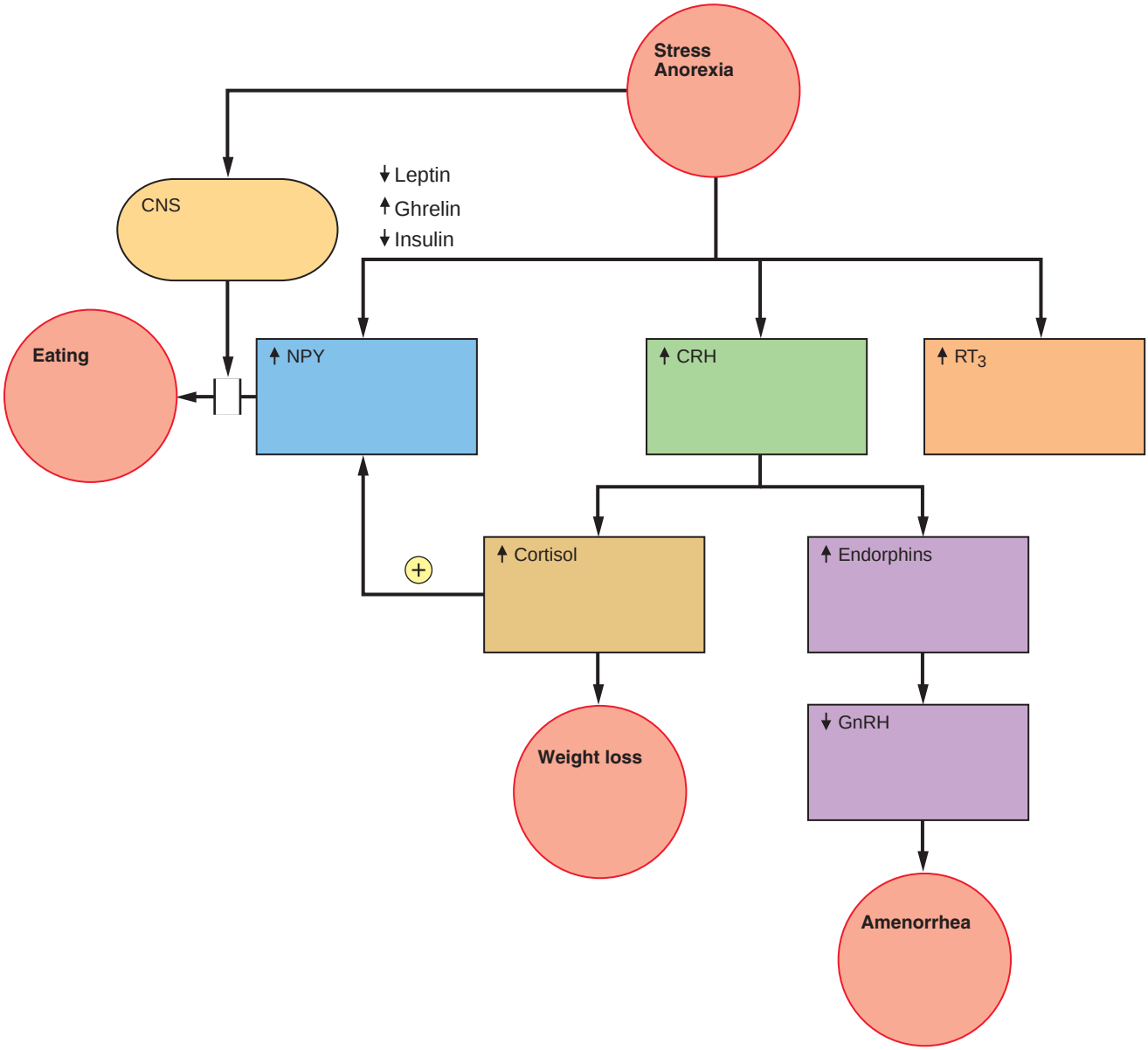
Is it possible that leptin has a target tissue role in reproduction? An isoform of leptin has been identified in the ovary, and leptin exerts specific actions on steroidogenesis as studied in vitro.[111, 112] Leptin inhibits the synergistic action of insulin-like growth factor-I (IGF-I) on FSH-stimulated estradiol production (but not progesterone) in rat granulosa cells, and also inhibits FSH stimulation of IGF-I production. In addition, leptin is expressed in human granulosa and cumulus cells, and is present in mature human oocytes and follicular fluid; thus, leptin appears to be secreted by the ovarian follicle.[113] A rise in maternal serum leptin levels after the administration of human chorionic gonadotropin and before ovum retrieval was correlated with a higher pregnancy rate.[113] Leptin, therefore, is involved in a multitude of important metabolic and developmental functions.

The leptin-ghrelin story has restored credibility to the critical weight hypothesis originally proposed by Rose Frisch in the 1970s.[114, 115] The critical weight hypothesis states that the onset and regularity of menstrual function necessitate maintaining weight above a critical level, and, therefore, above a critical amount of body fat. It was always a mystery how total body fat could talk with the brain. A mystery no longer! Fat talks to the brain via leptin, and the leptin system affects reproduction. Recombinant human leptin administration to women with hypothalamic amenorrhea secondary to exercise or weight loss was associated with an increase in the levels of gonadotropins, free thyroxine, and insulin-like growth factor, with a resumption of ovarian activity.[116] In addition, ghrelin secretion by the gastro-intestinal tract is involved in this balancing mechanism. Ghrelin levels are increased in anorectic patients, along with elevations in cortisol.[117]

There is a difference, however, between ordinary weight loss and stress-induced (e.g., exercise or psychologic problems such as anorexia) weight loss. In ordinary weight

loss, corticotropin-releasing hormone (CRH) secretion is reduced. In stress-induced weight loss, CRH secretion is increased.

CRH directly inhibits hypothalamic gonadotropin-releasing hormone (GnRH) secretion, probably by augmenting endogenous opioid secretion. Women with hypothalamic amenorrhea (including exercisers and women with eating disorders) demonstrate hypercortisolism (due to increased CRH and ACTH, perhaps augmented by NPY stimulation of ACTH secretion), suggesting that this is the pathway by which stress interrupts reproductive function.[118] The modulating effect of the leptin-ghrelin system appears to be mediated through CRH.[119] In addition, NPY and an endogenous melanocortin receptor antagonist, agouti-related peptide, directly inhibit gonadotropin release by suppressing GnRH.[120] In regard to reproduction, the final pathway is suppression of GnRH, a response to multiple inputs indicating the availability of metabolic fuel. The clinical presentation (inadequate luteal phase, anovulation, amenorrhea) depends on the degree of GnRH suppression. Nevertheless, stress and weight-loss induced suppression of reproductive function is associated with low leptin levels, and administration of leptin improves function.[116, 121] Unfortunately the administration of leptin in these patients further suppresses appetite and increases weight loss.[116]



A unifying hypothesis focuses on energy balance.[122] When available energy is excessively diverted, as in exercise, or when insufficient, as with eating disorders, reproduction is suspended in order to support essential metabolism for survival. Thus reproduction may not be directly affected by the level of body fat; rather, body fat is a marker of the metabolic energy state, and the extremely low leptin levels in anorexic patients are an appropriate attempt to restore appetite, an attempt that fails to overcome the stress-induced increase in CRH and its consequences. From a teleologic point of view, there is sense to these relationships; the responses that assist the body to withstand stress also inhibit menstrual function because a stressful period is not the ideal time for reproduction.

Leptin Summary

Because of the high levels of leptin present in overweight people, leptin function may be limited to an effect at low levels. A low circulating level of leptin may serve as a signal that fat stores are not sufficient for growth and reproduction. Thus low levels would ordinarily stimulate hyperphagia, reduce energy expenditure, and suppress gonadotropin secretion and reproduction. The high levels of leptin and the lack of leptin action associated with excess body weight and fat would then reflect resistance.

Although the leptin-ghrelin mechanism offers the potential for new treatments for obesity, it is not around the corner. Leptin, a polypeptide, cannot be administered orally, and in view of high levels in overweight people, another method must be found to attack the lack of effect, the apparent resistance to leptin. Thus far, it appears that genetic defects in the *LEP* gene are uncommon; nevertheless, for affected individuals, a leptin agonist is therapeutic. Even if this complex system yields new treatments, it is unlikely that we will be able to ignore eating appropriately and exercising adequately.

Inherited Aspects of Obesity

Fat cells develop from connective tissue early in fetal life. An important question is whether new fat cells are produced by metaplasia in the adult, or whether an individual achieves a total complement during a certain period of life. In other words, is excess fat stored by increasing the size of the fat cell, or by increasing the number of cells? The possibility arises that there is an inherited increase in the total number of fat cells, which just wait to be packed full of storage fat. Furthermore, the total number of fat cells may depend on an infant's nutritional state during the neonatal period and perhaps in utero as well.

Studies of fat obtained at surgery indicate that the mean fat cell volume is increased 3-fold in obese people, but an increase in the number of fat cells is seen only in the grossly obese. When patients diet, the fat cells decrease in size but not in number. Hypercellular obesity may be a more difficult problem to overcome, because an individual may be saddled with a permanent increase in fat cells.

Some researchers think that, at some period in a person's life, a fixed number of fat cells is obtained. Adolescence, infancy, and intrauterine life seem particularly critical.[123, 124] This premise is not solidly established, because there is no certain way to identify an empty fat cell, and potential fat cells cannot be recognized. Nevertheless, a hyperplastic type of obesity (more fat cells) may be associated with childhood and have a poor prognosis; a hypertrophic type (enlarged fat cells) that is responsive to dieting may occur in adults.

There certainly appears to be a genetic component. The weights of adopted children in Denmark correlated with the body weights of their biologic parents but not with their adoptive parents.[125] This would suggest that the genetic influence is even more important in childhood than environmental factors. Other work suggests that the familial occurrence of obesity can be attributed in part to a genetically related reduced rate of energy expenditure.[126] In studies of identical and fraternal twins reared apart, approximately 70% of the variance in the body mass index could be assigned to genetic influences and the remaining 30% to environmental effects.[127] After the age of 3, obesity in childhood predicts obesity in adulthood, and parental obesity doubles the risk of adult obesity in both obese and non-obese children.[128]

Some argue that each individual has a set point, a level regulated by a signal between the fat cells and the brain (the leptin system). According to this argument, previously obese people who have successfully lost weight have to maintain themselves in a state of starvation (at least as far as their fat cells are concerned).

Genetics and biochemistry are against many obese people. It is best to recognize that an obese individual who has suffered with the problem lifelong does have a disorder. However, for each individual, the extent to which the genetic predisposition is expressed depends on environmental influences. The prevalence of obesity is inversely related to the level of physical activity and education and directly related to parity.[129] Thus, socioeconomic and behavioral factors are important determinants of body weight, and surely each individual will reflect varying impacts of genetics and environment.

Endocrine Changes

The most important endocrine change in obesity is elevation of the basal blood insulin level. The circulating insulin level is proportional to the volume of body fat. Increases in body fat change the body's secretion and sensitivity to insulin, appropriately so because insulin acts to reduce food intake by inhibiting the expression of neuropeptide Y as well as affecting other agents that influence appetite.[57] The effect on neuropeptide Y is believed to be mediated via the leptin signaling system.

Overweight individuals are characterized by insulin resistance. The key factors that affect insulin resistance are the amount of fat tissue in the body, the caloric intake per day, the amount of carbohydrates in the diet, and the amount of daily exercise. At least one mechanism for the increased resistance to insulin observed with increasing weight is down-regulation of insulin receptors brought about by the increase in insulin secretion. The increase in insulin resistance affects the metabolism of carbohydrate, fat, and protein. Circulating levels of free fatty acids increase as a result of inadequate insulin suppression of the fat cell. Insulin resistance results in decreased catabolism of triglycerides, yielding a decrease in HDL-cholesterol and an increase in LDL-cholesterol. This, of course, is a major mechanism for the development of atherosclerosis. Hyperinsulinemia is also directly associated with hypertension. The hyperinsulinemia associated with obesity is reversible with weight loss. ***A significant improvement is achieved with a modest weight loss, only 5–10% of body weight.***

The metabolic syndrome is a serious complication of excess body weight; 60% of obese men and women have the metabolic syndrome.[11] Individuals with the metabolic syndrome have a very high risk of diabetes and cardiovascular disease. ***The diagnosis of the metabolic syndrome in an individual requires that three abnormal findings are present out of the five following clinical characteristics[130]:***

Hypertension	**—**	**130/85 or higher**
Triglyceride levels	**—**	**150 mg/dL or higher**
HDL-cholesterol levels	**—**	**less than 50 mg/dL**
Abdominal obesity	**—**	**greater than 35 inches (40 inches in men) waist circumference**
Fasting glucose	**—**	**110 mg/dL or higher**

Because of the variability associated with measurements of insulin, the fasting glucose to fasting insulin ratio is no longer recommended to establish the presence of insulin resistance; a 2-h oral glucose tolerance test is now the preferred method of assessment, with measurement of 2-h glucose and insulin levels after a 75-g glucose load.

Interpretation of the 2-h glucose response:

Normal	**—**	**less than 140 mg/dL**
Impaired	**—**	**140–199 mg/dL**
Noninsulin-dependent diabetes mellitus	**—**	**200 mg/dL and higher**

Interpretation of the 2-h insulin response:

Insulin resistance very likely	**—**	**100–150 U/mL**
Insulin resistance	**—**	**151–300 U/mL**
Severe insulin resistance	**—**	**greater than 300 U/mL**

It is impossible to predict exactly who eventually will develop diabetes because the tendency is recessive, and it will not develop in every generation in a family. But weight is a good tip-off. As weight increases, the frequency of occurrence of diabetes increases. Both gestational diabetes and insulin-dependent diabetes are more common in overweight pregnant patients.

Contrary to popular misconception, hypothyroidism does not cause obesity. Weight gain due to hypothyroidism is confined to the fluid accumulation of myxedema. There is no place, therefore, for thyroid hormone administration in the treatment of obesity when the patient is euthyroid.

Obese people are relatively unable to excrete both salt and water, especially while dieting. During dieting, this seems to be mediated by increased output of aldosterone and vasopressin. Because water produced from fat outweighs the fat, people on diets often show little initial weight loss. The early use of a diuretic may encourage a patient to persist with dieting.

The basic question is whether metabolic changes observed in obesity represent adaptive responses to a markedly enlarged fat organ or whether they are representative of a metabolic or hormonal defect. The former is true. These changes are secondary responses; they are totally reversible with weight loss. Four-year follow-up in a group of patients who did not regain their weight after dieting revealed persistently normal insulin and glucose responses; patients who regained their weight showed further deterioration in these metabolic factors.[131]

Anatomic Obesity

Gynoid obesity (the pear shape) refers to fat distribution in the lower body (femoral and gluteal regions), whereas android obesity (the apple shape) refers to central body distribution. Gynoid fat is more resistant to catecholamines and more sensitive to insulin than abdominal fat; thus, extraction and storage of fatty acids easily occur, and fat is accumulated more readily in the thighs and buttocks. This fat is associated with minimal fatty acid flux, and, therefore, the negative consequences of fatty acid metabolism are less. Gynoid fat is principally stored fat. The clinical meaning of all this is that women with gynoid obesity are less likely than women with android obesity to develop diabetes mellitus and coronary heart disease.[132]

During pregnancy, lipoprotein lipase activity increases in gynoid fat, further promoting fat storage and explaining the tendency for women to gain thigh and hip weight during pregnancy. Also, because this fat is more resistant to mobilization, it is harder to get rid of. This difficulty is related to the adrenergic receptor concentration in the fat cells, the regulation of which remains a mystery.

Android obesity refers to fat located in the abdominal wall and visceral-mesenteric locations. This fat is more sensitive to catecholamines and less sensitive to insulin and, thus, more active metabolically. It more easily delivers triglyceride to other tissues to meet energy requirements. This fat distribution is associated with hyperinsulinemia, impaired glucose tolerance, diabetes mellitus, an increase in androgen production rates, decreased levels of sex hormone-binding globulin, and increased levels of free testosterone and estradiol.[133,134] In addition, women with central obesity have decreased cortisol levels, a finding that would be consistent with increased leptin levels.[135, 136] These metabolic changes improve with weight loss.[136]

It is central body obesity that is associated with cardiovascular risk factors, including hypertension and adverse cholesterol-lipoprotein profiles.[137] The waist/hip ratio is a variable strongly and inversely associated with the level of HDL_2, the fraction of HDL-cholesterol most consistently linked with protection from cardiovascular disease.[138] The adverse impact of excess weight in adolescence can be explained by the fact that deposition of fat in adolescence is largely central in location.[14, 139] Weight loss in women with lower body obesity is mainly cosmetic, whereas loss of central body weight is more important for general health in that an improvement in cardiovascular risk is associated with loss of central body fat. At any given level of body mass index (BMI), an increase in central, android fat increases the risk of cardiovascular disease.

Estimating Central Body Obesity

The waist/hip ratio is a means of estimating the degree of upper to lower body obesity; the ratio accurately predicts the amount of intraabdominal fat (which is greater with android obesity).[140, 141] However, studies have demonstrated that the more easily determined circumference of the waist, measured just above the hip bones at end of expiration, is a better predictor of central, android abdominal fat.[142, 143] ***A waist circumference greater than 102 cm (about 40 inches) in men and 88 cm (about 35 inches) in women is predictive of abnormal endocrinologic and metabolic function and is associated with a major increase in risk of cardiovascular and metabolic diseases.[142] The risk begins to increase at threshold measurements of 94 cm for men and 80 cm for women.[144, 145]***

Management of Obesity

In addition to not smoking cigarettes, weight reduction is the most important health measure available for reducing the risk of cardiovascular disease. After adjusting for age and smoking, the Nurses' Health Study documented a 3-fold increase in risk for coronary disease among women with a body mass index of 29 or greater.[146] Even women who are mildly or moderately overweight have a substantial increase in coronary risk. In the Nurses' Health Study, 40% of coronary events could be attributed to excessive body weight, and in the heaviest women, 70%. But most importantly, weight loss is followed by a decrease in mortality from all causes, and, especially, a decrease in cardiovascular and cancer mortality.[147]

For most patients, after a routine evaluation to rule out pathology such as diabetes mellitus, the clinician is left with the frustrating task of prescribing a diet. But it is not enough to just prescribe a diet or prescribe an anorectic drug. An effective weight loss program requires commitment from both patient and clinician.

Clinician and patient should agree on the goal of a diet program. Although the clinician may want the patient to reach ideal weight, the patient may be satisfied with less. Motivation is improved when the goals meet both personal and medical objectives. It is realistic to lose 4–5 lb in the first month and 20–30 lb in 4–5 months. To achieve a respectable rate of weight loss, intake must be 500–1,000 calories below energy expenditure.[148] But as weight is lost, energy requirements decrease; therefore, unless energy intake decreases, the rate of weight loss will slow. Clinicians and patients need to establish reasonable goals, and only modest changes in diet and activity are necessary.

Despite various fads and diet books, the best diet continues to be a limitation of calories to between 900 and 1,200 calories/day, the actual amount depending on what the individual patient will accept and pursue. When energy intake is less than this, it is very difficult to obtain the recommended levels of vitamins and minerals. A daily vitamin and mineral supplement should be used with very low calorie diets. Randomized, controlled trials concluded that the low-carbohydrate, high-protein, high-fat diet produced a greater weight loss (only a 4-kg difference) in the first 6 months, but after a year there was no difference compared to a conventional low-calorie, low-fat diet.[149, 150] The low-carbohydrate diets produce ketosis, causing significant halitosis.

Ideal diet:	Carbohydrates	— 50%
	Protein	— 15–20%
	Fat	— less than 30%

The discouraging aspect is that to lose a pound of fat, the equivalent to a 3,500-calorie intake must be expended. Dieting has to be slow and steady to be effective. Successful programs include behavior modification, frequent visits to the clinician, and involvement of family members. Behavior modification starts with daily recording of activity and behavior related to food intake, followed by the elimination of inappropriate cues (other than hunger) that lead to eating.

Careful studies (performed in hospitalized subjects on metabolic wards) have indicated that the carbohydrate and fat composition of the diet has no effect on the rate of weight loss.[151] Restriction of calories remains the important principle, recognizing that reduction in fat intake is the most effective method of weight loss. The protein-sparing modified fast is a ketogenic regimen providing approximately 800 calories/day. Unsupplemented liquid protein diets have been associated with deaths due to cardiac arrhythmias. The low-calorie diets that use protein and carbohydrate supplemented with minerals and vitamins as the sole source of nutrition are safer but should be used only for severe obesity and under medical supervision.[152] These diets are still potentially dangerous. The other disadvantage to the semistarvation diet is that short-term success does not guarantee long-term weight maintenance. It is reported that at best only one-fourth to one-third of individuals who lose weight by a semistarvation ketogenic regimen plus behavior modification therapy have significant long-term weight reduction.[153] On the other hand, for that one-fourth to one-third, this represents a major accomplishment and is worth doing. Unfortunately, repeated dieting and recidivism have a negative impact. With each episode, the body learns to become more efficient, so that with each diet, weight comes off more slowly and is regained more rapidly.

It is not unusual to encounter patients who claim to be unable to lose weight despite following a diet with less than 1,200 calories/day. In a study of such patients, it was discovered that underreporting of actual food intake and overreporting of physical activity are both common.[154] While it may not be true for all patients, certainly some individuals do eat more

than they think and exercise less than they report to their clinicians. This is not a deliberate conscious attempt to deceive the clinician. These patients truly believe their resistance to weight loss is genetic and not due to their own personal behavior. They are astonished and distressed to learn the results of accurate recording of dietary intake and physical exercise. The help of a dietitian in recording a typical week's worth of eating and exercise is worthwhile. This kind of knowledge proves to be a powerful lever in providing the motivation to make the changes in lifestyle that can yield loss of weight.

Although most efforts yield short-term success, maintenance of weight loss is uncommon. Commercial organizations are no more successful than physician-directed programs or nonprofit self-help groups.[155, 156] However, randomized trials demonstrated that commercial programs or weight-loss medications that include group support and lifestyle changes are more effective, achieving modest but clinically significant improvements.[157–161] Approximately 90 to 95% of people who lose weight subsequently regain it.[83] Thus, it is obvious why gimmicks abound in this area of patient management. A more reasonable attitude is to emphasize how much can be gained with only a little weight loss. A weight loss of 5–10% of body weight produces beneficial effects on the risks of cardiovascular disease and diabetes mellitus.[147, 162] *Compliance and caloric intake are more important than the specific dietary composition or reputation.* In the U.S. Diabetes Prevention Program randomized clinical trial, intensive lifestyle interventions were more effective than prophylactic metformin treatment in reducing the incidence of diabetes.[163] Ten years after the initiation of the trial, the cumulative incidence of diabetes was reduced by 34% by lifestyle interventions and 18% in the metformin group.[164] These data demonstrate that there are substantial numbers of overweight individuals who can sustain long-term weight reduction when effective support and guidance are provided.

Unfortunately, smoking cessation is associated with an increase in being overweight.[165] And, of course, the link between smoking and weight control is exploited by the tobacco industry in its advertising. This should only increase the importance of our efforts to keep young people from starting smoking and to educate middle-aged people regarding the dangers of smoking.

Weight Loss Drugs

Sympathomimetic Noradrenergic Agents (amphetamine-related drugs):		
	Diethylpropion (Tenuate)	25 mg before meals; 75 mg slow release form in morning
	Phentermine	8 mg before meals; 15 or 37.5 mg in morning
	Phendimetrazine (Bontril)	35 mg before meals; 105 mg slow-release form daily
	Benzphetamine (Didrex)	25–50 mg 1–3 times daily
Noradrenergic and Serotonergic Agent (Serotonin and Norepinephrine Reuptake Inhibitor):		
	Sibutramine (Meridia)	10–15 mg daily
Serotonergic Agent (Serotonin Reuptake Inhibitor):		
	Fluoxetine	60 mg daily
Lipase Inhibitor		
	Orlistat (Xenical)	120 mg t.i.d
Cannabinoid Receptor Antagonist		
	Rimonabant	5 or 20 mg daily

Because of CNS and cardiovascular side effects (e.g., insomnia, nervousness, euphoria, hypertension, and tachycardia) and the lack of long-term data, the noradrenergic agents should be reserved for short-term use (3 months or less) for individuals who wish to lose

a small amount of weight. Tolerance and dependence are problems with continued use. Noradrenergic agents should not be used in individuals with cardiovascular disease. The popularity of the serotonergic agents received a setback with reports of increased risks due to high circulating levels of serotonin (even with short-term therapy) of the rare but life-threatening conditions of primary pulmonary hypertension and cardiac valvular disease, and two products, fenfluramine and dexfenfluramine, were withdrawn from the market.[166–168]

Sibutramine (10 or 15 mg daily) blocks the neuronal uptake of both norepinephrine and serotonin, and is an effective appetite suppressant with mild side effects (dry mouth, constipation, and insomnia), but it can cause hypertension, and close monitoring of blood pressure and pulse is necessary.[169] Sibutramine should not be used along with other serotonin reuptake inhibitors. Fluoxetine produces only a short-term weight loss.[170] Orlistat (120 mg t.i.d. with meals) acts only within the gastrointestinal tract, inhibiting pancreatic lipase and increasing fecal fat loss.[171] Orlistat is associated with annoying gastrointestinal side effects (increased flatulence, cramps, and defecation of oily stools). There is some loss of the fat-soluble vitamins, and a vitamin supplement should be taken at bedtime. Rimonabant, a cannabinoid receptor blocker, has produced a modest weight loss, sustained over 2 years.[172, 173]

Because weight is regained after discontinuing drug treatment, long-term therapy has been championed. However, very little information is available regarding the long-term use of appetite-suppressing drugs. Short-term studies indicate only a modest efficacy with variable responses.[83, 174] Most of the weight loss occurs in the first 6 months and is limited to 5–10 kg (11–22 lb). Nevertheless, long-term treatment may enable some individuals to sustain weight loss and to more effectively make beneficial changes in diet and lifestyle. It is recommended that treatment with appetite-suppressing drugs should be limited to individuals who have failed to lose weight with conventional methods and who demonstrate significant comorbidities, such as android obesity, coronary heart disease, insulin resistance, and hypertension.[83] *Attainment of a normal body weight is unlikely with drug treatment; however, a 5 to 10% weight loss will have an important beneficial impact on risk factors for disease*.[83] Nevertheless, lifestyle changes remain the most important and effective intervention for the treatment of obesity.

Surgical treatment and starvation should be reserved for patients who are morbidly obese. Both methods involve many potential problems and require close monitoring. Although there can be major complications, bariatric surgery, including laparoscopic procedures, produces a weight loss of 50 to 75% that is sustained for many years and is associated with a reversal of abnormal metabolic conditions in both adults and children with a reduction in mortality rate.[175–179]

Controlled studies have not demonstrated the effectiveness of thyroid preparations or human chorionic gonadotropin.[180] Indeed, adding thyroid hormone increases the loss of lean body mass rather than fat tissue. It is clear that adjunctive drug measures are not successful unless the patient is also motivated either to limit caloric intake or to increase the exercise level in what will be a lifelong battle.

A regular pattern of physical exercise reduces the risk of myocardial infarction in all people.[181] Both weight loss and increased physical activity lower the level of low density lipoprotein (LDL), and increase the level of high density lipoprotein (HDL).[182] A further benefit of strenuous or prolonged exercise is an inhibition of appetite that lasts many hours and that is associated with an increase in the resting metabolic rate for 2–48 h. The optimal program includes, therefore, a 1-hour, *daily* period of exercise (of moderate intensity, such as brisk walking), but even 2 or 3 times per week can be effective.[183] A combination of diet and exercise is better than either alone, and those who exercise are more successful in maintaining weight loss.[156, 184, 185] The best time for exercise is before meals or about 2 h after eating. Lifestyle changes (exercise, no smoking, healthy diet) lower the risk of coronary heart disease even in obese individuals.[186]

Unfortunately, one cannot burn up significant calories quickly; it takes 18 min of running or 2 h of walking to compensate for the average hamburger.[187]

Activity	Calories per Hour
Sleeping	90
Office work	240
Walking	240
Golf	300
Housework	300
Bicycling	360
Swimming	360
Tennis	480
Bowling	510
Running slowly	750 (ca. 120/mile)
Cross country skiing	840
Running fast	960 (ca. 160/mile)

Most frustrating is the problem of some patients who limit caloric intake yet do not lose weight. In fact, as the weights of certain patients increase, the number of calories required to remain in equilibrium decreases, due to a combination of reduced activity and a change in metabolism, now known to be appropriate consequences of changes in leptin levels. The Vermont study demonstrated that the normal person with induced obesity requires 2,700 calories to remain in equilibrium; spontaneously obese patients require only about 1,300 calories.[188] Others argue that virtually everyone can lose weight on a diet of 1,000 calories/day in that the maintenance requirement for a sedentary adult is about 1.5 times the resting metabolic rate (about 1,000–1,500 calories/day).[189] Nevertheless, an individual who has been overweight requires about 15% fewer calories to maintain weight than an individual who has never been obese.[190] The clinician must be careful to avoid a condemning or punitive attitude and understand that it is possible to significantly restrict caloric intake and not lose weight.

Patients struggle with frustration and despair unless the clinician can motivate them to increase physical activity. In all individuals, dieting is doomed to failure unless it is combined with physical exercise, but this is especially true in chronically obese patients. In other words, the lifestyle of an obese person must be changed to overcome the desire to be inactive (walk instead of riding). Only by significantly increasing caloric expenditure will the input-output equilibrium be disturbed. Unfortunately, only 26% of American adults are engaged in significant leisure time physical activity.[1]

The obese person feels trapped. Obesity leads to characteristic behavioral manifestations, including passive personality, frequent periods of depression, decreased self-respect, and a sense of being hopelessly overwhelmed by problems. But just as the endocrine and metabolic changes are secondary to obesity, many of the psychosocial attributes surrounding obesity are also secondary.

Maintenance of a newly gained lower weight requires constant preventive attention. Motivation to change and emotional support during the change are important. They can be provided by friends, relatives, clinicians, or self-help organizations. If the vicious circle of failed diets, resignation to fate, guilt, and shame can be broken, a more effective, happier person will emerge.

All references are available online at: http://www.clinicalgynendoandinfertility.com

20 Reproduction and the Thyroid

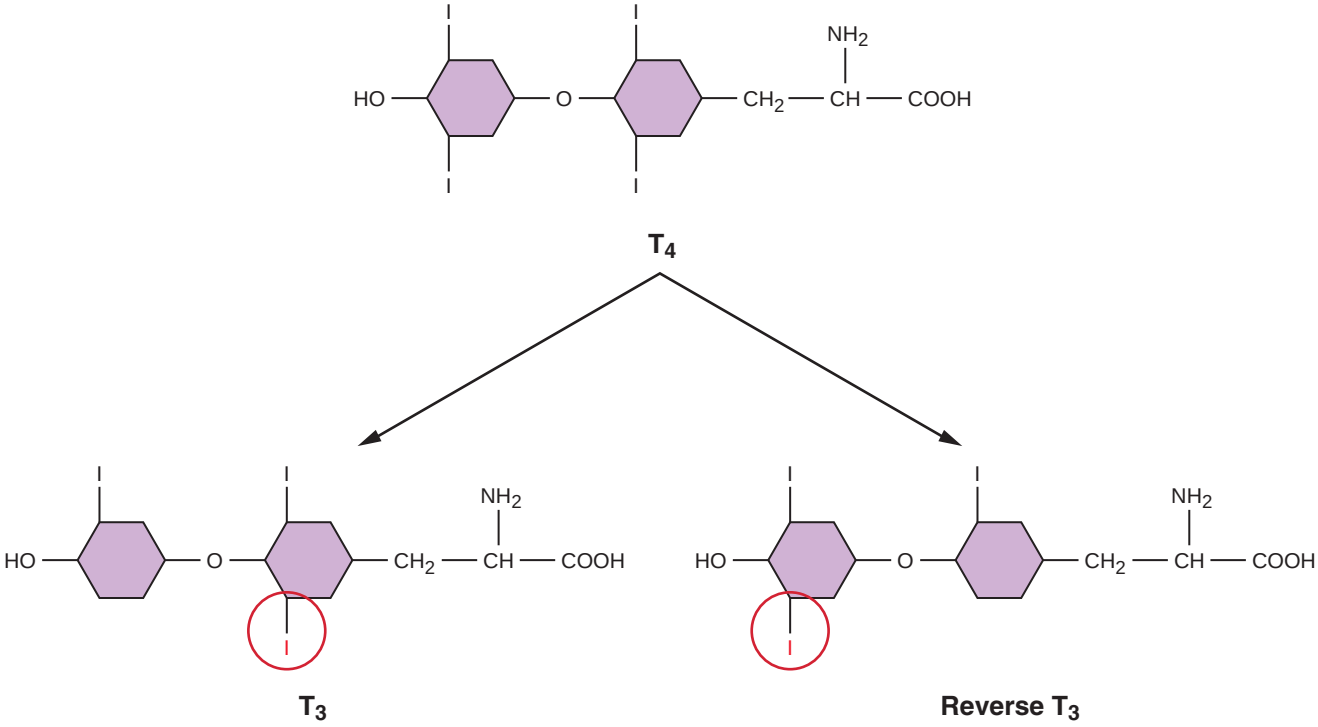
Thomas Wharton, in 1656, gave the thyroid gland its modern name (meaning oblong shield) because he believed the function of the thyroid was to fill vacant spaces and contribute to the shape and beauty of the neck, especially in women.[1] For unknown reasons, thyroid disease is more common in women than in men. Because most thyroid disease is autoimmune in nature, an increased susceptibility to autoimmune diseases, perhaps secondary to the female endocrine environment, is a likely contributing factor.

The clinical objective is to detect and treat thyroid disease before the symptoms and signs are significant and intense. Subtle thyroid disease is easily diagnosed by the sensitive laboratory assessments now available. Therefore, the key to early diagnosis is to maintain a high index of suspicion and to readily screen for the presence of abnormal thyroid function. There is growing support to institute routine thyroid screening in two female populations: older women and pregnant women (preconceptual screening would be even better).

Normal Thyroid Physiology

Thyroid hormone synthesis depends in large part on an adequate supply of iodine in the diet. In the small intestine, iodine is absorbed as iodide that is then transported to the

thyroid gland. Plasma iodide enters the thyroid under the influence of thyroid-stimulating hormone (TSH), the anterior pituitary thyrotropin hormone. Within the thyroid gland, iodide is oxidized to elemental iodine, which is then bound to tyrosine. Monoiodotyrosine and diiodotyrosine combine to form thyroxine (T4) and triiodothyronine (T3). These iodinated compounds are part of the thyroglobulin molecule, the colloid that serves as a storage depot for thyroid hormone. TSH induces a proteolytic process that results in the release of iodothyronines into the bloodstream as thyroid hormone.



Removal of one iodine from the phenolic ring of T4 yields T3, whereas removal of an iodine from the nonphenolic ring yields reverse T3 (RT3), which is biologically inactive. In a normal adult, about one third of the T4 secreted each day is converted in peripheral tissues to T3, and about 40% is converted to the inactive, reverse T3. About 80% of the T3 generated is derived outside the thyroid gland, chiefly in the liver and kidney. T3 is 3–5 times more potent than T4, and virtually all the biologic activity of T4 can be attributed to the T3 generated from it. Although T4 is secreted at 20 times the rate of T3, it is T3 that is responsible for most if not all of the thyroid action in the body. T3 is more potent than T4 because the nuclear thyroid receptor has a 10-fold greater affinity for T3 than T4. Although T4 may have some intrinsic activity of its own, it serves mainly as a prohormone of T3. It is hard to think of a body process or function that does not require thyroid hormone for its normal operation, not only metabolism but also development, steroidogenesis, and most specific tissue activities.

Carbohydrate calories appear to be the primary determinant of T3 levels in adults. A reciprocal relationship exists between T3 and RT3. Low T3 and elevated RT3 are seen in a variety of illnesses such as febrile diseases, burn injuries, malnutrition, and anorexia nervosa. The metabolic rate is determined to a large degree by the relative production of T3 and RT3. During periods of stress, when a decrease in metabolic rate would conserve energy, the body produces more RT3 and less T3, and metabolism slows. On recovery, this process reverses, and metabolic rate increases.

Circulating thyroid hormones are present in the circulation mainly bound to proteins. Approximately 70% of thyroid hormones are bound to thyroxine-binding globulin (TBG), which, therefore, is the major determining factor in the total thyroid hormone concentration in the circulation. The remaining 30% is bound to thyroxine-binding prealbumin and albumin. The binding proteins have a greater affinity for T4 and, thus, allow T3 to have greater entry into cells. TBG is synthesized in the liver, and this synthesis is increased by estrogens. The passage of thyroid hormones into and out of cells is regulated by cell membrane thyroid hormone transporters; mutations in a key transporter are associated with elevated T3 levels and psychomotor retardation.[2]

The nuclear receptor for thyroid hormone is a member of the super family that includes the steroid hormone receptors (Chapter 2).[3] The thyroid hormone receptor exists in several forms, the products of 2 genes located on different chromosomes. The α receptor gene is on chromosome 17, and the β receptor gene is on chromosome 3. The nuclear T3 receptor is truly ubiquitous, indicating the widespread actions of thyroid hormone throughout the body. Mutations in the gene for the thyroid receptor lead to the synthesis of a receptor that actually antagonizes normal receptors, a syndrome of thyroid resistance characterized by elevated thyroid hormone levels. TSH is elevated as well because of the impairment in thyroid hormone action.

The thyroid axis is stimulated by the hypothalamic factor, thyrotropin-releasing hormone (TRH) and inhibited by somatostatin and dopamine. Thyroid hormones regulate TSH by suppressing TRH secretion, but primarily by affecting the pituitary sensitivity to TRH (by reducing the number of TRH receptors). Pituitary secretion of TSH is very sensitive to changes in the circulating levels of thyroid hormone; a slight change in the circulating level of T4 will produce a many-fold greater response in TSH. TSH-secreting cells are regulated by T4, but only after the T4 is converted to T3 in the pituitary cells. Although modulation of thyroid hormone occurs at the pituitary level, this function is permitted by the hypothalamic releasing hormone, TRH. Although some tissues depend mainly on the blood T3 for their intracellular T3, the brain and the pituitary depend on their own intracellular conversion of T4. The measurement of T4 and TSH, therefore, provides the most accurate assessment of thyroid function.

The TSH response to TRH is influenced mainly by the thyroid hormone concentration in the circulation; however, lesser effects are associated with dopamine agonists (inhibition), glucocorticoids (inhibition), and dopamine antagonists (stimulation). Estrogen increases the TRH receptor content of the pituitary; hence, the TSH response to TRH is greater in women than in men, and greater in women taking estrogen-progestin contraceptives.

TRH also stimulates prolactin secretion by the pituitary. The smallest doses of TRH that are capable of producing an increase in TSH also increase prolactin levels, indicating a physiologic role for TRH in the control of prolactin secretion. However, except in hypothyroidism, normal physiologic changes as well as abnormal prolactin secretion can be understood in terms of dopaminergic inhibitory control, and TRH need not be considered.

Functional Changes with Aging

Thyroxine metabolism and clearance decrease in older people, and thyroxine secretion decreases in compensation to maintain normal serum thyroxine concentrations.[4] With aging, conversion of T4 to T3 decreases, and TSH levels increase. The TSH response to TRH is normal in older women. TBG concentrations decrease slightly in postmenopausal women but not enough to alter measurements in serum.

Thyroid Function Tests

Free Thyroxine (FT4)

Assays that measure free T4 are usually displacement assays using an antibody to T4. The result is not affected by changes in TBG and binding. The free T4 level has a different range of normal values from laboratory to laboratory, but is usually 0.8–2.0 ng/dL.

Total Thyroxine (TT4)

The total thyroxine, both the bound portion to TBG and the free unbound portion, is measured by displacement assays, and, in the absence of hormone therapy or other illnesses, it estimates the thyroxine concentration in the blood. However, the free T4 is unaffected by factors that influence TBG and is preferred.

Free Thyroxine Index (FTI or T7)

The free thyroxine index is calculated from the TT4 and the T3 resin uptake measurements. This test has been replaced by the free T4 assay.

Total T3 and Reverse T3

Both of these thyronines can be measured by sensitive immunoassays. However, in most clinical circumstances they add little to what is learned by the free T4 and TSH measurements. The clinical situations in which measurement will be useful are discussed under the specific diseases and indicated on the algorithm.

Thyroid-Stimulating Hormone (TSH)

TSH (also called thyrotropin) is measured by highly sensitive assays using monoclonal antibodies, usually in a technique that uses two antibodies, one directed at the α subunit and one directed at the β subunit of TSH. The normal levels vary from laboratory to laboratory, but the sensitive TSH assay can detect concentrations as low as 0.01 μU/mL, and the normal range is usually 0.45–4.5 μU/mL. White people and the elderly normally have slightly higher TSH levels (individuals 80 years and older may have an upper limit as high as 7.5 μU/mL), making the upper limit for the normal range more difficult to interpret.[5,6] TSH is a very sensitive indicator of thyroid hormone action at the tissue level because it is dependent on the pituitary exposure to T4. In the absence of hypothalamic or pituitary disease, the sensitive TSH assay will provide the best indication of excess or deficient thyroxine; slight changes in T4 are reflected in a many-fold greater response in TSH. Nearly all women with elevated TSH levels have hypothyroidism. Transient changes in TSH can be caused by systemic illnesses, major psychiatric disturbances, and pharmacologic treatment with glucocorticoid agents or dopamine.

TSH-Receptor Antibody

Antibodies that compete with TSH for its receptor are collectively known as TSH-receptor antibodies, known as TRAb. The assays measure the inhibition of TSH binding, but the percentage binding usually correlates with activation of the receptor by stimulating antibodies. Most patients with Graves' disease will have detectable TRAb, TSH-receptor antibodies. This is an essential test in patients with hyperthyroidism to differentiate Graves' disease from non-autoimmune hyperthyroidism, specifically toxic multinodular goiter, in order to select appropriate therapy. This test is not needed to monitor treatment.

Radioactive Iodine Uptake Scan

Because the thyroid gland is the only tissue that utilizes iodine, radioisotopes of iodine can be used as a measure of thyroid gland activity and to localize activity within the gland. The scan is obtained 24 h after the administration of I^{123} given orally.

The Laboratory Evaluation

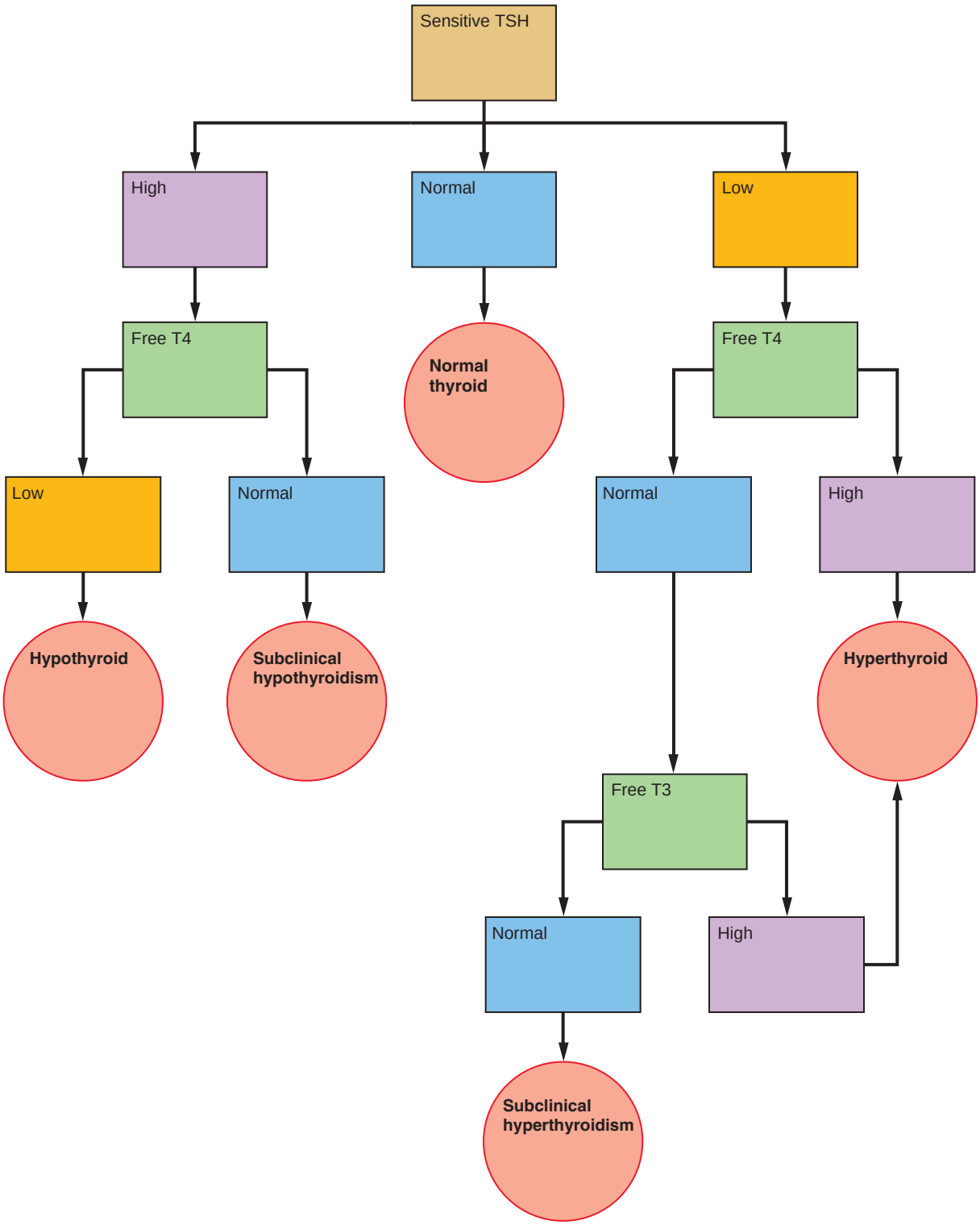
The algorithm represents a cost-effective and accurate clinical strategy. For screening purposes, or when there is a relatively low clinical suspicion of thyroid disease, the initial step is to measure TSH. A normal TSH essentially excludes hypothyroidism or hyperthyroidism. A high TSH requires the measurement of free T4 to confirm the diagnosis of hypothyroidism.

If the initial TSH is low, especially less than 0.08 μU/mL, then measurement of a high T4 will confirm the diagnosis of hyperthyroidism. If the T4 is normal, the T3 level is measured, because some patients with hyperthyroidism will have predominantly T3 toxicosis. If the T3 is normal, this compensated state is called subclinical hyperthyroidism. Some of these patients will eventually have increased T4 or T3 levels with true hyperthyroidism.

Hypothyroidism

In most cases of hypothyroidism, a specific cause is not apparent. It is believed that hypothyroidism is usually secondary to an autoimmune reaction, and, when goiter formation is present, it is called Hashimoto's thyroiditis.[7] Unless abnormal thyroid function can be documented by specific laboratory assessment, empiric treatment with thyroid hormone is not indicated, and it is especially worth emphasizing that thyroid hormone treatment does not help infertility in euthyroid women. Hypothyroidism can be a cause of recurrent miscarriages, and an assessment of thyroid function is worthwhile in these patients.

Hypothyroidism increases with aging and is more common in women.[8] Up to 45% of thyroid glands from women older than age 60 show evidence of thyroiditis.[9] In women admitted to geriatric wards, 2–4% have clinically apparent hypothyroidism. ***Therefore, hypothyroidism is frequent enough to warrant consideration in most older women, justifying screening even in asymptomatic older women. We recommend that older women be screened with the TSH assay every 5 years beginning at age 35, then every 2 years beginning at age 60, or with the appearance of any symptoms suggesting hypothyroidism.***[10]



Menstrual irregularities and bleeding problems are common in hypothyroid women. Amenorrhea can be a consequence of hypothyroidism, either with TRH-induced increases in prolactin or with normal prolactin levels. Other clinical manifestations of hypothyroidism include constipation, cold intolerance, psychomotor retardation, carpal tunnel syndrome, and decreased exercise tolerance. However, patients often are asymptomatic. Close evaluation can reveal mental slowness, decreased energy, fatigue, poor memory, somnolence, slow speech, a low-pitched voice, water retention, periorbital edema, delayed reflexes, or a low body temperature and

bradycardia. Hypothyroidism can cause hypertension, cognitive abnormalities, pericardial effusion, asymmetric septal myocardial hypertrophy, myopathy, neuropathy, ataxia, anemia, elevated cholesterol and LDL-cholesterol, or hyponatremia. Myxedematous infiltration can produce enlarged, cystic ovaries.[11] The increase in cholesterol is due to impaired LDL-cholesterol clearance secondary to a decrease in cell membrane LDL receptors. The mechanism for this LDL effect is attributed to a thyroid response element in the LDL receptor gene.[12]

Serum enzymes may be elevated because of decreased clearance, including creatine phosphokinase (CPK), aspartate aminotransferase (AST, SGOT), alanine aminotransferase (ALT, SGPT), lactic dehydrogenase (LDH), and alkaline phosphatase, triggering a fruitless search for other organ disease. ***It is worth screening for hypothyroidism in any women with abnormal menses or with complaints of fatigue and depression. In addition, patients should be screened who have elevated levels of cholesterol and LDL-cholesterol.***

Diagnosis of Hypothyroidism

With primary thyroid failure, the circulating thyroid hormone levels fall, stimulating the pituitary to increase TSH output. Elevated TSH and low T4 confirm the diagnosis. Hypothyroidism can occur due to pituitary failure in which case the TSH will be inappropriately low for the T4. The most common cause is autoimmune thyroid disease (elevated titers of antithyroid antibodies) in areas with normal iodine intake. However, making an etiologic diagnosis in women adds little to the clinical management.

Subclinical Hypothyroidism

In early hypothyroidism, with undetectable symptoms or signs, a compensated state can be detected by an elevated TSH (greater than the upper limit of the normal range of 0.45–4.5 μU/mL) and normal T4 (called subclinical hypothyroidism). The prevalence is greater in women. About 2 to 5% *each year* will eventually become clinically hypothyroid with low T4 concentrations.[13, 14] Subclinical hypothyroidism is present in 4 to 8.5% of U.S. adults, is less common in blacks, and increases with age, present in up to 20% of women over age 60.[14]

A good reason to treat subclinical hypothyroidism is to avoid the development of a goiter. Furthermore, some patients in retrospect (after treatment) recognize improved physical and mental well-being. Patients with subclinical hypothyroidism have alterations in energy metabolism in skeletal muscle.[15] An improvement in impaired cognitive function and emotional behavior has been documented with thyroxine treatment of subclinical hypothyroidism.[16] In those patients who are asymptomatic, it is worth measuring antithyroid antibodies. A positive test identifies those who are more likely to become clinically hypothyroid, at a rate of approximately 20% per year. With only very slight elevations of TSH (less than 10 μU/mL), it is reasonable not to treat asymptomatic patients, especially those over age 80, and to check thyroid function every 6 months to detect further deterioration; however symptomatic patients may benefit from treatment.[13] Patients with an abnormal cholesterol-lipoprotein profile can show a rapid improvement with thyroxine treatment.[17–19] Subclinical hypothyroidism is a strong risk factor for coronary heart disease.[20] In addition, iron deficiency anemia is common in patients with subclinical hypothyroidism, and responds better when levothyroxine is added to iron treatment.[21]

The risk of pregnancy loss is increased for women with uncorrected overt or even subclinical hypothyroidism. The results of a study of pregnancy outcomes in women with hypothyroidism challenged the notion that subclinical hypothyroidism has no impact on pregnancy.[22]

The incidence of pregnancy loss was very low in treated hypothyroid women having normal thyroid indices, but markedly increased in women with elevated TSH levels, including both women with untreated subclinical disease and those with overt disease who received inadequate exogenous thyroid hormone replacement. ***These observations indicate that subclinical hypothyroidism is not entirely benign and further justify recommendations to include TSH screening in the evaluation of women with recurrent pregnancy loss.***

Treatment of Hypothyroidism

Initial therapy is straightforward with synthetic thyroxine, T4, given daily. Mixtures of T4 and T3, such as desiccated thyroid, provide T3 in excess of normal thyroid secretion. It is better to provide T4 and allow the peripheral conversion process to provide the T3.[23, 24] "Natural" thyroid preparations are not better, and in fact are potentially detrimental. Patients taking biologic preparations should be switched to synthetic thyroxine. Studies have documented that adding T3 to T4 does not improve treatment outcomes.[25–27] Because of a risk of coronary heart disease in older women, the initial dose should be 25–50 μg/day for 4 weeks, at which time the dose is increased by 25 μg daily every 4 weeks according to the clinical and biochemical assessment. Usually the dose required will be close to 1.5 μg/lb body weight, but it may be less in very old women.[28] The average final dose required in the elderly is approximately 70% of that in younger patients. ***Patients who have been on thyroid hormone for a long time may have their medication discontinued. Recovery of the hypothalamic-pituitary axis usually requires 8 weeks at which time the TSH and free T4 levels can be measured.***

Evaluation of Therapy

When the patient appears clinically euthyroid, evaluation of TSH levels will provide the most accurate assessment of the adequacy of thyroid hormone replacement. The goal is to maintain the TSH in the *lower half* of the normal range, between 0.45 and 2.0 μU/mL.[7, 29] Thyroid hormone requirements decrease with age. A patient being treated with thyroid hormone should be evaluated once every year with the TSH assay, and each patient should consistently remain on the same levothyroxine product. If the TSH level is low, then the free T4 should be measured to help adjust the thyroxine dose.[30] ***The full response of TSH to changes in T4 is relatively slow; a minimum of 8 weeks is necessary between changes in dosage and assessment of TSH.***

Hyperthyroidism

The two primary causes of hyperthyroidism are Graves' disease (toxic diffuse goiter) and Plummer's disease (toxic nodular goiter).[31] Plummer's disease is usually encountered in postmenopausal women who have had a long history of goiter. Twenty percent of hyperthyroid patients are older than 60, and 25% of older women with hyperthyroidism present with an apathetic or atypical syndrome.

Graves' disease, about 5 to 10 times more common in women than men, is characterized by the triad of hyperthyroidism, ophthalmopathy, and pretibial myxedema and is caused by autoantibodies that have TSH properties and, therefore, bind to and activate the TSH receptor. The measurement of TSH-receptor antibodies is essential to distinguish Graves' disease

from toxic goiter. Menstrual changes associated with hyperthyroidism are unpredictable, ranging from amenorrhea to oligomenorrhea to normal cycles (hence, the amenorrhea in a thyrotoxic woman can be due to pregnancy).

The classic symptoms of thyrotoxicosis are nervousness, disturbed sleep, heat intolerance, weight loss, sweating, palpitations, and diarrhea. These symptoms are associated with typical findings on physical examination: proptosis, lid lag, tachycardia, tremor, warm and moist skin, and goiter. Women in the reproductive years usually present with the classic picture. In postmenopausal women, symptoms are often concentrated in a single organ system, especially the cardiovascular or central nervous system. Goiter is absent in 40%. Sinus tachycardia occurs in less than half, but atrial fibrillation occurs in 40% and is resistant to cardioversion or spontaneous reversion to sinus rhythm. In old women, there is often a coexistent disease, such as an infection or coronary heart disease that dominates the clinical picture.

Hyperthyroidism in older women is sometimes described as “apathetic hyperthyroidism” because the clinical manifestations are different. The triad of weight loss, constipation, and loss of appetite, suggesting gastrointestinal malignancy, occurs in about 15% of older patients with hyperthyroidism. Ophthalmopathy is rare in older patients. The clinician should consider the diagnosis in older patients with “failure to thrive”; in patients who are progressively deteriorating for unexplained reasons; and in patients with heart disease, unexplained weight loss, and mental or psychological changes.

Psychological changes are not unusual in hyperthyroid women. Women who complain of emotional lability and nervousness should be screened for hyperthyroidism.

Diagnosis of Hyperthyroidism

The diagnosis of hyperthyroidism requires laboratory testing. A suppressed TSH (below 0.4 μU/mL) with a high T4 or a high T3 confirms the diagnosis. About 2% of U.S. adults will have subclinical hyperthyroidism, more common in women and blacks.[14] Progression to overt hyperthyroidism is essentially limited to patients with a TSH level lower than 0.1 μU/mL. Graves’ disease is associated with the presence of TSH receptor autoantibodies, TRAb. The measurement of TRAb in all patients with hyperthyroidism is important in order to confirm a diagnosis of Graves’ disease.[32] Most patients should have a radioactive iodine thyroid uptake and scan after laboratory confirmation of the diagnosis. If the uptake is suppressed then drug therapy is indicated. The scan will indicate whether the patient has a diffuse toxic goiter, a solitary hot nodule, or a hot nodule in a multinodular gland. Toxic multinodular goiters occur more frequently in the elderly. TSH hypersecretion as a cause of hyperthyroidism is extremely rare; the combination of a normal or elevated TSH and elevated thyroid hormone will be the clue to this possibility.

Subclinical Hyperthyroidism

By definition, patients with subclinical hyperthyroidism have normal T4 and T3 levels, but subnormal concentrations of TSH. TSH levels can be suppressed to 0.1–0.5 μU/mL by general illnesses and drugs such as glucocorticoids, dopamine, and anticonvulsants; however, this suppression does not extend below 0.1 μU/mL. Values below 0.1 μU/mL are regarded as nondetectable, and patients with overt hyperthyroidism usually have undetectable TSH. Subclinical hyperthyroidism is half as common in older people as subclinical hypothyroidism (excluding the most common cause, treatment with excessive doses of thyroxine). Keep in mind that the dose of thyroxine required to treat hypothyroidism declines

with age (because of the decrease in metabolic clearance with age); all patients being treated with thyroid hormone should have their TSH levels assessed every year. Atrial fibrillation is a common cardiovascular problem associated with subclinical hyperthyroidism, especially in older women when the TSH is less than 0.1 μU/mL.[14, 33] Progression to overt hyperthyroidism is uncommon. ***Therefore, TSH levels less than 0.1 μU/mL should be treated to avoid bone loss and atrial fibrillation in older women, or in those at risk for osteoporosis and heart disease. With TSH levels of 0.1–0.4 μU/mL, treatment is indicated only in older patients but TSH follow-up in younger women is warranted every 6 months.***[14, 34, 35]

Treatment of Hyperthyroidism

There are multiple objectives of therapy: control of thyroid hormone effects on peripheral tissues by pharmacologic blockade of beta-adrenergic receptors, inhibition of thyroid gland secretion and release of thyroid hormone, and specific treatment of nonthyroidal systemic illnesses that can exacerbate hyperthyroidism or be adversely affected by hyperthyroidism.[36] Antithyroid drugs are administered first to achieve euthyroidism before definitive therapy is accomplished by radioactive iodine treatment only when symptoms are severe or when radioactive iodine therapy must be delayed. Of course, it is important to ensure that a woman is not pregnant before treatment with radioactive iodine, and pregnancy should be postponed for several months after treatment. Monitoring treatment response requires a full 8-week interval for stabilization of the hypothalamic-pituitary-thyroid system.

Antithyroid Drugs

The drug of choice in most circumstances (except in pregnant women, as discussed later) is methimazole because it has fewer adverse effects. The drug inhibits organification of iodide and decreases production of T4 and T3. The oral dose is 10–20 mg daily. The onset of effect takes 2–4 weeks. Remember that the half-life of thyroxine is about 1 week, and the gland usually has large stores of T4. Maximal effect occurs at 4–8 weeks. The dose can be titrated down once the disease is controlled to a maintenance dose of 5–10 mg daily. The major side effects are rash, gastrointestinal symptoms, and agranulocytosis (an idiosyncratic reaction). Propranolol and other beta-blockers are effective in rapidly controlling the effects of thyroid hormone on peripheral tissues. The dose is usually 20–40 mg, every 12 h orally, and the dose is titrated to maintain a heart rate of about 100 beats/min. The drug may cause bronchospasm, worsening congestive heart failure, fatigue, and depression. Rarely inorganic iodine is needed to block release of hormone from the gland. Lugol's solution, 2 drops in water daily, is sufficient. The onset of effect is 1–2 days, with maximal effect in 3–7 days. There may be an escape from protection in 2–6 weeks, and the drug can cause rash, fever, and parotitis. Iodine precludes radioiodine administration for several months.

Most patients to be treated with radioactive iodine are not pretreated with antithyroid drugs. Some patients with hot nodules in multinodular glands require surgery because of the size of the gland and because the hyperthyroidism tends to recur in new nodules after the ablation of the original hot nodule. This can result in repetitive treatments with substantial doses of radioactive iodine, and surgery may be preferable. All patients definitively treated for hyperthyroidism must be monitored for the onset of hypothyroidism.

Osteoporosis and Excessive Thyroxine

Because postmenopausal women are at increased risk for osteoporosis and frequently develop hyperthyroidism or receive levothyroxine treatment for hypothyroidism, it is important to understand how thyroid hormone affects bones. Thyroid hormone excess alters bone integrity via direct effects on bone and gut absorption of calcium and indirectly through the effects of vitamin D, calcitonin, and parathyroid hormone.[37]

Thyroid hormone increases bone mineral resorption. In addition, total and ionized calcium increase in hyperthyroid women, leading to increases in serum phosphorus; alkaline phosphatase; and bone Gla protein (osteocalcin), a marker of bone turnover. Parathyroid hormone decreases in response to the increased serum calcium, and this results in decreased hydroxylation of vitamin D. Intestinal calcium and phosphate absorption decrease, while urinary hydroxyproline and calcium excretion increase. The net effect of excessive thyroid hormone is increased bone resorption and a subsequent decrease in bone density—osteoporosis.[38]

These effects become more clinically important in prolonged exposure to excessive thyroid hormone.[39] Women who have had hyperthyroidism are at greater risk for fractures and experience postmenopausal fractures earlier than usual.[40, 41] Postmenopausal women with subclinical hyperthyroidism have lower bone density measurements and experience a higher rate of fractures.[42, 43]

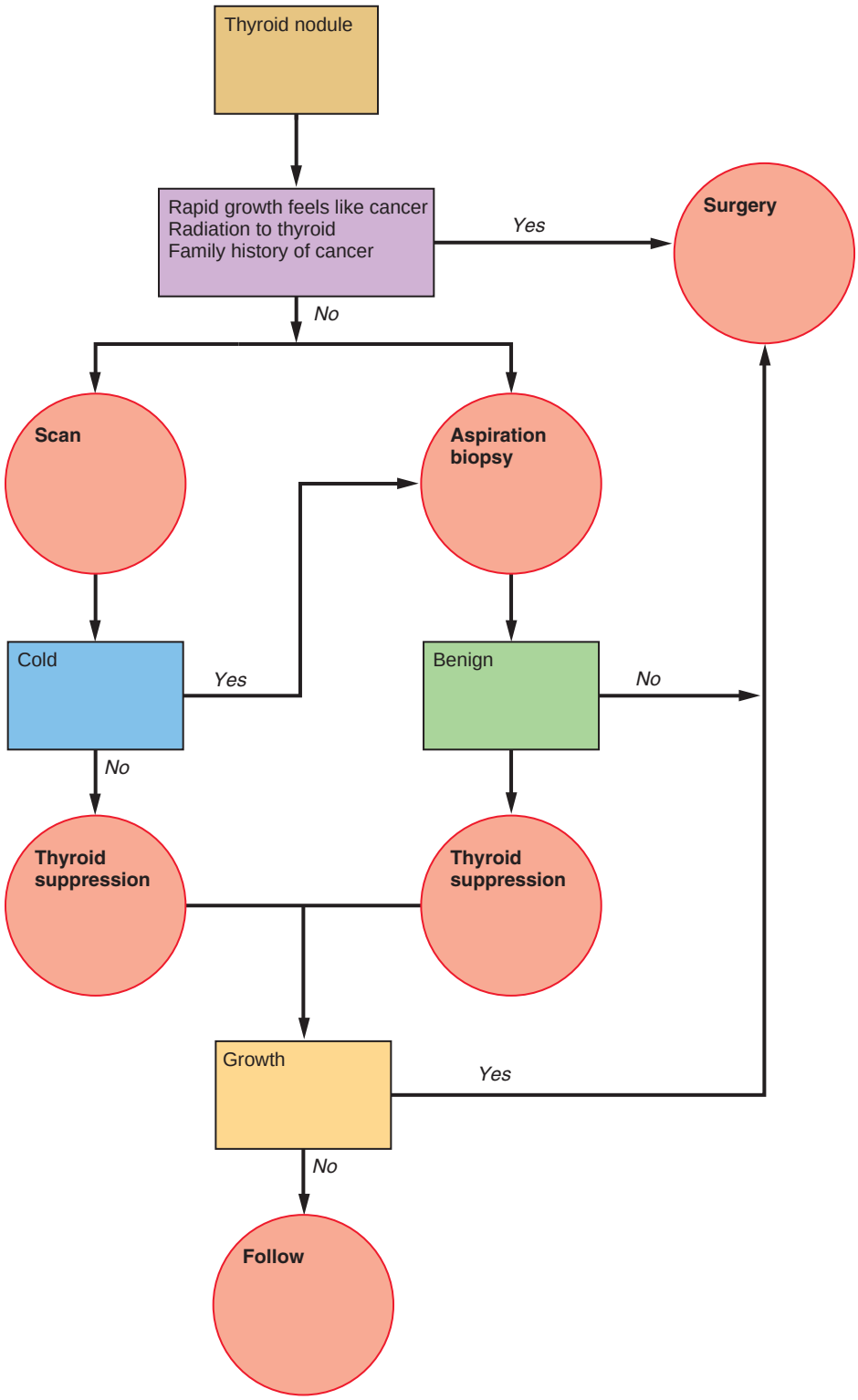
A major concern is that mild chronic excess thyroid hormone replacement, especially in postmenopausal women, might increase the risk of osteoporosis, and indeed this has been documented.[44] Bone density has been found to be reduced (9%) in premenopausal women receiving enough thyroxine to suppress TSH for 10 years or more.[45] A meta-analysis of the literature on this subject concluded that premenopausal women treated for long durations did not suffer a clinically significant loss of bone (probably because of the protective presence of estrogen); nevertheless, postmenopausal women lose an excess of bone if thyroid treatment results in TSH levels below the normal range.[46] Case-control and cohort studies, however, have been unable to detect an increase in fractures associated with thyroid administration.[47–49]

Because this is a problem easily avoided, we believe exposure to excessive thyroxine must be added to the risk factors for osteoporosis. It makes sense to monitor patients (both premenopausal women and especially postmenopausal women) receiving thyroxine with TSH levels to ensure that levothyroxine doses are “physiologic.” Some patients who require TSH suppressive doses of thyroxine, such as patients with cancer, must be considered at increased risk of osteoporosis. It would be wise to assess bone density in women on long-term thyroid treatment and in women receiving high-dose thyroxine suppression of TSH. The use of hormone therapy, exercise programs, and possibly bisphosphonate treatment must be seriously considered for these patients. In a cross-sectional study of elderly women, the bone loss associated with long-term thyroid treatment was avoided in those women also taking estrogen.[50]

Thyroid Nodules

The major concern with thyroid nodules is the potential for thyroid cancer.[51] Single nodules are 4 times more common in women, and carcinoma of the thyroid is nearly 3 times more

common in women than in men. The incidence rises steadily after the age of 55. Mortality from thyroid cancer occurs predominantly in the middle-aged and the elderly. There are 4 major types of primary thyroid carcinoma: papillary, follicular, anaplastic, and medullary. In solitary nodules that are "cold" (those that do not take up radioactive iodine or pertechnetate on thyroid scan), 12% prove to be malignant. This also means that the majority are



benign. Surgical excision of nodules can result in vocal cord paralysis, hypoparathyroidism, and other complications. Therefore, the goal is to select patients for curative surgery who have the greatest likelihood of having cancer in the nodule.

Epidemiologic and Clinical Data

The major risk factors for thyroid cancer are family history of this disease or a history of irradiation to head or neck. In those who have received thyroid irradiation, about one-third will have thyroid abnormalities, and about one-third of those with abnormalities will have thyroid cancer (about 10% overall). The carcinogenic risk has been estimated to be 1% per 100 rads in 20 years. A rapidly growing nodule, a hard nodule, the presence of palpable regional lymph nodes, or vocal cord paralysis greatly increase the probability of thyroid cancer.

Thyroid nodules in multinodular thyroid glands, not previously exposed to thyroid irradiation, have no greater risk of thyroid carcinoma than normal glands. Therefore, predominant thyroid nodules in multinodular glands should be followed, and, if a nodule grows, then biopsy or surgery should be considered.

Diagnostic Strategy

In patients with a thyroid nodule, laboratory assessment of thyroid function is essential. When abnormal thyroid function is present, the nodule is almost always benign. Ultrasonography of the thyroid is necessary to identify the presence of nonpalpable nodules and to guide fine-needle biopsy. Detection of a thyroid nodule is followed by clinical characterization of the nodule; examination of the lymph nodes; and inquiry regarding rapid growth, family history, and history of thyroid irradiation. In the presence of any positive findings, surgery is recommended for excision of the nodule. If none of these is present, proceed directly to fine-needle aspiration biopsy. Suppression with levothyroxine treatment is no longer recommended for two reasons: poor efficacy and an inability to differentiate benign lesions from thyroid cancer.

Although accounting for only 5% of thyroid cancers, medullary carcinoma is associated with early spread and poor survival rates. Medullary carcinoma of the thyroid is unique in having serum calcitonin as a very sensitive and specific tumor marker. Measuring serum calcitonin is recommended for older patients with solitary nodules to achieve earlier diagnosis of medullary carcinoma.[52]

Thyroid nodules discovered during pregnancy are not managed differently. Pregnancy does not affect the course of thyroid malignancy, and low-grade tumors can be treated after delivery.[53] Radioactive iodine, of course, must not be administered during pregnancy.

Fine-Needle Aspiration

Fine-needle aspiration biopsy has an 83% sensitivity and 92% specificity in diagnosing thyroid malignancy.[54] When “indeterminate” by biopsy, about one-third prove to be malignant at thyroidectomy. If the fine-needle aspiration biopsy reveals suspicious cells or is indeterminate, a subtotal thyroidectomy should be performed for diagnosis and treatment. If the aspiration biopsy is benign some would repeat the biopsy in 1 year to avoid false negatives. Growth indicates the need for biopsy or surgery. In some cases, espe-

cially in older women, nodules can be followed with little risk. Clinical experience with levothyroxine suppression has been disappointing and is no longer recommended.

The Thyroid Gland and Pregnancy

In response to the metabolic demands of pregnancy, there is an increase in the basal metabolic rate (which is mainly due to fetal metabolism), iodine uptake, and the size of the thyroid gland caused by hyperplasia and increased vascularity.[55] However, despite this increase in thyroid activity, a pregnant woman is euthyroid with levels of TSH, free T4, and free T3 remaining within the normal range; thyroid nodules and goiter require evaluation. During pregnancy, iodide clearance by the kidney increases. For this reason (plus the iodide losses to the fetus), the prevalence of goiter is increased in areas of iodine deficiency.[56] This is not a problem in the U.S., and any goiter should be regarded as pathologic. In many parts of the world, iodine is not sufficiently available in the environment, and pregnancy increases the risk of iodine deficiency.

The increase in thyroid activity in pregnancy is accompanied by a marked increase in the circulating levels of TBG in response to estrogen; therefore, a new equilibrium is reached with an increase in the bound portion of the thyroid hormone. The mechanism for the estrogen effect on TBG is an increase in hepatic synthesis and an increase in glycosylation of the TBG molecule that leads to decreased clearance.

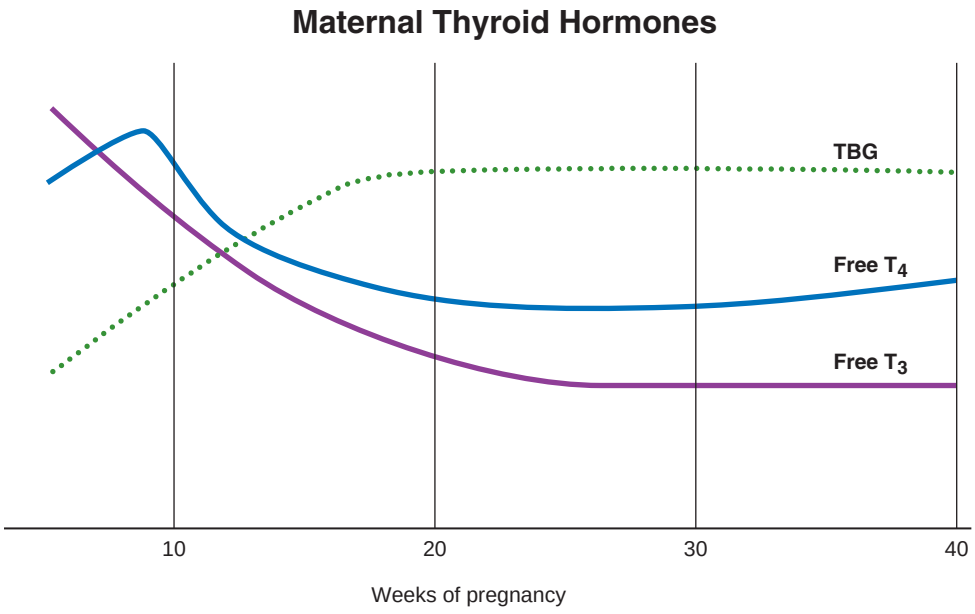
The increase in thyroid activity is attributed to the thyrotropic substances secreted by the placenta: a chorionic thyrotropin and the thyrotropic activity in human chorionic gonadotropin (hCG).[57, 58] It has been calculated that hCG contains approximately 1/4,000th of the thyrotropic activity of human TSH. In conditions with very elevated HCG levels, the thyrotropic activity can be sufficient to produce hyperthyroidism (gestational hyperthyroidism), and this can even be encountered in normal pregnancy.[59]

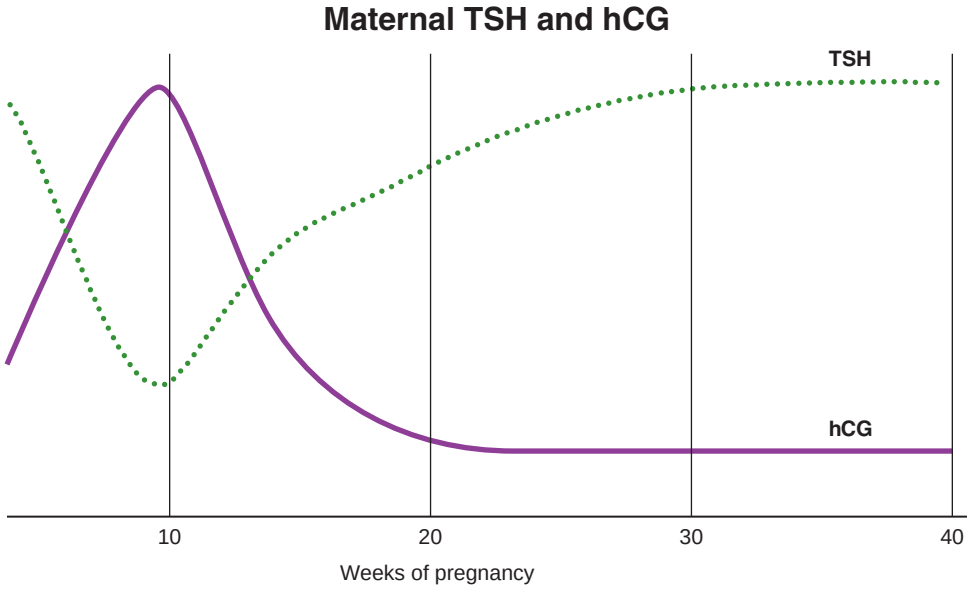
TBG levels reach a peak (twice nonpregnant levels) at about 15 weeks, which is maintained throughout the rest of pregnancy.[60] T4 undergoes a small increase in the first trimester, but T3 increases more markedly. Because of the increase in TBG, free T4 and T3 levels then decrease, although they remain within the normal range.[61] There is an inverse relationship between maternal circulation levels of TSH and hCG.[60] TSH reaches a nadir at the same time that hCG reaches a peak at 10 weeks of pregnancy. TSH levels then increase as hCG levels drop to their stable levels throughout the rest of pregnancy. ***Thus, the range of normal for TSH levels change with each trimester.[62] The lower limit of TSH in the first and second trimesters is 0.03 and 0.13 μU/L in the third trimester. The upper limit of normal in the first trimester is 2.3 and 3.5 μU/L in the second and third trimesters.***

These changes support a role for hCG stimulation of the maternal thyroid gland, especially during early pregnancy, providing a small but important increase in maternal thyroid hormones for the fetus until fetal thyroid function is sufficient to serve fetal needs.[58, 60, 63] It is well recognized that patients who have conditions associated with very high levels of hCG (trophoblastic disease, hCG-secreting cancers) can develop hyperthyroidism. The thyroid-stimulating activity of hCG is explained by the molecular homology between hCG and TSH, and between their receptors.

In normal pregnancies, placental transfer of TSH, T4, and T3 is limited in both directions. Slight, but significant, transfer of T4 and T3 does occur, however, when maternal levels are very high or when fetal levels are substantially lower than the maternal levels. Therefore, in the early weeks of pregnancy, before the fetal thyroid gland becomes active, the fetal

brain is dependent on the placental transport of maternal T4.[64] Both overt and subclinical maternal hypothyroidism are associated with increased risks of miscarriage, preeclampsia, low birth weight, premature delivery, and a decrease in intelligence in the children.[22, 65–69]



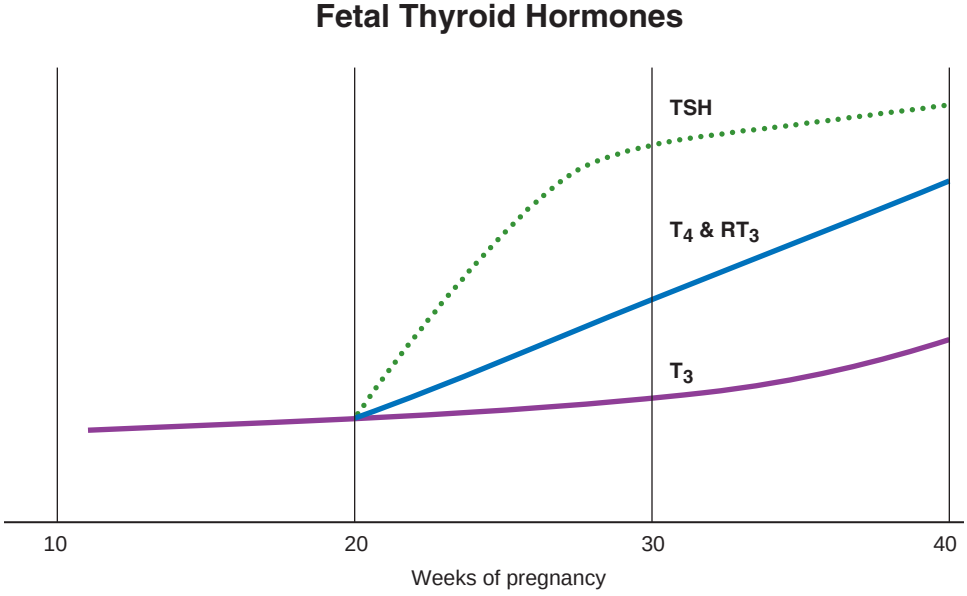


The majority of patients with hyperemesis gravidarum have laboratory values consistent with hyperthyroidism, and the severity of the hyperemesis correlates with the degree of hyperthyroidism.[70, 71] These patients have higher levels of hCG, and the transient hyperthyroidism and severity of the hyperemesis may be mediated by the thyrotropic and steroidogenic activity of the hCG. These clinical manifestations in normal pregnancies may be linked to a specific subpopulation of hCG molecules with greater thyrotropic bioactivity (because highly purified, standard hCG has only trivial TSH-like activity).[72] Specifically, hCG with reduced sialic acid content is increased in pregnant patients with hyperemesis and hyperthyroidism.[73] Women with hyperemesis gravidarum do not require antithyroid drug treatment unless they exhibit symptoms of hyperthyroidism, supported by appropriate laboratory results.

Thyroid Physiology in the Fetus and the Neonate

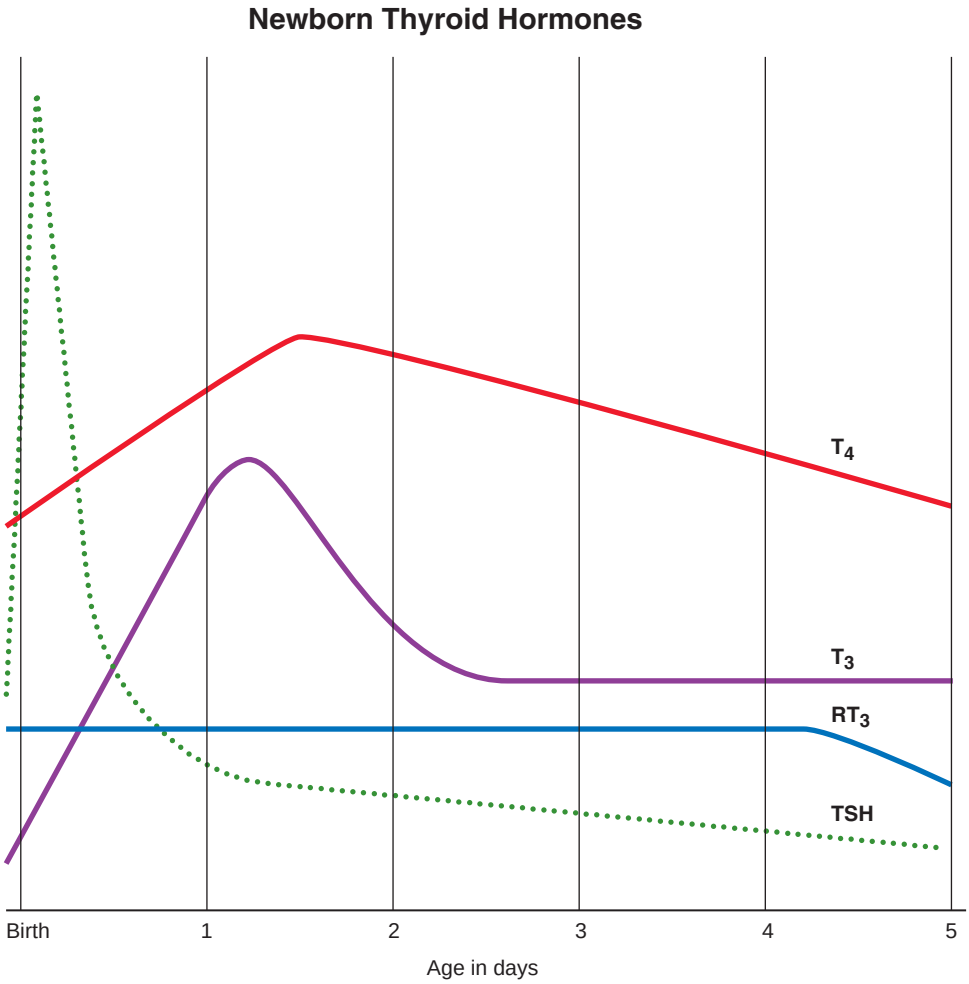
The human fetal thyroid gland develops the capacity to concentrate iodine and synthesize hormone between 8 and 10 weeks of gestation, at the same time that the pituitary begins to synthesize TSH.[74, 75] Some thyroid development and hormone synthesis are possible in the absence of the pituitary gland, but optimal function requires TSH. By 12–14 weeks, development of the pituitary-thyroid system is complete. Function is minimal, however, until an abrupt increase in fetal TSH occurs at 20 weeks. As with gonadotropin and other pituitary hormone secretion, this thyroid function correlates with the maturation of the hypothalamus and the development of the pituitary portal vascular system, which makes releasing hormone available to the pituitary gland.

Fetal TSH increases and reaches a plateau at 28 weeks and remains at relatively high levels to term. The free T4 concentration increases progressively. At term, fetal T4 levels exceed maternal levels. Thus, a state of fetal thyroidal hyperactivity exists near term.[76]



The major thyroid hormone secreted by the fetus is T4. However, total T3 and free T3 levels are low throughout gestation, and levels of RT3 are elevated, paralleling the rise in T4. Like T3, this compound is derived predominantly from conversion of T4 in peripheral tissues. The increased production of T4 in fetal life is compensated by rapid conversion to the inactive RT3, allowing the fetus to conserve its fuel resources. However, there is some evidence that RT3, as well as T4, through nongenomic actions, regulate fetal brain development.[77]

The fetus is iodine deficient when a mother's iodine intake is low. Supplying the mother adequate iodine is an important problem in many parts of the world.[78] A fetal goiter is occasionally detected by ultrasonography at a size that could impede normal delivery. The fetus can be treated, with regression of the goiter, by the administration of levothyroxine into the amniotic fluid.[79] However, the fetus is usually protected by the transplacental transfer of maternal T4, and, therefore, fetal treatment of hypothyroidism can usually be postponed until delivery. Fetal hyperthyroidism occurs in association with maternal Graves' disease, discussed later.



With delivery, the newborn moves from a state of relative T3 deficiency to a state of T3 thyrotoxicosis. Shortly after birth serum TSH concentrations increase rapidly to a peak at 30 min of age. They fall to baseline values by 48–72 h. In response to this increase in TSH, total T4 and free T4 increase to peak values by 24–48 h of age. T3 levels increase even more, peaking by 24 h of age. By 3–4 weeks, the thyroidal hyperactivity has disappeared.

The postnatal surge in TSH is accompanied by a prolactin surge, suggesting that both are increased in response to TRH. The TRH surge is thought to be a response to rapid neonatal cooling. A puzzle is the fact that the early increase in T3 is independent of TSH and is tied in some way to cutting of the umbilical cord. Delaying cord cutting delays the increase in T3, but TSH levels still reach their peak at 30 min. Probably by changing liver perfusion, cord cutting augments peripheral (largely liver) conversion of T4 to T3. The later increases in T3 and T4 (after 2 h) are due to increased thyroid gland activity. These thyroid changes after birth probably represent defense mechanisms against the sudden entry into the cold world. The high RT3 levels during pregnancy continue during the first 3–5 days of life, and then fall gradually to normal levels by 2 weeks.

Summary of Fetal and Newborn Thyroid Changes

1. TSH and T4 appear in the fetus at 10–13 weeks. Levels are low until an abrupt rise at 20 weeks.
2. T4 rises rapidly and exceeds maternal values at term.
3. T3 levels rise, but concentrations are relatively low, similar to hypothyroid adults.
4. RT3 levels exceed normal adult levels.
5. The fetal pattern of low T3 and high RT3 is similar to that seen with calorie malnutrition.
6. After delivery, TSH peaks at 30 min of age, followed by a T3 peak at 24 h and a T4 peak at 24–48 h. The T3 increase is independent of the TSH change.
7. High RT3 levels persist for 3–5 days after delivery, then reach normal values by 2 weeks.

Newborn Screening for Hypothyroidism

The incidence of neonatal hypothyroidism is about 1 in 4,000 live births, and newborn screening programs exist in most of the world.[80] The problem is that congenital hypothyroidism is not clinically apparent at birth. Fortunately, infants with congenital hypothyroidism have low T4 and high TSH concentrations easily detected in blood, and early high-dose treatment before 3 months of age is usually associated with normal mental development,[81] but persistent impaired mental performance has been observed in long-term follow-up studies.[82, 83] Less than normal development can be a consequence of a delay in treatment or inadequate dosage, but despite early treatment with high doses, extremely low thyroid hormone production in the fetus is associated with deficits later in life.[84] Recently, a failure to develop normally in a small percentage of individuals despite early diagnosis and treatment has been linked to genetic defects in transcription factors (FOXE-1, NKX2.1, PAX8) that are important in both the thyroid gland and the central nervous system.[85, 86]

There is a familial tendency for hypothyroidism, and if the diagnosis is made in the antepartum period, intraamniotic injections of thyroxine can raise fetal levels of thyroid hormone.[87] A early clue to the presence of congenital hypothyroidism is a reduction in baseline variability in fetal heart rate tracings.[88] Ultrasonographic examination of patients with polyhydramnios should include a search for a fetal goiter. In addition, the fetus should be monitored for goiter formation in women treated with antithyroid drugs for hyperthyroidism during pregnancy. Amniotic fluid iodothyronines and TSH reflect fetal plasma levels, and abnormal values may allow prenatal diagnosis of fetal hypothyroidism by amniocentesis.[87, 89] However, fetal cord blood sampling is advocated for accurate diagnosis.[90] Treatment of fetal hypothyroidism is important because prenatal hypothyroidism can affect some aspects of development, e.g., the full function of physical skills. Although transfer of thyroid hormone from mother to fetus is limited, even a small amount provides protection, especially to the brain.

Hyperthyroidism in Pregnancy

Untreated thyrotoxicosis in pregnancy is associated with a higher risk of preeclampsia, heart failure, intrauterine growth retardation, preterm birth, and stillbirth.[91, 92] Heart failure is a consequence of the demands of pregnancy superimposed on the hyperdynamic cardiovascular state induced by increased thyroid hormone.[93]

The most common cause of thyrotoxicosis in pregnancy is Graves' disease. Most patients with Graves' disease will have detectable levels of TSH-receptor antibodies, TRAb. The clinician should always keep in mind that trophoblastic disease can cause hyperthyroidism due to the TSH property inherent in human chorionic gonadotropin. The maternal changes with pregnancy can make diagnosis difficult. Tachycardia on awakening from sleep and a failure to gain weight should make a clinician suspicious. Hyperemesis gravidarum is a common presentation of hyperthyroidism in pregnancy. Laboratory assessment is unaffected by pregnancy and should follow our algorithm. TRAb will be undetectable in gestational hyperthyroidism. ***Remember that the range of normal for TSH levels change with each trimester.[62] The lower limit of TSH in the first and second trimesters is 0.03 and 0.13 μU/L in the third trimester.***

Subclinical hyperthyroidism need not be treated as it is not associated with adverse obstetrical and neonatal complications.[94, 95] Women with gestational hyperthyroidism associated with hyperemesis gravidarum seldom have clinical signs of hyperthyroidism and usually experience a rapid, spontaneous recovery within a few weeks. A few women with gestational hyperthyroidism will require antithyroid drug treatment, therefore screening all women with hyperemeis for thyroid function is strongly recommended.

A clinical triad usually distinguishes the presence of Graves' disease: a goiter, the presence of thyroid antibodies, and the presence of TSH-receptor antibodies. The choice of treatment is between surgery and antithyroid drugs. Most women can be successfully treated with thioamide drugs.[91] Propylthiouracil, carbimazole, and methimazole are equally effective for pregnant women; however, although the data are controversial, methimazole and carbimazole have been associated with teratogenesis.[53] Propylthiouracil is preferred in breastfeeding patients because it is less concentrated in breast milk.[96]

The aim of treatment should be to maintain mild hyperthyroidism in the mother to avoid thyroid dysfunction in the fetus. Treatment of maternal hyperthyroidism with propylthiouracil, even with moderate doses of 100–200 mg daily, suppresses T4 and increases TSH levels in newborns.[97] The infants are clinically euthyroid, however, and their laboratory measurements are normal by the 4th to 5th day of life. In addition, follow-up assessment has indicated unimpaired intellectual development in children whose mothers received propylthiouracil during pregnancy.[98] Nevertheless, pregnant women with thyrotoxicosis should be treated with as low a dose as possible of the preferred drug, propylthiouracil. With proper antithyroid drug treatment, very few, if any, deleterious effects are experienced by mother, fetus, or neonate.[99] Although small amounts of antithyroid drugs are transmitted in breast milk, the amount has no impact on neonatal thyroid function, and breastfeeding should be encouraged. Poor control of maternal hyperthyroidism is associated with increased risks of preeclampsia and low-birth-weight infants.[92]

Maternal TSH-like autoantibodies (stimulating TRAb, TSH-receptor antibodies) can cross the placenta and cause fetal thyrotoxicosis and demise. Some have advocated fetal cord blood sampling in women with Graves' disease who are euthyroid but who have positive titers of TRAb to assess the fetal thyroid status.[100] However, the size of the fetal thyroid can be monitored by ultrasonography and be used to monitor treatment of maternal hyperthyroidism.[101, 102] Fetal cord blood sampling is rarely needed. An increase in fetal thyroid

size on ultrasonography indicates excessive treatment, and, when dosage reductions are ineffective in reducing size, fetal thyrotoxicosis should be suspected because of the transplacental passage of TSH-like antibodies. Close monitoring of the fetus is not necessary if the maternal TRAb measurements are negative.[102] The fetus can be treated by treating the mother. Neonates have to be observed closely until antithyroid drugs are cleared (a few days) and the true thyroid state can be assessed.

Thyroid Storm

Thyroid storm is a life-threatening augmentation of thyrotoxicosis that is usually precipitated by stress such as labor, cesarean section, or infection. Stress should be limited as much as possible in patients with uncontrolled thyrotoxicosis.

Hypothyroidism in Pregnancy

Serious hypothyroidism is rarely encountered during pregnancy. Patients with this degree of illness usually do not get pregnant. Many patients with mild hypothyroidism never have a laboratory assessment for thyroid function during pregnancy and go undetected. Preeclampsia, intrauterine growth retardation, fetal distress, and fetal death are more frequent in women with significant hypothyroidism or with subclinical hypothyroidism.[22, 65–69, 103, 104] There is also reason to believe that patients with hypothyroidism, even subclinical hypothyroidism, have an increased rate of spontaneous miscarriage.[22, 103–105] The mechanism may be impaired function of important organs such as the endometrium, the corpus luteum, and especially the placenta. The children born to women with hypothyroidism have deficits in intelligence; remember that in the first half of pregnancy, the fetus is dependent on maternal thyroid hormones.[106–110] For all of these reasons, a strong case can be made for detecting and treating hypothyroidism in early pregnancy; preconceptual screening is ideal with establishment of the euthyroid state (a TSH not higher than 2.5 μU/L) before pregnancy. ***It is best to screen women in early pregnancy with both TSH and free T4 measurements.*** Keep in mind that the high hCG levels in early pregnancy can reduce maternal TSH concentrations. After the initial screening, TSH levels should be measured every 2 months during each pregnancy. ***The upper limit of normal in the first trimester is 2.3 and 3.5 μU/L in the second and third trimesters.*** There is no debate in regard to treating pregnant women who are hypothyroid; however, detecting and treating women with elevated TSH and normal T4 levels (subclinical hypothyroidism) is controversial.

Universal screening of pregnant women will detect both subclinical hypothyroidism and subclinical hyperthyroidism, each present in a substantial number (around 2% each) of asymptomatic women.[68, 111, 112] Advocacy for universal screening is not an established recommendation. A major prospective U.S. study could not detect adverse outcomes associated with maternal thyroid hypofunction, although only 247 women had low thyroid function out of a population of 10,990 patients.[112] The conservative approach, as expressed by the American College of Obstetricians and Gynecologists, is to await the results of clinical trials, and especially, the outcome of a national study on pediatric neurodevelopment.[113] Because there is a consensus to treat subclinical hypothyroid function in nonpregnant women, in our view, it is only logical to apply the same standard to pregnant women. This includes measuring antithyroid antibodies in women with elevated TSH levels because a positive test identifies those who are likely to become clinically hypothyroid. Women with

thyroid antibodies have a significant risk of becoming hypothyroid as pregnancy progresses and also an increased risk of postpartum thyroiditis.[114]

Women with positive thyroid antibodies undergoing in vitro fertilization experience a lower pregnancy rate and an increase in spontaneous miscarriage, which can be prevented with levothyroxine treatment.[115, 116] The experience with assisted reproduction emphasizes the importance of detecting and treating less than normal thyroid function. Pregnant women treated for hypothyroidism have been reported to have an increase in preeclampsia, diabetes, and pre-term birth.[117, 118] ***The obstetrical complications in treated women may reflect a lack of appreciation for the change in thyroid hormone requirement during pregnancy.***

Women being treated for hypothyroidism require an increase (20–50%) in thyroxine during pregnancy, beginning as early as the 5th week of pregnancy.[119–121] The reasons for this increase include the estrogen-induced increase in thyroid-binding globulin, the dilutional effect of the increase in vascular volume, and the increase in placental transport and metabolism. When previously diagnosed hypothyroid women become pregnant, it is best to empirically increase the levothyroxine dose by about 30% as soon as pregnancy is diagnosed, with further adjustments according to the TSH levels.[121, 122] ***TSH should be monitored monthly and again in the postpartum period, and dosage should be adjusted to keep the TSH level in the lower half of the normal range, less than 2.5 μU/mL in the first trimester and less than 3.0 μU/ mL in the rest of pregnancy.*** Postpartum, the dose should be immediately decreased to the prepregnancy level. The need for proper monitoring and adequate treatment cannot be overemphasized; surveys of pregnant women with hypothyroidism find a substantial number with either excessive suppression of TSH or elevated TSH levels.[103, 123] Women with thyroid autoimmunity (presence of thyroid antibodies) must be monitored with TSH levels for at least 6 months after delivery because of their increased risk for postpartum thyroiditis.

Postpartum Thyroiditis

Autoimmune thyroid disease is suppressed to some degree by the immunologic changes of pregnancy. Thus, there is a relatively high incidence of postpartum thyroiditis (5–10%), 1–6 months after delivery (most commonly at 3 months), manifested by either hyperthyroidism or hypothyroidism, although usually transient hyperthyroidism (lasting 1–2 months) is followed by hypothyroidism.[124, 125] This condition is due to a destructive thyroiditis associated with thyroid microsomal autoantibodies.[126] A possible etiologic mechanism is fetal microchimerism, the presence of fetal cells in the maternal thyroid gland.[127] Women at high risk for postpartum thyroiditis are those with a personal or family history of autoimmune disease, and those with a previous postpartum episode. Women with insulin-requiring diabetes mellitus are at particularly high risk.[128] Postpartum thyroiditis can also be triggered by an early miscarriage.[129] ***Women at high risk for postpartum thyroiditis, especially those with thyroid antibodies, should have TSH measurements at 3 and 6 months postpartum.***

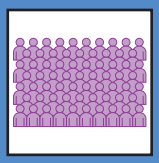
Most importantly, the symptoms in these women are often attributed to anxiety or depression, and the obstetrician must have a high index of suspicion for hypothyroidism. The symptoms usually last 1–3 months, and almost all women return to normal thyroid function. Postpartum thyroiditis tends to recur with subsequent pregnancies, and eventually hypothyroidism remains.[130] Because spontaneous remission is common, patients who are treated for hypothyroidism should be reassessed 1 year after gradual withdrawal of thyroxine. ***Patients who return to normal should undergo annual laboratory surveillance of their thyroid status; at least 20% and probably more will develop hypothyroidism in the next 5–10 years.***[131]

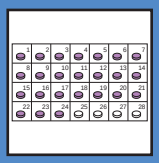
All references are available online at: http://www.clinicalgynendoandinfertility.com

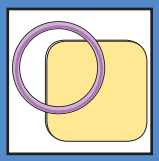
SECTION

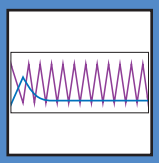
III

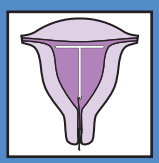
CONTRACEPTION

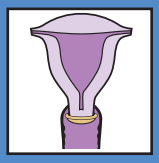




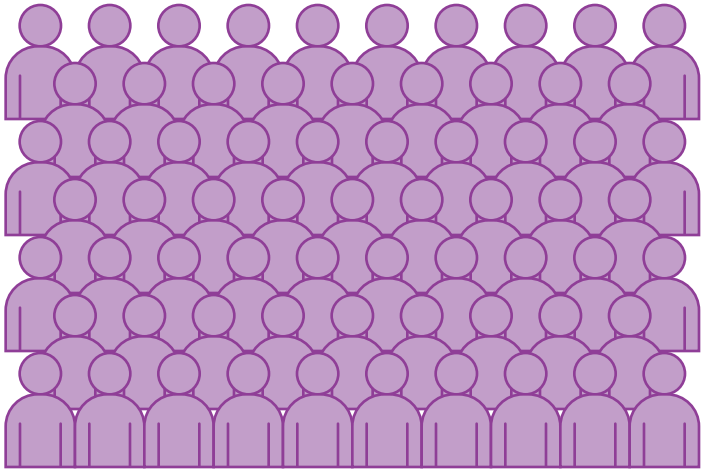








21 Family Planning, Sterilization, and Abortion



As societies become more affluent, fertility decreases. This decrease is a response to the use of contraception and induced abortion. During her reproductive lifespan, the average !Kung woman, a member of an African tribe of hunter-gatherers, experienced 15 years of lactational amenorrhea, 4 years of pregnancy, and only 48 menstrual cycles.[1] In contrast, a modern urban woman will experience 420 menstrual cycles. Contemporary women undergo earlier menarche and start having sexual intercourse earlier in their lives than in the past. Even though breastfeeding has increased in recent years, its duration is relatively brief, and its contribution to contraception in the developed world is trivial. Therefore, it is more difficult today to limit the size of a family unless some method of contraception is used.

Contraception is not new, but its widespread development and application are new. The era of modern contraception dates from 1960 when intrauterine devices were reintroduced and oral contraception was first approved by the U.S. Food and Drug Administration. For the first time, contraception did not have to be a part of the act of coitus. However, national family planning services and research were not funded by the U.S. Congress until 1970, and the last U.S. law prohibiting contraception was not reversed until 1973.

Today, more women younger than age 25 in the U.S. become pregnant than do their contemporaries in other Western countries.[2–4] The U.S. teenage pregnancy rates are twice as high as those in England, Wales, and Canada, and 8 times as high as those of the Netherlands and Japan. The differences disappear almost completely after age 25. This is largely because American men and women after age 25 utilize surgical sterilization at a great rate.

It is not true that young American women want to have these higher pregnancy rates. About 82% of all pregnancies among American teenagers are unintended.[5] Increasing effective contraceptive use among young Americans began to have an impact in 1991. In the 1990s,

the teenage pregnancy rate reached the lowest rate since estimates began in 1976, a 21% decline from 1991 to 1997 for teenagers 15–17 years and a 13% decline for older teenagers.[6] Overall, there was a 17% decline in teenage birth rates and a 12.8% decline in teenage induced abortions from 1991 through 1999. From 1995 to 2002, 14% of the decline in teen pregnancy was a consequence of decreased sexual activity among U.S. teenagers; however, 86% of the decline was attributed to an increase in the use of effective contraception.[7] In 2004, the proportion of induced abortions in the U.S. obtained by teens reached a low of 17%.[8]

After a 14-year 34% decline, birth rates for teenagers began to increase in 2005, the first increase since 1991. The rate increased 5% between 2005 and 2008.[9] There is appropriate concern that this increase reflects difficulties in contraceptive access, affordability, and correct use. In addition, in recent years, fewer teens have received instruction regarding contraception.[10] The evidence overwhelmingly indicates that abstinence programs have *not* had a positive impact on teen sexual behavior, including the delay of the initiation of sex or the number of sexual partners.[11] In contrast, comprehensive sex education programs that include contraception are effective and do not increase the frequency of sex or hasten the initiation.[12]

Nearly half of all pregnancies (49%) in the U.S. are unplanned, and about 40% of these are aborted.[5, 13] American teenagers abort nearly half of their pregnancies, and this proportion is similar to that seen in other countries.[13] But older American women, aged 20–34, have the highest proportion of pregnancies aborted compared with other countries, indicating an unappreciated, but real, problem of unintended pregnancy existing beyond the teenage years. In fact, American women older than age 40 have had for the last two decades a high ratio of abortions per live births, a ratio very similar to that of teenagers.[8]

Delaying marriage prolongs the period in which women are exposed to the risk of unintended pregnancy. This, however, cannot be documented as a major reason for the large differential between young adults in Europe and the U.S. The available evidence also indicates that a difference in sexual activity is not an important explanation. The major difference between American women and European women is that American women under age 25 are less likely to use any form of contraception. Significantly, the use of oral contraceptives (the main choice of younger women) is lower in the U.S. than in other countries.

Why are Americans different? The cultures in areas such as the U.K. and the Scandinavian countries are certainly very similar with similar rates of sexual experience. A major difference must be attributed to the availability of contraception. In the rest of the world, contraceptive services can be obtained from more accessible resources and relatively inexpensively. Major American problems are the enormous diversity of people and the unequal distribution of income in the U.S. These factors influence the ability of our society to effectively provide education regarding sex and contraception and to effectively make contraception services available.

Efficacy of Contraception

A clinician's anecdotal experience with contraceptive methods is truly insufficient to provide the accurate information necessary for patient counseling. The clinician must be aware of the definitions and measurements used in assessing contraceptive efficacy and must draw on the talents of appropriate experts in this area to summarize the accurate and comparative failure rates for the various methods of contraception. The publications by Trussell et al., summarized below, accomplish these purposes and are highly recommended.[14–18]

Definition and Measurement

Contraceptive efficacy is generally assessed by measuring the number of unplanned pregnancies that occur during a specified period of exposure and use of a contraceptive method. The two methods that have been used to measure contraceptive efficacy are the Pearl index and life-table analysis.

The Pearl Index

The Pearl index, created by Raymond Pearl in 1933, is defined as the number of failures per 100 woman-years of exposure.[19] The denominator is the total months or cycles of exposure from the onset of a method until completion of the study, an unintended pregnancy, or discontinuation of the method. The quotient is multiplied by 1,200 if the denominator consists of months or by 1,300 if the denominator consists of cycles.

With most methods of contraception, failure rates decline with duration of use. The Pearl index is usually based on a lengthy exposure (usually 1 year) and, therefore, fails to accurately compare methods at various durations of exposure. This limitation is overcome by using the method of life-table analysis.

Life-Table Analysis

Life-table analysis calculates a failure rate for each month of use. A cumulative failure rate can then compare methods for any specific length of exposure. Women who leave a study for any reason other than unintended pregnancy are removed from the analysis, contributing their exposure until the time of the exit.

Contraceptive Failures

Contraceptive failures do occur and for many reasons. Thus, "method effectiveness" and "use effectiveness" have been used to designate efficacy with correct and incorrect use of a method. It is less confusing to simply compare the very best performance (the lowest expected failure rate) with the usual experience (typical failure rate) as noted in the table of failure rates during the first year of use. The lowest expected failure rates are determined in clinical trials, in which the combination of highly motivated subjects and frequent support from the study personnel yields the best results. Contraceptive typical failure rates have been estimated using the data from the 1995 and 2002 U.S. National Survey of Family Growth, correcting for the underreporting of induced abortion.[17, 18, 20]

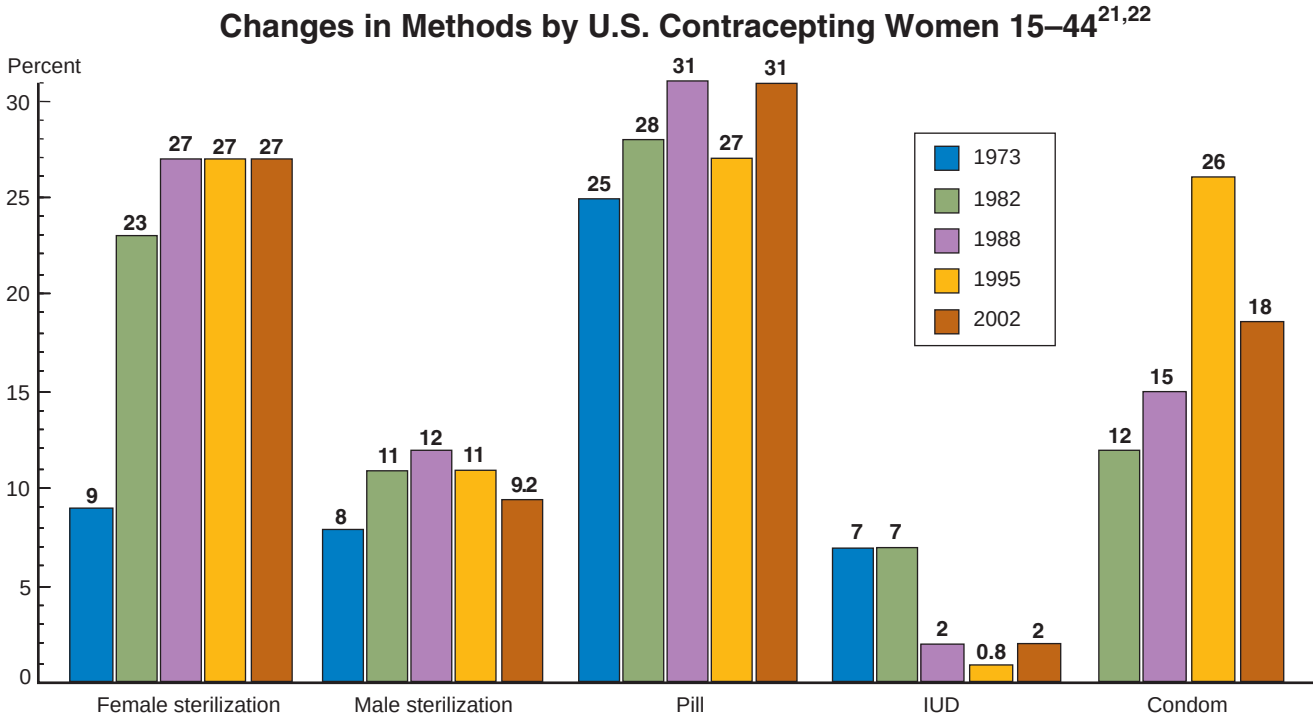
The 2002 estimates of failure were not significantly different compared with the previous estimates from the 1995 national survey. Women over the age of 30 were less likely to experience failure than young women; teens were more than twice as likely to experience a failure than older women. Hispanic women and even more so, black women, experienced higher failure rates. Groups that were less likely to experience contraceptive failure were women who did not intend to have a subsequent birth and women who had no previous births. Married women experienced the lowest failure rates and cohabiting women the highest. The most important determinants of pill failure, therefore, were: age, intention toward a future birth, parity, and marital status. Interestingly, once these factors were

accounted for, duration of use, race, ethnicity, and poverty status no longer affected the risk of pill failure. The same factors influence condom use, but when corrected for these factors, race, ethnicity, and poverty affected the risk of condom failure.

This is a subject of great interest because the rate of unintended pregnancies in the U.S. continues to be high. About one-half (over 3 million) of all pregnancies in the U.S. are unintended, and in 2002 about 53% of those occurred in women using a method of contraception.[5, 13, 21] Here is a more striking statistic: one of every two American women aged 15–44 has experienced an unintended pregnancy.[13]

Failure Rates During the First Year of Use, United States[17, 18, 20]

	Percent of women with pregnancy	
Method	*Lowest Expected*	*Typical*
No method	85%	85%
Combination Pill	0.3%	8.7%
Progestin only	0.5%	3.0%
IUDs:		
Levonorgestrel IUD	0.1%	0.1%
Copper T 380A	0.6%	1.0%
Implant	0.05%	1.0%
Injectable		
3-month	0.3	0.3%
1-month	0.05	3.0%
Patch	0.3	8.0%
Vaginal ring	0.3	8.0%
Female sterilization	0.5%	0.7%
Male sterilization	0.1%	0.2%
Spermicides	18.0%	29.0%
Periodic abstinence		25.3%
Calendar	9.0%	
Ovulation method	3.0%	
Symptothermal	2.0%	
Post-ovulation	1.0%	
Withdrawal	4.0%	18.4%
Cervical cap		
Parous women	26.0%	32.0%
Nulliparous women	9.0%	16.0%
Sponge:		
Parous women	20.0%	32.0%
Nulliparous women	9.0%	16.0%
Diaphragm and spermicides	6.0%	16.0%
Condom		
Male	2.0%	17.4%
Female	5.0%	27.0%



Contraceptive Use in the United States

The National Survey of Family Growth is conducted by the National Center for Health Statistics of the Centers for Disease Control and Prevention. Data are available from 1972, 1976, 1982, 1988, 1995, and 2002.[21–26] The samples are very large; therefore, the estimates are very accurate.

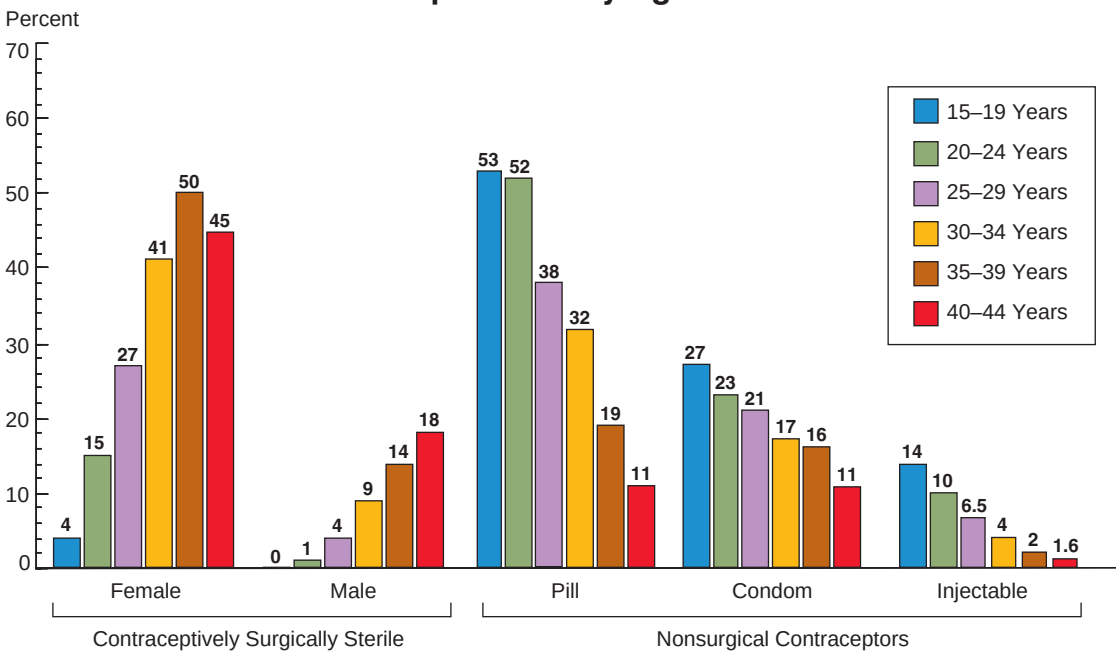
Pregnancy rates in the 1990s declined for women younger than age 30 years and increased in older women. From 1990 to 1997, the decrease in women in their early twenties was 8%, and the increase in women in their early thirties was 3%. The percent of married couples using sterilization as a method of contraception more than doubled from 1972 to 1988, and has remained stable since then. The use of oral contraception reached a high in 1992, slightly decreased in 1995, especially among Hispanic and black Americans, and returned to 31% of contracepting women in 2002. Approximately 10.7 million American women used oral contraceptives in 1988, and 11.6 million in 2002. Among never married women and women under age 25, oral contraception is the leading method of birth control. About 53% of contracepting women under age 25 were using oral contraception in 2002. From 1988 to 2002, oral contraception rose among women aged 30–44 to 32% of contraceptors aged 30–34, and 11% aged 40–44. About 5.3% of contraceptors in 2002 were using the 3-month injectable method, and 1.2% transdermal, vaginal ring, and implant methods.

In the 1990s, there was an increase in condom use by never married and formerly married women, women younger than 25, black women, and Hispanic women. These changes reflected the concern regarding sexually transmitted infections, including human immunodeficiency virus (HIV). But in 2002, the use of condoms alone returned to the level observed in the 1980s, probably because of the use of transdermal, vaginal ring, implant, and injectable methods. About one-third of condom users in 2002 were using more than one method, especially younger and never married women, including use of an oral contraceptive and a condom in 14% at first intercourse! Most importantly, the percentage of

women who used a contraceptive method at their first premarital intercourse increased from 43% before 1980 to 79% in 2002. Condom use at first intercourse increased from 22% before 1980 to 67% in 2002.

In 1982, 56% of U.S. women, 15–44 years of age, were using contraception, and this has increased to 62% (about 40 million women). In 2002, contraceptive sterilization (male and female) was utilized by 36% of these women (the next leading method was oral contraception, 31%). The number of reproductive-aged women using the intrauterine device (IUD) decreased by two-thirds from 1982 to 1988 and further decreased in 1995, from 7.1% to 2% to 0.8%, respectively, but rose to 5% in 2008. IUD use is concentrated in the U.S. in married women older than age 35. In 1982 more than 2 million women (about 8% of contraceptors) used the diaphragm, but use of the diaphragm has nearly disappeared in the U.S. (0.3% of contracepting women in 2002).

Contraceptive Use by Age in 2002 [21,22]



The oral contraceptive (53%) and condoms (27%) are the most popular methods among teenagers. However, studies have repeatedly documented that the use of the implant and injectable methods is associated with lower discontinuation rates and a lower rate of repeat pregnancies following delivery.[27, 28] This warrants continuing efforts to extend the use of these methods.

In 2002, 62% of all women 15–44 years of age were using some method of contraception, whereas 38% of women of reproductive age were not using a method of contraception for the following reasons:

18.1%	—	Not sexually active.
9.5%	—	Pregnant or trying to get pregnant.
1.6%	—	Male and female sterility.
1.5%	—	Sterilized for medical reasons.
7.4%	**—**	***At risk for an unintended pregnancy.***

The women at risk for an unintended pregnancy increased by 1.43 million women (2.2%) from 1995 to 2002, and the increase was in all age groups. These women accounted for more than half of unintended pregnancies in the U.S.; of the rest, about 43% are a consequence of incorrect contraceptive use; only 5% can be attributed to method failure.[13, 18] This increase alone in women at risk and not using contraception, therefore, would amount to about 500,000 unintended pregnancies and 270,000 induced abortions in 2002. In our view, these numbers reflect problems of contraceptive access, affordability, and correct use in the U.S. The number of unintended pregnancies is highest among low-income women, women who have not completed high school, women aged 18 to 24, unmarried, especially cohabiting, women, and members of racial or ethnic minority groups.[5]

U.S. couples have made up for the lack of contraceptive effective use and availability by greater reliance on voluntary sterilization. Between 1973 and 1982, oral contraception and sterilization changed places as the most popular contraceptive method among women over the age of 30. Approximately one-half of American couples choose sterilization within 15 to 20 years of their last wanted birth. During the years of maximal fertility, oral contraceptives are the most common method peaking at age 20–24. The use of condoms is the second most widely used method of reversible contraception, rising from about 9% in the mid-1980s to approximately 26% of contracepting women in 1995, decreasing to 18% in 2002.[21, 22]

Overall use of contraception among women at risk of unintended pregnancy decreased from 92.5% in 1995 to 89.3% in 2002.[18] The use of contraception among poor women at risk of pregnancy decreased from 92.1% in 1995 to 86.3% in 2002. For various reasons, American women have had increasing difficulty in obtaining effective contraception.

Women at each end of the economic spectrum, the poorest and the wealthiest, experienced a decrease in failure rates from 1995 to 2002, although the poorest women continued to have a higher failure rate than did the better-off women. Also, although the difference in overall failure rate was not statistically significant comparing 1995 and 2002, there was about a 2.5% improvement; this missed mathematical significance but it may reflect a meaningful change in our population. This change is probably due to an increase in pill and injectable methods and a decrease in condom use during this period of time. Women living in poverty who must rely on condoms or withdrawal (male-dependent methods) have about a 2-fold increase in failure rates, but if they can use the pill, their failure rates are the same as better-off women. The message is clear: we need to make the more effective methods available for poor women.

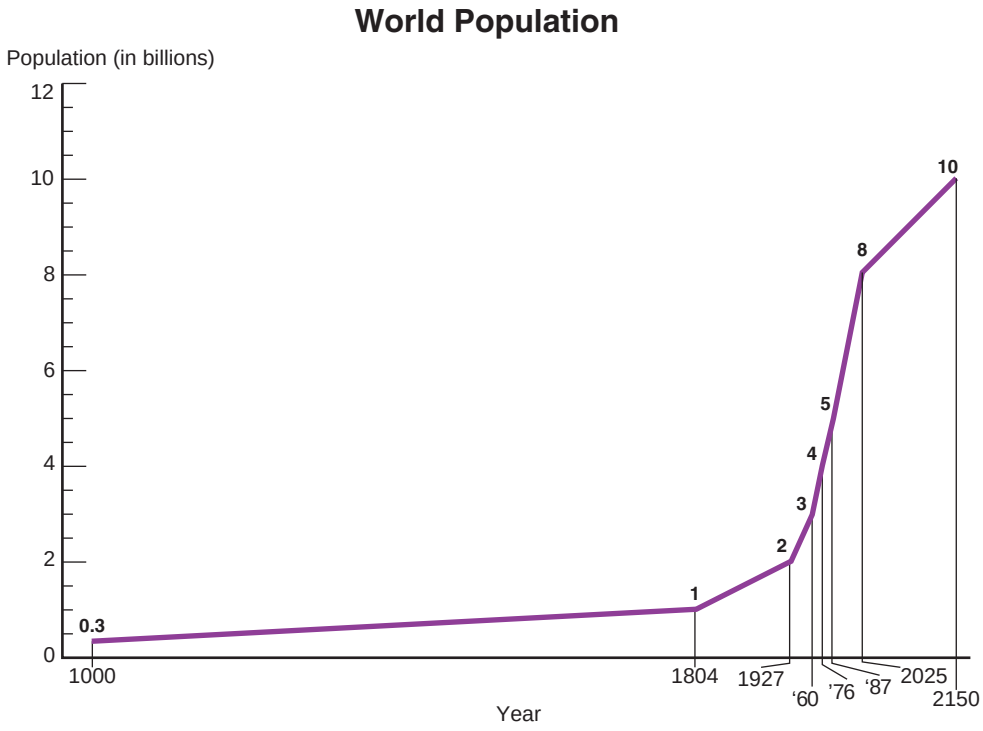
What do women have to do to achieve good contraceptive efficacy, and if they are already using a method, to switch to a more effective one? Choices must be available for various methods. The technique of using a method must be compatible with an individual and her lifestyle. Some methods require partner cooperation. Once chosen and obtained, the individual must exert dedication to its use. The failure to substantially improve contraceptive failure rates from 1995 to 2002 indicates that we are not making enough progress with each of these variables.

It is not enough to say the obvious—that we need greater education—but we need to learn where and when education is most effective, where is money best spent, and how to maximize the choices available for all women. This isn't a task just for professional health care providers; it is a widespread social problem that requires policy and budgeting decisions. The problems *are* more sociologic, such as cost and insurance coverage (and the ridiculous insurance practice of providing pills only one month at a time). These are reasons why other countries have lower percentages of women at risk for unintended pregnancies.

The pattern of contraceptive use in Canada is similar to that of the U.S., with a similar percentage of oral contraceptive use (about 43% of women 15–44 years of age) and a slightly lower use of sterilization.[29, 30] Canada, too, has seen an increase in condom use and a decrease in use of the IUD. In England, the primary method of contraception is oral contraception (28%) followed by condoms (24%), the IUD (4%), and injectable methods (3%); 7% of the reproductive-aged women and 10% of the men have been sterilized.[31] In France, 49% of reproductive-aged women use oral contraceptives, and although IUD use has slightly decreased (only among younger women), French women use the IUD at a rate that is more than 16-fold greater compared with North American women.[32, 33] Most French women use oral contraceptives when young and then turn to the IUD in their older years (only 4.1% of French women relied on sterilization; male sterilization is virtually nonexistent).

The Impact of the Worldwide Use of Contraception

The world population is expected to stabilize at above 10 billion after 2180, with a fertility rate of 2.1 children per woman.[34] Approximately 96% of the population growth now occurs in developing countries, so that by 2050, 10% of the population will live in developed countries, a decrease from the current 25%. Today, the fertility rate is about 1.6 children per women in China, Eastern and Western Europe, North America, Japan, Australia, and New Zealand.[34] Some time after 2020, *all* of the growth in global population will occur in developing countries.



WORLD POPULATION

1 billion - achieved in 1804
2 billion - achieved in 1927
3 billion - achieved in 1960
4 billion - achieved in 1974
5 billion - achieved in 1987
6 billion - achieved in 1999
8 billion - in 2025
9 billion - in 2050

Throughout the world, 45% of married women of reproductive age practice contraception. However, there is significant variation from area to area; e.g., more than 70% in the U.S. and China but only 6% in Nigeria.[35] About 71 million married women living in developing countries are at risk of an unplanned pregnancy.[36] Less than 15% of women of reproductive age in the world are using oral contraceptives, and more than half live in the U.S., Brazil, France, and Germany.

The 76% of the world's population living in developing countries account for:

−85% of all births,
−95% of all infant and childhood deaths,
−99% of all maternal deaths.

The problem in the developing world is self-evident. The ability to regulate fertility has a significant impact on infant, child, and maternal mortality and morbidity. A pregnant woman has a 200 times greater chance of dying if she lives in a developing country rather than in a developed country.[37] The health risks associated with pregnancy and childbirth in the developing world are far, far greater than risks secondary to the use of modern contraception.[38] To meet the projected growth in the world's population, the number of women using family planning will need to increase substantially from 1998 to 2025; for example, 40 million more women in India will need to use some method of contraception![35] In the developing world, about 140 million women who don't want to get pregnant are not using contraception.

In recent years, there has been an appropriate shift from a narrow focus on contraception to a broader view that encompasses the impact of poverty, emphasizes overall well-being and the rights of individuals, endorses gender equality, and examines the interactions among these issues.[39] It is not enough to simply limit fertility; contraception is only one component of reproductive health.

The Impact of Use and Nonuse

Inadequate access to contraception is associated with a high induced abortion rate. Effective contraceptive use largely, although not totally, replaces the resort to abortion. The combination of restrictive abortion laws and the lack of safe abortion services continues to make unsafe abortion a major cause of morbidity and mortality throughout the world, especially in many developing countries where abortion services are illegal.[40] Both safe and unsafe abortions can be minimized by maximizing contraceptive services. However, the

need for safe abortion services will persist because contraceptive failures account for about half of the 1.2 million annual induced abortions in the U.S.[41]

In the U.S., money spent on public funding for family planning saves money spent on medical, welfare, and nutritional services.[42] States with higher family planning expenditures have fewer induced abortions, low-birth weight newborns, and premature births.[43] The investment in family planning leads to short-term reductions in expenditures on maternal and child health services and, after 5 years, a reduction in costs for education budgets. Cutting back on publicly funded family planning services largely affects poor women, increasing the number of unintended births and abortions.

Sexually Transmitted Infections and Contraception

The interaction between clinician and patient for the purpose of contraception provides an opportunity to control sexually transmitted infections (STIs). The modification of unsafe sexual practices reduces the risk of unplanned pregnancy and the risk of infections of the reproductive tract. A patient visit for contraception is an excellent time for STI screening; if an infection is symptomatic, it should be diagnosed and treated during the same visit in which contraception is requested. A positive history for STIs should trigger both screening for asymptomatic infections and counseling for safer sexual practices. Attention should be given to the contraceptive methods that have the greatest influence on the risk of STIs.

Global Warming and Contraception

In the midst of politics and philosophy heavily promoting a "green" effort to limit global warming, a very important point is being ignored. ***Even small increases in population have a major impact on the global environment, including excessive consumption of resources in affluent societies.***

Thomas Robert Malthus, an English clergyman, mathematician, and political economist, published six editions of his famous book, *An Essay on the Principle of Population*, between the years 1798 and 1826. The Malthusian Hypothesis can be expressed very simply: the human population will outgrow the world's resources needed for its support. Malthus argued that population could be controlled only by a high death rate or a low birth rate. But because he didn't approve of birth control, he concluded that a high death rate would be necessary, caused by misery, in the form of wars, famine, and disease, and vice (contraception was in this category, along with murder). Without misery and vice, overpopulation, therefore, would lead to poverty, an animalistic competition for food, and a general loss of civilization.

The Malthusian Hypothesis has been resurrected in recent times. There is a growing awareness that our planet is running out of clean air, potable water, and specific agricultural and mineral commodities. Optimists look to the power of technology and human creativity to solve this Malthusian problem, but the acute need for effective contraception cannot be ignored, as it is by most economists. Effective family planning programs not only benefit individuals, but also national economies and the global environment. The need

and demand for family planning are extant in every part of our world, although greatest in the developing countries. Lacking is the required political and financial commitment. An appreciation for the impact on global warming can provide added motivation.

Contraception and Litigation

Clinicians are concerned about the prospect of bad outcomes associated with contraceptive use leading to litigation. Multimillion dollar verdicts and settlements in favor of plaintiffs who have used products as innocent as spermicides capture national attention. Actually, these events are very unusual compared with the widespread use of contraception.

The best way to avoid litigation is good patient communication. Patients who sue usually claim there were contraindications or risks that were not conveyed by the clinician. The best way to influence litigation is to keep good records. Good clinician's records are the most formidable weapon for the defense. Documentation is vital, but it is useless without thorough history taking. Good records and good history taking put the responsibility on the patient's honesty in response to the clinician.

Document that the risks and benefits of all methods were discussed.
Document a plan for follow-up.
Document all interactions with the patient, including phone calls.

The Future

From 1970 to 1986, the number of births in women older than 30 quadrupled; from 1990 to 2005, the fertility rate among women older than 30 remained relatively stable, but in 2005 and 2006, the birth rates for women over 30 and for women over 40 increased.[9, 44, 45] As couples deferred pregnancy until later in life, the use of sterilization under age 35 declined, and the need for reversible contraception increased.

Until 2005, the highest number of births in the U.S. occurred between 1947 and 1965—the post-World War II baby boom (a demographic phenomenon shared by all parts of the developed world). The entire cohort of women born in this period will have reached their 45th birthday around 2010. We have experienced, therefore, an unprecedented number of women in the later childbearing years. This group of women not only increased in number but changed its fertility pattern.

The deferment of marriage is a significant change in our society. However, only a small percentage of the decline in the total fertility rate is accounted for by the increase in the average age at first marriage. Most of the decline in total fertility rate is accounted for by changes in marital fertility rates. In other words, postponement of pregnancy in marriage is the more significant change. This combination of increasing numbers, deferment of marriage, and postponement of pregnancy in marriage is responsible for the fact that we are seeing more and more older women who need reversible contraception. In short, there will continue to be longer duration of contraceptive use in younger women and greater use in older women, a pattern that began in 1990.

Change in U.S. Female Demographics 1985–2000[46]					
Age	*1985*	*1990*	*1995*	*2000*	*% Change 1985–2000*
15–24	19.5 mill.	17.4 mill.	16.7 mill.	17.7 mill.	–9.2%
25–29	10.9	10.6	9.3	8.6	–21.1
30–34	10.0	11.0	10.8	9.4	–6.0
35–44	16.2	19.1	21.1	21.9	+35.2
Total 15–44	56.6	58.1	57.9	57.6	+1.8

Fortunately, clinicians and patients recognized that low-dose steroid contraception is very safe for healthy, nonsmoking, older women. Between 1988 and 1995, the use of oral contraceptives doubled among women aged 35–39, and increased 6-fold in women older than age 40.[26] However, as the previously mentioned statistics indicate, its use is still not sufficient to meet the need. In addition to fulfilling a need, this population of women has a series of benefits to be derived from steroid contraception that tilt the risk/benefit ratio to the positive side (Chapter 22).

The growing need for reversible contraception would also be served by increased use of the IUD. The decline in IUD use in the U.S. was in direct contrast to the experience in the rest of the world, a complicated response to publicity and litigation. An increased risk of pelvic infection with contemporary IUDs in use is limited to the act of insertion and the transportation of pathogens to the upper genital tract. This risk is effectively minimized by careful screening with preinsertion cultures and the use of good technique. A return to IUD use by American couples is both warranted and desirable.

A major problem in the U.S. is the prevalence of misconceptions. More than half of women, even well-educated women, are not accurately aware of the efficacy or the benefits and side effects associated with contraception.[47–49] Unfortunately, a significant percentage of women still do not know that there are many health benefits with the use of steroid contraception. Misconceptions regarding contraception have, in many instances, achieved the stature of myths. Myths are an obstacle to good utilization and can only be dispelled by accurate and effective educational efforts.

Contraceptive advice is a component of good preventive health care. The clinician's approach is a key. This is an era of informed choice by the patient. Patients deserve to know the facts and need help in dealing with the state of the art and the uncertainty. But there is no doubt that patients, especially young patients, are influenced in their choice by their clinicians' advice and attitude. Although the role of a clinician is to provide the education necessary for the patient to make proper choices, one should not lose sight of the powerful influence exerted by the clinician in the choices ultimately made. In the 1970s, we approached the patient with great emphasis on risk. In the 1990s, studies effectively documented the risks and benefits of contraception. In the new century, the approach should be different, highlighting the benefits and the greater safety of appropriate contraception. If one attempts to sum the impact of the benefits of contraception on public health, as some have done with models focusing on hospital admissions, there is no doubt that the benefits outweigh the risks. The impact can be measured in terms of both morbidity and mortality. However, the impact on public health is of little concern during the clinician–patient interchange in the medical office. Here personal risk is paramount, and compliance with effective contraception requires accurate information presented in a positive, effective fashion.

Sterilization

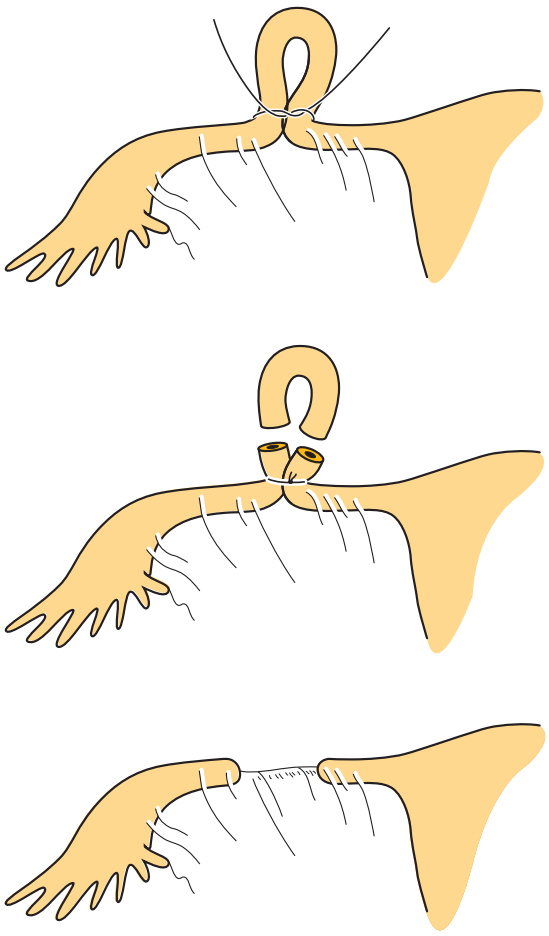
Contraceptive methods today are very safe and effective; however, we remain decades away from a perfect method of contraception for either women or men. Because reversible contraceptive methods are not perfect, more than a third of American couples use sterilization instead, and sterilization is now the predominant method of contraception in the world.

Over the past 20 years, over 1 million Americans each year have undergone a sterilization operation, and recently, more women than men. Currently 39% of reproductive-aged American women rely on contraceptive sterilization: 27% undergo tubal occlusion (11 million women), and 11% depend on their partners' vasectomies (4 million men).[26] This same trend has occurred in Great Britain, where by age 40, more than 20% of men and women have had a sterilization procedure.[50] In Spain and Italy, sterilization rates are very low, but the use of oral contraceptives and the IUD is very high.[51]

History

James Blundell proposed in 1823, in lectures at Guy's Hospital in London, that tubectomy ought to be performed at cesarean section to avoid the need for repeat sections.[52] He also proposed a technique for sterilization, which he later described so precisely that he must

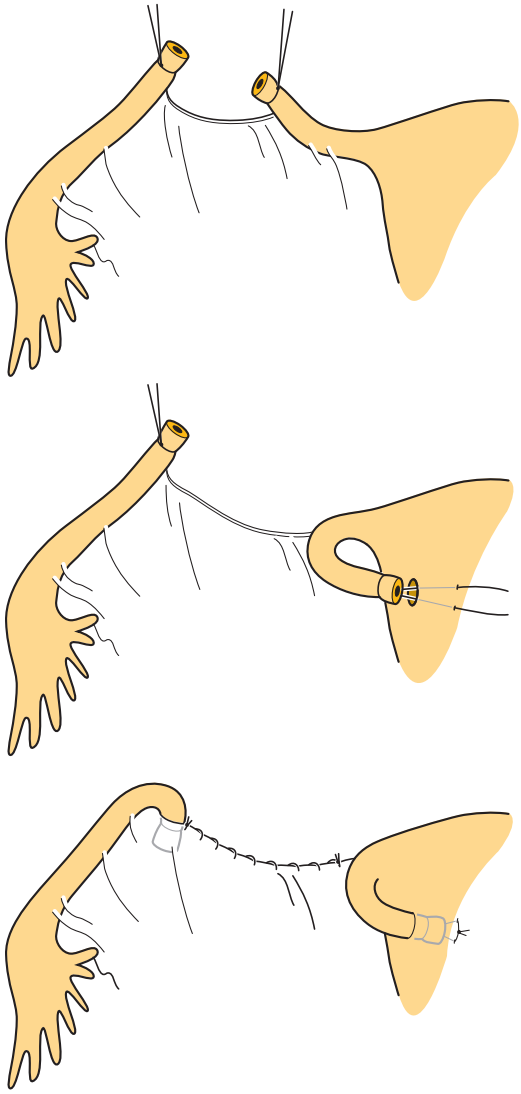
The Pomeroy



actually have performed the operation, although he never wrote about it. The first report was published in 1881 by Samuel Lungren of Toledo, Ohio, who ligated the tubes at the time of cesarean section, as Blundell had suggested 58 years earlier.[53] The Madlener procedure was devised in Germany in 1910 and reported in 1919. Because of many failures, the Madlener technique was supplanted in the U.S. by the method of Ralph Pomeroy, a prominent physician in Brooklyn, New York. This method, still popular today, was not described to the medical profession by Pomeroy's associates until 1929, 4 years after Pomeroy's death. Frederick Irving of the Harvard Medical School described his technique in 1924, and the Uchida method was not reported until 1946.

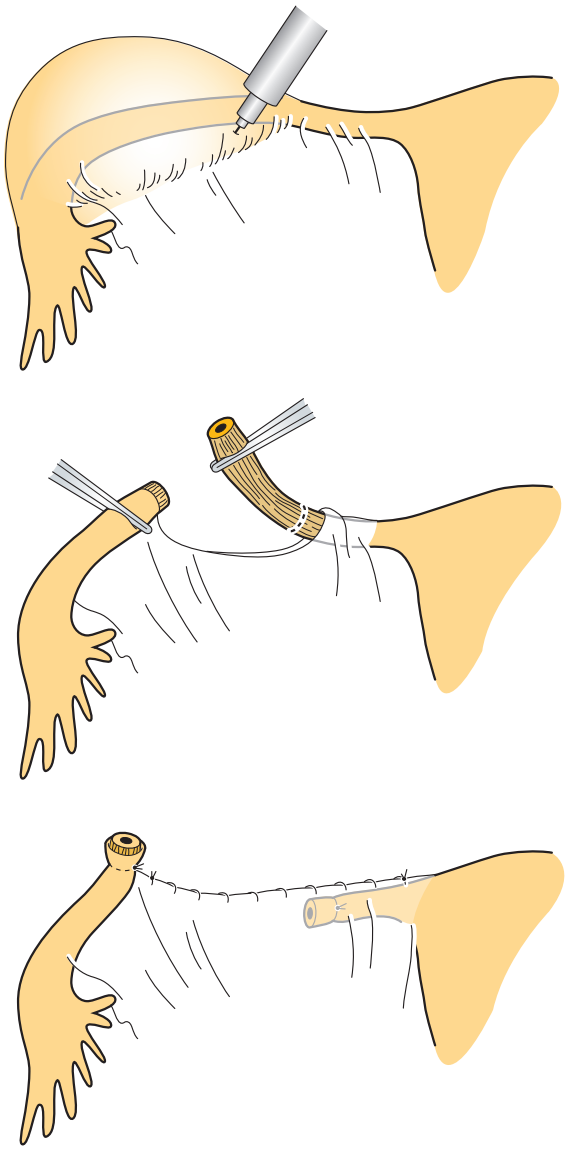
Few sterilizations were performed until the 1930s when "family planning" was first suggested as an indication for surgical sterilization by Baird in Aberdeen. He required women to be older than 40 and to have had eight or more children. Mathematical formulas of this kind persisted through the 1960s. In 1965, Sir Dugald Baird delivered a remarkable lecture, entitled "The Fifth Freedom," calling attention to the need to alleviate the fear of unwanted pregnancies and the important role of sterilization.[54] By the end of the 1960s, sterilization was a popular procedure.

The Irving



Laparoscopic methods were introduced in the early 1970s. The annual number of vasectomies began to decline, and the number of tubal occlusion operations increased rapidly. By 1973, more sterilization operations were performed for women than for men. This is accurately attributed to dramatic decreases in costs, hospital time, and pain because of the introduction of laparoscopy and minilaparotomy methods. The use of laparoscopy for tubal occlusion increased from only 0.6% of sterilizations in 1970 to more than 35% by 1975.[55] Since 1975, minilaparotomy, a technique popular in the less developed world, has been increasingly performed in the U.S. These methods have allowed women to undergo sterilization operations at times other than immediately after childbirth or during major surgery.

The Uchida



Laparoscopy and minilaparotomy have led to a profound change in the convenience and cost of sterilization operations for women. In 1970, the average woman stayed in the hospital 6.5 days for a tubal sterilization. By 1975, this had declined to 3 days, and today, women rarely remain in the hospital overnight. The shorter length of stay achieved from 1970 to 1975 represented a savings of more than $200 million yearly in health care costs and a

tremendous increase in convenience for women eager to return to work and their families.[56] Unlike some advances in technology, laparoscopy and minilaparotomy sterilization are technical innovations that have resulted in large savings in medical care costs.

The great majority of sterilization procedures are accomplished in hospitals by physicians in private practice, but a rapidly increasing proportion is performed outside of hospitals in ambulatory surgical settings, including physicians' offices. In either hospital or outpatient settings, female sterilization is a very safe operation. Deaths specifically attributed to sterilization now account for a fatality rate of only 1.5 per 100,000 procedures, a mortality rate that is lower than that for childbearing (about 8 per 100,000 births in the U.S.).[57, 58] When the risk of pregnancy from contraceptive method failure is taken into account, sterilization is the safest of all contraceptive methods.

Vasectomy has long been more popular in the U.S. than anywhere else in the world, but why don't more men use it? One explanation is that women have chosen laparoscopic sterilization in increasing numbers. Another is that men have been frightened by reports, often from animal data, of associations with autoimmune diseases, atherosclerosis, and, most recently, prostatic cancer. Large epidemiologic studies have failed to confirm any definite adverse consequences.[59] When patients consider sterilization, we can assure them that vasectomy has not been demonstrated to have any harmful effects on men's health.[60] In addition, vasectomy is less expensive than tubal sterilization, morbidity is less, and mortality is essentially zero.

Efficacy of Sterilization

Laparoscopic and minilaparotomy sterilizations are not only convenient, they are almost as effective at preventing pregnancy as were the older, more complex operations. Vasectomy is also highly effective once the supply of remaining sperm in the vas deferens is exhausted. Approximately 50% of men will reach azoospermia at 8 weeks, but the time to achieve azoospermia is highly variable, reaching only about 60% to 80% after 12 weeks.[61, 62]

Failure Rates During the First Year of Use, United States[17, 18, 20]

	Percent of women with pregnancy	
Method	*Lowest Expected (%)*	*Typical (%)*
Female sterilization	0.5	0.7
Male sterilization	0.1	0.2

In addition to the specific operation used, the skill of the operator and characteristics of the patient make important contributions to the efficacy of female sterilization. Up to 50% of failures are due to technical errors. The methods using complicated equipment, such as spring-loaded clips and silastic rings, fail for technical reasons more commonly than do simpler procedures such as the Pomeroy tubal ligation.[63] Minilaparotomy failures, therefore, occur much less frequently from technical errors.

It is hardly surprising that more complicated techniques of tubal occlusion have higher technical failure rates. What is surprising is the finding that characteristics of the patient influence the likelihood of failure even when technical problems are controlled for in

analytical adjustments. In a careful study of this issue, two patient characteristics, age and lactation, demonstrated a significant impact.[64] Patients younger than 35 years were 1.7 times more likely to become pregnant, and women who were not breastfeeding following sterilization were 5 times more likely to become pregnant. These findings probably reflect the greater fecundity of younger women and the contraceptive contribution of lactation.

Significant numbers of pregnancies after tubal occlusion are present before the procedure. For this reason, some clinicians routinely perform a uterine evacuation or curettage prior to tubal occlusion. It seems more reasonable (and cost-effective) to exclude pregnancy by careful history taking, physical examination, and an appropriate pregnancy test prior to the sterilization procedure.[65]

Because method, operator, and patient characteristics all influence sterilization failures, it is difficult to predict which individual will experience a pregnancy after undergoing a tubal occlusion. Therefore, during the course of counseling, all patients should be made aware of the possibility of failure as well as the *intent* to cause permanent, irreversible sterility. It is important to avoid giving patients the impression that the tubal occlusion procedure is foolproof or guaranteed. Individual clinicians must be cautious judging their own success in accomplishing sterilization because failure is infrequent and many patients who become pregnant after sterilization never reveal the failure to the original surgeon.

Ectopic pregnancies can occur following tubal occlusion, and the incidence is much higher with some types of tubal occlusion.[66–68] Bipolar tubal coagulation is more likely to result in ectopic pregnancy than is mechanical occlusion.[63, 69, 70] The probable explanation is that microscopic fistulae in the coagulated segment connecting to the peritoneal cavity permit sperm to reach the ovum. Ectopic pregnancies following tubal ligation are more likely to occur 3 or more years after sterilization, rather than immediately after. The proportion of ectopic pregnancies is 3 times as high in the fourth through the tenth years after sterilization as in the first 3 years.[70] For laparoscopic methods, the cumulative rate of ectopic pregnancy continues to increase for at least 10 years after surgery, reaching 17 per 1,000 for bipolar coagulation.[70] Overall, however, the risk of an ectopic pregnancy in sterilized women is lower than if they had not been sterilized. Nevertheless, approximately one-third of the pregnancies that occur after tubal sterilization are ectopic.[70]

Vaginal procedures have higher failure rates than laparoscopy or minilaparotomy, but the principal disadvantage is a higher rate of infection. Intraperitoneal infection is a rare complication of minilaparotomy or laparoscopic techniques, but in vaginal procedures, abscess formation approaches 1%.[71] This risk can be reduced by the use of prophylactic antibiotics administered intraoperatively, but open laparoscopy is usually easier and safer than vaginal sterilization even in obese women.

Sterilization and Ovarian Cancer—A Benefit of Sterilization

Serous ovarian cancer, the most common ovarian cancer, originates in the fimbriae of the fallopian tubes.[72–74] Evidence consistently indicates that tubal sterilization is associated with a major reduction in the risk of ovarian cancer.[75–79] Evidence from the Nurses' Health Study indicated that tubal sterilization was associated with a 67% reduced risk of ovarian cancer.[75] In the prospective mortality study conducted by the American Cancer Society, women who had undergone tubal sterilization experienced about a 30% reduction in the risk of fatal ovarian cancer.[76]

Female Sterilization Techniques

Because laparoscopy permits direct visualization and manipulation of the abdominal and pelvic organs with minimal abdominal disruption, it offers many advantages. Hospitalization is not required; most patients return home within a few hours, and the majority return to full activity within 24 h. Discomfort is minimal, the incision scars are barely visible, and sexual activity need not be restricted. In addition, the surgeon has an opportunity to inspect the pelvic and abdominal organs for abnormalities. The disadvantages of laparoscopic sterilization include the cost; the expensive, fragile equipment; the special training required; and the risks of inadvertent bowel or vessel injury.

Laparoscopic sterilization can be achieved with any of these methods:

1. Occlusion and partial resection by unipolar electrosurgery.
2. Occlusion and transection by unipolar electrosurgery.
3. Occlusion by bipolar electrocoagulation.
4. Occlusion by mechanical means (clips or silastic rings).

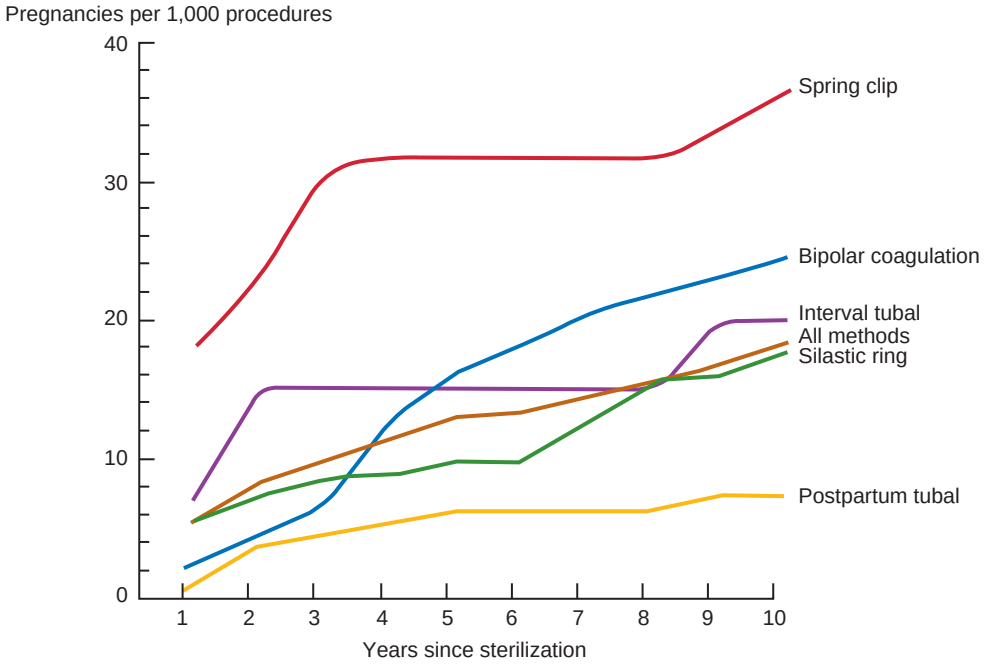
All of these methods can use an operating laparoscope alone, the diagnostic laparoscope with operating instruments passed through a second trocar, or both the operating laparoscope and secondary puncture equipment. All can be used with the "open" laparoscopic technique in which the laparoscopic instrument is placed into the abdominal cavity under direct vision to avoid the risk of bowel or blood vessel puncture on blind entry. Patient acceptance and recovery are approximately the same with all methods.

It is now apparent that the long-term failure rates for all methods are higher than previous estimates; overall, 1.85% of sterilized American women experience a failure within 10 years.[80] *As much as one-third of these failures are ectopic pregnancies.*[70] The higher failure rates with silastic rings, the Hulka-Clemens clip, and bipolar coagulation reflect the greater degree of skill required for these methods. Because of the effect of declining fecundity with increasing age, younger sterilized women are more likely to have a failure, including ectopic pregnancy, compared with older women. For these reasons, younger women seeking sterilization should consider the use of the IUD or implants, reversible methods that offer very low failure rates.

Female Tubal Sterilization Methods 10-Year Cumulative Failure Rates[80, 81]

Unipolar coagulation	-	**0.75%**
Postpartum tubal excision	-	**0.75%**
Filshie clip	-	**0.56%**
Silastic (Falope or Yoon) ring	-	**1.77%**
Interval tubal excision	-	**2.01%**
Bipolar coagulation	-	**2.48%**
Hulka-Clemens clip	-	**3.65%**

Life Table Cumulative Probability of Pregnancy[80]



Tubal Occlusion by Electrosurgical Methods

If electrons from an electrosurgical generator are concentrated in one location, heat within the tissue increases sharply and desiccates the tissue until resistance is so high that no more current can pass. Unipolar methods of sterilization create a dense area of current under the grasping forceps of the unipolar electrode. To complete the circuit, however, these electrons must spread through the body and be returned to the generator via a return electrode (the ground plate) that has a broad surface to minimize the density of the current to avoid burns as the electrons leave the body. "Unipolar" refers to the method that requires the patient ground plate.

Unipolar electrosurgery can create a unique electrical "capacitance" problem. A capacitor is any device that can hold an electric charge and can exist wherever an insulated material separates two conductors that have different potentials. This property of capacitance explains some of the inadvertent bowel burns that occurred with laparoscopic sterilization.[82] The operating laparoscope is a hollow metal tube surrounding an active electrode, the forceps used to grasp and coagulate the tubes. When current passes through the active electrode, the laparoscope itself becomes a capacitor. Up to 70% of the current passed through the active electrode can be induced into the laparoscope. If bowel or other structures touch a laparoscope, which is insulated from the abdominal incision (e.g., by a plastic cannula), the stored electrons will be discharged at high density directly into the vital organ. This potential hazard is eliminated by using a metal trocar sleeve rather than a nonconductive sleeve. Because there is little pressure behind the electrons from a low-voltage generator, not enough heat is generated to burn the skin as the capacitance current leaks out into the patient's body through the metal sleeve. Even if the active electrode comes in direct contact with the laparoscope, as when a two-incision technique is used, the current will leak harmlessly through the metal trocar sleeve. The risk of inadvertent coagulation of bowel or other organs cannot be completely eliminated because all body surfaces offer a path back to the ground plate.

The unipolar electrosurgical technique is straightforward. The isthmic portion of the fallopian tube is grasped and elevated away from the surrounding structures, and the electrical energy is applied until the tissue blanches, swells, and then collapses. The tube is then grasped, moving toward the uterus, recoagulated, and the steps repeated until 2–3 cm of tube have been coagulated. Some surgeons advise against cornual coagulation for fear it may increase the risk of ectopic pregnancy due to fistula formation.

The coagulation and transection technique is performed in a similar fashion with the same instruments. In order to transect the tube, however, an instrument designed to cut tissue must be used. The transection of tissue increases the risk of possible bleeding and does not, by itself, reduce the failure rate over coagulation alone. The specimens obtained by this method are usually coagulated beyond microscopic recognition and, therefore, will not provide pathologic evidence of successful sterilization.

The bipolar method of sterilization eliminates the ground plate required for unipolar electrosurgery and uses a specially designed forceps. One jaw of the forceps is the active electrode, and the other jaw is the ground electrode. Current density is great at the point of forceps contact with tissue, and the use of a low-voltage, high-frequency current prevents the spread of electrons. By eliminating the return electrode, the chance of an aberrant pathway through bowel or other structures is greatly reduced. There is, however, a disadvantage with this technique. Because electron spread is decreased, more applications of the grasping forceps are necessary to coagulate the same length of tube than with unipolar coagulation. As desiccation occurs at the point of high current density, tissue resistance increases, and the coagulated area eventually provides resistance to flow of the low-voltage current. Should the resistance increase beyond the voltage's capability to push electrons through the tissue, incomplete coagulation of the endosalpinx can result.[83] ***Bipolar coagulation is very effective only if three or more sites are coagulated on each tube.***[84]

Bipolar cautery is safer than unipolar cautery with regard to burns of abdominal organs, but most studies indicate higher failure rates. Although the bipolar forceps will not burn tissues that are not actually grasped, care must be taken to avoid coagulating structures adherent to the tubes. For example, the ureter can be damaged when the tube is adherent to the pelvic side wall.

Tubal Occlusion with Clips and Rings

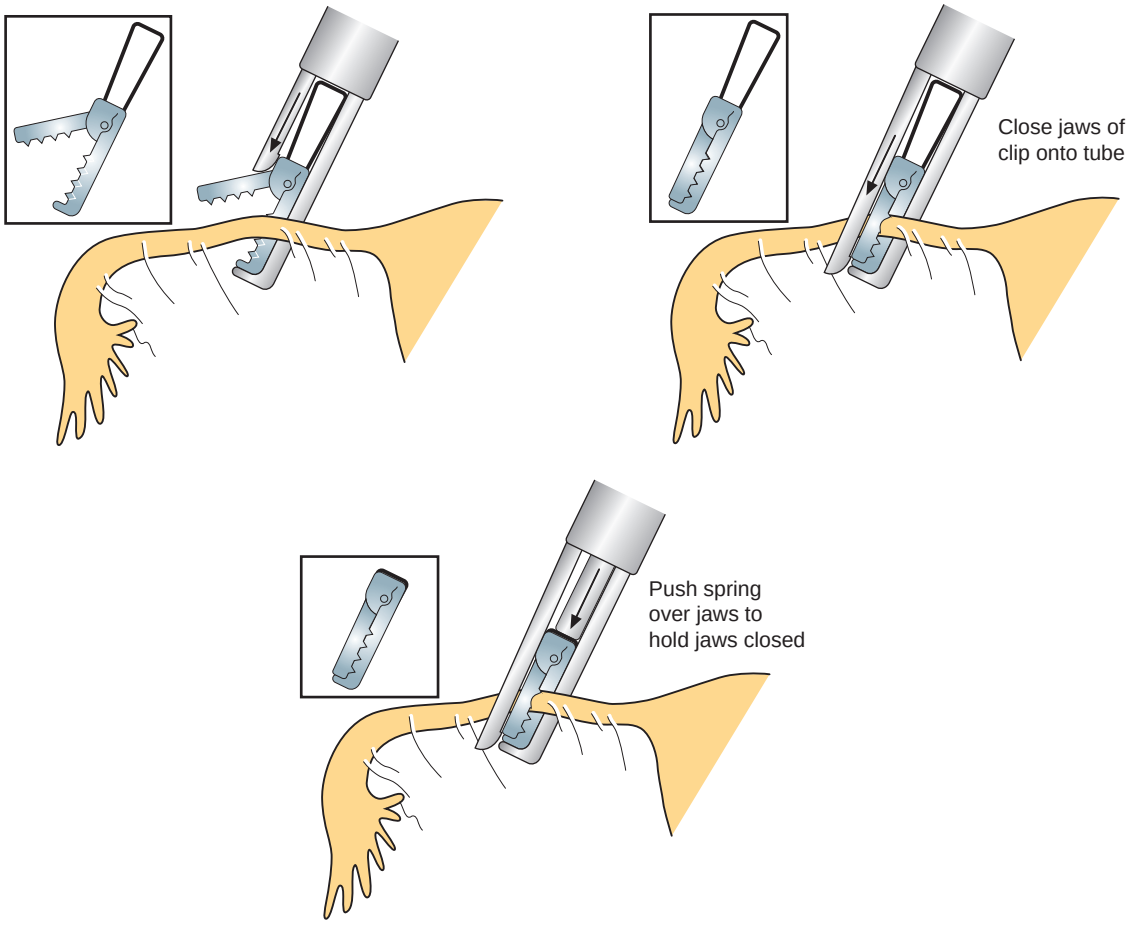
Female sterilization by mechanical occlusion eliminates the safety concerns with electrosurgery. However, mechanical devices are subject to flaws in material, defects in manufacturing, and errors in design, all of which can alter efficacy. Three mechanical devices have been widely used and have low failure rates with long-term follow-up: the Hulka-Clemens (spring) clip, the Filshie Clip, and the silastic (Falope or Yoon) ring. Each of the three requires an understanding of its mechanical function, a working knowledge of the intricate applicator necessary to apply the device, meticulous attention to maintenance of the applicators, and skillful tubal placement. These devices are less effective when used immediately postpartum on dilated tubes.

Hulka-Clemens Spring Clip

The spring clip consists of two plastic jaws made of Lexan, hinged by a small metal pin 2 mm from one end. Each jaw has teeth on the opposed surface, and a stainless steel spring is pushed over the jaws to hold them closed over the tube. A special laparoscope for

one-incision application is most commonly used, although the spring clip can also be used in a two-incision procedure. The spring clip is applied at a 90-degree angle to include some mesosalpinx at the proximal isthmus of a stretched fallopian tube. The spring clip destroys 3 mm of tube and has 1-year pregnancy rates of 2 per 1,000 women but the highest 10-year cumulative failure rate.[63, 80]

The Hulka-Clemens Spring Clip



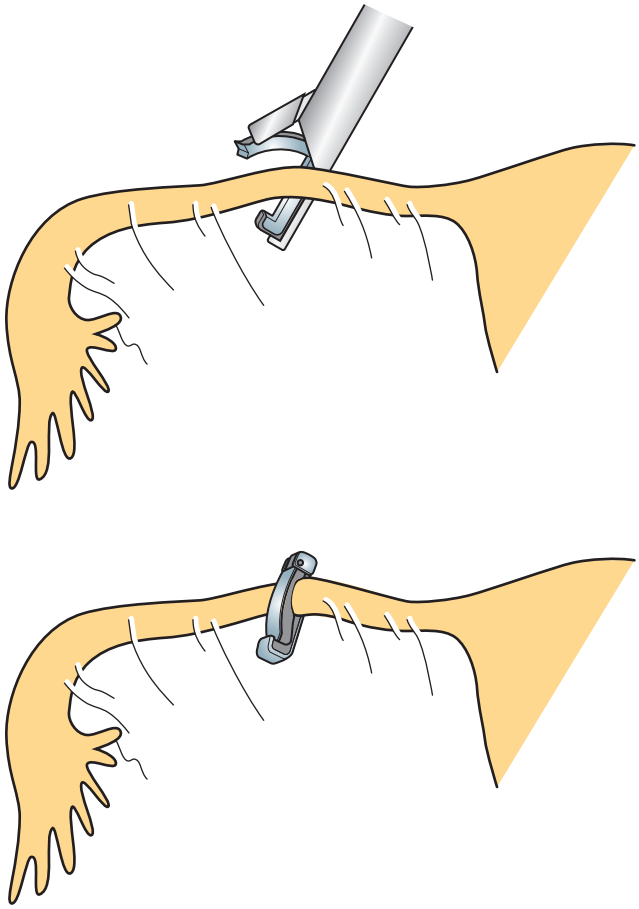
Complications unique to spring clip sterilization result from mechanical difficulties. If the clip is dislodged or dropped into the abdomen during the procedure, it should be retrieved. Usually, it can be removed laparoscopically, but sometimes laparotomy is necessary. Should incomplete occlusion or incorrect alignment of the clip occur, a second clip can be applied without hazard. This clip offers a good chance for reanastomosis, better than electrosurgical methods that destroy more tube.

Filshie Clip

The Filshie clip is made of titanium lined with silicone rubber. The hinged clip is locked over the tube using a special applicator through a second incision or operating laparoscope. The rubber lining of the clip expands on compression to keep the tube blocked. Only 4 mm of the tube are destroyed. Failure rates 1 year after the procedure approximate 1 per 1,000

women.[68] A 15-year follow-up study in Quebec reported a cumulative failure rate of 9 per 1,000 women, whereas the 10-year failure rate in the U.K. was 5.6 per 1,000 woman.[81, 85] Because the Filshie clip is longer, it is reported to occlude dilated tubes more readily than does the spring clip. Both the spring clip and the Filshie clip provide good chances for tubal reanastomosis.

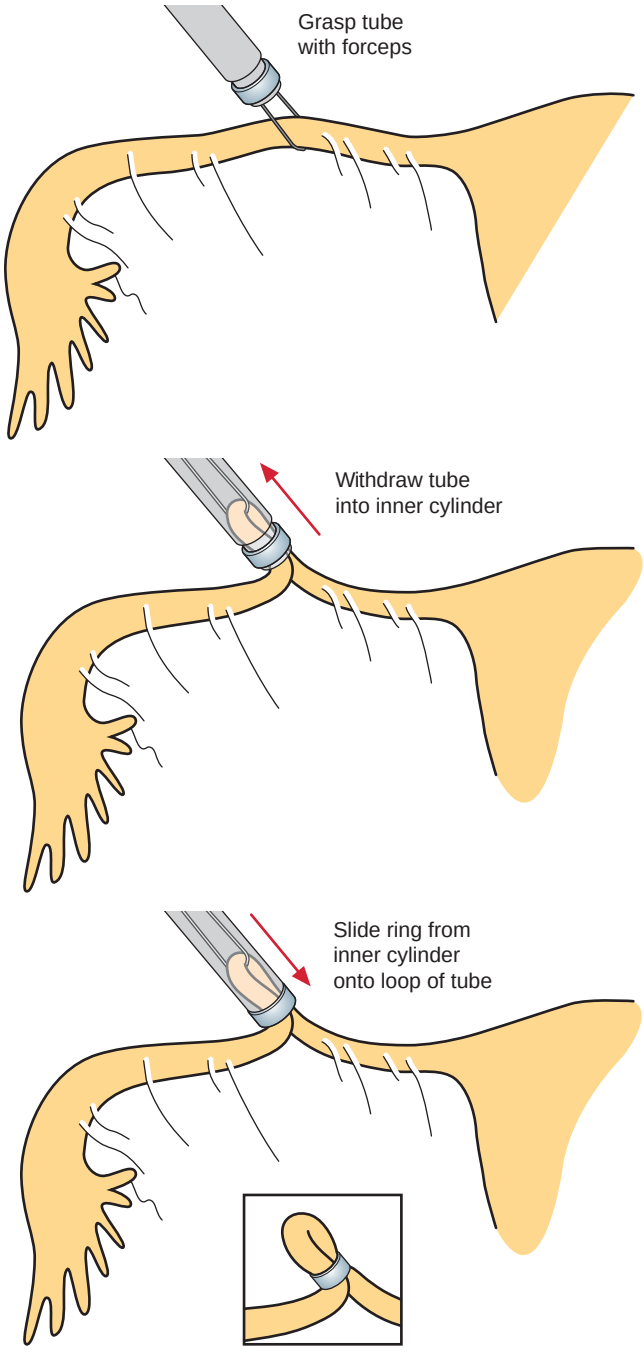
The Filshie Clip



Silastic (Falope or Yoon) Ring

This nonreactive silastic rubber band has an elastic memory of 100% if stretched to no more than 6 mm for a brief time (a few minutes at most). A special applicator, 6 mm in diameter, can be placed through a second cannula or through a standard offset operating laparoscope. The applicator is designed to grasp a knuckle of tube and release the silastic band onto a 2.5 cm loop of tube. The avascular loop of tube can be resected with biopsy forceps to provide a pathology specimen, but this is rarely performed (it does not increase efficacy). Ten percent to 15% of patients experience severe postoperative pelvic cramping from the tight bands (which can be alleviated by the application of a local anesthetic to the tube before or after banding).

The Silastic (Falope-Yoon) Ring



As with application of clips, the ring should be placed at the junction of the proximal and middle third of each fallopian tube. Necrosis occurs promptly and a 2–3 cm segment of the tube is destroyed. Failure rates are about 1% after 2 years, and the 10-year cumulative rate is better only with unipolar coagulation, postpartum tubal excision, and the Filshie clip.[80]

Mesosalpingeal bleeding is the most common complication of silastic ring application. It usually occurs when the forceps grabs not only the tube but also a vascular fold of mesosalpinx. The mesosalpinx can also be torn on the edge of the stainless steel cylinder as the

tube is drawn into the applicator. If bleeding is noted, application of the silastic band often controls it. If the placement of additional bands or electrocoagulation fails to stop bleeding, laparotomy may be required.

Silastic rings are occasionally placed on structures other than the tube. If this mistake is recognized, the band can usually be removed from the round ligament or mesosalpingeal folds by grasping the band with the tongs of the applicator and applying gradual, increasing traction. If a gentle attempt fails, removal is not necessary. If rings are inadvertently discharged into the peritoneal cavity, they can safely be left behind.

Patients should be prepared for the use of electrosurgical instruments in case bands or clips cannot be applied (because of adhesions or bleeding).

Minilaparotomy

Tubal ligation, accomplished through a small suprapubic incision, *minilaparotomy*, is the most frequent method of interval female sterilization around the world. In the U.S. and most of the developed world, laparoscopy is more popular, but minilaparotomy is gaining in favor because of its safety, simplicity, and adaptability to ambulatory surgical settings (particularly when local anesthesia is used).[86, 87]

The fallopian tubes can be occluded through the minilaparotomy incision with bands or clips, but a simple Pomeroy tubal ligation is the method most commonly used. Patient characteristics, such as obesity, previous pelvic infection, or previous surgery, are the principal determinants of complications.[88]

Minilaparotomy is accomplished through an incision that usually measures 2–4 cm in length. Tubal ligation through a suprapubic incision can be accomplished for obese patients, but the incision will necessarily exceed the usual length. Forceful retraction increases the pain associated with the procedure and the time of recovery. For these reasons, we believe that minilaparotomy for ambulatory tubal occlusion should be limited to patients who are not obese (usually less than 150–160 lb, 70 kg).

Tubal occlusion is difficult to accomplish through a minilaparotomy if the uterus is immobile. Laparoscopic tubal occlusion, on the other hand, does not require extreme uterine elevation or rotation and is a better choice for a patient with a uterus fixed in position.

The Transcervical Approach

Although current methods of sterilization are safe and effective, they require skillful surgeons and, in the case of laparoscopic operations, elaborate and expensive equipment. Simpler approaches could make sterilization available and acceptable to more women. Transcervical methods have used electrocoagulation, cryosurgery, or laser to destroy the interstitial portion of the tube, to inject sclerosing agents or tissue adhesives (Femcept) through the tubal ostia, and to mechanically obstruct the tubal lumen. Most of these methods, and the formed-in-place silicone plugs applied hysteroscopically are either too complicated or have high failure rates. The most practical approach is the application of sclerosing agents to the tubal openings using cannulae or an intrauterine device. Transcervical insertion of quinicrine pellets during the proliferative phase of the menstrual cycle occludes the tubes and is the most promising of the “non-surgical” approaches, but long-term safety and efficacy have not been assessed.[89]

Essure is a metal coil device with polyester fibers (polyethylene terephthalate), which when placed hysteroscopically within the proximal segment of the fallopian tube spanning the utero-tubal junction, expands when released, anchoring itself in place. The polyester fibers in the device stimulate a tissue reaction, which is fibrotic and occlusive. After backup contraception for 3 months, a hysterosalpingogram is performed to confirm occlusion. If occlusion is not present at 3 months, contraception is continued and hysterosalpingography is repeated 3 months later. The procedure is quick, performed in the outpatient setting, often without analgesia, and considerably less expensive than laparoscopy. Effective permanent sterilization was initially achieved in about 85–90% of women.[90] With greater experience, placement can reach a 96% rate.[91, 92] The risk of pregnancy after 5 years was 2.6 per 1,000 woman, with no ectopic pregnancies.[93]

Adiana is a two-step procedure. Through a hysteroscope, radiofrequency energy is delivered to the fallopian tube to remove a thin layer of cells and stimulate tissue response. A soft silicone insert that is smaller than a grain of rice is implanted at the site. Tissue growth around the implant creates permanent blockage, confirmed by hysterosalpingography 3 months after the procedure. Successful placement is achieved in 92–95% of women, with a 1-year pregnancy rate of 1.1%.[94, 95]

The Vaginal Approach

Although vaginal techniques are still used for tubal sterilization, high rates of infection and occasional pelvic abscesses following these operations have caused most clinicians to abandon them.[71] An apparent advantage in obese patients is sometimes deceptive because omental fat can block access to the fallopian tubes. Open laparoscopy is usually easier and safer in obese women.

Counseling for Sterilization

All patients undergoing a surgical procedure for permanent contraception should be aware of the nature of the operation, its alternatives, efficacy, safety, and complications. The operation can be described using drawings or pelvic models, slides, or recordings. The description of the operation should emphasize its similarities to and differences from laparoscopy and pelvic surgery, especially hysterectomy or ovariectomy that may be confused with simple tubal ligation. Women who undergo tubal sterilization by any method are 4–5-fold more likely to have a hysterectomy; no biologic explanation is apparent, and this may reflect patient attitudes toward surgical procedures.[96] Alternatives, including vasectomy, steroid contraception, long-acting hormone methods, barrier methods, and IUDs, should be reviewed. It is important to emphasize to the patient that tubal ligation is not intended to be reversible, that it cannot be guaranteed to prevent intrauterine or ectopic pregnancy, and that failures can occur long after the sterilization procedure. Informed consent is best obtained at a time when a patient is not distracted or distraught; e.g., not immediately before or after an induced abortion.

Sexuality

There is no detrimental effect on sexuality specifically due to sterilization procedures.[97, 98] Indeed, sexual life is usually positively affected. Many couples are less inhibited and more spontaneous in lovemaking when they do not have to worry about an unwanted pregnancy.

Menstrual Function

The effects of tubal sterilization on menstrual function have been confusing and, therefore, difficult to explain, but the issue is now resolved. The first well-controlled studies of this issue demonstrated no change in menstrual patterns, volume, or pain.[99, 100] Subsequently these same authors reported an increase in dysmenorrhea and changes in menstrual bleeding.[101, 102] However, authors often failed to agree in their findings (a change found by one group was not confirmed by the other). Adding to the confusion, the incidence of hysterectomy for bleeding disorders in women after tubal sterilization was reported to be increased by some,[103] but not by others.[104] In a large cohort of women in a group health plan, hospitalization for menstrual disorders was significantly increased; however, the authors believed this reflected bias by patient and physician preference for surgical treatment.[105] In the U.S. prospective long-term follow-up study of sterilization, the increased risk of hysterectomy after sterilization was concentrated in women who were treated for gynecologic disorders before tubal sterilization.[96] These discordant reports did not make patient counseling about the long-term effects of tubal sterilization an easy task.

It was initially speculated that extensive electrocoagulation of the fallopian tubes can cause ovarian tissue damage, changing ovarian steroid production. This was suggested as the reason why menstrual changes were detected with longer (4 years) follow-up, whereas no changes had been noted with the use of rings or clips.[105–107] However, attempts to relate poststerilization menstrual changes with extent of tissue destruction failed to find a correlation, and an increase in hospitalization for menstrual disorders after unipolar cautery could not be documented.[105, 107] A long-term follow-up study (3–4.5 years) failed to document any significant changes in menstrual cycles.[108] Follow-up studies at 3 months, 6 months, 2 years, and 5 years after sterilization by bipolar coagulation or Hulka clip could detect no changes in day 3 measures of ovarian reserve (gonadotropins, estradiol, and inhibin) or luteal phase levels of estradiol and progesterone.[109–112]

The U.S. Collaborative Review of Sterilization, the largest and most comprehensive assessment of sterilization, could find no evidence that tubal sterilization is followed at 2 years and again at 5 years by a greater incidence of menstrual changes or abnormalities.[107, 113] The evidence indicates that tubal sterilization does not cause menstrual abnormalities.

Reversibility

An important objective of counseling is to help couples make the right decision about an irreversible decision to become sterile. The active participation of both partners is a critical factor.[114] Not all couples are pleased following sterilization; in one series, 2% of U.S. women expressed regret 1 year later and 2.7% after 2 years.[115] At the 2-year mark, the main factors associated with regret were age less than 30 and sterilization at the convenient time of a cesarean section. In another long-term follow-up study of U.S. women, women younger than the age of 30 at the time of sterilization were most likely to express regret, but no differences were observed comparing timing after cesarean delivery, vaginal delivery, or a year later.[116] In the 1995 National Survey of Family Growth, nearly 25% of U.S. women with a tubal ligation expressed a desire for reversal, by either one of the partners or both.[117] The U.S. Collaborative Review of Sterilization reported that only 14.3% requested reversal information after 14 years of follow-up, but the percentage was nearly 4 times higher in women aged 18 to 24 at sterilization.[118] Overall, however, only 1.1% of women obtained reversal. In Canada, 1% of men and women obtained a reversal within 5 years

after sterilization; in the U.S. reversal within 5 years was obtained by 0.2% of women and 0.4% of men, a difference that could reflect the lack of insurance coverage for this procedure in the U.S.[85, 119]

In Europe where tubal sterilization is less common, the most important risk factor for regret was an unstable marriage.[120] A change in marital status is undoubtedly an important reason for a desire to reverse sterilization.[121]

Young couples in unstable relationships need special attention in counseling. Furthermore, for many couples tubal occlusion at the time of cesarean section or immediately after a difficult labor and delivery is not the best time for the procedure. It is important to know that sterilized women have not been observed to develop psychological problems at a greater than expected rate.[122, 123]

Microsurgery for tubal reanastomosis is associated with excellent results if only a small segment of the tube has been damaged. Pregnancy rates correlate with the length of remaining tube; a length of 4 cm or more is optimal. Thus, the pregnancy rates are lowest with electrocoagulation and reach 70–80% with clips, rings, and surgical methods such as the Pomeroy.[124, 125] About 2 per 1,000 sterilized women will eventually undergo tubal reanastomosis.[121]

Male Sterilization: Vasectomy

Vasectomy is safer, easier, less expensive, and has a lower failure rate than female sterilization.[126] The operation is almost always performed under local anesthesia, usually by a urologist in a private office.[127] Surgeons who do more than 10 operations yearly have lower complication rates.[128]

Hematomas and infection occur rarely and are easily treated with heat, scrotal support, and antibiotics. Most men will develop sperm antibodies following vasectomy, but no long-term sequelae have been observed, including no increased risk of immune-related diseases or cardiovascular disease.[60, 129–132] Adverse psychological and sexual effects have not been reported.[133] Because the other constituents of semen are made downstream from the testes, men do not notice a decreased volume or velocity of ejaculate.

Prostate cancer is the most frequent cancer among men, with a lifetime risk of 1 in 8 in the United States. An increased risk of prostate cancer after vasectomy was reported in several cohort and case-control studies.[134–137] However, there was disagreement because other studies could not support an association between prostate or testicular cancer risk and vasectomy.[131, 138–143] In a very large mixed racial/ethnic (black and white; Chinese-Americans and Japanese-Americans) case-control study of prostate cancer, no statistically significant increase in risk could be identified after vasectomy, including no effect of age at vasectomy or years since vasectomy.[144] Reviews of 6 cohort studies and 5 case-control studies concluded that there is no increased risk of cancer of the testis following vasectomy, and consideration of the studies examining the possible association between prostate cancer and vasectomy (6 cohort and 7 case-control studies) found the evidence to be equivocal and weak.[145, 146] A meta-analysis of the literature concluded that there is no increased risk of prostate cancer in men who have undergone vasectomy.[147] A case-control study from New Zealand found no increase in prostate cancer even after more than 25 years since the vasectomy.[148] Observational studies cannot totally avoid potential biases, and the disagreement regarding prostate cancer is consistent with either no effect or an effect too small to escape

confounding biases. It is worth noting that the countries with the highest vasectomy rates (China and India) do not have the highest rates of prostate cancer. Screening for prostate cancer should be no different in men who have had a vasectomy.

Animal studies had indicated that vasectomy accelerates the process of atherosclerosis. In the U.S. Physicians' Heath Study (a large prospective cohort study), no increase in the risk of subsequent cardiovascular disease could be detected following vasectomy.[130] Indeed, vasectomy has not been demonstrated to have any adverse consequences or harmful effects on men's health.[59, 60]

Vasectomy does not change the secretion of human immunodeficiency virus (HIV) into semen, and vasectomy should not change the risk of HIV transmission.[149]

Vasectomy reversal is associated with pregnancy rates as high as 70–80%.[150] The prospect for pregnancy diminishes with time elapsed from vasectomy, decreasing significantly to 30% after 10 years; the best results are achieved when reversal is performed within 3 years after vasectomy.[151] In most cases, sperm can be collected at the time of the reversal procedure and frozen for potential future intracytoplasmic sperm injection.

Medical Methods for the Male

A reversible method of contraception for men has been sought for years. Hormonal contraception for men is inherently a difficult physiologic problem because, unlike cyclic ovulation in women, spermatogenesis is continuous, dependent upon gonadotropins and high levels of intratesticular testosterone.[152] Investigational approaches to inhibit production of sperm include the administration of sex steroids, the use of GnRH analogs, and the administration of gossypol, a derivative of cottonseed oil.[153, 154]

The sex steroids reduce testosterone synthesis, which leads to loss of libido and development of female secondary sexual characteristics. Furthermore, despite the use of large doses, sperm counts are not adequately reduced in all subjects. Levonorgestrel, cyproterone acetate, and medroxyprogesterone acetate all have been studied combined with testosterone, given intramuscularly to provide the desired systemic androgen effects. GnRH analogs also decrease the endogenous synthesis of testosterone, and supplemental testosterone must be provided. The overall metabolic and health consequences of these approaches have not been assessed, and frequent injections are required.

Gossypol effectively decreases sperm counts to contraceptive levels, apparently by incapacitating the sperm producing cells. Experience in China revealed that a substantial number of men remain sterile after exposure to gossypol, and animal studies in the U.S. indicated that gossypol, or contaminants of the preparation, were toxic; work on gossypol was discontinued.[155] Analogs of gossypol may offer potential but are years away from development.

Induced Abortion

Contraception is more effective and convenient than ever, but even the most conscientious couples can experience contraceptive failure. In the past, failure of contraception

meant another, sometimes unwanted, birth or recourse to dangerous, secret abortion. The most ancient medical texts indicate that abortion has been practiced for thousands of years. Induced abortion did not become illegal until the 19th century, as a result of changes in the teachings of the Catholic Church (life begins at fertilization) and in the U.S., the efforts of the American Medical Association to have greater regulation of the practice of medicine.

In the 1950s, vacuum aspiration led to much safer abortion, and beginning in Asia, induced abortion was gradually legalized in the developed countries of the world. This trend reached the U.S. from Western Europe in the late 1960s when California, New York, and other states rewrote their abortion laws. The U.S. Supreme Court followed the lead of these states in 1973 in the "Roe versus Wade" decision that limited the circumstances under which "the right of privacy" could be restricted by local abortion laws. By 1980, legal abortion became the most common surgical procedure performed in the U.S. The average cost of a nonhospital abortion with local anesthesia in 2005 was $413.[156]

The number of abortions performed in the United States has been decreasing since a peak of 1.6 million was reached in 1981, declining to 1.33 million in 1993 and 1.18 million in 1997, with the greatest decrease among teenagers.[157–159] This is partly because the number of pregnancies in the U.S. has been decreasing and the proportion of reproductive-aged women younger than age 30 is also decreasing.[160] However, better use of effective contraception made a major contribution to the decline in the abortion rate. Accounting for underreporting, a more accurate estimate indicated about 1.36 million induced abortions in 1996, 1.31 million in 2000, and 1.21 million in 2005, the lowest number since 1976.[8, 156, 161, 162] In 2004 and 2005, 57% of induced abortions were obtained by women in their 20s, and 17% by women younger than 20. ***The number of births in the U.S., including teenage births, began to increase in 2005,[9, 45] and it is anticipated abortion numbers will parallel this recent change.***

Overall, a little over 3 million (49%) of American pregnancies each year are unintended, but the percentage is only 40% among white women in contrast to 54% among Hispanics and 69% among blacks.[5, 8] Each year, 42% of unintended pregnancies are terminated by induced abortions, and 60% of these abortions are obtained by women who have one or more children. The rate of unintended pregnancies and abortions is about 4 times higher among poor women.

Worldwide, about 22% of all pregnancies end in induced abortion.[163] The number of induced abortions has declined in developed countries to about 7 million annually. Most induced abortions occur in developing countries, about 35 million annually, where more than half are unsafe, illegal abortions. Notably, Western Europe with good contraceptive education and accessibility has an abortion rate that is almost half that of North America. It is also worth emphasizing that in countries where there are legal restrictions on abortion, the abortion rates are not lower compared with areas where abortion is legally permitted; however, these illegal abortions are associated with infection and hemorrhage, accounting for 13% of maternal deaths worldwide.[164] ***Abortion restrictions do not reduce the rate of abortions, but do make the procedure less safe.***

American teenagers are especially dependent on abortion compared with their European counterparts who are better educated about sex and use contraception more often and more effectively. In 2005, 17% of women who obtained legal abortions were adolescents.[5, 156] In addition, from ages 20–34, American women have the highest proportion of pregnancies aborted compared with other countries, indicating a high rate of unintended pregnancy occurring beyond the teenage years. The lack of perfect contraception and imperfect use of contraception will keep abortion with us.

Safety of Induced Abortions

Public health authorities have demonstrated that the legalization of abortion reduced maternal morbidity and mortality more than any single development since the advent of antibiotics to treat puerperal infections and blood banking to treat hemorrhage. The number of American women reported as dying from abortion declined from nearly 300 deaths in 1961, to only 6 in 1985, 10 in 1992, and 4 in 1999, or about *0.6 deaths for every 100,000 legal abortions*.[165, 166] For comparison, in 1990, the maternal death rate for childbirth in the U.S. was 10 per 100,000 births and for ectopic pregnancy it was approximately 50 per 100,000 cases,[167–169] and, in 1992, 17 deaths were associated with spontaneous miscarriage.[165]

The most important determinants of abortion mortality are duration of gestation and type of anesthesia: later abortions and general anesthesia are more hazardous.[170–172] As with mortality, morbidity rates vary primarily with duration of pregnancy, but other factors are important as well, including type of operation, age of patient, type of anesthesia, operator's skill, and method of cervical dilatation. More experienced surgeons and younger, healthier women are less likely to have complications.

Major and minor complications in a series of 170,000 first-trimester abortion patients were as follows[173]:

Major Complications (Hospitalization Required)

Retained tissue	— 27.7 per 100,000 induced abortions
Sepsis	— 21.2
Uterine perforation	— 9.4
Hemorrhage	— 7.1
Inability to complete	— 3.5
Intrauterine plus tubal pregnancy	— 2.4

Minor Complications (Managed in Clinic or Office)

Mild infection	— 462.0 per 100,000 induced abortions
Reaspiration same day	— 180.8
Reaspiration later	— 167.8
Cervical stenosis	— 16.5
Cervical tear	— 10.6
Underestimated gestation	— 6.5
Convulsive seizure	— 4.0

The possibility that abortion can result in longer-term complications has been examined in more than 150 studies.[174] There is no evidence for any adverse consequences of vacuum aspiration abortion for subsequent fertility,[175, 176] future pregnancies,[177, 178] or increased risk for ectopic pregnancy.[179, 180] Second-trimester abortions do not increase the rate of preterm delivery or midtrimester fetal losses.[181] Multiple induced abortions do not increase the risk of a subsequent ectopic pregnancy but may increase the rate of preterm delivery in subsequent pregnancies.[182, 183] The long-term effects of second-trimester abortion may depend on the method used.[184] A French study disagrees with these conclusions, finding a slightly increased risk of ectopic pregnancy in women with a prior induced abortion and no previous ectopic pregnancy, and Chinese and Danish studies found a small increase in the risk of spontaneous miscarriage following surgically induced abortions.[185–187]

The psychological sequelae of elective abortion have been studied and debated. The evidence indicates that depression is less frequent among women postabortion compared

with postpartum; that women denied abortion experience resentment for years; and that the children born after abortion is denied have social, occupational, and interpersonal difficulties lasting into early adulthood.[188] Extensive reviews of the evidence, including one by the American Psychological Association, concluded that first-trimester induced abortions do not increase the risk for mental health problems.[189, 190] Adverse psychological responses following induced abortion are influenced by a need for secrecy, perceptions of associated stigma, inadequate social support, and troubling personality characteristics such as low self esteem or improper coping strategies, but the most important predictor of a negative mental reaction is a prior history of mental problems.[190]

Conflicting results were reported in over 20 studies examining the risk of breast cancer associated with the number of abortions (especially induced abortions) experienced by individual patients.[191, 192] Concern for an adverse effect has been based on the theoretical suggestion that a full-term pregnancy protects against breast cancer by invoking complete differentiation of breast cells, but abortion increases the risk by allowing breast cell proliferation in the first trimester of pregnancy but not allowing the full differentiation that occurs in later pregnancy. In these studies there was a major problem of recall bias; women who develop breast cancer are more likely to truthfully reveal their history of induced abortion than healthy women. ***In studies that avoided recall bias (e.g., by deriving data from national registries instead of personal interviews), the risk of breast cancer was identical in women with and without induced abortions.[193, 194] More careful case-control and cohort studies, including the Nurses' Health Study, also failed to link a risk of breast cancer with induced or spontaneous abortions.[195–197]***

Safe abortion is still unavailable to many women in parts of Asia, Africa, and Latin America.[198] Therefore, many women resort to clandestine, unsafe abortions, accounting for about 13% of the world's maternal mortality. These deaths are preventable. Family planning services that provide effective contraceptive choices as well as access to safe abortion early in pregnancy are essential in order for societies to achieve desired fertility rates and a healthy female population.

Pretreatment Care of Abortion Patients

Approximately 90% of the 1.2 million induced abortions performed in the U.S. yearly are performed during the first trimester of pregnancy, and a growing percentage is accounted for by medical abortion.[8] During the first trimester, abortion morbidity and mortality rates are less than one-tenth those of abortions performed in the later midtrimester.[167] The vast majority of these procedures occur in free-standing abortion clinics, although in recent years, physicians have performed larger numbers in their offices where women are less subject to the harassment that has plagued clinics.[5, 157] The safety of outpatient abortion surgery under local anesthesia is well established.

The care of the patient who has decided to terminate a pregnancy begins with the diagnosis of intrauterine pregnancy and an accurate estimate of gestational age. Failure to accomplish this is the most common source of abortion complications and subsequent litigation. Tests for pregnancy, including vaginal ultrasound, should be used when accuracy is difficult.

Nearly all women who want to terminate a pregnancy in the first trimester are good candidates for an outpatient surgical procedure under local anesthesia. Possible exceptions include patients with severe cardiorespiratory disease, severe anemias or coagulopathies,

mental disorders severe enough to preclude cooperation, and excessive concern about operative pain that is not alleviated by reassurance.

Surgical abortions should not be undertaken for women who have known uterine anomalies or leiomyomas or who have previously had difficult first-trimester abortion procedures, unless ultrasonography is immediately available and the surgeon is experienced in its intraoperative use. Previous cesarean section or other pelvic surgery is not a contraindication to outpatient first-trimester surgical abortion.

Counseling Abortion Patients

Counseling has played a critical role in the development of efficient and acceptable abortion services.[199] Whether abortion is accomplished in a clinic, a physician's office, or a surgical center, the functions of a counselor must be fulfilled to ensure quality patient care. These include helping with decision-making; providing information about the procedure; obtaining informed consent; providing emotional support for the patient and her family before, during, and after the procedure; and providing information about contraception.[200] Referral opportunities should be provided for prenatal care or adoption for women who choose to carry an unplanned pregnancy to term. These responsibilities can be performed by a physician, nurse, psychologist, social worker, or a trained lay person. An informed consent document should unequivocally state the possibilities of common adverse outcomes, such as incomplete abortion, infection, uterine perforation, the need for laparotomy, ectopic pregnancy, and failed abortion. The counselor should document that all preoperative responsibilities have been discharged.

Nearly half of induced abortions in the U.S. are repeat abortions.[201] Repeat elective abortions are more common in older women, in women using a method of contraception, and in women reporting alcohol or drug abuse. Counseling after elective abortions should emphasize effective, long-term methods of contraception such as the IUD, implants, or sterilization.

Methods for First-Trimester Abortions

The most widely used technique for first-trimester abortions is vacuum curettage.[162, 165, 202] The procedure is performed using local anesthesia (a paracervical block). Cervical dilation is accomplished with tapered Pratt dilators. Some surgeons recommend the preoperative insertion of cervical tents. These are osmotic dilators of dried seaweed or synthetic hydrophilic substances that are left in place from a few hours (synthetic) to overnight (seaweed).[203] Mifepristone (RU 486), the progesterone antagonist, produces preoperative cervical dilation equally effectively, and the ease of its single oral dose makes it a more attractive choice, but oral mifepristone requires a long pretreatment (24–36 h).[204] Buccal misoprostol (400 μg) dilates the cervix as effectively as laminaria when given 4 hours prior to the procedure, and it is relatively inexpensive.[205] After the procedure, the patient is observed for 1–2 h before returning home.

Aspiration surgical abortion is safe and effective, but it is not available everywhere, and some women find it difficult to undergo a surgical procedure or to go to a clinic where they may be subject to loss of privacy or harassment. Nonsurgical methods make abortion available to more women and improve the circumstances under which pregnancies are terminated. The progesterone antagonist mifepristone (RU 486) and the antimetabolite

methotrexate have both been demonstrated to effectively induce abortion early in pregnancy when combined with a prostaglandin. Both medical and surgical first-trimester abortions do not increase risks in future pregnancies, including ectopic pregnancy, spontaneous miscarriage, preterm birth, or low birthweight.[206] Although both medial and surgical abortion methods have a very low rate of serious complications, medical abortion is associated with more discomfort, bleeding, and incomplete abortion.[207, 208]

France and China were the first countries to approve the marketing of the medical abortifacient mifepristone, a synthetic relative of the progestational agents in oral contraceptives. Mifepristone acts primarily, but not totally, as an antiprogestational agent. Both progesterone and mifepristone form hormone-responsive element-receptor complexes that are similar, but the mifepristone complex has a slightly different conformational change (in the hormone-binding domain) that prevents full gene activation. The agonistic activity of this progestin antagonist is due to its ability to activate certain, but not all, of the transcription activation functions on the progesterone receptor. The dimethyl (dimethylaminophenyl) side chain at carbon 11 is the principal factor in its antiprogesterone action. There are three major characteristics of its action that are important: a long half-life, high affinity for the progesterone receptor, and active metabolites.

A single oral dose of mifepristone has been followed a day later by the administration of a prostaglandin analogue. Several analogues have been used, but the most widely available and best tolerated was misoprostol, 800 μg administered vaginally.[209] The combination allowed a reduction in dose of both agents. When administered in the first 8 weeks of pregnancy, this medical termination carried success and complication rates similar to that achieved with vacuum curettage.[210, 211]

Misoprostol is a stable, orally active synthetic analogue of prostaglandin E_1, available commercially for the treatment of peptic ulcer. Combined with mifepristone, it provides an effective, simple, inexpensive method that can be administered at home.[211–215] In the large U.S. trial of 600 mg mifepristone followed by 400 μg misoprostol orally, there was a 1% failure rate under 7 weeks of pregnancy and 9% from 8 weeks to 9 weeks.[216] Termination occurred in 50% of the women within 4 h after misoprostol administration and 75% within 24 h.

Based on worldwide experience, the regimen with the least side effects and cost, but equally good efficacy, was a combination of a lower dose of oral mifepristone (200 mg), followed 48 h later by the vaginal administration of 800 μg misoprostol at home.[211, 214, 215, 217–220] The vaginal administration of misoprostol allowed medical abortion up to 63 days after the last menstrual period.[211] Repeated doses of misoprostol have been recommended for the management of delayed expulsion. ***However, because of a problem with infection, the protocol for medical abortion changed.***

In 2005, the Centers for Disease Control and Prevention reported four cases of fatal toxic shock syndrome in California associated with *Clostridium sordellii* that occurred within one week after medical abortions (induced with 200 mg of oral mifepristone and 800 μg of vaginal misoprostol).[221] This prompted many clinicians and the Planned Parenthood Federation to switch to buccal administration of misoprostol, recognizing that *Clostridium* species are present in a small percentage of vaginal cultures.

Beginning in 2006, Planned Parenthood Federation clinics changed to a protocol that administered 800 μg misoprostol buccally 24–48 h after 200 mg mifepristone and either to screen with vaginal cultures, especially for STIs, and treat appropriately or to to provide prophylactically doxycycline 100 mg b.i.d. for 7 days beginning on the day of mifepristone. In 2008, the protocol was again changed, requiring prophylactic doxycycline treatment. The routine use of antibiotics was associated with a greater reduction in serious infection compared with the method of screen and treat. Following these changes the

number of serious infections decreased from 93 to 6 cases in 100,000 medical abortions, a decline of 93%.[222] ***The buccal misoprostol protocol provides comparable efficacy up to 63 days gestation, an expected outcome because the pharmacokinetics of misoprostol are essentially the same with either buccal or vaginal administration.[223, 224] The buccal route of administration of misoprostol is now the recommended method.***

Fatalities with *Clostridium* species have been associated with spontaneous miscarriages, the postpartum period, after trauma or surgical procedures, and even when buccal misoprostol was used.[225] The best prevention of fatal toxic shock with this rare infection is awareness of the possibility and early recognition. Abdominal cramping as a presenting complaint makes the diagnosis difficult because this is a common symptom following medical abortion. Unique characteristics include: the absence of fever, markedly elevated white counts, fluid effusions sufficient to produce hemoconcentration, and eventually tachycardia and hypotension. Specific antibiotics with demonstrated efficacy against *Clostridium sordelli* have not been identified, although doxycycline inhibits *Clostridium* growth in vitro. Early recognition of this rare infection would mandate consideration of aggressive surgery with hysterectomy, a lesson learned from the experience with septic abortions in the years before legalized abortion.

It is likely that abortion with mifepristone is the result of multiple actions. Although mifepristone does not induce labor, it does open and soften the cervix (this may be an action secondary to endogenous prostaglandins). Its major action is its blockade of progesterone receptors in the endometrium. This leads to a disruption of the embryo and the production of prostaglandins. The disruption of the embryo and perhaps a direct action on the trophoblast lead to a decrease in human chorionic gonadotropin (hCG) and a withdrawal of support from the corpus luteum. The success rate is dependent on the length of pregnancy—the more dependent the pregnancy is on progesterone from the corpus luteum, the more likely that the progesterone antagonist, mifepristone, will result in abortion. The combined mifepristone-prostaglandin analogue method is usually restricted to pregnancies that are not beyond 9 weeks' gestation. However, a regimen using a higher dose of misoprostol (administered vaginally) achieved a 95% complete abortion rate in women at 9–13 weeks' gestation.[226] Other progesterone antagonists have been developed, but only mifepristone has undergone extensive abortion trials.

Mifepristone is most noted for its abortifacient activity and the political controversy surrounding it. However, the combination of its agonistic and antagonistic actions can be exploited for many uses, including contraception, therapy of endometriosis, induction of labor, treatment of Cushing's syndrome, and, potentially, treatment of various cancers. Doses of 2–5 mg/day inhibit ovulation and produce amenorrhea in over 90% of cycles, and in a pilot study of 50 women, there were no pregnancies.[227] A clinical trial indicated that a daily dose of 5 mg mifepristone would be an effective oral contraceptive.[228]

Methotrexate was tested as an abortifacient in the same dose used to treat ectopic pregnancy, 50 mg intramuscularly per square meter of body surface area.[229] Later, a single 75-mg intramuscular dose was demonstrated to be as effective.[230] Methotrexate has also been administered orally in doses of 25 or 50 mg.[231] As with mifepristone, a prostaglandin is added to promote expulsion of the uterine contents. The first trials demonstrated that if the prostaglandin (800 μg misoprostol vaginally) was given a week after the injection of methotrexate, this method could be almost as, *but not as* effective as mifepristone.[232] Efficacy diminishes with advancing gestation beyond 7 weeks since the last menstrual period.[233–235] Because methotrexate takes longer to act than mifepristone, the prostaglandin is used a week after the initial treatment, and is repeated a day later if expulsion has not occurred. Methotrexate is easily available and inexpensive. It has been used in low doses to treat psoriasis and rheumatoid arthritis, as well as ectopic pregnancy, without adverse effects. It is, however, a known teratogen that can be deadly in high doses, and its use as an abortifacient results in prolonged bleeding and a prolonged time to abortion (up to a month

in some cases). Mifepristone is preferred by clinicians who have experience with both methods, but there are no direct comparison studies of methotrexate and mifepristone.

Another approach uses the combination of tamoxifen and misoprostol. The administration of tamoxifen (20 mg daily for 4 days) followed by misoprostol (800 μg vaginally, with a second dose if necessary 24 h later) was associated with a 92% rate of complete abortion in 100 women with pregnancies less than 9 weeks gestational age.[236] Similar good results were obtained in a comparison of tamoxifen (20 mg b.i.d. for 4 days) with methotrexate.[237]

The use of prostaglandin alone has also been pursued.[238] Relatively high success rates have been reported with multiple dosing,[239] but the most effective regimen and the best method of administration remain to be determined.[240] The administration of 800 μg misoprostol daily for 3 days has been reported to be very effective late in the first trimester (10–12 weeks).[241] In very early gestation, a single vaginal dose of 800 μg misoprostol or multiple doses within 24 h is as effective as the usual combination of mifepristone and oral misoprostol.[239, 242]

One word of caution regarding misoprostol, the synthetic prostaglandin E_1 analogue: it is now recognized that infants born to pregnant women exposed to misoprostol have an increased risk of abnormal vascular development resulting in Möbius's syndrome (congenital facial paralysis with or without limb defects) and defects such as equinovarus and arthogryposis.[243–245] Although the risk is low, this possibility must be considered in decision-making when the various methods for first-trimester abortion are considered.

Careful prospective follow-up assessments can detect no health differences in women who have medical abortions compared with women who have abortions by vacuum aspiration.[246] Although women having medical abortions experience more bleeding and cramping, with appropriate counseling and support, women are equally satisfied with surgical and medical abortions.[247]

Complications of Abortions

Postoperative complications of elective abortions are classified as either immediate or delayed. Uterine perforation and uterine atony are examples of immediate complications. Delayed complications can occur several hours to several weeks after the operation. These usually present according to the major complaint: bleeding, pain, and continuing symptoms of pregnancy.

Bleeding

By far the most common cause of unusually heavy postabortal bleeding is retained products of conception. Rates in large series vary from 0.2 to 0.6%.[173] Patients with retained products of conception occasionally present several weeks after an abortion, but most report excessive bleeding within 1 week. Severe pain or pelvic tenderness suggests that infection is also present. Treatment is prompt aspiration of the uterus with the largest cannula that will pass the cervix.

Infection

Infection is sometimes marked by uterine bleeding; although without retained products of conception, the volume of blood loss is usually modest. Fever and uterine tenderness are the most common signs of postabortal endometritis, occurring in about 0.5% of cases.[173] Some studies indicate that prophylactic antibiotics reduce the risk of surgical postabortal infection.[248, 249]

Most clinicians agree that women at risk of pelvic infection benefit from the use of prophylactic antibiotics prior to induced abortion; others state that women who have not had a previous delivery should receive prophylaxis, and still others believe that all abortion patients would benefit from prophylactic antibiotics.[250, 251] A meta-analysis of antibiotics at the time of surgically-induced abortion unequivocally concluded that prophylactic antibiotics should be routinely used without exceptions, and, as noted previously, prophylactic antibiotics are now advocated for medical abortions.[222, 252] Because both gonorrhea and chlamydia, as well as other organisms, can cause postabortion infections, a tetracycline seems the best drug for prophylaxis. Doxycycline, 100 mg an hour before the surgical abortion and 200 mg 30 min afterward, is the most convenient and comprehensive regimen.[253] Tetracycline, 500 mg once before and once after the operation, is also acceptable. Metronidazole, 400 mg an hour before and 4–8 h afterward, has been tested and is effective treatment for patients with bacterial vaginosis detected at the time of abortion.[254, 255] The treatment of choice for medical abortions is doxycycline 100 mg b.i.d. for 7 days beginning on the day of mifepristone.

Patients who present with uterine tenderness, fever, and bleeding require uterine aspiration and antibiotic treatment. Patients who have fevers higher than 38°C (101°F), signs of peritoneal inflammation, and uterine tenderness require hospitalization and intravenous antibiotics active against anaerobes, gonorrhea, and chlamydia. Outpatient treatment with doxycycline, 100 mg b.i.d. for 14 days, should be reserved for patients whose signs and symptoms are confined to the uterus.

Dysfunctional Uterine Bleeding Following Abortion

Women may present with uterine bleeding but without signs or symptoms of retained products of conception or infection. When these two diagnoses have been ruled out by absence of fever, a closed cervix, and a nontender uterus, the bleeding itself can be treated hormonally. Curettage is rarely necessary unless bleeding is excessive.

Ectopic Pregnancy

Failure to diagnose ectopic pregnancy at the time of induced abortion can cause a patient to return with complaints of persistent bleeding with or without pelvic pain. Careful examination of the uterine aspirate for villi at the time of abortion should make a missed ectopic pregnancy an unusual cause of delayed bleeding. If, however, a patient presents with this possibility, quantitative measurement of chorionic gonadotropin and vaginal ultrasonography should be used for accurate diagnosis and management.

Cervical Stenosis

Patients who experience amenorrhea or hypomenorrhea and cyclic uterine pain after first-trimester abortion may have stenosis of the internal os. This condition occurs in about 0.02% of cases and is more common among women whose abortions are performed in the early first trimester with a minimum of cervical dilatation and a small diameter, flexible plastic cannula. Possibly, the tip of this type of cannula abrades the internal os, and the minimal dilatation allows the abraded areas to heal in contact. The condition is easily treated with cervical dilatation with Pratt dilators under paracervical block.

Other Late Complications

Amenorrhea, usually without pain, can be caused by Asherman's syndrome, destruction and scarification of the endometrium. This condition is very rare and usually follows endometrial infection. This problem is best diagnosed and treated at hysteroscopy.

Sensitization of Rh-negative women should be prevented. Approximately 4% of these women become sensitized following an induced abortion (the later the abortion the higher the proportion). Subsequent hemolytic disease of the newborn can be prevented by administering 50 µg (250 IU) Rh immunoglobulin to all Rh-negative, Du-negative women undergoing early abortion. The standard calculated dose is administered for second-trimester abortion.

Abortion in the Second Trimester

The great majority (90%) of induced abortions occur in the first trimester. Of the 10% performed in the second trimester, only 1.5% or less of induced abortions occur after 20 weeks gestation, most at 20–21 weeks.[8] The most common reasons for delays in seeking abortion are a failure to recognize pregnancy, a poor understanding of payment mechanisms, and awkward or confusing referral processes.[256]

Second-trimester abortions can be accomplished surgically or medically. After about 14 weeks of pregnancy, the surgical procedure is termed dilation and evacuation (D & E). Several approaches have been used for the medical termination of pregnancy. These include the vaginal, intramuscular, or intra-amniotic administration of prostaglandins and the intra-amniotic injection of hypertonic saline or urea, but these medications have been replaced by misoprostol and mifepristone. The D & E procedure has been considered safer and less expensive than the medical methods and better tolerated (and thus preferred) by patients.[257–261]

With surgical abortion in the second trimester, the training, experience, and skills of the surgeon are the primary factors that limit the gestational age at which abortion can be safely performed. Advanced gestational age by itself incurs increased risks for all types of complications. These are multiplied when the duration of pregnancy is discovered, after beginning uterine evacuation, to be beyond the experience and skill of the surgeon or capacity of the equipment. Uterine perforation, infection, bleeding, amniotic fluid embolism, and anesthetic reactions are increased as gestational age increases.[257]

When errors in estimating gestational age require the surgeon to use unfamiliar instruments or techniques that are not frequently practiced, the increased duration of the procedure can cause problems. Efforts to sedate or relieve pain by administering additional drugs increase the risk of toxic reactions or overdosage. If a change from local to general anesthesia is undertaken, the patient is at much greater risk of anesthetic complications. Finally, if complications caused by advanced gestational age necessitate transfer of the patient to physicians who are not familiar with uterine evacuation techniques, the patient may undergo unnecessarily extensive surgery, such as hysterectomy, with all the risks inherent in emergency procedures.

Preoperative cervical dilation with osmotic dilators makes first-trimester abortion safer and easier and is essential for second-trimester abortion. Local anesthesia instead of general

anesthesia also makes abortion safer.[262, 263] Some patients are not good candidates for surgical procedures of any kind under local anesthesia, and others may have special reasons to prefer that an abortion be performed under general anesthesia. Patient requests should be seriously considered, but the clinician also has a responsibility to inform the patient of the risks and benefits of local versus general anesthesia.

The medical method of induced abortion in the first trimester is also effective for second-trimester abortions. A combination of the progesterone antagonist, mifepristone, (a single oral 200-mg dose of mifepristone administered 36 h before prostaglandin treatment) and 800-mg misoprostol administered orally or vaginally is highly effective.[264, 265] The treatment usually includes amniotomy and oxytocin infusion.

Post Abortion Contraception

Despite improvements in contraceptive technology, unintended pregnancies lead to millions of spontaneous and voluntary abortions, unwanted births, and ectopic pregnancies in the U. S. and worldwide.[5] Social and economic disparities are associated with high rates of unintended pregnancies and induced abortions, especially among adolescents.[266] Contraception plays an important role in rectifying disparities and protecting women's health; at no time is access to effective methods more important than immediately after induced abortion. Adolescents are especially receptive to contraception counseling that is provided immediately after an induced abortion.[267]

Barriers to effective contraceptive use vary around the world. In the U.S., where nearly half of induced abortions are repeat procedures, failure to receive contraception and alcohol and drug use are strongly associated with an increased risk of repeat abortion.[201] Comprehensive counseling about contraceptive options is not always incorporated into abortion care, and this failure explains many repeat abortions.[268] Studies indicate that family planning counseling in an abortion clinic is well accepted and promotes modern contraceptive use.[268, 269]

Women who terminate pregnancies are at high risk of having another unintended pregnancy.[270] Clinicians should seize the opportunity to assist women who are highly motivated to implement contraception at this point in time. An ideal time to provide contraception is promptly after treatment for an unsafe abortion, especially in women at high risk of subsequent unintended pregnancies because of lack of access to contraception.[271] Post abortion contraceptive counseling and initiation of contraceptive methods on the day of abortion are associated with increased continuation rates.[268]

The safety of immediate placement of an intrauterine device (IUD), injection of depot medroxyprogesterone acetate, and initiation of oral contraceptives following first- and second-trimester abortions is well-documented. Data regarding other methods (progestin-only pills, Norplant, Implanon, transdermal, and vaginal methods) can be extrapolated from studies of postpartum women.[272] Studies regarding Implanon insertion following abortion are underway.

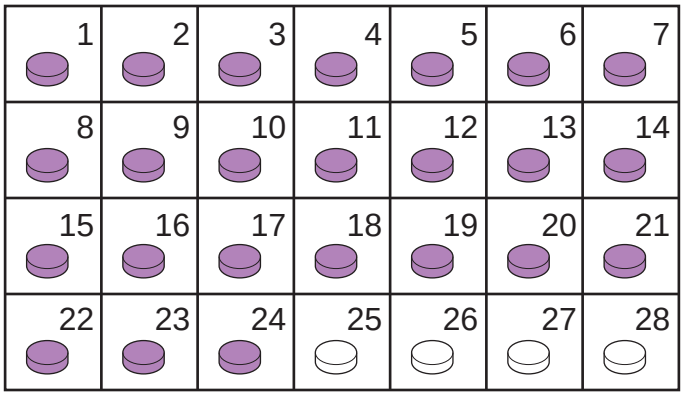
Immediate post first-trimester abortion IUD placement is known to be safe and effective, but information about the second trimester is more limited. A prospective cohort study of 256 women demonstrated a low rate of expulsion (3%) and discontinuation (8.3%), with a high rate of IUD acceptance, following a second-trimester surgical abortion.[273] The rate of infection was not significantly increased. An older, 1983, study by the World Health Organization reported a much higher expulsion rate of 20%, but this study included multiple

sites with multiple IUD types, and did not utilize ultrasound guidance in placement.[274] A Cochrane meta-analysis of post abortion IUD placement (9 randomized trials including first- and second-trimester abortions), concluded that expulsion rates were higher with all IUD types than in interval placement.[275] More recently, a randomized trial of IUD insertion after first-trimester surgical abortion indicated no major difference in expulsion rates comparing immediate insertion with delayed insertion, and no increase in the risk of infection.[276, 277] Of note is the study indicating that the risk of repeat abortion is decreased among IUD users compared with other methods.[278]

In summary, IUD insertion and the immediate start of depot medroxyprogesterone acetate, progestin-only pills, and barrier methods is indicated following either a first- or second-trimester abortion. After the elective termination of a pregnancy of less than 12 weeks, estrogen-progestin contraception can be started immediately. After a pregnancy of 12 or more weeks, the third postpartum week rule should be followed to avoid the postpartum risk of venous thromboembolism. Information about the rates of subsequent unintended pregnancies for newer methods of contraception following second trimester abortion is limited; however, initiation of all hormonal methods of contraception immediately following a second-trimester abortion is endorsed by the World Health Organization and reduces the risk of repeat abortion. The vaginal ring and the contraceptive patch are assumed to have the same efficacy and safety as oral contraceptives. The abortion setting is an ideal time to provide a contraceptive method, like IUDs and implants, which requires a skilled clinician.

All references are available online at: http://www.clinicalgynendoandinfertility.com

22 Oral Contraception



Contraception is commonly viewed as a modern event, a recent development in human history. On the contrary, efforts to limit reproduction predate our ability to write about it. It is only contraception with synthetic sex steroids that is recent.

The History of Oral Contraception

It wasn't until the early 1900s that inhibition of ovulation was observed to be linked to pregnancy and the corpus luteum. Beginning in 1920, Ludwig Haberlandt, professor of physiology at the University of Innsbruck, Austria, demonstrated that ovarian extracts given orally could prevent fertility in mice. Haberlandt is acknowledged as the first to perform experiments with the aim of producing a method of hormonal contraception; he called it "hormonal sterilization."[1] In the 1920s, a Viennese gynecologist, Otfried Otto Fellner, conducting experiments in his spare time, and administering ovarian and placental extracts to a variety of animals, also reported hormonal sterilization.[2] By 1931, Haberlandt proposed the administration of hormones for birth control. An extract named Infecundin was produced in collaboration with the Hungarian pharmaceutical company Gideon Richter, but Haberlandt's early death of a heart attack in 1932, at age forty-seven, brought an end to this effort. Fellner disappeared after the annexation of Austria to Hitler's Germany.

The concept was annunciated by Haberlandt, but steroid chemistry wasn't ready. The extraction and isolation of a few milligrams of the sex steroids required starting points measured in gallons of urine or thousands of pounds of organs. Edward Doisy processed 80,000 sow ovaries to produce 12 mg of estradiol.

Russell Marker

The supply problem was solved by a cantankerous iconoclast, Russell E. Marker, who completed his thesis, but not the course work, for his Ph.D. The following story is derived from Marker's own words, in an autobiographical article and from a two-hour interview for the oral history archives of the Chemical Heritage Foundation in Philadelphia.[3, 4]

Marker, born in 1902 in a one-room log cabin on a farm near Hagerstown, Maryland, received his bachelor's degree in organic chemistry and his master's degree in colloidal chemistry from the University of Maryland. Although he had completed his work for a Ph.D., his supervisor, Morris S. Kharasch, announced that Marker still lacked some required chemistry courses. Considering the courses a waste of time, Marker said, "The hell with it," and abruptly left.

After leaving the University of Maryland, Marker worked first in the laboratory of the Naval Powder Factory, then with the Ethyl Gasoline Corporation, where in 1926 he developed the system of octane rating of gasoline. Frank Whitmore, dean of Pennsylvania State College, now Pennsylvania State University, visited Marker at Ethyl. Impressed with his work, Whitmore said, "If you're ever looking for a job, let me know."

From 1927 to 1934, Marker worked at the Rockefeller Institute, publishing a total of 32 papers on configuration and optical rotation as a method of identifying compounds. He became interested in steroid chemistry, but he was told to continue with his work in optical technology. Instead, Marker called Dean Whitmore at Penn State.

In September 1935, Marker moved to Penn State at a reduced salary, from $4,400 per year at Rockefeller to $1,800, but with the freedom to pursue any field of research. His work was supported mainly by research grants from the Parke-Davis pharmaceutical company. At that time, it required the ovaries from 2,500 pregnant pigs to produce 1 mg of progesterone. Marker decided to pursue the goal of an abundant and inexpensive supply of progesterone, and for several years he concentrated on urine from pregnant animals. Then in 1939, Marker devised the method, called the Marker degradation, to convert a sapogenin molecule into a progestin.

Marker was convinced that the solution to the problem of obtaining large quantities of steroid hormones was to find plants in the family that includes the lily, the agave, and the yam that contained sufficient amounts of diosgenin, a plant steroid, a sapogenin, that could be used as a starting point for steroid hormone production. He discovered that a species of *Trillium*, known locally as Beth's root, was collected in North Carolina for the preparation of Lydia Pinkham's Compound, popular at the time to relieve menstrual discomfort. A principal ingredient in Beth's root was diosgenin, but the rhizome was too small to provide sufficient amounts for commercial production.

Marker's search for an appropriate plant took him to California, Arizona, and Texas. Spending his summer vacations in the Southwest and Mexico collecting sapogenin-containing plants, Marker's laboratory analyzed more than a hundred thousand pounds of over four hundred different species of plants. Marker discovered that the roots of the *Dioscorea* plant (a wild yam) were the richest source of sapogenins.

On a visit to Texas A & M University, Marker found a picture of a large *Dioscorea* (*Dioscorea mexicana)* in a book that he just happened to pick up and browse through while spending the night at the home of a retired botanist who was helping him collect diosgenin-containing plants. After returning to Pennsylvania, he traveled by train for 3 days to search for this *Dioscorea* in Mexico.

Marker first went to Mexico City in November 1941, but his effort was blocked by the lack of a plant-collecting permit from the Mexican government. He returned in January 1942, and the American Embassy arranged for a Mexican botanist who had a collecting permit to accompany Marker to Veracruz. Marker rented a truck with a driver, and when the botanist arrived at Marker's hotel, he was accompanied by his girlfriend and her mother, who served as the girl's chaperone. Marker was forced to take the entire group. They covered 80 miles the first day, staying overnight in Puebla. The next day, the drive to Tehuacan was a shorter trip, but the botanist insisted on a 2-day stay devoted to his own collection of specimens. Then next morning, the botanist refused to go any further, claiming that the natives had discovered Marker was American and wanted nothing to do with him. They turned around, managed to overcome a breakdown of the truck near Puebla, and made it back to Mexico City 5 days after starting, with nothing to show for the trip.

The next day, a Monday morning, Marker reported to the American Embassy and was advised to leave Mexico. It was just after Pearl Harbor and Mexico was being courted by Germany. The Embassy was concerned for the safety of Americans traveling in Mexico. Instead of returning home, Marker took an overnight bus to Puebla, arriving after midnight, and boarded a second bus that already held pigs and chickens in addition to a few passengers. He arrived in Orizaba the next morning, and fortunately there was a small hotel next to the bus terminal. Marker remembered that the botany book in which he first read a description of *D. mexicana* indicated that the plant, a wild yam vine that grows up trees in the mountains of southern Mexico, could be found along a stream that crossed the road between Orizaba and Cordoba. He climbed aboard the local bus to Cordoba, which he stopped and disembarked when the bus drove through a large stream crossing the road about ten miles after leaving Orizaba. He found a small country store next to the road, owned by an Indian named Alberto Moreno.

Moreno did not speak English; Marker did not speak Spanish. But somehow, Marker conveyed his desire to obtain the *Dioscorea* that was known locally as "cabeza de negro," black tubers. Moreno in turn somehow made Marker understand that he should return the next morning. And there in the store, the next morning, were two plants, each in a bag that Moreno placed on the roof of the next bus back to Orizaba. Each tuber was nine to twelve inches long and consisted of white material like a turnip; it was used by local Mexicans as soap and as a poison to catch fish. When Marker got off the bus in Orizaba, both bags were missing. A policeman was there, but it became apparent he was there to collect a fee for the return of the bags. Marker gave him what he had, a ten-dollar bill, but that only retrieved one bag, which he managed to smuggle back to Pennsylvania.

Marker used only a portion of the plant to isolate diosgenin. In February 1942, he took the remainder to the Parke-Davis chemists in Detroit. Demonstrating his process for obtaining diosgenin, Marker convinced the director of research, Oliver Kamm, that he was on to something, a source for raw material that could provide for the commercial production of hormones. Unfortunately, they could not convince the president of Parke-Davis, nor could Marker convince anyone at several other companies.

Unable to obtain support from the pharmaceutical industry, Marker, drew on half of his life savings and returned to Mexico in October 1942. He arranged with Albert Moreno to collect the roots of the Mexican yam. Marker paid Mexican medical students to collect the yams. The students were arrested when farmers reported that their yams were being stolen, but not before Marker had enough to prepare a syrup.

Back in the U.S. with his syrup, Marker arranged to work in the New York laboratory of a friend, Norman Applezweig, an organic chemist involved in steroid research, in return for one-third of whatever progesterone his syrup could yield.[5] He isolated diosgenin and synthesized 3 kg of progesterone, the largest lot of progesterone ever produced. United States

pharmaceutical companies still refused to back Marker, and even his university refused, despite Marker's urging, to patent the process.

Before Marker left Mexico, he looked through the yellow pages in a Mexico City telephone directory and found something he recognized, a company called "Laboratorios Hormona," owned by a lawyer who was a Hungarian immigrant, Emeric Somlo, and a German immigrant who had both a medical degree and a Ph.D. in chemistry, Frederick A. Lehman.

> … when the phone rang. A distant voice asked in barely comprehensible Spanish if he {Frederick Lehman} spoke English.
> "Yes, of course."
> "I found your company's name in the telephone book, since I recognized two words, 'Laboratories' and 'Hormones.' I have something you may be interested in: a cheap source for progesterone."
> "Who are you?"
> "I am Marker, a steroid chemist."[6]

Visiting the company, Marker met Lehman, the minority owner of Laboratorios Hormona, who had the good sense to see where this was going. From his reading of the literature, he knew who Marker was; he knew the value of steroids; and he was a businessman. Lehman called his partner who was visiting New York and convinced him to return as soon as possible. The three men agreed to form a Mexican company for the production of hormones, and Marker returned to the U.S., leaving behind a list of equipment and chemicals to be ordered.

Marker returned to Mexico in spring 1943 to collect plants and to check on progress at Laboratorios Hormona. He just happened to mention to Lehmann that he had 2 kg of progesterone. As soon as Marker returned to Pennsylvania, he received a phone call from Somlo who said that if Marker still had those 2 kg of progesterone he sure would like to see it; could he meet him in New York? Over dinner at the Waldorf-Astoria, Somlo offered Marker 40% of their new company in exchange for the progesterone, with a share in future profits. Marker arranged for a friend to deliver the progesterone to Somlo in New York. Somlo had a small company in New York called Chemical Specialties, and the progesterone used in the first studies leading to oral contraception were obtained from this Syntex subsidiary.

In December 1943, Marker resigned from Pennsylvania State College and went to Mexico where he collected the roots of *D. mexicana*—ten tons worth! Marker chopped them up with a machete, and left the pieces to dry in the sun across from Moreno's store in a small structure for drying coffee. It took 2 months of work in an old pottery shed in Mexico City to prepare several pounds of progesterone, worth $160,000, with the help of several young women who had little education and spoke no English.

Somlo suggested calling their new company Synthesis, but Marker insisted on some link to Mexico, and the three partners formed Syntex (from *synthesis* and *Mexico*), incorporated in March 1944. Marker moved into a new four-room laboratory, and over the next year, produced over 30 kg of progesterone and 10 kg of dehydroepiandrosterone. The price of progesterone fell from $200 to $50 a g.

During this time, Marker received expenses, but he was not given his share of the profits or the 40% share of stock due to him. In March 1945, Somlo claimed there were no profits, but then admitted that the profits had been paid to the two partners in Mexico as salaries. Failing to reach a settlement, Marker left Syntex in May 1945, took some of his young female workers with him, and started a new company in Texcoco, called Botanica-Mex. He changed to *Dioscorea barbasco*, which gave a greater yield of diosgenin, and the price of progesterone dropped to $10 a g, and later to $5.

> After I broke up with Lehmann and Somlo, I chose a place east of Mexico City (Texcoco), where labor and water were plentiful. I there repeated my simple procedure of converting diosgenin into progesterone. My workers were happy but one day they came to me and said, "We all live on this dry-lake bed, and we come from very far away. If you want us to go on working for you, we need bicycles." "Sure," said Marker, "I'll buy them for you, and you will pay them back from your salary." The workers, happy with this offer, and the image of a white man with promise, celebrated drunkenly one evening. Late at night they went to a nearby quarry where a great effigy of the Aztec rain god was still attached by its back to the bedrock (It wasn't moved to the museum until 1964). They then began chiseling my name over Tláloc's right eyebrow, but were interrupted by angry villagers and had to run away after having carved only the first two letters.[3]

The volcanic stone monolith of Tláloc the rain god was carved in a horizontal position sometime in the period of 400 B.C. to 200 A.D. On April 16, 1964, the unfinished statue was detached and transported on a day's journey to Mexico City, and placed in a vertical position at the road entrance to the Museo Nacional de Antropologia, an imposing 168 tons, 23 feet high. The initials "MA" can be easily discerned at the right edge of the headdress; Marker's workers obviously intended to place his full name across the entire width. The evening arrival of the rain god was greeted by a crowd of 25,000 people. Despite the fact that it was the dry season, a record rainfall fell on the day the statue arrived![7]

Marker's new company was allegedly harassed, legally and physically, by Syntex, and in 1946 it was sold to Gideon Richter, which moved it to Mexico City and renamed it Hormosynth. Eventually it came under the ownership of Organon of Holland, which still uses it under the name of Quimica Esteroides. By the 1960s, several pharmaceutical companies were benefiting from the root-gathering operations in Mexico, closely regulated by the Mexican government that imposed annual quotas, about 43 thousand tons, to balance harvesting with the new annual growth. Mexican yams provided the starting material for the manufacture of oral contraceptives for about 15 years, giving way to other sources, such as soya beans, methods for total synthesis, or microbial fermentation.[8]

In 1949, Marker retired to Pennsylvania to devote the rest of his life to traveling, and in 1959 he began an association with a French silversmith who had emigrated to Mexico City, and then with his son, Pedro Leites. After 1970, Marker turned to collecting paintings by Mexican artists. The artwork and the replicas of antique works in silver were successful businesses that allowed him, in the 1980s, to endow scientific lectureships at both Pennsylvania State University and the University of Maryland. In 1970, the Mexican government honored Marker and awarded him the Order of the Aztec Eagle; staying true to his irascible nature, he declined. In 1984, Pennsylvania State University established the annual Marker Lectures in Science and, in 1987, the Russell and Mildred Marker Professorship of Natural Product Chemistry. In 1987, Marker was granted an honorary doctorate in science from the University of Maryland, the degree he failed to receive in 1926.

In 1990, Marker was planning on a quiet visit to Mexico to present a plaque made in his honor by Pennsylvania State University to Adolfina Moreno, the daughter of Alberto, the owner of the small country store whom Marker met in 1942. Mexican scientists and pharmaceutical people learned of the visit, and that summer a chartered busload of fifty people retraced Marker's trip from Mexico City to Orizaba.[6] Marker rode in a car with Frederico Lehman's son, Pedro, who had become a distinguished chemist. Meeting in an auditorium at the University of Veracruz, Marker was honored by speeches and an engraved silver tray. After lunch at a local brewery, nearly 100 people made their way to the bridge over the Mezcala River. Marker entered the living quarters behind the store now owned by Adolfina. She tearfully thanked him and pointed to a nearby photo, her marriage picture from fifty years ago, with Marker in the wedding group. At the age of 92, Russell Earl Marker died in Wernersville, Pennsylvania, in 1995, from complications after a broken hip.

The Race for Cortisone

When Marker left Syntex, he took his know-how with him. Fortunately for Syntex, there still was no patent on his discoveries. George Rosenkranz left his native Hungary to study chemistry in Switzerland under the renowned steroid chemist Leopold Ruzicka, who was awarded the 1939 Nobel Prize in Chemistry.[9] On the day Pearl Harbor was attacked, Rosenkranz was in Havana waiting for a ship to Ecuador where the chair in organic chemistry awaited him at the University of Quito. The ship never showed. Rebuffed by the national university in Cuba, Rosenkranz took a job with a local pharmaceutical firm for $25 per week. Because of his success in developing new products, he was soon earning $1,000 per month and directing a research program with Ph.D. candidates from the university. He was also learning how to be a business man; for example, he organized the shark-fishing business in Cuba in order to produce vitamin A from shark liver oil.[10]

The Rosenkranz laboratory was following Marker's published techniques and making small amounts of progesterone and testosterone from sarsaparilla roots imported from Mexico. The news of this activity led to an invitation from Syntex to take over for Marker, with an option of buying 15% of Syntex stock, although the company was currently practically bankrupt.

Rosenkranz's task was complicated by Marker's secretiveness. He found reagents labeled with code words; Marker's workers identified solvents by their weight and smell. Rosenkranz gave up on reconstructing Marker's process, and worked out his own commercial manufacture of progesterone and testosterone from Mexican yams, and soon Syntex was making large profits providing the sex hormones as raw material to other pharmaceutical companies. Rozenkranz now had a large active laboratory that attracted a young chemist, Carl Djerassi. These men knew each other, meeting and interacting with each other at the Laurentian Hormone Conference, the annual meeting organized and directed by Gregory Pincus.

The Djerassi family lived in Bulgaria for hundreds of years after escaping Spain during the Inquisition.[11] Carl Djerassi, the son of a Bulgarian physician, was born in Vienna, as was his physician mother. Djerassi, age 16, and his mother fled the Nazi Anschluss and emigrated to the U.S. in 1939. A Jewish refugee aid organization placed Djerassi with a family in Newark, New Jersey. With a scholarship to Tarkio College in Tarkio, Missouri, he was exposed to Middle America, where he earned his way giving talks to church groups about Bulgaria and Europe. His education was further supported by another scholarship from Kenyon College in Ohio, where he pursued chemistry. After a year working for CIBA, Djerassi received his graduate degree from the University of Wisconsin. Returning to CIBA and being somewhat unhappy, he responded to an invitation to visit Syntex. Rosenkranz proposed that Djerassi head a research group to concentrate on the synthesis of cortisone. Djerassi's initial reaction was that "the location of Syntex in the chemical desert of Mexico made the offer seem ludicrous."[12] But the 26-year-old Djerassi, impressed by Rosenkranz and excited by the challenge to develop a method to synthesize cortisone, accepted the position and moved to Mexico City in the fall of 1949.

Earlier in 1949, Philip S. Hench, a Mayo Clinic rheumatologist, showed a movie at a medical meeting documenting crippled arthritic patients before treatment and the same patients active, even dancing, after daily injections with cortisone. Cortisone can be converted to the more active cortisol (also called hydrocortisone), the major product of the adrenal cortex. Cortisone is produced by hydroxylation, which converts the oxygen attached at the 11 position to a hydroxyl group by adding a hydrogen.

Hench had obtained the very expensive cortisone through a biochemist at the Mayo Foundation, Edward C. Kendall, the discoverer of the thyroid hormone, thyroxine, who had been working with Lewis H. Sarett at Merck & Company to determine the structures of

compounds isolated from extracts of the adrenal cortex and from cattle bile; cortisone was known as Kendall's Compound E. Hench reported good results in 14 patients; his movie received a standing ovation,[13] and in 1950, Hench and Kendall were awarded the Nobel Prize in Physiology or Medicine. It was recognized that continuing regular treatment would be necessary, and the race was on to develop an easy and cheap method to synthesize cortisone and related drugs.

In Mexico City, Carl Djerassi was using the plant steroid diosgenin from the Mexican yam as the starting point. In 2 years time, Syntex achieved the partial synthesis of cortisone, reported in 1951.[14] The Syntex method never reached commercialization, however, because a more efficient process was developed by the Upjohn Company. Djerassi's productivity at Syntex, 60 publications, attracted a job offer from Wayne State University.[15] Wanting all along to be in the academic world, Djerassi moved to Detroit in January 1951. Five years later, he took a leave of absence to return to Syntex, now American-owned and a public company. Syntex's topical corticoid anti-inflammatory products, Synalar and Neosynalar, came from Djerassi's laboratory. Djerassi maintained his laboratory at Wayne State, and in 1959, when W.S. Johnson at Wisconsin moved to head the chemistry department at Stanford University, Djerassi joined him—a professorial position he held for the next 25 years.

The Upjohn Company and G.D. Searle & Company joined the competition to synthesize cortisone, with Upjohn, the bigger company, devoting over 150 scientists and technicians to the task. Upjohn leadership assigned a symbol to represent the project, a blow torch, making it clear that this was a heated race they wished to win.[16] G.D. Searle was a smaller company, but its participation in this race would cement a long-term relationship with Gregory Pincus.

G.D. Searle was founded in 1888 by Gideon Daniel Searle, a pharmacist in Indiana, to provide elixirs, syrups, and drugs directly to clinicians. Searle's son, Claude, graduated from Rush Medical College in 1898 and developed a large, successful practice in Sabula, Iowa. In 1909, when his father suffered a stroke, the son returned to Chicago to manage the company, setting up a research department that developed new products. His son, Jack Searle, graduated from the University of Michigan with a degree in pharmacy, and succeeded his father as president of the company in 1936. He recruited Albert L. Raymond from the Rockefeller Institute to serve as director of research, working in new laboratories in Skokie, Illinois. Dramamine, to prevent motion sickness, and Banthine, to treat peptic ulcers, came from these laboratories.

By 1949, Raymond and the G.D. Searle company were supporting steroid research at the Worcester Foundation for Experimental Biology in Massachusetts, and Gregory Pincus, the cofounder of the Worcester Foundation, was a Searle consultant.[17] Pincus and Oscar Hechter had developed a perfusion method, pumping blood, serum, or a serum-like solution through fresh endocrine glands (adrenal glands, testicles, or ovaries) held in a glass apparatus and collecting the perfused fluid. Using the enzymes in the glands, precursors in the perfusing fluid were converted to the final products, hydrocortisone or the sex steroids. This was a method that could be used to produce commercial amounts of cortisone products.

The round-faced, balding, acerbic Oscar Hechter came to the Worcester Foundation in 1944 on a fellowship funded by G.D. Searle. Pincus assigned him the task of perfusing adrenal glands, with the aim of identifying the products of adrenal secretion and the hope of creating a system for commercial production. Five years later, Hechter presented the first positive results at a conference in Detroit in 1949.[18] At that same meeting, Hechter saw Hench's movie and listened to his results. Hechter returned to the Foundation and urged that his project be given top priority. Pincus's enduring relationship with Searle that yielded research support and new steroid compounds for almost never-ending testing began in earnest with the race for cortisone and his development of the perfusion system to use animal glands for the synthesis of steroid drugs. The perfusion system was complicated. It required the development of methods to maintain the animal organs, a web of glassware to

infuse and collect appropriate perfusing solutions, and the separation and identification of the steroid products. At the moment of its coveted value in 1946, Pincus chose to sell his rights to Searle for only one dollar, allowing Searle to patent the process.[19] In return, Pincus obtained and tested steroids that could yield products for clinical use.

Responding to Pincus and Hechter's success, the Searle company constructed rows of perfusion systems in their Skokie plant. Each contained a periodically replaced fresh beef adrenal gland, producing every few hours a large volume of perfused solution. The long-term plan was to engineer a more economical and profitable system. But in the meantime, Searle was able to provide substantial amounts of cortisone to clinical researchers throughout the U.S.

At the same time, Merck ramped up Sarett's 36-step synthesizing process from bile acids, and by the end of 1950, they were selling cortisone acetate to clinicians for a price that had been reduced from $200 per gram to $35. In Kalamazoo, Michigan, Upjohn chemists were pursuing a method based on the process used to make penicillin, conversion of precursors by microbes to the desired product. The work was headed by Durey H. Peterson, the son of Swedish immigrants. Peterson supported his education by playing semi-professional baseball.[16] Early in his career, he developed nylon surgical suturing material as well as "Toni," a product for home permanents to create curly hair. Peterson joined Upjohn in 1946 to work on antibiotics, but he almost immediately became part of the race to synthesize cortisone. Peterson believed that lower microorganisms might possess the same enzymes used by adrenal glands to make cortisone, especially the difficult step of introducing an oxygen molecule to the structure. When told this could not be done, Peterson said, "The microorganisms do not know this."[16]

Using paper chromatography methods developed by Alejandro Zaffaroni, Peterson and H.C. Murray attacked the problem, beginning in 1949. First they needed a microorganism. This they acquired, a fungus of the *Rhizopus* species, by leaving an agar plate on the window sill of the "oldest and dirtiest laboratory at the Upjohn Company."[16] In one year's time, the two chemists proved the value of microorganisms in chemical synthesis. Their method used *Rhizopus nigricans* to covert progesterone to 11-hydroxyprogesterone, that could in turn be processed into hydrocortisone, also called cortisol, the major corticosteroid secreted by the adrenal cortex.

By 1955, Upjohn had become the market leader, and Searle shut down its perfusion cells and quit the race. Upjohn's commercialization of the methods developed by Peterson and Murray led to popular and successful products. But the Searle people had gained valuable experience that would eventually pay off with other synthesized hormones and products.

The Upjohn method used progesterone as the starting point, available in the early 1950s only from Syntex. George Rosenkranz's laboratory at Syntex was also pursuing the industrial synthesis of cortisone, and in July 1951, Syntex was about to sign a contract with a large chemical firm to begin production. This never happened because of a phone call. Rosenkranz told the story: "I received a phone call from Upjohn asking me whether we would be able to accept an order for ten tons of progesterone at forty-eight cents a gram."[10] The quantity was unheard of, and Upjohn's order remained a puzzle until the microfermentation method was published. Rosenkranz accepted the order, and Syntex found itself as the key supplier of progesterone to other companies.

The Synthetic Progestational Drugs, Norethindrone and Norethynodrel

Djerassi and other Syntex chemists turned their attention to the sex steroids. They discovered that the removal of the 19-carbon from yam-derived progesterone increased the

progestational activity of the molecule. The clue for this work came from Maximilian Ehrenstein at the University of Pennsylvania, who reported in 1944 that a potent progestational compound he had produced appeared to be progesterone without its carbon at the 19 position; henceforth the 19-nor family of compounds indicated steroid chemical structures without the carbon atom at the 19 position.[20] Chemists at Schering A.G. in Berlin had produced orally active versions of estradiol and testosterone in 1938, by substituting an acetylene group in the 17-position of the parent compounds. The resulting ethinyl estradiol later became the estrogen component in oral contraceptives. The ethinyl testosterone product was known as ethisterone, marketed in 1941, and the Syntex chemists reasoned that removal of the 19-carbon would increase the progestational potency of this orally active compound.

On October 15, 1951, norethindrone was synthesized at Syntex; the final steps were performed by Luis Miramontes, working on his undergraduate thesis in chemistry under Djerassi's supervision.[12] The patent application was filed 6 weeks later on November 22, 1951, and the work was presented in April 1952 at the annual meeting of the American Chemical Society and published in 1954.[21] The greater potency of norethindrone, achieved by removing the 19-carbon of ethinyl testosterone, compared with progesterone was demonstrated in monkeys and then 4 women at the National Institutes of Health, reported in 1953, 1956, and 1957.[22–24] Syntex supplied norethindrone to many investigators, including Gregory Pincus. Edward T. Tyler first reported its clinical use in 1955 for the treatment of menstrual disorders.[25]

Frank Colton, a chemist at G.D. Searle & Company, filed a patent for norethynodrel, a compound closely related to norethindrone, differing only in the position of the double bond, on August 31, 1953. The Polish-born Colton received his Ph.D. in chemistry from the University of Chicago. From 1949 to 1951, he was a research fellow working with Edward Kendall at the Mayo Foundation on the synthesis of cortisone. Colton joined Searle in 1951, along with Byron Riegel, to develop steroid drugs, succeeding with Nilevar, the first commercial anabolic agent marketed in 1956 and Aldactone, the anti-aldosterone antihypertensive agent introduced in 1959.

Norethynodrel was the result of a deliberate and planned program to create orally active agents with progestational activity. Later, Colton pointed out that although the Syntex and Searle chemists followed a similar path, they were independently pursuing the trail blazed by previous scientists.[17] Along the way, hundreds of compounds were sent to Pincus at the Worcester Foundation to test for ovulation inhibition in rabbits. Their best drug, norethynodrel, assigned the number SC-4642, was synthesized at Searle in a process that was considered to be significantly different from the Syntex method.[17]

Djerassi urged legal proceedings for patent infringement, claiming that norethynodrel was converted to Syntex's compound, norethindrone, by gastric acid, but Parke-Davis, the American company licensing norethindrone, did not want to make waves presumably because Parke-Davis was supplying the antihistamine component of Searle's best-selling product for motion sickness, Dramamine.[12] Pincus would ultimately choose the Searle compound, norethynodrel for clinical testing as an oral contraceptive, and Syntex, not having marketing capability, licensed norethindrone to other pharmaceutical companies. Norethindrone was tested as a contraceptive by Edward Tyler in Los Angeles and Joseph Goldzieher in San Antonio, Texas, but Parke-Davis chose not to pursue government approval, probably fearing religious reactions. Subsequently, Syntex turned to the Ortho division of Johnson & Johnson. By 1964, Ortho, Parke-Davis, and Syntex (now in California) were marketing oral contraceptives containing norethindrone or its acetate.

The creation of norethindrone and norethynodrel by the chemists was essential in the development of oral contraception because the natural hormone progesterone is relatively impotent given orally, requiring very large doses that even then do not achieve a uniform response. The synthetic progestational agents are very active when administered orally, producing reliable effects with small doses.

A Wall Street entrepreneur, Charles Allen, acquired Syntex in 1956 for $2 million cash and a loan of $2 million to be paid from future profits.[9] Rosenkranz became president and CEO, Alejandro Zaffaroni, an Italian who emigrated from Montevideo, Uruguay, executive vice president. Zaffaroni obtained his Ph.D. in 1949 in biochemistry from the University of Rochester, developing a paper chromatography system that soon became a principal method of studying steroid hormones.[26] Rosenkranz met Zaffaroni at the Laurentian Hormone Conference in 1951. Their aim was to develop a pharmaceutical company on a foundation of research. Carl Djerassi, who had left for an academic position at Wayne State University, was recruited back to the company. Rosenkranz said, "We were the brilliant amateurs with a 'can do anything' attitude. We were like stem cells (though then none of us really knew the concept). We could differentiate into anything we desired. Production, finance, sales, marketing—all held no fear for us."[9]

In 1961, the company moved to Palo Alto, California, influenced by Djerassi who was teaching at Stanford University. The growth of the company was meteoric, with blockbuster hits like Synalar, a topical corticoidsteroid for the treatment of psoriasis, and Naproxen, a nonsteroid, anti-inflammatory drug. Much of this success was to an innovative philosophy in the pharmaceutical business, "patent and publish."[9] The Syntex scientists were encouraged to promptly publish their results, gaining the peer recognition that is such a motivating force for basic scientists. In 1994, Roche Holdings acquired Syntex for $5.3 billion.

Djerassi eventually left Syntex to become a full-time professor at Stanford University, and is now a playwright and novelist living in San Francisco. Zaffaroni started his own company in 1968, ALZA (after his own name), dedicated to new methods of drug delivery, such as a skin patch. ALZA was acquired by Johnson & Johnson in 2000.

Gregory Pincus

Gregory Goodwin (Goody) Pincus was born in 1903 in New Jersey, the son of Russian Jewish immigrants who lived on a farm colony founded by a Jewish philanthropic organization.[19] Pincus was the oldest of six children and grew up in a home of intellectual curiosity and energy, but even his family regarded him as a genius.

Pincus graduated from Cornell and went to Harvard to study genetics, joining Hudson Hoagland and B.F. Skinner as graduate students of W.J. Crozier in physiology, receiving degrees in 1927. Crozier's hero was Jacques Loeb who discovered artificial parthenogenesis working with sea urchin eggs. Most importantly, Loeb was a strong believer in applying science to improve human life. Thus, Crozier, influenced by Loeb, taught Pincus, Hoagland, and Skinner (respectively, in reproductive biology, neurophysiology, and psychology) to apply science to human problems. This was to be the cornerstone of Pincus's own philosophy.

Hoagland, after a short stay at Harvard, spent a year in Cambridge, England, and then moved to Clark University in Worcester, Massachusetts, to be the chair of biology at the age of 31. Pincus went to England and Germany, and returned to Harvard as an assistant professor of physiology.

Pincus performed pioneering studies of meiotic maturation in mammalian oocytes, in both rabbit and human oocytes. In 1934, Pincus reported the achievement of in vitro fertilization of rabbit eggs, earning him a headline in the *New York Times* that alluded to Haldane and Huxley. An article in *Colliers* depicted him as an evil scientist. By 1936, Harvard had cited Pincus's work as one of the university's outstanding scientific achievements of all time, but Harvard denied him reappointment in 1937.

At Clark University, Hudson Hoagland was in constant conflict with the president of the university, Wallace W. Atwood, the senior author of a widely used textbook on geography. In 1931, the Department of Biology consisted of one faculty member and his graduate student, and their chair, Hudson Hoagland. Hoagland, upset and angry over Harvard's refusal to grant reappointment to his friend (suspecting that this was because of anti-Semitism), invited Pincus to join him. Hoagland secured funds for Pincus from philanthropists in New York City, enough for a laboratory and an assistant. This success impressed the two men, especially Hoagland, planting the idea that it would be possible to support research with private money.

Min-Chueh Chang was born in Tai Yuan, China, on October 10, 1908. In 1933, he earned a bachelor's degree in animal psychology from the Tsing Hua University in Peking, and stayed at the university as a teacher. Chang won a national competition in 1938 that funded study abroad. He chose to study agricultural science at Edinburgh University. After one year, he was pleased to receive an invitation from Arthur Walton to study the physiology of sheep sperm at The University of Cambridge, and he promptly accepted.

Chang received his Ph.D. in animal breeding under the direction of Walton and Sir John Hammond at the University of Cambridge in 1941. It was virtually impossible to leave England during the early years of World War II, and Chang continued to work at the University. In 1944, Chang planned to return to China, but first he wanted to spend a year in the U.S. He wrote three letters to American scientists, and only Pincus answered, offering a fellowship at Clark University. Chang mistakenly assumed that a fellowship in the U.S. was the same as at the University of Cambridge where a Fellow was assured of a lifetime income. The successful recruitment of M-C Chang by Pincus was to pay great dividends. Years later, Chang would direct the testing of new progestins to effectively inhibit ovulation in animals.

Soon Hoagland had put together a group of outstanding scientists, but because of his ongoing antagonism with President Atwood, the group was denied faculty status. Working in a converted barn, they were totally supported by private funds. By 1943, 12 of Clark's 60 faculty were in the Department of Biology.

Frustrated by the politics of academia, Hoagland and Pincus (who both enjoyed stepping outside of convention) had a vision of a private research center devoted to their philosophy of applied science. Indeed, the establishment of the Worcester Foundation for Experimental Biology, in 1944, can be attributed directly to Hoagland and Pincus, their friendship for each other, and their confidence, enthusiasm, ambition, and drive. It was their spirit that turned many members of Worcester society into financial supporters of biologic science. Hoagland and Pincus accomplished what they set out to do. They created and sustained a vibrant, productive scientific institution in which it was a pleasure to work.

Although named the Worcester Foundation for Experimental Biology, the Foundation was located in the summer of 1945 across Lake Quinsigamond in a house on an estate in Shrewsbury. From 1945 to the death of Pincus in 1967, the staff grew from 12 to 350 (scientists and support people), 36 of whom were independently funded and 45 were postdoctoral fellows. The annual budget grew from $100,000 to $4.5 million. One hundred acres of adjoining land were acquired, and the campus grew to 11 buildings. In its first 25 years, approximately 3,000 scientific papers were published.

But in those early years, Pincus was the animal keeper, Mrs. Hoagland the bookkeeper, M-C. Chang was the night watchman, and Hoagland mowed the lawn. During the years of World War II, Pincus and Hoagland combined their interests in hormones and neurophysiology to focus on stress and fatigue in industry and the military.

Katharine Dexter McCormick (1875–1967) was a trained biologist, an early suffragist, and rich, inheriting millions from her mother and a McCormick fortune from her husband.

She was the second woman to graduate from the Massachusetts Institute of Technology, socially conscious, and a generous contributor to family planning efforts. Her intervention with money, energy, incisive thinking, and persistent dedication was instrumental in the development of oral contraception. In 1904, she married Stanley McCormick, the son of Cyrus McCormick, the founder of International Harvester. Katharine's husband suffered from schizophrenia, and she established the Neuroendocrine Research Foundation at Harvard to study schizophrenia. This brought her together with Hoagland, who told her of the work being done by Chang and Pincus who were seeking orally active progestins to inhibit ovulation.

Pincus attributed his interest in contraception to his growing appreciation for the world's population problem, and to a 1951 visit in New York with Margaret Sanger, at that time president of the Planned Parenthood Federation of America. Sanger promised a small amount of money and expressed hope that a method of contraception could be derived from the laboratory work being done by Pincus and Chang. During this meeting, Pincus formulated his thoughts derived from his mammalian research. He envisioned a progestational agent in pill form as a contraceptive, acting like progesterone in pregnancy.

Margaret Sanger brought Pincus and Katharine McCormick together. On June 7, 1953, when 78-year-old Katharine met with 50-year-old Pincus at the Worcester Foundation and wrote him a check for $20,000; she promised him another $20,000. A week later, Pincus and Hoagland met with Katharine and her lawyer. They signed a contract outlining the goals, the decision-making process, and the timetable. Pincus received a second check for $20,000, and Katharine agreed to fund laboratory improvements, which ended up as the completion of a new building in 1955.

Katharine's contract with the Worcester Foundation stipulated that Pincus would provide written reports every 2 weeks. In addition, Pincus and John Rock, the Boston gynecologist performing the initial oral contraceptive studies in his patients, made many visits to Katharine's home office on Beacon Street across the street from the Harvard Club. Katharine had Sara De Laney, her secretary, take careful notes in shorthand, and at the next visit De Laney read the transcribed notes to her boss so that she would be prepared. Periodically the principals met at the Worcester Foundation. Katharine peppered Pincus, Chang, and Rock with questions and urged them to stop wasting time. She found Pincus "imaginative and inspirational; Rock was informative and very realistic about medical work." By now everyone was familiar with Katharine's methods. She had earned their respect, and detailed reports on laboratory results, clinical planning, and budgets were immediately forthcoming. Time and time again, Katharine proved that she handled delays poorly, but she approached each meeting with an eagerness that slowly but surely was rewarded with success after 7 years and an expenditure of about $2 million of Katharine's money.

In her last years, Katharine continued to support the work of Pincus and Chang. When testing the hundreds of compounds that yielded the progestational agents in birth control pills, Chang observed that some of them prevented implantation of fertilized eggs in rabbits. From 1962 to 1966, Chang and Pincus were pursuing a drug that could prevent pregnancy with one administration, a day or two after sexual intercourse. With Pincus's death, this project was abandoned. It is not certain whether Chang and Pincus coined the phrase the "morning after" pill, but it is accurate to state that the concept came from Chang.

When Pincus and Chang began their studies, the focus was on inhibition of ovulation, first by progesterone, and then by synthetic progestins. Chang's contribution was easy to overlook. Chang worked away in his laboratory, and it was Pincus who was highly visible, raising the money and providing direction. Chang started by repeating the experiments reported by Makepeace in 1927, documenting that progesterone could inhibit ovulation. The first experiment was on April 25, 1951, and Chang quickly moved to testing the newly synthesized progestins from Searle and Syntex.

By December 1953, three synthetic progestins were selected as the most potent and effective in inhibiting ovulation: norethindrone from Syntex, and Searle's norethynodrel and norethandrolone. The animal and human results were published in *Science* in 1956.[27, 28] In 1957, these three compounds were approved for the treatment of menstrual disorders with the trade names of Norlutin, Enovid, and Nilevar, respectively.

It was Pincus who made the decision to involve a physician because he knew human experiments would be necessary. John Rock, chief of gynecology and obstetrics at Harvard, met Pincus at a scientific conference and discovered their mutual interest in reproductive physiology. Rock and his colleagues pursued Pincus's work. Using oocytes from oophorectomies, they reported in vitro fertilization in 1944, the first demonstration of fertilization of human oocytes in vitro. Rock was interested in the work with progestational agents, not for contraception, however, but because he hoped the female sex steroids could be used to overcome infertility.

In their first collaborative study, Pincus and Rock administered oral progesterone, 300 mg/day. Pincus suggested a 20-day regimen beginning on day 5 of the menstrual cycle.[29] He had two reasons for choosing this regimen: (1) it covered the time period during which nearly all, if not all, ovulations occurred, and (2) the withdrawal menstrual bleed at the conclusion of the treatment period would mimic the timing of a normal menstrual cycle and reassure the women that they were not pregnant. The first study involved 33 volunteers who ovulated regularly but had been infertile for 2 years. The women were treated for 1 to 3 cycles after a baseline control month. About 85% of the treated women did not ovulate during the treatment cycles. Not one became pregnant during treatment, pleasing Pincus who all along was aiming for contraception, and four became pregnant after treatment, pleasing Rock who initially was motivated by his pursuit of the "rebound" phenomenon for the treatment of infertility.

Sanger and McCormick needed some convincing that Rock's Catholicism would not be a handicap, but they were eventually won over because of his stature. Rock was a physician who literally transformed his personal values in response to his recognition of the problems secondary to uncontrolled reproduction. With the help of Luigi Mastroianni, the first administration of synthetic progestins to women was to Rock's patients in 1954. Of the first 50 patients to receive 10–40 mg of synthetic progestin (a dose extrapolated from the animal data) for 20 days each month, all failed to ovulate during treatment (causing Pincus to begin referring to the medication as "the pill"), and 7 of the 50 became pregnant after discontinuing the medication, again pleasing Rock, who all along was motivated to treat his infertile patients.

In 1956, with Celso-Ramon Garcia and Edris Rice-Wray, working in Puerto Rico, the first human trial was performed. The initial progestin products were contaminated with about 1% mestranol. In the amounts being used, this added up to 50–500 μg of mestranol, a sufficient amount of estrogen to inhibit ovulation by itself. When efforts to provide a more pure progestin lowered the estrogen content and yielded breakthrough bleeding, it was decided to retain the estrogen for cycle control, thus establishing the principle of the combined estrogen-progestin oral contraceptive. Early clinical trials were also conducted by J.W. Goldzieher in San Antonio and E.T. Tyler in Los Angeles.

Pincus, a longtime consultant to Searle, picked the Searle compound for extended use, and with great effort, convinced Searle that the commercial potential of an oral contraceptive warranted the risk of possible negative public reaction. Pincus also convinced Rock, and together they pushed the U.S. Food and Drug Administration for acceptance of oral contraception. In 1957, Enovid was approved for the treatment of miscarriages and menstrual disorders, and on June 23, 1960, for contraception. Neither Pincus nor the Worcester Foundation got rich on the pill; alas, there was no royalty agreement.

The pill did bring Pincus fame and travel. There is no doubt that he was very much aware of the accomplishment and its implications. As he traveled and lectured in 1957, he said:

"How a few precious facts obscurely come to in the laboratory may resonate into the lives of men everywhere, bring order to disorder, hope to the hopeless, life to the dying. That this is the magic and mystery of our time is sometimes grasped and often missed, but to expound it is inevitable."[30]

Pincus was the perfect person to bring oral contraception into the public world, at a time when contraception was a private, suppressed subject. Difficult projects require people like Pincus. A scientific entrepreneur, he could plow through distractions. He could be hard and aggressive with his staff. He could remain focused. He hated to lose, even in meaningless games with his children. Yet he combined a gracious, warm, charming manner with his competitive hardness. He was filled with the kind of self-confidence that permits an individual to forge ahead, to translate vision into reality. Pincus died in 1967 (as did Katharine McCormick at the age of 92), of myeloid metaplasia. Rock died in 1984, at the age of 94. Chang died in 1991 at the age of 82, and was buried in Shrewsbury, near his laboratory and close to the grave of Pincus.

Pincus wrote his book, *The Control of Fertility*, in 1964–1965, because "a break came in the apparent dam to publication on reproductive physiology and particularly its subdivisions concerned with reproductive behavior, conception, and contraception."[30]

> "We have conferred and lectured in many countries of the world, seen at first hand the research needs and possibilities in almost every European, Asiatic, Central, and South American country. We have faced the hard fact of overpopulation in country after country, learned of the bleak demographic future, assessed the prospects for the practice of efficient fertility control. This has been a saddening and a heartening experience; saddening because of the sight of continuing poverty and misery, heartening because of the dedicated colleagues and workers seeking to overcome the handicap of excess fertility and to promote healthy reproductive function. Among these we have made many friends, found devoted students."[30]

Syntex, a wholesale drug supplier, was without marketing experience or organization. By the time Syntex had secured arrangements with Ortho for a sales outlet, Searle marketed Enovid in 1960 (150 μg mestranol and 9.85 mg norethynodrel). Ortho-Novum, using norethindrone from Syntex, appeared in 1962. Wyeth Laboratories introduced norgestrel in 1968, the same year in which the first reliable prospective studies were initiated. It was not until the late 1970s that a dose-response relationship between problems and the amount of steroids in the pill was appreciated. Health care providers and patients, over the years, have been confronted by a bewildering array of different products and formulations. The solution to this clinical dilemma is relatively straightforward, the theme of this chapter: use the lowest doses that provide effective contraception.

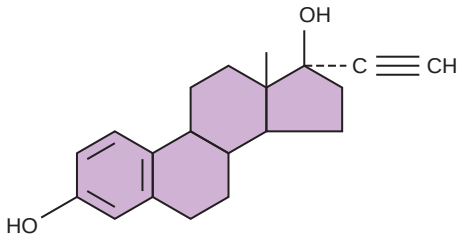
Pharmacology of Steroid Contraception

The Estrogen Component of Combination Oral Contraceptives

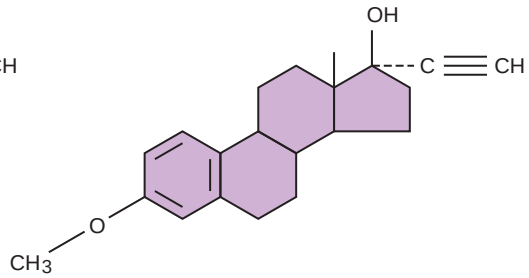
Estradiol is the most potent natural estrogen and is the major estrogen secreted by the ovaries. The major obstacle to the use of sex steroids for contraception was reduced activity of the compounds when given orally. A major breakthrough occurred in 1938 when

it was discovered that the addition of an ethinyl group at the 17 position increased oral activity. Ethinyl estradiol is a very potent oral estrogen and is the form of estrogen in most oral contraceptives. Another estrogen, present in older products, was the 3-methyl ether of ethinyl estradiol, mestranol.

Mestranol and ethinyl estradiol are different from natural estradiol and must be regarded as pharmacologic drugs. Animal studies suggested that mestranol is weaker than ethinyl estradiol, because mestranol must first be converted to ethinyl estradiol in the body. Indeed, mestranol will not bind to the cellular estrogen receptor. Therefore, unconjugated ethinyl estradiol is the active estrogen in the blood for both mestranol and ethinyl estradiol. In the human body, differences in potency between ethinyl estradiol and mestranol do not appear to be significant, certainly not as great as indicated by assays in rodents. This is now a minor point because, with the exception of a contraceptives with estradiol valerate or estradiol, all of the low-dose oral contraceptives contain ethinyl estradiol.



Ethinyl estradiol



Mestranol

The metabolism of ethinyl estradiol (particularly as reflected in blood levels) varies significantly from individual to individual, and from one population to another.[31, 32] There is even a range of variability at different sampling times within the same individual. Therefore, it is not surprising that the same dose can cause side effects in one individual and not in another.

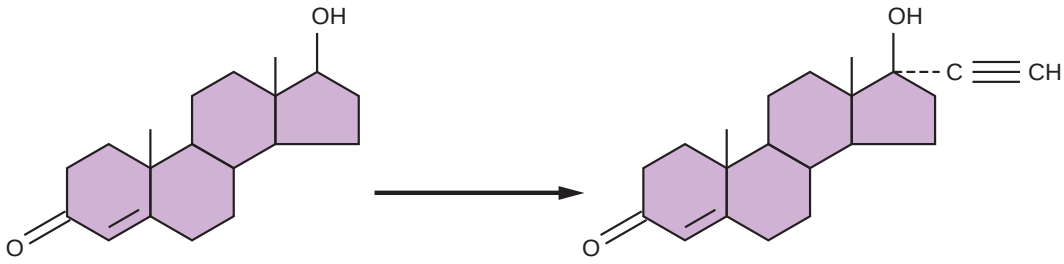
Estradiol valerate is an esterified form of estradiol, allowing oral administration with significant potency. The ester is rapidly hydrolyzed to estradiol. Combinations of several progestins with estradiol valerate have demonstrated good contraceptive efficacy.

The estrogen content (dosage) of the pill is of major clinical importance. Thrombosis is one of the most serious side effects of the pill, playing a key role in the increased risk of death (in the past with high doses) from a variety of circulatory problems. This side effect is related to estrogen, and it is dose related. Therefore, the dose of estrogen is a critical issue in selecting an oral contraceptive.

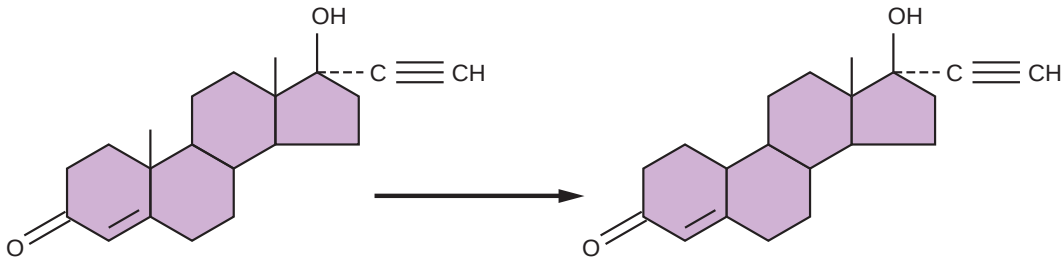
The Progestin Component of Combination Oral Contraceptives

The discovery of ethinyl substitution and oral potency led (at the end of the 1930s) to the preparation of ethisterone, an orally active derivative of testosterone. In 1951, it was demonstrated that removal of the 19-carbon from ethisterone to form norethindrone did not destroy the oral activity, and most importantly, it changed the major hormonal effect from that of an androgen to that of a progestational agent. Accordingly, the progestational derivatives of testosterone were designated as 19-nortestosterones (denoting the

missing 19-carbon). The androgenic properties of these compounds, however, were not totally eliminated, and minimal anabolic and androgenic potential remains within the structure.



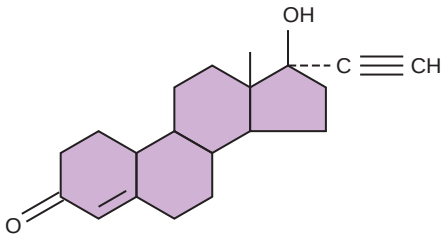
Testosterone **Ethisterone**



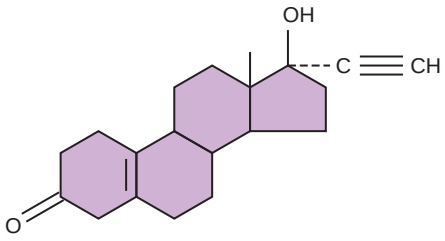
Ethisterone **Norethindrone**

The "impurity" of 19-nortestosterone, i.e., androgenic as well as progestational effects, was further complicated in the past by a belief that they were metabolized within the body to estrogenic compounds. This question was restudied, and it was argued that the previous evidence for metabolism to estrogenic compounds was due to an artifact in the laboratory analysis. More recent studies indicate that norethindrone can be converted to ethinyl estradiol; however, the rate of this conversion is so low that insignificant amounts of ethinyl estradiol can be found in the circulation or urine following the administration of the commonly used doses of norethindrone.[33] Any estrogenic activity, therefore, would have to be due to a direct effect. In animal and human studies, however, only norethindrone, norethynodrel, and ethynodiol diacetate have estrogen activity, and it is very slight due to weak binding to the estrogen receptor.[34] Clinically, androgenic and estrogenic activities of the progestin component, therefore, are insignificant due to the low dosage in the current oral contraceptives. As with the estrogen component, serious side effects have been related to the high doses of progestins used in old formulations, and routine use of oral contraceptives should now be limited to the low-dose products.

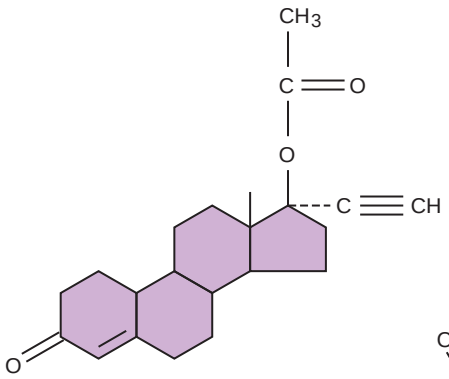
The norethindrone family contains the following 19-nortestosterone progestins: norethindrone, norethynodrel, norethindrone acetate, ethynodiol diacetate, lynestrenol, norgestrel, norgestimate, desogestrel, and gestodene.



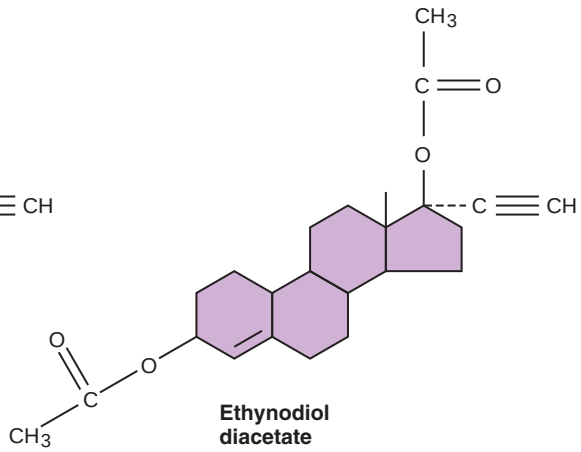
Norethindrone



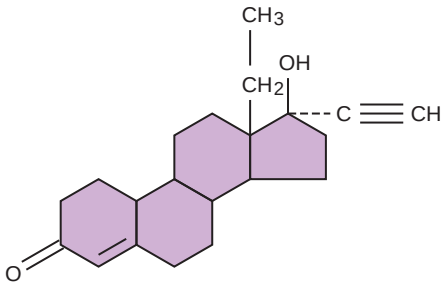
Norethynodrel



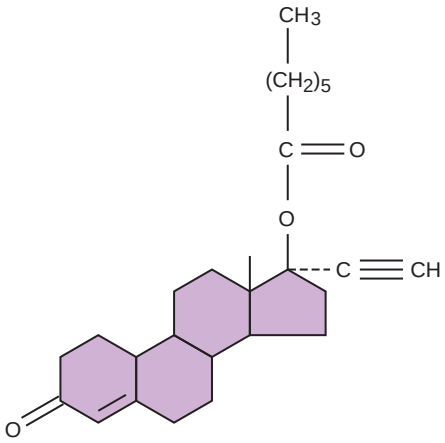
Norethindrone acetate



Ethynodiol diacetate



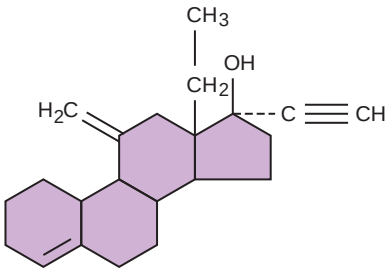
Levonorgestrel



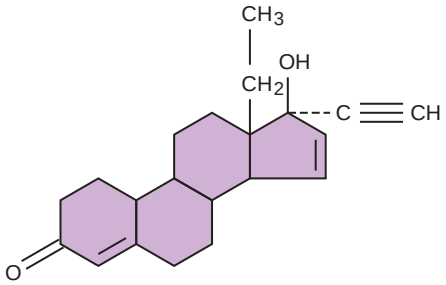
Norethindrone enanthate

Most of the progestins closely related to norethindrone are converted to the parent compound. Thus the activity of norethynodrel, norethindrone acetate, ethynodiol diacetate, and lynestrenol is due to rapid conversion to norethindrone.

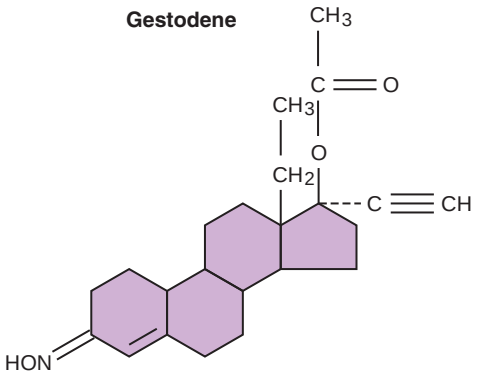
Norgestrel is a racemic equal mixture of the dextrorotatory enantiomer and the levorotatory enantiomer. These enantiomers are mirror images of each other and rotate the plane of polarized light in opposite directions. The dextrorotatory form is known as d-norgestrel, and the levorotatory form is l-norgestrel (known as levonorgestrel). Levonorgestrel is the active isomer of norgestrel.



Desogestrel



Gestodene



Norgestimate

Desogestrel undergoes two metabolic steps before the progestational activity is expressed in its active metabolite, 3-keto-desogestrel, known as etonogestrel. This metabolite differs from levonorgestrel only by a methylene group in the 11 position. Gestodene differs from levonorgestrel by the presence of a double bond between carbons 15 and 16; thus, it is Δ-15 gestodene. It is metabolized into many derivatives with progestational activity, but not levonorgestrel. Several metabolites have the potential to contribute to the activity of norgestimate. Although norgestimate is a "new" progestin, epidemiologists included it in the oral contraceptive second-generation family because its activity was believed to be largely due to levonorgestrel and levonorgestrel metabolites.[35, 36] Almost all of the biologic effects are attributed to the 17-deacetylated metabolite, now known as norelgestromin; the levonorgestrel metabolites are tightly bound to sex hormone-binding globulin (unlike norelgestromin) severely limiting their biologic activity.[37]

Definitions Used In Epidemiologic Studies

Low-Dose Oral Contraceptives — Products containing less than 50 μg ethinyl estradiol

First-Generation Oral Contraceptives — Products containing 50 μg or more of ethinyl estradiol

Second-Generation Oral Contraceptives — Products containing levonorgestrel, norgestimate, and other members of the norethindrone family and 20, 30, or 35 μg ethinyl estradiol

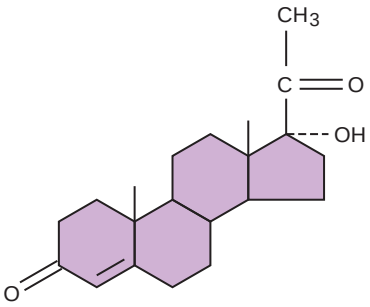
Third-Generation Oral Contraceptives — Products containing desogestrel or gestodene with 20, 25, or 30 μg ethinyl estradiol

Fourth-Generation Oral Contraceptives — Products containing drospirenone, dienogest, or nomegestrol acetate.

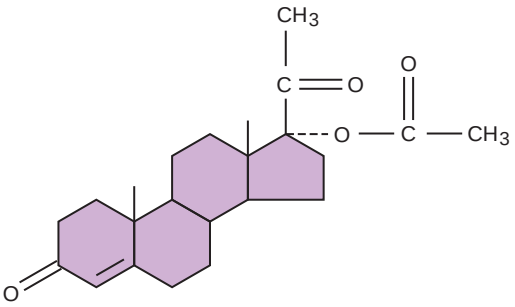
Probably the greatest influence on the effort that yielded the new progestins was the belief throughout the 1980s that androgenic metabolic effects were important, especially in terms of cardiovascular disease. Cardiovascular side effects are now known to be due to a dose-related stimulation of thrombosis by estrogen and not secondary to metabolic effects such as lipid changes. In the search to find compounds that minimize androgenic effects, however, the pharmaceutical companies succeeded.

The new progestins include desogestrel, gestodene, and norgestimate.[38] In regard to cycle control (breakthrough bleeding and amenorrhea), the new formulations are comparable with previous low-dose products. All progestins derived from 19-nortestosterone have the potential to decrease glucose tolerance and increase insulin resistance. The impact on carbohydrate metabolism of the previous low-dose formulations was very minimal, and the impact of the new progestins is negligible. Most changes are not statistically significant, and when they are, they are so subtle as to be of no clinical significance. The decreased androgenicity of the progestins in the newer products is reflected in increased sex hormone-binding globulin and decreased free testosterone concentrations to a greater degree than the older oral contraceptives. This difference could be of greater clinical value in the treatment of acne and hirsutism, but comparative clinical studies have indicated similar effects for all oral contraceptives.[39]

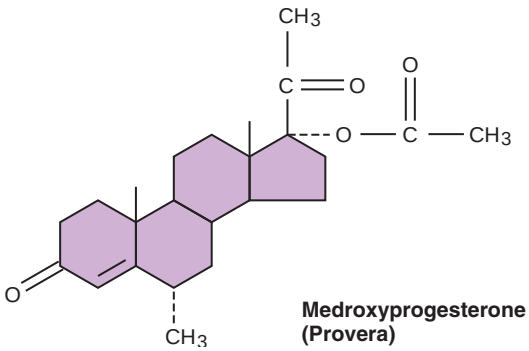
The new progestins, because of their reduced androgenicity, predictably do not adversely affect the cholesterol-lipoprotein profile. Indeed, the estrogen-progestin balance of combined oral contraceptives containing one of the new progestins even promote favorable lipid changes. Thus, the new formulations have the potential to offer protection against cardiovascular disease, an important consideration as we enter an era of women using oral contraceptives for longer durations and later in life. But one must be cautious regarding the clinical significance of subtle changes, and it is unlikely there will be a major impact.



17α-Hydroxyprogesterone



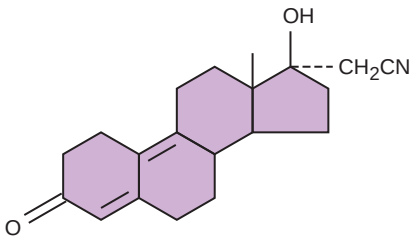
17-Acetoxy progesterone



Medroxyprogesterone acetate (Provera)

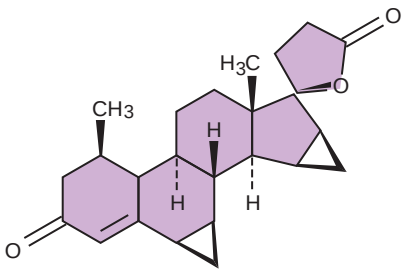
A second group of progestins became available for use when it was discovered that acetylation of the 17-hydroxy group of 17-hydroxyprogesterone produced an orally active but weak progestin. An addition at the 6 position is necessary to give sufficient progestational strength for human use, probably by inhibiting metabolism. Derivatives of progesterone with substituents at the 17 and 6 positions include the widely used medroxyprogesterone acetate. Chlormadinone and cyproterone acetate are progesterone derivatives with a 17α-acetoxy group, marketed in a combination with ethinyl estradiol.

Dienogest is a 19-nortestosterone that has a cyanomethyl group instead of an ethinyl group in the 17 position and an additional double bond, combining the properties of both the 19-nortestosterone family and the derivatives of progesterone.[40] It exerts antiandrogenic activity and is used in a 2-mg dose combined with 30 μg ethinyl estradiol or estradiol valerate as an oral contraceptive.[41–43] The combination oral contraceptive with estradiol valerate (Qlaira, Natazia) uniquely consists of 4 phases, with the estradiol valerate dose decreasing from 3 mg to 1 mg over 26 days, and dienogest provided in a dose of 2 mg on days 3–7 and 3 mg on days 8–24. This phasic administration of estradiol valerate and dienogest provides ovulation inhibition and a bleeding profile comparable to that of a standard, low-dose, monophasic oral contraceptive.[42–44]



Dienogest

Drospirenone is a progestin that is an analogue of spironolactone. Its biochemical profile is very similar to progesterone, including a high affinity for the mineralocorticoid receptor that produces an antimineralocorticoid effect.[45, 46] Contraceptive efficacy equal to that of other formulations is achieved in the combination of 3.0 mg drospirenone and 30 μg ethinyl estradiol (Yasmin). Because drospirenone is spironolactone-like with antiandrogenic and antimineralocorticoid activity, caution is recommended in regard to serum potassium levels, avoiding its use in women with abnormal renal, adrenal, or hepatic function. However, hyperkalemia and its complications have not been a clinical problem encountered with the use of a drospirenone-containing contraceptive in the general population.[47]



Drospirenone

It has been suggested that the oral contraceptive that contains drospirenone is effective for treating premenstrual syndrome/premenstrual dysphoric disorder (PMDD). In an open-label, 1-year study of 326 women, Yasmin was associated with a significant reduction in scores assessing negative affect, water retention, and increased appetite during the

premenstrual and menstrual phases of their cycles.[48] A similar effect was observed in new users and in those who switched from other oral contraceptives. We have learned over the last decade that treatments for premenstrual syndrome must be studied in comparison with a placebo because of the powerful placebo response associated with this disorder. In a double-blind, placebo-controlled randomized trial, 82 women with established diagnoses of PMDD were assessed using the Calendar of Premenstrual Experiences scale.[49] A statistically significant reduction associated with Yasmin treatment was achieved in only one category, that measuring acne, appetite, and food cravings. The overall result was essentially not significant.

In a multicenter 2-year study in Europe of 900 women, Yasmin was compared to Marvelon (the same dose of ethinyl estradiol and 150 μg desogestrel).[50] Marvelon was associated with a small increase in body weight after the fifth cycle; the average body weight associated with Yasmin remained throughout the 2 years below the baseline level at the beginning of the study, but increased to a level above the baseline at the end of the study. The early weight loss amounted to only 1% of body weight and may reflect diuretic action. This study also observed a small reduction in premenstrual symptoms with Yasmin. The evidence, therefore, for a beneficial impact on PMDD with Yasmin is minimal.

The 24-day version of this oral contraceptive, Yaz (3 mg drospirenone and 20 μg ethinyl estradiol), demonstrated in a 3-month multicenter, double-blind, randomized trial symptomatic improvement in 450 women with PMDD.[51] The magnitude of the treatment response compared with placebo amounted to a need to treat 8 women to achieve at least a 50% reduction in symptom severity in a single patient. ***These results suggest that modest beneficial results in treating PMDD may be achieved with any oral estrogen-progestin contraceptive administered in an extended regimen. We would expect similar results with the daily, continuous administration of any oral contraceptive.***

Nomegestrol acetate (NomAc), derived from progesterone with the elimination of the 19 carbon, is used in Europe in postmenopausal hormone therapy and recently combined with estradiol for contraception. The first 24 pills of a package contain 2.5 mg NomAc and 1.5 mg estradiol, and the last four are placebos in an extended regimen typical of newer oral contraceptives. NomAc has potent inhibitory effects on gonadotropin secretion, and no androgenic activity (in fact, it is somewhat antiandrogenic).[52] Unlike drospirenone, NomAc has no glucocorticoid or mineralocorticoid activity.[53] Its relative lack of endometrial effects has been associated with an increase in irregular bleeding in some clinical trials.

Different Formulations

The multiphasic preparation alters the dosage of both the estrogen and progestin components periodically throughout the pill-taking schedule. The aim of these new formulations is to alter steroid levels in an effort to achieve lesser metabolic effects and minimize the occurrence of breakthrough bleeding and amenorrhea, while maintaining efficacy. However, metabolic studies with the multiphasic preparations indicate no differences or very slight improvements over the metabolic effects of low-dose monophasic products.

An estrophasic approach (Estrostep) combines a continuous low dose of a progestin with a low, but gradually increasing dose of estrogen.[54] This approach minimizes estrogen exposure at the beginning of the cycle, yielding a low rate of side effects such as nausea. The increasing estrogen results in a marked increase in sex hormone-binding globulin that produces a very low androgenic state by reducing the bioavailability of circulating free androgens, and this formulation is very effective for treating acne.[55,56]

Extended Regimens

Two clinical concerns prompted the development of an oral contraceptive regimen with a reduction in the pill-free interval: (1) breakthrough bleeding, and (2) ovarian activity during pill taking that could lead to ovulation and failure of contraception. Extending the active pill cycle by several days is aimed at decreasing breakthrough bleeding and spotting and reducing the length of withdrawal bleeding without compromising efficacy or safety, and perhaps increasing contraceptive protection by a greater suppression of ovarian activity. This strategy has produced several new 24-day products: Loestrin 24 Fe (1 mg norethindrone acetate/ethinyl estradiol 20 μg with 4 iron-containing placebo pills), Yaz (3 mg drospirenone/ethinyl estradiol 20 μg), and Minesse (60 μg gestodene/ethinyl estradiol 15 μg).

The traditional combination oral contraceptive pill, consisting of estrogen and progestin components, is given daily for 3 of every 4 weeks, for a total of 21 days. Despite multiple contraceptive actions, there has been concern that the current lower-dose products allow follicular development in some individuals, especially in those who metabolize and clear steroid hormones rapidly.[57] Even with greater follicular activity with the lowest-dose oral contraceptives; however, ovulation is still effectively prevented in most women.[58] Nevertheless, recognition of follicular growth emerging during the standard pill-free interval with oral contraceptives and awareness that ovarian activity is greater with the lowest-dose estrogen formulations, along with the problem of breakthrough bleeding, provided the motivation to shorten the pill-free interval.

A move to low doses of estrogen in combined oral contraceptives has been fueled by a desire to minimize estrogen-linked, serious cardiovascular side effects. Breakthrough bleeding rates are higher with the lower-dose (20 μg ethinyl estradiol) oral contraceptives, although not dramatically.[59–61] Breakthrough bleeding is higher in women who smoke and in smokers who use formulations with 20 μg ethinyl estradiol.[62] Breakthrough bleeding gives rise to fears and concerns; it is aggravating, and even embarrassing. These are reasons why although breakthrough bleeding during oral contraceptive use is considered a *minor* side effect, it can have a *major* consequence: interruption of adherence to therapy resulting in unwanted pregnancies. A nationwide survey identified irregular bleeding as the primary reason for discontinuation of oral contraception.[63] It is important to emphasize that there is no evidence that the onset of bleeding is associated with decreased efficacy, no matter what oral contraceptive formulation is used, even the lower-dose products. Indeed, in a careful study, breakthrough bleeding did not indicate decreases in the contraceptive blood levels of the estrogen and progestin components.[64]

The most frequently encountered breakthrough bleeding occurs in the first few months of use as the endometrium adjusts to the pharmacologic impact of the oral contraceptive. The incidence is greatest in the first 3 months, ranging from 10 to 30% in the first month to less than 10% in the third. However, the differences among the various 21-day formulations containing 20 μg ethinyl estradiol are of minimal clinical significance. For this reason, the new approach evolved, increasing the number of days with active drug treatment to 24.

Ovarian follicles begin to grow during the 7-day pill-free interval in traditional regimens of oral contraceptives because FSH levels begin to rise after 4 pill-free days.[65] These follicles can reach impressive size; ovarian follicles more than 10 mm in diameter have the potential to continue growth as a dominant follicle.[66] The successful conversion to a dominant follicle marks the "selection" of a follicle destined to ovulate, the process whereby, with rare exception, only a single follicle succeeds.[67, 68] However, follicles that reach sizes consistent with dominant follicles, even preovulatory follicles, are not assured of ovulation in women using oral contraceptives in a compliant fashion. Some stop growing, while others continue to grow but fail to ovulate, almost assuredly because of suppression of the LH surge.[69–71] Nevertheless, evidence indicates that dominant follicles can emerge, secrete preovulatory levels of estradiol, and ovulate, and in this case, contraceptive efficacy requires the other progestational actions.[58, 71, 72]

Clinical Studies with Extended Regimens

The major clinical study with Loestrin 24 Fe was a 6-month, open-label, randomized, active-controlled study in 32 centers in the U.S.[73] Nine hundred and thirty eight patients were randomized to either the 24-day product or the 21-day formulation (1 mg norethindrone acetate/ethinyl estradiol 20 μg) in a 4:1 ratio that yielded 705 women in the treated group and 181 in the comparison group available for analysis. A strength of this study was that it compared the 24-day product to an identical 21-day formulation. The study was not powered to determine a significant difference in efficacy comparing the two products. The Pearl index for the 24-day schedule was 1.82 (1.78 for subjects 35 years old and younger). The cumulative pregnancy rate for 6 months was 0.9%. These numbers are typical for all oral contraceptives.

The number of days with breakthrough bleeding or spotting was comparable in both groups, but the 24-day group demonstrated a steady decline in breakthrough bleeding/spotting days, so that in cycle 6 the mean number of bleeding days was significantly lower in the 24-day group (0.95 vs. 1.63). Among the women in the 24-day group, those who switched from another oral contraceptive had a lower mean number of bleeding days compared to new users, probably reflecting suppression of endometrial growth by the previous use. Each cycle with the 24-day product demonstrated a shorter duration of withdrawal bleeding (bleeding beginning after the last day of active drug intake), achieving statistical significance in the second cycle. Combining breakthrough bleeding and withdrawal bleeding, the total number of days over the entire 6 treatment cycles with bleeding was significantly less in the 24-day group: 18.6 for the 24-day schedule compared with 23.2 for the 21-day regimen.

A reasonable concern with extending the days of active treatment is the resulting increase in overall hormone exposure. The accumulative dosages over 6 cycles, however, were not notably different. The 24-day schedule totaled 144 mg of norethindrone acetate and 2.88 mg ethinyl estradiol, compared with 126 mg norethindrone acetate and 2.52 mg ethinyl estradiol in the 21-day group. There were no demonstrable differences in adverse events.

A 21-day product has been compared with extending the schedule to 23 days, using 75 μg gestodene/20 μg ethinyl estradiol.[74–76] The 23-day regimen produced a greater suppression of ovarian activity as measured by lower estradiol levels and less follicular activity as assessed by ultrasonography; however, the incidence of bleeding and spotting days was similar in the two treatment groups. The 23-day regimen was associated with shorter withdrawal bleeding periods compared with the 21-day schedule.

Ovarian activity was compared in a group of women using 60 μg gestodene/ethinyl estradiol 15 μg for 24 days compared to a group using the same product on the standard 21-day regimen.[77] The women using the 21-day regimen experienced greater follicular activity with larger follicles and higher estradiol levels. Breakthrough bleeding was more prevalent with the 24-day schedule; however, the number of treatment cycles in this small study was not large enough to assess bleeding control. A larger study compared the 24-day regimen of 60 μg gestodene/ethinyl estradiol 15 μg with a 21-day regimen using 150 μg desogestrel/ethinyl estradiol 20 μg and reported a greater incidence of breakthrough bleeding with the 24-day regimen; however the length of bleeding was shorter and the intensity of bleeding was reduced.[78] The 1-year overall incidence of breakthrough bleeding with this 24-day, 15 μg ethinyl estradiol product has been reported to be 19.3%.[79]

In a short study of only three treatment cycles, the 24-day 3 mg drospirenone/ethinyl estradiol 20 μg product was compared to a 21-day regimen of the same formulation.[80] The 24-day schedule was associated with greater follicular suppression and only one ovulation in cycle 3 compared with 4 ovulations in the 21-day regimen when the initial 3 tablets in the third month were substituted with placebos. A study of 12 women using this formulation for either 23 days or 24 days documented greater suppression of FSH, LH, inhibin B, and estradiol during the pill-free interval when compared to women using the same product for 21 days.[65]

Clinical Recommendation

The 24-day regimen of low-dose oral contraceptives achieves its goals. Both bleeding and ovarian activity are reduced. Indeed, the two are related. ***Diminished ovarian follicular activity is responsible for less fluctuation in endogenous estrogen levels, resulting in a more quiescent and stable endometrium.*** Extended (and continuous dosing) regimens compared with the standard 21-day regimen are associated with a decrease in menstrual discomfort, headaches, and bloating.[81–83]

Another clinical advantage of the 24-day regimen is a reduction in the risk of "escape" follicular activity if a patient inadvertently starts a new package 1 or 2 days late. Randomized studies that extended the pill-free interval by 2 or 3 days observed that women taking a 20 μg ethinyl estradiol formulation had a greater increase in follicular activity compared with women using a 35 μg ethinyl estradiol product.[84, 85] Ovarian follicular activity is greater with products containing 20 μg ethinyl estradiol, and the maximal follicular size reached is larger. In one study, a greater proportion of women on a 20 μg product, around 30%, achieved follicular diameters of 15 mm or greater, compared with a 35 μg formulation when the pill-free interval was extended from 7 to 9 days.[85] Once follicles achieve a diameter greater than 10 mm, an increasing percentage go on to ovulate even in the presence of oral contraceptive treatment.[86] The lower-dose formulations produce less suppression of gonadotropin secretion, documented in these studies by higher FSH, LH, and estradiol blood levels among the users of the 20 μg ethinyl estradiol product.

Not only does the 24-day product allow a day or two grace period, but the extended hormone exposure suppresses gonadotropin and follicular activity to a greater degree. Thus, even in patients with good compliance, a greater reduction in follicular activity can reduce the possibility of breakthrough ovulations and contraceptive failure. This would be difficult and expensive to document because it would require a clinical trial with a very large number of patients.

A regimen is available that supplies a package containing the number of pills required for 84 days of daily administration, a reduction of menstrual frequency to 4 per year.[87] This approach includes Seasonale (20 μg ethinyl estradiol and 100 μg levonorgestrel), with 7 placebo pills after 84 active pills, and Seasonique, with 7 pills of 10 μg ethinyl estradiol after 84 active pills. The combination with 7 days of estrogen was a response to the discovery that 84 active pill days are rapidly followed by a rise in FSH with stimulation of follicular growth.[88] A slighter higher rate of breakthrough bleeding with this regimen improves with time, but overall, bleeding is less with Seasonique because of better FSH suppression at the end of the 84-day combination estrogen-progestin period.[89, 90]

Steroid contraception in the currently used low doses has been demonstrated to be very safe for healthy women. Efforts to improve steroid contraception are now focusing on maximizing adherence to treatment and minimizing pregnancies from contraceptive failures. The 24-day regimen offers clinicians and patients the important advantage of reduced bleeding and the possible advantage of greater efficacy because of better compliance as well as a reduction in ovarian activity.

Continuous Dosing

More and more women are embracing the idea that fewer menstrual periods provide a welcome relief from bleeding and menstrual symptoms. Clinicians for years have prescribed unlimited daily oral contraceptives to treat conditions such as endometriosis, bleeding disorders, menstrual seizures, and menstrual migraine headaches, even to avoid bleeding in athletes and busy individuals. Many women do not require the periodic experience of

vaginal bleeding to assure themselves they are not pregnant. And of course, modern society is long past the notion that menstrual bleeding is a cleansing event, a detoxification. It is not necessary for women using oral contraceptives to experience any withdrawal bleeding. Monthly bleeding, periodic bleeding, or no bleeding—this is an individual woman's choice. Any combination oral contraceptive can be used on a daily basis; even the lowest estrogen dose formulations provide excellent bleeding and side effect profiles in a continuous regimen.[82, 91–93]. ***Oral contraceptives with the lowest doses of estrogen should be used for overweight and older women, but progestin-only methods are an even better choice.***

Eliminating a pill-free interval also reduces symptoms associated with menstruation, such as headaches, dysmenorrhea, and bloating.[94–96] As with the extended regimen, continuous dosing provides greater ovarian suppression, reducing the potential for breakthrough and escape ovulations.[97] A further benefit of continuous use is simplification of the pill-taking schedule with the potential of better compliance and a lower failure rate. Continuous dosing can also be achieved with the contraceptive vaginal ring and the contraceptive patch. The return of ovulation and achievement of pregnancy are not delayed after discontinuation of continuous dosing.[97, 98]

Generic Products

Generic products are therapeutically equivalent drugs, containing the same amount of active ingredients in the same concentration and dosage form. These products are less expensive, marketed by pharmaceutical companies after patent expiration of the original drug. Generic oral contraceptives need only meet the test of bioequivalence; studies to demonstrate efficacy, side effects, and safety are not required. Meeting the test of bioequivalence requires demonstration in a small number of subjects that absorption, concentrations, and time curves are comparable to the reference drug. The generic product will be approved if the bioequivalence testing ranges from 80 to 125% of the values for the reference drug (differences no greater than 20% lower or 25% higher). Approved, patented products must not vary more than ±10%; therefore a generic oral contraceptive could contain only 70% of the standard dose. In the lowest-dose oral contraceptives, this could impair efficacy. However, we should hasten to point out that there has been no evidence or even anecdotal suggestions that generic oral contraceptives have reduced efficacy or cause more side effects such as breakthrough bleeding. Patients should be forewarned that generic products differ in shape, packaging, and color.

Off Label Uses of Steroid Contraception

Steroid contraception is often used for noncontraceptive purposes. The list is long, including treatment of acne, dysmenorrhea, heavy or irregular vaginal bleeding, menses-associated mood changes, the polycystic ovary syndrome, and endometriosis. For most of the oral contraceptive's 50-year history, all of these have been "off label" applications, but recently pharmaceutical companies have conducted trials to obtain label "indications" to use in advertising directed to both clinicians and consumers. In order to acquire such an indication, the company simply has to demonstrate that their formulation is better than a placebo at, for example, improving acne or relieving the symptoms of premenstrual dysphoric disorder. Because these trials usually compare a product to a placebo or just to another contraceptive formulation, the studies do not reveal whether the product receiving approval for an "indication" is really better than others. Prices and formularies restrict patient access to the full range of oral contraceptives[99]; therefore, clinicians must make judgements by comparing findings from unrelated studies and experience to decide which pill to use for a specific purpose in an individual patient. In most cases, as we will emphasize, it is unlikely that there are major differences among similar products.

Potency

For many years, clinicians, scientists, medical writers, and even the pharmaceutical industry attempted to assign potency values to the various progestational components of oral contraceptives. An accurate assessment, however, has been difficult to achieve for many reasons. Progestins act on numerous target organs (e.g., the uterus, the mammary glands, and the liver), and potency varies depending on the target organ and end point being studied. In the past, animal assays, such as the Clauberg test (endometrial change in the rabbit) and the rat ventral prostate assay, were used to determine progestin potency. Although these were considered acceptable methods at the time, a better understanding of steroid hormone action and metabolism and a recognition that animal and human responses differ have led to greater reliance on data collected from human studies.

Historically, this has been a confusing issue because publications and experts used potency ranking to provide clinical advice. There is absolutely no need for confusion. Oral contraceptive progestin potency is no longer a consideration when it comes to prescribing oral contraception, because the potency of the various progestins has been accounted for by appropriate adjustments of dose. In other words, the biologic effect (in this case the clinical effect) of the various progestational components in current low-dose oral contraceptives is approximately the same. The potency of a drug does not determine its efficacy or safety, only the amount of a drug required to achieve an effect.

Clinical advice based on potency ranking is an artificial exercise that has not stood the test of time. There is no clinical evidence that a particular progestin is better or worse in terms of particular side effects or clinical responses. Thus oral contraceptives should be judged by their clinical characteristics: efficacy, side effects, risks, and benefits. Our progress in lowering the doses of the steroids contained in oral contraceptives has yielded products with little serious differences.

Mechanism of Action

The combination pill, consisting of estrogen and progestin components, prevents ovulation by inhibiting gonadotropin secretion via an effect on both pituitary and hypothalamic centers. The progestational agent in the pill primarily suppresses luteinizing hormone (LH) secretion (and thus prevents ovulation), while the estrogenic agent suppresses follicle-stimulating hormone (FSH) secretion (and thus prevents the emergence of a dominant follicle). Therefore, the estrogenic component significantly contributes to the contraceptive efficacy. However, even if follicular growth and development were not sufficiently inhibited, the progestational component would prevent the surge-like release of LH necessary for ovulation.

The estrogen in the pill serves two other purposes. It provides stability to the endometrium so that irregular shedding and unwanted breakthrough bleeding can be minimized; and the presence of estrogen is required to potentiate the action of the progestational agents. The latter function of estrogen has allowed reduction of the progestational dose in the pill. The mechanism for this action is probably estrogen's effect in increasing the concentration of intracellular progestational receptors. Therefore, a minimal pharmacologic level of estrogen is necessary to maintain the efficacy of the combination pill.

Because the effect of a progestational agent will always take precedence over estrogen (unless the dose of estrogen is increased many, many-fold), the endometrium, cervical mucus, and perhaps tubal function reflect progestational stimulation. The progestin in the combination pill produces an endometrium that is not receptive to ovum implantation, a decidualized bed with exhausted and atrophied glands. The cervical mucus becomes thick and impervious to sperm transport. It is possible that progestational influences on secretion

and peristalsis within the fallopian tubes provide additional contraceptive effects. Even if there is some ovarian follicular activity (especially with the lowest dose products), these actions serve to ensure good contraceptive efficacy.[100]

Efficacy

In view of the multiple actions of oral contraceptives, it is hard to understand how the omission of a pill or two can result in a pregnancy. Indeed, careful review of failures suggests that pregnancies usually occur because initiation of the next cycle is delayed allowing

Failure Rates During the First Year of Use, United States[103–105]

	Percent of women with pregnancy	
Method	*Lowest Expected*	*Typical*
No method	85%	85%
Combination Pill	0.3%	8.7%
Progestin only	0.5%	3.0%
IUDs:		
Levonorgestrel IUD	0.1%	0.1%
Copper T 380A	0.6%	1.0%
Implant	0.05%	1.0%
Injectable		
3-month	0.3	0.3%
1-month	0.05	3.0%
Patch	0.3	8.0%
Vaginal ring	0.3	8.0%
Female sterilization	0.5%	0.7%
Male sterilization	0.1%	0.2%
Spermicides	18.0%	29.0%
Periodic abstinence		25.3%
Calendar	9.0%	
Ovulation method	3.0%	
Symptothermal	2.0%	
Post-ovulation	1.0%	
Withdrawal	4.0%	18.4%
Cervical cap		
Parous women	26.0%	32.0%
Nulliparous women	9.0%	16.0%
Sponge:		
Parous women	20.0%	32.0%
Nulliparous women	9.0%	16.0%
Diaphragm and spermicides	6.0%	16.0%
Condom		
Male	2.0%	17.4%
Female	5.0%	27.0%

escape from ovarian suppression. Strict adherence to 7 pill-free days is critical in order to obtain reliable, effective contraception. For this reason, the 28-day pill package, incorporating 7 pills that do not contain steroids, is a very useful aid to ensure adherence to the necessary schedule. Even better, the use of extended regimens or continuous dosing offers the potential to minimize, if not eliminate, pill failures.

The most prevalent problems that can be identified associated with apparent oral contraceptive failures are vomiting and diarrhea.[101, 102] ***Even if no pills have been missed, patients should be instructed to use a backup method for at least 7 days after an episode of gastroenteritis. An alternative is to place the pill in the vagina during the illness (discussed later).***

The contraceptive effectiveness of the new progestin oral contraceptives, multiphasic formulations, and lowest estrogen dose products are unequivocally comparable with older low-dose (less than 50 μg estrogen) and higher dose monophasic combination birth control pills.[100]

Metabolic Effects of Oral Contraception

Cardiovascular Disease

The Coagulation System

Thrombosis can be divided into two major categories, venous thromboembolism and arterial thrombosis. Venous thromboembolism includes both deep vein thrombosis and pulmonary embolism. Arterial thrombosis includes acute myocardial infarction and stroke.

The goal of the clotting mechanism is to produce thrombin, which converts fibrinogen to a fibrin clot. Thrombin is generated from prothrombin by factor Xa in the presence of factor V, calcium, and phospholipids. The vitamin K-dependent factors include factors VII, IX, and X, as well as prothrombin. Antithrombin III is one of the body's natural anticoagulants, an irreversible inhibitor of thrombin and factors IXa, Xa, and XIa. Protein C and protein S are two other major inhibitors of coagulation and are also vitamin K-dependent. Protein C, and its helper, protein S, inhibit clotting at the level of factors V and VIII. Tissue plasminogen activator (t-PA) is produced by endothelial cells and released when a clot forms. Both t-PA and plasminogen bind to the fibrin clot. The t-PA converts the plasminogen to plasmin which lyses the clot by degrading the fibrin. Deficiencies of antithrombin III, protein C, and protein S are inherited in an autosomal dominant pattern, accounting for 10–15% of familial thrombosis. The most common inherited causes of venous thromboembolism are the factor V Leiden mutation, followed distantly by a mutation in the prothrombin gene.[106]

Coagulation Factors:
- **Factors that favor clotting when increased**
  - **Fibrinogen**
  - **Factors VII, VIII, X**
- **Factors that favor clotting when decreased**
  - **Antithrombin III**
  - **Protein C**
  - **Protein S**

Fibrinolysis Factors:
- **Factors that favor clotting when increased**
  - **Plasminogen activator inhibitor-1 (PAI-1)**
- **Factors that favor clotting when decreased**
  - **Antiplasmin**

An inherited resistance to activated protein C has been identified as the basis for about 50% of cases of familial venous thrombosis, due in almost all cases to a gene alteration recognized as the factor V Leiden mutation.[107, 108] The factor V Leiden mutation is found in approximately 30% of individuals who develop venous thromboembolism.[109] Activated protein C inhibits coagulation by degrading factors V and VIII. One of the three cleavage sites in factor V is the precise site of a mutation known as the factor V Leiden mutation that substitutes glutamine instead of arginine at this site (adenine for guanine at nucleotide 1691 in the gene).[109] This mutation makes factor V resistant to degradation and activation in fibrinolysis. The entire clotting cascade is then resistant to the actions of the protein C system.

Heterozygotes for the factor V Leiden mutation have an 8-fold increased risk of venous thrombosis, and homozygotes have an 80-fold increased risk, and this risk is further enhanced by oral contraceptive use. The highest prevalence (3–4% of the general population) of factor V Leiden is found in Europeans, and its occurrence in populations not of European descent is very rare, perhaps explaining the low frequency of thromboembolic disease in Africa, Asia, and in Native Americans.[110] The mutation is believed to have arisen in a single ancestor approximately 21,000 to 34,000 years ago.[111] It has been suggested that this was a useful adaptation in heterozygotes in response to life-threatening bleeding, such as with childbirth.

The next most common inherited disorder after the factor V Leiden mutation is a mutation, a guanine to adenine change, in the gene encoding prothrombin.[106, 112] The prevalence of this abnormality in the white population is estimated to range from 0.7 to 4%.[113] Oral contraceptive use has been reported to markedly increase the risk of venous thrombosis in carriers of the prothrombin mutation.[114] Perhaps other unidentified disorders make a contribution because an increased risk of venous thrombosis with oral contraceptives has been reported in women with elevated prothrombin levels despite an absence of the prothrombin gene mutation.[115]

The administration of pharmacologic amounts of estrogen as in high-dose oral contraceptives causes an increase in the production of clotting factors such as factor V, factor VIII, factor X, and fibrinogen.[116] The progestin component also influences the clotting factor responses.[117] Some studies of the blood coagulation system have concluded that both monophasic and multiphasic low-dose oral contraceptives have no significant clinical impact on the coagulation system. Slight increases in thrombin formation are offset by increased fibrinolytic activity.[118, 119] Other studies of formulations containing 30 and 35 μg of ethinyl estradiol indicate an increase in clotting factors associated with an increase in platelet activity.[120] However, these changes are essentially all within normal ranges and their clinical significance is unknown.[117]

Smoking produces a shift to hypercoagulability.[121] A 20 μg estrogen formulation has been reported to have no effect on clotting parameters, even in smokers.[121, 122] One study comparing a 20 μg product with a 30 μg product found similar mild procoagulant and fibrinolytic activity, although there was a trend toward increased fibrinolytic activity with the lower dose.[123] These mixed reports make it essential to base clinical decisions on the epidemiologic studies of clinical events.

There is no evidence of an increase in risk of cardiovascular disease among past users of oral contraception.[124–126] In the Nurses' Health Study, the Royal College of General Practitioners' Study, and the Oxford Family Planning Association Study, long-term past use of oral contraceptives was not associated with an increase in overall mortality.[127–129]

Part of the concern for a possible lingering effect of oral contraceptive use was based on a presumed adverse impact on the atherosclerotic process, which would then be added to the effect of aging and, thus, would be manifested later in life. Instead, the findings have been consistent with the contention that cardiovascular disease due to oral contraception is secondary to acute effects, specifically estrogen-induced thrombosis, a dose-related event.

Venous Thromboembolism

Older epidemiologic evaluations of oral contraceptives and vascular disease indicated that venous thrombosis was an effect of estrogen, limited to current users, with a disappearance of the risk by 3 months after discontinuation.[130, 131] Thromboembolic disease was believed to be a consequence of the pharmacologic administration of estrogen, and the level of risk was believed to be related to the estrogen dose.[132–134] Smoking was documented to produce an additive increase in the risk of arterial thrombosis, [135–137] but had no effect on the risk of venous thromboembolism.[138, 139]

Is there still a risk of venous thromboembolism with the current low-dose (less than 50 μg ethinyl estradiol) formulations of oral contraceptives? In the first years of oral contraception, the available products, containing 80 and 100 μg ethinyl estradiol (an extremely high dose), were associated with a 6-fold increased risk of venous thrombosis.[140] Because of the increased risks for venous thrombosis, myocardial infarction, and stroke, lower dose formulations (less than 50 μg estrogen) came to dominate the market, and clinicians became more careful in their screening of patients and prescribing of oral contraception. Two forces, therefore, were at work simultaneously to bring greater safety to women utilizing oral contraception: (1) the use of lower dose formulations, and (2) the avoidance of oral contraception by high-risk patients. Because of these two forces, the Puget Sound study in the United States documented a reduction in venous thrombosis risk to 2-fold.[141] The new studies also reflect the importance of these two forces, but they still indicate an increased risk.

The World Health Organization (WHO) Collaborative Study of Cardiovascular Disease and Steroid Hormone Contraception was a hospital-based, case-control study with subjects collected from 21 centers in 17 countries in Africa, Asia, Europe, and Latin America.[142] As part of this study, the risk of *idiopathic* venous thromboembolism associated with a formulation containing 30 μg ethinyl estradiol and levonorgestrel (doses ranging from 125 to 250 μg) was compared with the risk with preparations containing 20 or 30 μg ethinyl estradiol and either desogestrel or gestodene (data from 10 centers in 9 countries).[143] The users of the levonorgestrel formulations had an increased odds ratio of 3.5 compared with nonusers. Current users of a desogestrel product had an increased risk of 9.1 compared with nonusers, and with gestodene, the odds ratio was also 9.1. Thus, the increased risk for desogestrel and gestodene was 2.6 times that of levonorgestrel, when adjusted for body weight and height.

The Transnational Study on Oral Contraceptives and the Health of Young Women analyzed 471 cases of deep vein thrombosis and/or venous thromboembolism from the United Kingdom and Germany.[144] Second-generation oral contraceptives were defined as products containing 35 μg or less of ethinyl estradiol and a progestin other than desogestrel or gestodene. Comparing users of second-generation products to nonusers, the odds ratio was 3.2. Comparing users of desogestrel and gestodene products to users of second-generation oral contraceptives, the risk of venous thromboembolism was 1.5-fold greater.

A third major study was from Boston University, but the data were derived from the General Practice Research Database, a computerized system involving the general practitioners in the U.K.[145] Using this cohort, the death rate from pulmonary embolism, stroke, and acute myocardial infarction was calculated in the users of levonorgestrel, desogestrel, and gestodene low-dose oral contraceptives. Over a 3-year period, they collected a total of 15 unexpected idiopathic cardiovascular deaths in users of these products, a *nonsignificant change*, and no

difference in the risk comparing desogestrel and gestodene with levonorgestrel. The risk estimates for venous thromboembolism (adjusted for smoking and body size) were about 2 times greater for desogestrel and for gestodene, compared with levonorgestrel uses. In an updated analysis from this same group and database, the findings were unchanged, except that smoking was found to be a risk factor for venous thromboembolism.[146] An American case-control study concluded that norgestimate- and levonorgestrel-containing oral contraceptives had similar risks for venous thromboembolism, but there was a small increase in risk associated with desogestrel.[147]

Similar results were reported when women with deep vein thrombosis in the Leiden Thrombophilia Study in the Netherlands were reanalyzed for their use of oral contraceptives.[148] As expected, the risk of deep vein thrombosis was markedly higher in women who were carriers of the factor V Leiden mutation and in women with a family history of thrombosis.

In Denmark, Lidegaard and colleagues performed a hospital-based, case-control study of women with confirmed diagnoses of venous thromboembolism in 1994 and 1995 (in Denmark, all women with this diagnosis are hospitalized, and therefore, very few, if any, cases were missed).[149] A 2-fold increased risk of venous thromboembolism was found in current users of oral contraceptives, regardless of estrogen doses ranging from 20 to 50 μg. The increased risk was concentrated in the first year of use. *Because there were more short-term users of the new progestins and more long-term users of the older progestins, adjustment for duration of use resulted in no significant differences between the different types of progestins*. Those factors associated with an increased risk of thromboembolism included coagulation disorders, treated hypertension during pregnancy, family history of venous thromboembolism, and an increasing body mass index. Notably, conditions not associated with an increased risk of venous thromboembolism included smoking, migraine, diabetes, hyperlipidemia, parity, or age at first birth. There was still insufficient strength in this study to establish the absence or presence of a dose-response relationship comparing the 20 μg estrogen dose to higher doses; however, a 5-year update reported the following useful information:[150]

- The risk of venous thrombosis associated with current use of oral contraceptives declined with increasing duration of use.
- The risk was slightly greater with desogestrel or gestodene.
- Smoking more than 10 cigarettes per day increased the risk.
- Oral contraceptives with 20 μg estrogen had a lower risk than products with 30–40 μg.
- Progestin-only contraceptive products did not increase the risk.

Case-control studies using cases of venous thromboembolism derived from the computer records of general practices in the U.K. concluded that the increased risk associated with oral contraceptives was the same for all types, and that the pattern of risk with specific oral contraceptives suggested confounding because of "preferential prescribing" (defined later).[151, 152] *In these studies, matching cases and controls by year of birth eliminated differences between different types of oral contraceptives*. A similar analysis based on 42 cases from a German database again found no difference between new progestin and older progestin oral contraceptives.[153] Thus, in these two studies, more precise adjustments for age eliminated a confounding bias. An assessment of the incidence of venous thromboembolism in the U.K. before and after the decline in third-generation progestin use could detect no impact on the statistics (neither an increase nor a decrease).[154]

A reanalysis of the Transnational Case-Control Study considered the duration and patterns of oral contraceptive use.[155, 156] This reanalysis focused on first-time users of second- and third-generation oral contraceptives. *Statistical analysis with adjustment for duration of use in 105 cases who were first-time users could find no differences between second- and*

third-generation products. A similar reanalysis of the U.K. General Practice Database could demonstrate no difference between different oral contraceptive formulations.[157]

A case-control study in Germany assessed the outcome when the cases were restricted to hospitalized patients compared to results when all cases, both in-hospital and out-of-hospital, were considered.[158] The conclusion indicated that hospital-based studies overestimated the risk of venous thromboembolism, and that there was no difference comparing progestins when all cases were included.

Former users discontinue oral contraceptives for a variety of reasons, and often are switched to what clinicians perceive to be "safer" products, a practice called "*preferential prescribing.*"[159–161] Individuals who do well with a product tend to remain with that product. Thus, at any one point in time, individuals on an older product will be relatively healthy and free of side effects—the "*healthy user effect.*" This is also called *attrition of susceptibles* because higher risk individuals with problems are gradually eliminated from the group.[162] *Comparing users of older and newer products, therefore, can involve disparate cohorts of individuals*.

Because desogestrel- and gestodene-containing products were marketed as less androgenic and therefore "better" (a marketing claim not substantiated by epidemiologic studies), clinicians chose to provide these products to higher risk patients and older women.[159, 160] In addition, clinicians switched patients perceived to be at greater risk for thrombosis from older oral contraceptives to the newer formulations with desogestrel and gestodene. Furthermore, these products were prescribed more often to young women who were starting oral contraception for the first time (these young women will not have experienced the test of pregnancy or previous oral contraceptive use to help identify those who have an inherited predisposition to venous thrombosis). These changing practice patterns exert different effects over the lifetime of a product, and analytical adjustments are extremely difficult.

The initial studies were impressive in their agreement. All indicated increased relative risks associated with desogestrel and gestodene compared with levonorgestrel. Nevertheless, all of the early studies, somewhat similar in design, were influenced by the same unrecognized biases. *Persistent errors will produce consistent conclusions*.

Forty cases of venous thrombosis in drospirenone (Yasmin) users (two of which were fatal) were reported in Europe in 2002.[163] The Dutch College of General Practitioners issued a statement encouraging clinicians not to prescribe Yasmin. However, this is the similar story we experienced with "third-generation" progestins, only to learn that preferential prescribing and the healthy user effect probably biased the early case-control studies. In postmarketing surveillance, only one case of venous thrombosis occurred in a million cycles of Yasmin compared with five among users of other oral contraceptives.[163] In a subsequent monitoring study, the incidence of venous thromboembolism in new users was comparable to that seen with other low-dose oral contraceptives.[164] The European Active Surveillance Study (EURAS) was a large cohort study that enrolled *only new users* of oral contraceptives containing a variety of progestins, including drospirenone and levonorgestrel.[165] The incidence of cardiovascular events was similar for all progestins. An American cohort study also focused on new users of oral contraceptives, and thromboembolism occurred at a similar low rate comparing drospirenone users with other oral contraceptives.[166]

The Danish investigators continued their interest in hormonal contraception and venous thrombosis, performing a national cohort study using the reliable Danish national registries of events from 1995 to 2005.[167] As in the earlier Danish case-control study,[150] the risk of venous thrombosis in current users of oral contraceptives decreased with duration of use and with the dose of estrogen, and was slightly higher with products containing desogestrel, gestodene, drospirenone, and cyproterone. Did this study escape the problems of preferential prescribing and the healthy user effect (attrition of susceptibles)? The incidence of thrombotic events in the comparator group (levonorgestrel users) was lower than that

reported in other studies, suggesting that this group did demonstrate a healthy user effect. The study was not limited to new users, a requirement in order to avoid the confounding due to attrition of susceptibles. The authors argued that preferential prescribing was not prevalent in Denmark after 1995 (but they offered no proof of this contention), and they further argued that the fact that the use of other medications was similar comparing levonorgestrel and drospirenone suggested a similar level of health and a lack of preferential prescribing. This study was unable to control for body mass index (BMI) or family history of thrombosis, two important markers for women at high risk of venous thrombosis. Preferential prescribing remains a possible confounder in the Danish study; however, the problem of a healthy user effect is even more likely.

A case-control study from the Netherlands also reported higher risks of venous thrombosis in users of desogestrel, gestodene, drospirenone, and cyproterone compared with levonorgestrel users.[168] The authors supported their results by citing findings from their own institution that users of oral contraceptives containing drospirenone and cyproterone have lower levels of free protein S and free tissue factor pathway inhibitor associated with greater resistance to activated protein C compared with levonorgestrel users.[169] The relative risks in this study were surprisingly high, higher than all other reports involving low-dose oral contraceptives. Once again, the healthy user effect was a likely confounder in that the study was not limited to new users. The authors claimed to compensate for the attrition of susceptibles by analyzing only short-term users. Even though the validity of this approach can be debated, the results indicated nonsignificant increased risks with drospirenone and cyproterone compared with levonorgestrel, and any conclusion was limited by a small number of short-term users. In this study, the risk associated with products containing 20 μg ethinyl estradiol was not increased.

An international study is underway, the International Active Surveillance study of women taking Oral Contraceptives (INAS-OC), designed to record cardiovascular events in a cohort of more than 80,000 oral contraceptive users.[170] Two-year and 5-year follow-up reports are anticipated.

The risk of venous thrombosis associated with modern oral estrogen-progestin contraceptives is increased about 2-fold, but manifested primarily in the first years of use and concentrated in overweight women.[165, 171, 172] The risk, which increases with increasing body weight and age, is influenced in a major way by the estrogen dose, and the difference among progestin products is small, either real but not meaningful clinically or a reflection of biases and confounders. The impact of smoking on the risk of venous thrombosis is less than that on the risk of arterial thrombosis, but smoking, especially heavy smoking, may act synergistically with oral contraceptives.[173]

The risk of venous thromboembolism in the general population is now considered to be higher than previously estimated because of the prevalence of modern diagnostic methods. The new studies also indicate that the risk associated with low-dose oral contraceptives is lower than previously reported, and more prevalent in high-risk individuals

Relative Risk and Actual Incidence of Venous Thromboembolism[165, 171, 172]

Population	*Relative Risk*	*Incidence*
Young women—general population	1	5–10/10,000/year
Pregnant women	12	60–120
High-dose oral contraceptives	6–10	30–100
Low-dose oral contraceptives	2	10–20
Leiden mutation carrier	6–8	30–80
Leiden carrier and oral contraceptives	10–15	50–100
Leiden mutation—homozygous	80	400–800

(obesity and inherited or acquired thrombophilias). Oral contraceptives with the lowest doses of estrogen should be used for overweight and older women, but progestin-only methods are an even better choice.

Venous Thromboembolism and Thrombophilias

An inherited resistance to activated protein C, the factor V Leiden mutation, may account for a significant portion of the patients who experience venous thrombosis while taking oral contraceptives.

The factor V Leiden mutation is the most common inherited coagulation problem, transmitted in an autosomal-dominant fashion.[107, 174] Heterozygotes have a 6- to 8-fold increased risk of venous thromboembolism, and homozygotes an 80-fold increased risk. Oral contraceptive users who have this mutation have been reported to have a 30-fold increased risk of venous thrombosis.[175, 176] Some have effectively argued, however, that this increase has been overestimated, and it is closer to 10–15-fold.[177] The risk of developing venous thrombosis is greatest in the initial months of use, and it has been suggested that venous thrombosis occurring in the first month of exposure should make the clinician suspect the presence of a clotting disorder.[178]

Should screening for the factor V Leiden mutation (or for other inherited clotting disorders) be routine prior to prescribing contraceptives? The carrier frequencies of the Leiden mutation in the American population (the percentages are similar in men and women) are as follows:[179]

Caucasian Americans	5.27%
Hispanic Americans	2.21%
Native Americans	1.25%
Black Americans	1.23%
Asian Americans	0.45%

These estimates are consistent with European assessments, indicating that this is a trait carried in people of European origin. About 1 in 5,000 individuals is homozygous for the Leiden mutation. In the United States, of the approximately 12 million women currently using oral contraceptives, about 540,000 are likely to carry the factor V Leiden mutation. However, because the incidence rate of venous thromboembolism is so low (5–10 per 10,000 young women per year), the number of women required to be screened to prevent one event is prohibitively large. The prevalence of all deficiencies is only about 0.5% in the asymptomatic population, and only one-third of patients at risk are detected by the present tests.[180]

Furthermore, because only a small number of women even with the Leiden mutation (less than 1 in 1,000) have a clinical event (99.85% of the individuals who test positive will NOT have a clinical event!), the finding of a positive screening test, especially considering the high rate of false-positive tests, would be a barrier to the use of oral contraceptives, and a subsequent increase in unwanted pregnancies would likely follow. Pregnancy, of course, has a much greater risk of venous thromboembolism than oral contraceptives. ***Most experts believe that screening for inherited disorders should be pursued only in women with a previous episode of venous thromboembolism or a close positive family history (parent or sibling) of venous thrombosis.***

The second most prevalent inherited thrombophilia is the prothrombin gene *20210A* mutation. A combination of the prothrombin gene mutation and the Leiden mutation is found

in about 2% of venous thromboembolism cases.[181] Less frequent are genetic defects in coagulation inhibitors (antithrombin, protein C, and protein S), but these defects carry a substantial increase in risk.

Acquired thrombophilias include the presence of antiphospholipid antibodies (lupus anticoagulant and anticardiolipin) usually associated with autoimmune diseases.

The inherited and acquired thrombophilias predispose to venous thromboembolism in a synergistic manner with estrogen-containing contraceptives. However, the actual incidence of events is low, and identification of a thrombophilia does not predict a clinical event.

Arterial Thrombosis

Because the incidence of cerebral thrombotic attacks (thrombotic strokes and transient ischemic attacks) among young women is higher than venous thromboembolism and myocardial infarction, and death and disability are more likely, cerebral arterial thrombosis is the most important possible side effect. Nevertheless, the incidence is low, and a small increase in the very low incidence of stroke in young women carries with it little increase in absolute risk. Because the incidence of cerebral thrombotic attacks is higher in women over age 40, we should do our best, as the following discussion will indicate, to make sure estrogen-progestin contraceptive users over age 40 are in good health and without significant risk factors for cardiovascular disease (especially hypertension, migraine with aura, and smoking).

Although it has been difficult to establish arterial thrombosis dose-response relationships with estrogen because these events are so rare, modern, lower estrogen doses are related to the risk of myocardial infarction and thrombotic strokes in case-control studies.[182, 183] Thus, a rationale for advocating low-dose estrogen oral contraceptives continues to be valid.

Arterial Thrombosis—Myocardial Infarction

A population-based, case-control study analyzed 187 cases of myocardial infarction in users of low-dose oral contraceptives in the Kaiser Permanente Medical Care Program.[184] *There was no statistically significant increase in the odds ratio for myocardial infarction in current oral contraceptive users compared with past or never users.*

In the Transnational case-control study of myocardial infarctions collected from 16 centers in Austria, France, Germany, Switzerland, and the U.K. cigarette smoking carried a higher risk for myocardial infarction than oral contraceptives, and *nonsmoking users of oral contraceptives had no evidence of an increased risk.*[185, 186] In addition, there was an indication that patient screening is important in minimizing the impact of hypertension on the risk of myocardial infarction. Similar results were reported in a case-control study based on subjects in England, Scotland, and Wales, and another in America.[187, 188]

In the WHO multicenter study, there were 368 cases of acute myocardial infarction.[189] Factors associated with an increased risk of myocardial infarction included smoking, a history of hypertension (including hypertension in pregnancy), diabetes, rheumatic heart disease, abnormal blood lipids, and a family history of stroke or myocardial infarction. Duration of use and past use of oral contraceptives did not affect risk. Although there was about a 5-fold overall increased odds ratio of myocardial infarction in current users of oral contraceptives, essentially all cases occurred in women with cardiovascular risk factors. There was no apparent effect of increasing age on risk; however, there were only 12 cases among oral contraceptives users less than 35 years old. There was no apparent relationship

with estrogen dose, and there was no apparent influence of type or dose of progestin, but the rare occurrence of this condition produced such small numbers that there was insufficient statistical power to accurately assess the effects of progestin type, and estrogen and progestin doses. *The conclusion of this study was that the risk of myocardial infarction in women who use oral contraceptives is increased only in smokers.*

In a Danish case-control study of acute myocardial infarction in young women, a statistically significant increase in risk was noted only in current users of 50 µg ethinyl estradiol.[183] There was a progressive increase in risk with the number of cigarettes smoked, (accounting for 80% of the acute myocardial infarctions in young women), increasing body mass index, treated hypertension, treated hypertension in pregnancy, diabetes mellitus, hyperlipidemia, frequent migraine, and family history of myocardial infarction. However, only family history of myocardial infarction and smoking affected the risk associated with oral contraceptives; no influence on oral contraceptive risk was apparent with diabetes, hypertension, and heart disease. No differences could be demonstrated according to type of progestin.

A case-control study from the Netherlands found that the risk of myocardial infarction was highest among users of oral contraceptives who smoked, had diabetes mellitus, or who were hypercholesterolemic.[190] The risk of myocardial infarction was not affected by the presence of the factor V Leiden mutation or the prothrombin gene mutation.

The Women's Lifestyle and Health Cohort Study is a prospective cohort study of 106,841 Norwegian and Swedish women, started in 1991, specifically designed to assess the long-term health effects of hormonal contraceptives.[191] Low-dose (less than 50 µg ethinyl estradiol) oral contraceptives were not associated with an increased risk of myocardial infarction. All previous cohort studies date back to oral contraceptive use with higher doses of estrogen in the 1970s and 1980s. For example, in the report from the Nurses' Health Study in 1988, an increased risk of myocardial infarction was found in current users.[192]

Case-control studies of low-dose estrogen oral contraceptives have concluded that an increased risk of arterial disease occurs only in women who have hypertension or are smokers.[124, 184, 185, 189, 190] The cohort studies don't help us with this important issue because the numbers are too small for definitive analyses of subgroups. Nevertheless, British and Finnish cohorts were reported to have increased risks of developing myocardial infarction in oral contraceptive users who smoked.[193, 194]

Incidence of Myocardial Infarction in Reproductive Age Women[189]

Overall Incidence[195]	*5/100,000/year*
Women Less Than Age 35	
Nonsmokers	4/100,000/year
Nonsmokers & OCs	4/100,000/year
Smokers	8/100,000/year
Smokers & OCs	43/100,000/year
Women 35-Years-Old and Older	
Nonsmokers	10/100,000/year
Nonsmokers & OCs	40/100,000/year
Smokers	88/100,000/year
Smokers & OCs	485/100,000/year

NOTE: The above incidences are estimates based on oral contraceptive use paired with cardiovascular risk factors prevalent in the general population. Effective screening would produce smaller numbers. The increased risks in the smokers and OC groups reflect the impact of undetected cardiovascular risk factors, especially hypertension.

Arterial Thrombosis—Stroke

Older case-control and cohort studies indicated an increased risk of cerebral thrombosis among current users of high-dose oral contraceptives.[196–198] However, thrombotic stroke did not appear to be increased in healthy, nonsmoking women with the use of oral contraceptives containing less than 50 μg ethinyl estradiol.[197, 198] A case-control study of all 794 women in Denmark who suffered a cerebral thromboembolic attack during 1985–1989 concluded that there was an almost 2-fold increased relative risk associated with oral contraceptives containing 30–40 μg estrogen, and the risk was significantly influenced by both smoking and the dose of estrogen in additive (not synergistic) fashion.[137] A case-control analysis of data collected by the Royal College of General Practitioners' Oral Contraception Study concluded that current users were at increased risk of stroke (with a persisting effect in former users); however, this outcome was limited mainly to smokers and to formulations with 50 μg or more of estrogen.[198]

A population-based, case-control study of 408 strokes from the California Kaiser Permanente Medical Care Program found no increase in risk for either ischemic stroke or hemorrhagic stroke.[199] The identifiable risk factors for ischemic stroke were smoking, hypertension, diabetes, elevated body weight, and low socioeconomic status. The risk factors for hemorrhagic stroke were the same plus greater body mass and heavy use of alcohol. *Current users of low-dose oral contraceptives did not have an increased risk of ischemic or hemorrhagic stroke compared with former users and with never users*. There was no evidence for an adverse effect of increasing age or for smoking (for hemorrhagic stroke, there was a suggestion of a positive interaction between current oral contraceptive use and smoking, but the numbers were small, and the result was not statistically significant).

The Transnational study analyzed their data for ischemic stroke in a case-control study of 220 ischemic strokes in the U.K., Germany, France, Switzerland, and Austria.[200] Overall, there was a 3-fold increase in the risk of ischemic stroke associated with the use of oral contraceptives, with higher risks observed in smokers (more than 10 cigarettes per day), in women with hypertension, and in users of higher dose estrogen products. No differences were observed comparing second- and third-generation progestins. A Dutch case-control study also found no differences comparing second and third generation progestins.[201] A case-control study from the state of Washington concluded that there is no increased risk of stroke in current users of low-dose oral contraceptives.[202] A pooled analysis of the data from California and Washington concluded that low-dose oral contraceptives are not associated with an increase in the risk of stroke.[203]

The World Health Organization data on stroke come from the same collaborative study that yielded the publications on venous thromboembolism. The results with stroke were published as two separate reports, one on ischemic stroke and the other on hemorrhagic stroke.[204, 205]

This hospital-based, case-control study from 21 centers in 17 countries accumulated 697 cases of ischemic stroke, 141 from Europe and 556 from developing countries.[204] The overall odds ratio for ischemic stroke indicated about a 3-fold increased risk. In Europe, however, the risk was statistically significant only for higher-dose products, and *NOT* statistically significant for products with less than 50 μg ethinyl estradiol. In developing countries, there was no difference in risk with low-dose and higher dose oral contraceptives. This is believed to be due to the strong influence of hypertension. In Europe, it was uncommon for women with a history of hypertension to be using oral contraceptives; however, this was not the case in developing countries. Duration of use and type of progestin had no impact, and past users did not have an increased risk, but smoking 10 or more cigarettes daily exerted a synergistic effect with oral contraceptives, increasing the risk

of ischemic stroke, approximating the effect of hypertension and oral contraceptives. The risk was greater in women 35 years and older; however, this, too, was believed to be due to an effect of hypertension. *Thus, the conclusion of this study was that the risk of ischemic stroke is extremely low, concentrated in those who use higher dose products, smoke, or have hypertension*.

In the WHO study on hemorrhagic stroke, there were 1,068 cases.[205] Current use of oral contraceptives was associated with a slightly increased risk of hemorrhagic stroke only in developing countries, not in Europe. This again probably reflects the presence of hypertension, because the greatest increased risk (about 10- to 15-fold) was identified in current users of oral contraceptives who had a history of hypertension. Current cigarette smoking also increased the risk in oral contraceptive users, but not as dramatically as hypertension. For hemorrhagic stroke, the dose of estrogen had no effect on risk, and neither did duration of use or type of progestin. *This study concluded that the risk of hemorrhagic stroke due to oral contraceptives is increased only slightly in older women, probably occurring only in women with risk factors such as hypertension*.

A second Danish case-control study included thrombotic strokes and transitory cerebral ischemic attacks analyzed together as cerebral thromboembolic attacks.[182] In this study, the 219 cases during 1994 and 1995 included 146 cases of cerebral infarction and 73 cases of transient ischemic attacks. There was a dose-response relationship with estrogen in the dose ranges of 20, 30–40, and 50 μg ethinyl estradiol, although the number of 20 μg users (5 cases, 22 controls) was not sufficient to establish a lower risk at this lower dose. This analysis claimed a reduced risk associated with desogestrel and gestodene; however, the odds ratio did not achieve statistical significance. An updated 5-year report of the Danish case-control study indicated that the odds ratio of cerebral thrombosis decreased from a high of 4.5 with 50 μg estrogen pills to 1.6 with 20–40 μg pills.[206] Hypertension increased the risk 5-fold, migraine 3.2 times, diabetes 5.6 times, hyperlipidemia and coagulation disorders about 12-fold.

The Norwegian-Swedish Women's Lifestyle and Health Cohort Study supports the conventional wisdom of the last decade that low-dose oral contraceptives do not increase the myocardial infarction or stroke in healthy, nonsmoking women, regardless of age.[207] Screening for hypertension is especially important in that it is a major risk factor for stroke associated with oral contraceptive use.

In the absence of hypertension, the effect of smoking in women under age 35 is too small to be measured. In the presence of hypertension, it is currently believed that with medical control of blood pressure and close follow-up (blood pressure monitoring every 3 months), nonsmoking women under age 35 and otherwise healthy can use low-dose oral contraception.

Incidence of Stroke in Reproductive Age Women[195, 199, 204, 205]	
Incidence of Ischemic Stroke	5/100,000/year
	1–3/100,000/year in women under age 35
	10/100,000/year in women over age 35
Incidence of Hemorrhagic Stroke	6/100,000/year
Excess Cases Per Year Due to OCs, Including Smokers and Hypertensives	2/100,000/year in low-dose OC users
	1/100,000/year in low-dose OC users under age 35
	8/100,000/year in high-dose users

Conclusion—Myocardial Infarction and Stroke

Oral contraceptives containing less than 50 μg ethinyl estradiol do not increase the risk of myocardial infarction or stroke in healthy, nonsmoking women, regardless of age. The effect of smoking in women under age 35 is, as we have long recognized, not detectable in the absence of hypertension. After age 35, the subtle presence of hypertension makes analysis difficult, but the Kaiser study indicates that increasing age and smoking by themselves have little impact on the risk of stroke in low-dose oral contraceptive users. The screening of patients in the Kaiser program was excellent, resulting in few women with hypertension using oral contraceptives. ***There is no reason to doubt that these conclusions apply as well to the transdermal and vaginal methods of steroid contraception.***

Epidemiologic studies fail to find any substantial risk of ischemic or hemorrhagic stroke with low-dose oral contraceptives in healthy, young women. The WHO study did find evidence for an adverse impact of smoking in women under age 35; the Kaiser study did not. This difference is explained by the confounding effect of hypertension, the major risk factor identified. In the WHO study, a history of hypertension was based on whether a patient reported ever having had high blood pressure (other than in pregnancy) and not validated by medical records. In the Kaiser study, women were classified as having hypertension if they reported using antihypertensive medication (less than 5% of oral contraceptive users had treated hypertension, and there were no users of higher dose products). In the WHO study, the effect of using oral contraceptives in the presence of a high-risk factor is apparent in the different odds ratios when European women who received good screening from clinicians were compared with women in developing countries who received little screening; therefore, more women with cardiovascular risk factors in developing countries were using oral contraceptives. ***The studies indicate that hypertension should be a major concern, especially in regards to the risk of stroke***.

Over the years, there has been recurring discussion over whether to provide oral contraceptives over-the-counter on a nonprescription basis. The data in the WHO study make an impressive argument against such a move. The increased risk of myocardial infarction was most evident in developing countries where 70% of the cases received their oral contraceptives from a nonclinical source. Deprived of screening, women with risk factors in developing countries were exposed to greater risk.

Smoking

Smoking continues to be a difficult problem, not only for patient management, but for analysis of data as well. In large U.S. surveys in 1982 and 1988, the decline in the prevalence of smoking was similar in users and nonusers of oral contraception; however, 24.3% of 35- to 45-year-old women who used oral contraceptives were smokers![208] In this group of smoking, oral contraceptive-using women, 85.3% smoked 15 or more cigarettes per day (heavy smoking). Despite the widespread teaching and publicity that smoking is a contraindication to oral contraceptive use over the age of 35, more older women who use oral contraceptives smoke and smoke heavily, compared with young women. This strongly implies that older smokers are less than honest with clinicians when requesting oral contraception, and this further raises serious concern over how well this confounding variable can be controlled in case-control and cohort studies. Studies documenting the effects of smoking bans on hospital admissions for myocardial infarctions indicate a rapid reduction in 6 months to a year.[209, 210] ***A former smoker must have stopped smoking for at least 6 consecutive months and preferably 12 months to be regarded as a nonsmoker. Women who have nicotine in their bloodstream obtained from patches or gum should be regarded as smokers.***

Lipoproteins and Oral Contraception

The balance of estrogen and progestin potency in a given oral contraceptive formulation can potentially influence cardiovascular risk by its overall effect on lipoprotein levels. Oral contraceptives with relatively high doses of progestins (doses not used in today's low-dose formulations) do produce unfavorable lipoprotein changes.[211] The levonorgestrel triphasic exerts no significant changes on HDL-cholesterol, LDL-cholesterol, apoprotein B, and no change or an increase in apoprotein A. The monophasic desogestrel and desogestrel pills have a favorable effect on the lipoprotein profile, while the triphasic norgestimate and gestodene pills produce beneficial alterations in the LDL/HDL and apoprotein B/apoprotein A ratios.[212–215] Like the triphasic levonorgestrel pills, norethindrone multiphasic pills have no significant impact on the lipoprotein profile over 6–12 months.[216] *In summary, studies of low-dose formulations indicate that the adverse effects of progestins are limited to the fixed-dose combination with a dose of levonorgestrel that exceeds that in the multiphasic formulation or in the low-dose products*. The formulation that contains 100 μg levonorgestrel and 20 μg ethinyl estradiol produces short-term changes in the lipid profile that are similar to those seen with other low-dose oral contraceptives, and with long-term use, the levels revert to those observed at baseline before treatment.[217]

An important study in monkeys indicated a protective action of estrogen against atherosclerosis, but by a mechanism independent of the cholesterol-lipoprotein profile. Oral administration of a combination of estrogen and progestin to monkeys fed a high-cholesterol, atherogenic diet decreased the extent of coronary atherosclerosis despite a reduction in HDL-cholesterol levels.[218–220] In somewhat similar experiments, estrogen treatment markedly prevented arterial lesion development in rabbits.[221–223] In considering the impact of progestational agents, lowering of HDL is not necessarily atherogenic if accompanied by a significant estrogen impact. These animal studies help explain why older, higher dose combinations, which had an adverse impact on the lipoprotein profile did not increase subsequent cardiovascular disease.[124, 127] The estrogen component provided protection through a direct effect on vessel walls, especially favorably influencing vasomotor and platelet factors such as nitric oxide and prostacyclin.

This conclusion is reinforced by angiographic and autopsy studies. Young women with myocardial infarctions who have used oral contraceptives have less diffuse atherosclerosis than nonusers.[224, 225] Indeed, a case-control study indicated that the risk of myocardial infarction in patients taking older, high-dose levonorgestrel-containing formulations is the same as that experienced with pills containing other progestins.[124]

In the past two decades, we were subjected to considerable marketing hype about the importance of the impact of oral contraceptives on the cholesterol-lipoprotein profile. If indeed certain oral contraceptives had a negative impact on the lipoprotein profile, one would expect to find evidence of atherosclerosis as a cause of an increase in subsequent cardiovascular disease. There is no such evidence. Thus, the mechanism of the cardiovascular complications is undoubtedly a short-term acute mechanism—thrombosis (an estrogen-related effect).

Remember too that the new progestins, because of their reduced androgenicity, predictably do not adversely affect the cholesterol-lipoprotein profile. The estrogen-progestin balance of combined oral contraceptives containing one of the new progestins even promotes favorable lipid changes.[38] Thus, the new formulations have the potential to offer protection against cardiovascular disease, an important consideration as we enter an era of women using oral contraceptives for longer durations and later in life. But one must be cautious regarding the clinical significance of subtle changes, and as with the older progestins and

adverse changes in lipoproteins, it is unlikely that oral contraceptives will have a clinically meaningful beneficial effect on the incidence of coronary heart disease.

Hypertension

Oral contraceptive-induced hypertension was observed in approximately 5% of users of higher dose pills. More recent evidence indicates that small increases in blood pressure can be observed even with 30 μg estrogen, monophasic pills, including those containing the new progestins. However, an increased incidence of clinically significant hypertension has not been reported.[226–229] The lack of clinical hypertension in most studies may be due to the rarity of its occurrence. The Nurses' Health Study observed an increased risk of clinical hypertension in current users of low-dose oral contraceptives, providing an incidence of 41.5 cases per 10,000 women per year.[230] Therefore, an annual assessment of blood pressure is still an important element of clinical surveillance, even when low-dose oral contraceptives are used. Postmenopausal women in the Rancho Bernardo Study who had previously used oral contraceptives (probably high-dose products) had slightly higher (2–4 mm Hg) diastolic blood pressures.[231] Because past users do not demonstrate differences in incidence or risk factors for cardiovascular disease, it is unlikely this blood pressure difference has an important clinical effect.

Variables such as previous toxemia of pregnancy or previous renal disease do not predict whether a woman will develop hypertension on oral contraception.[232] Likewise, women who have developed hypertension on oral contraception are not more predisposed to develop toxemia of pregnancy. Overall, there is no evidence that previous oral contraceptive users have an increased risk of hypertension during a subsequent pregnancy.[233, 234] The exception is the Nurses' Health Study, which indicated that recent users for a long duration (8 or more years) have a 2-fold increased risk of preeclampsia, a finding that was based on a small number of cases.[235] These epidemiologic associations are hard to establish because of the role of underlying hypertension in pregnancy-induced hypertension and the difficulty in assessing the efficacy of hypertension screening in oral contraceptive users.

The mechanism for an effect on blood pressure is thought to involve the renin angiotensin system. The most consistent finding is a marked increase in plasma angiotensinogen, the renin substrate, up to 8 times normal values (on higher dose pills). In nearly all women, excessive vasoconstriction is prevented by a compensatory decrease in plasma renin concentration. If hypertension does develop, the renin-angiotensinogen changes take 3–6 months to disappear after stopping combined oral contraception.

One must also consider the effects of oral contraceptives in patients with preexisting hypertension or cardiac disease. Women on oral contraceptives and with uncontrolled hypertension have an increased risk of arterial thrombosis.[189, 204, 205] Women with treated hypertension using oral contraceptives have been reported to have poor control of blood pressure with higher diastolic pressures.[236] ***In our view, with medical control of the blood pressure and close follow-up (at least every 3 months), the clinician and the nonsmoking patient who is under age 35 and otherwise healthy may choose low-dose oral contraception.*** Close follow-up is also indicated in women with a history of preexisting renal disease or a strong family history of hypertension or cardiovascular disease. It seems prudent to suggest that patients with marginal cardiac reserve should utilize other means of contraception. Significant increases in cardiac output and plasma volume have been recorded with oral contraceptive use (higher dose pills), probably a result of fluid retention.

Cardiovascular Disease—Summary

The outpouring of epidemiologic data in the last two decades allows the construction of a clinical formulation that is evidence-based. The following conclusions are consistent with the recent reports.

Oral Contraceptives and Thrombosis

- Pharmacologic estrogen increases the production of clotting factors.
- Progestins have no significant impact on clotting factors.
- Past users of oral contraceptives do not have an increased incidence of cardiovascular disease.
- All low-dose oral contraceptives, regardless of progestin type, have an increased risk of venous thromboembolism, concentrated in the first 1–2 years of use. The actual risk of venous thrombosis with low-dose oral contraceptives is lower in the new studies compared with previous reports. Some have argued that this is due to preferential prescribing and the healthy user effect. However, it is also logical that the lower risk reflects better screening of patients and lower estrogen doses. The risk increases with increasing age and body weight.
- Smoking has a lesser effect on the risk of venous thrombosis compared with arterial thrombosis.
- Smoking and estrogen have an additive effect on the risk of arterial thrombosis. Why is there a difference between venous and arterial clotting? The venous system has low flow with a state of high fibrinogen and low platelets, in contrast to the high-flow state of the arterial system with low fibrinogen and high platelets. Thus, it is understandable why these two different systems can respond in different ways.
- Hypertension is a very important additive risk factor for stroke in oral contraceptive users.
- Low-dose oral contraceptives (less than 50 μg ethinyl estradiol) do not increase the risk of myocardial infarction or stroke in healthy, nonsmoking women, regardless of age.
- Almost all myocardial infarctions and strokes in oral contraceptive users occur in users of high-dose products, or users with cardiovascular risk factors over the age of 35. In the Oxford Family Planning Association cohort, cardiac deaths occurred only in women who smoked 15 or more cigarettes per day.[129]
- Arterial thrombosis (myocardial infarction and stroke) has a dose-response relationship with the dose of estrogen, but there are insufficient data to determine whether there is a difference in risk with products that contain 20, 30, or 35 μg ethinyl estradiol.

Epidemiologic studies reinforce the belief that the risks of arterial and venous thrombosis are a consequence of the estrogen component of combination oral contraceptives. Current evidence does not support an advantage or disadvantage for any particular formulation, except for the greater safety associated with any product containing less than 50 μg ethinyl estradiol. Although it is logical to expect the greatest safety with the lowest dose of estrogen, the rare occurrence of arterial and venous thrombosis in healthy women makes it unlikely that there will be any measurable differences in the attributable incidence of clinical events with all low-dose products.

The epidemiologic data emphasize the importance of good patient screening. The occurrence of arterial thrombosis is essentially limited to older women who smoke or have cardiovascular risk factors, especially hypertension. The impact of good screening is evident in the repeated failure to detect an increase in mortality due to myocardial infarction or stroke in healthy, nonsmoking women.[129, 145, 195] Although the risk of venous thromboembolism is

slightly increased, the actual incidence is still relatively rare. The overall mortality rate with venous thromboembolism is about 1%, probably less with oral contraceptives, because most deaths from thromboembolism are associated with trauma, surgery, or a major illness. The minimal risk of venous thrombosis associated with oral contraceptive use does not justify the cost of routine screening for coagulation deficiencies. Nevertheless, the importance of this issue is illustrated by the increased risk of a very rare event, cerebral sinus thrombosis, in women who have an inherited predisposition for clotting and use oral contraceptives.[106, 237]

If a patient has a close family history (parent or sibling) or a previous episode of idiopathic thromboembolism, an evaluation to search for an underlying abnormality in the coagulation system is warranted. It has been reported that family history of venous thromboembolism has low predictive value.[238] Another study indicated that testing for thrombophilia did not allow for prediction of recurrent events, but risk factors such as family history did provide prediction.[239] The conservative recommendation for a high-risk woman considering exposure to exogenous estrogen stimulation is to rule out an underlying thrombophilia. The following measurements are recommended, and abnormal results require consultation with a hematologist regarding prognosis and prophylactic treatment. The list of laboratory tests is long, and because this is a dynamic and changing field, the best advice is to consult with a hematologist. If a diagnosis of a congenital deficiency is made, screening should be offered to other family members.

Hypercoaguable Conditions	*Thrombophilia Screening*
Antithrombin III deficiency	Antithrombin III
Protein C deficiency	Protein C
Protein S deficiency	Protein S
Factor V Leiden mutation	Activated protein C resistance ratio
Prothrombin gene mutation	Activated partial thromboplastin time
Antiphospholipid syndrome	Hexagonal activated partial thromboplastin time
	Anticardiolipin antibodies
	Lupus anticoagulant
	Fibrinogen
	Prothrombin G mutation (DNA test)
	Thrombin time
	Homocysteine level
	Complete blood count

Combination estrogen-progestin contraception is contraindicated in women who have a history of idiopathic venous thromboembolism, and also in women who have a close family history (parent or sibling) of idiopathic venous thromboembolism. These women will have a higher incidence of congenital deficiencies in important clotting measurements, especially antithrombin III, protein C, protein S, and resistance to activated protein C.[240] Such a patient who screens negatively for an inherited clotting deficiency might still consider the use of oral contraceptives, but this would be a difficult decision with unknown risks for both patient and clinician, and it is more prudent to consider other contraceptive options. Other risk factors for thromboembolism that should be considered by clinicians include an acquired predisposition such as the presence of lupus anticoagulant or malignancy, and immobility or trauma. Varicose veins are not a risk factor unless they are very extensive.[140] ***Progestin-only methods, including implants, depot medroxyprogesterone acetate, and the levonorgestrel-releasing IUS are recommended for high risk women and for women who are anticoagulated.***

The conclusion once again is that low-dose oral contraceptives are very safe for healthy, young women. By effectively screening older women for the presence of smoking and cardiovascular risk factors, especially hypertension, we can limit, if not eliminate, any increased risk for arterial disease associated with low-dose oral contraceptives. And it is very important to emphasize that there is no increased risk of cardiovascular events associated with duration of use (long-term). In large cohort studies, the risk of overall mortality comparing

users and nonusers of oral contraceptives is identical.[127–129] It is worth noting again that these conclusions likely also apply to transdermal and vaginal estrogen-progestin contraception.

Carbohydrate Metabolism

With the older high-dose oral contraceptives, an impaired glucose tolerance test was present in many women. In these women, plasma levels of insulin as well as the blood sugar were elevated. Generally, the effect of oral contraception is to produce an increase in peripheral resistance to insulin action. Most women can meet this challenge by increasing insulin secretion, and there is no change in the glucose tolerance test, although 1-h values may be slightly elevated.

Insulin sensitivity is affected mainly by the progestin component of the pill.[241] The derangement of carbohydrate metabolism may also be affected by estrogen influences on lipid metabolism, hepatic enzymes, and elevation of unbound cortisol. The glucose intolerance is dose-related, and once again effects are less with the low-dose formulations. ***Insulin and glucose changes with low-dose monophasic and multiphasic oral contraceptives are so minimal, that it is now believed they are of no clinical significance.***[229, 242–245] This includes long-term evaluation with hemoglobin A1c.

The observed changes in studies of oral contraception and carbohydrate metabolism are in the nondiabetic range. In order to measure differences, investigators have resorted to analysis by measuring the area under the curve for glucose and insulin responses during glucose tolerance tests. A highly regarded cross-sectional study utilizing this technique reported that even lower dose formulations have detectable effects on insulin resistance.[241] The reason this is important is that it is now recognized that hyperinsulinemia due to insulin resistance is a contributor to cardiovascular disease.

Because long-term, follow-up studies of large populations have failed to detect any increase in the incidence of diabetes mellitus or impaired glucose tolerance (even in past and current users of high-dose pills),[231, 246, 247] the concern now appropriately focuses on the slight impairment as a potential risk for cardiovascular disease. If slight hyperinsulinemia were meaningful, wouldn't you expect to see evidence of an increase in cardiovascular disease in past users who took oral contraceptives when doses were higher? As we have emphasized before, there is no such evidence. The data strongly indicate that the changes in lipids and carbohydrate metabolism that have been measured are not clinically meaningful.

It can be stated definitively that oral contraceptive use does not produce an increase in diabetes mellitus.[246–249] The minor hyperglycemia associated with oral contraception is not deleterious and is completely reversible. Even women who have risk factors for diabetes in their history are not affected. In women with recent gestational diabetes, no significant impact on glucose tolerance could be demonstrated over 6–13 months comparing the use of low-dose monophasic and multiphasic oral contraceptives with a control group, and no increase in the risk of overt diabetes mellitus could be detected with long-term follow-up.[250, 251] A high percentage of women with previous gestational diabetes develop overt diabetes and associated vascular complications. Until overt diabetes develops, it is appropriate for these patients to use low-dose oral contraception.

In clinical practice, it may, at times, be necessary to prescribe oral contraception for the overt diabetic. No effect on insulin requirement is expected with low-dose pills.[252] According to the older epidemiologic data, the use of oral contraceptives increased the risk of thrombosis in women with insulin-dependent diabetes mellitus; therefore,

women with diabetes have been encouraged to use other forms of contraception. However, this effect in women under age 35 who are otherwise healthy and nonsmokers is probably very minimal with low-dose oral contraception, and reliable protection against pregnancy is a benefit for these patients that outweighs the small risk. A case-control study could find no evidence that oral contraceptive use by young women with insulin-dependent diabetes mellitus increased the development of retinopathy or nephropathy.[253] In a 1-year study of women with insulin-dependent diabetes mellitus who were using a low-dose oral contraceptive, no deterioration could be documented in lipoprotein or hemostatic biochemical markers for cardiovascular risk.[254] And finally, no effect of oral contraceptives on cardiovascular mortality could be detected in a group of women with diabetes mellitus.[255]

The Liver

The liver is affected in more ways and with more regularity and intensity by the sex steroids than any other extragenital organ. Estrogen influences the synthesis of hepatic DNA and RNA, hepatic cell enzymes, serum enzymes formed in the liver, and plasma proteins. Estrogenic hormones also affect hepatic lipid and lipoprotein formation, the intermediary metabolism of carbohydrates, and intracellular enzyme activity. Nevertheless, an extensive analysis of the prospective cohorts of women in the Royal College of General Practitioners' Oral Contraception Study and the Oxford Family Planning Association Contraceptive Study could detect no evidence of an increased incidence or risk of serious liver disease among oral contraceptive users.[256]

The active transport of biliary components is impaired by estrogens as well as some progestins. The mechanism is unclear, but cholestatic jaundice and pruritus were occasional complications of higher dose oral contraception, and are similar to the recurrent jaundice of pregnancy, i.e., benign and reversible. The incidence with lower dose oral contraception is unknown, but it must be a very rare occurrence.

The only absolute hepatic contraindication to steroid contraceptive use is acute or chronic cholestatic liver disease. Cirrhosis and previous hepatitis are not aggravated. Once recovered from the acute phase of liver disease (normal enzyme levels), a woman can use steroid contraception.

Data from the Royal College of General Practitioners' prospective study indicated that an increase in the incidence of gallstones occurred in the first years of oral contraceptive use, apparently due to an acceleration of gallbladder disease in women already susceptible.[257] In other words, the overall risk of gallbladder disease was not increased, but in the first years of use, disease was activated or accelerated in women who were vulnerable because of asymptomatic disease or a tendency toward gallbladder disease. The mechanism appears to be induced alterations in the composition of gallbladder bile, specifically a rise in cholesterol saturation that is presumably an estrogen effect.[258] The Nurses' Health Study reported no significant increase in the risk of symptomatic gallstones among ever-users, but slightly elevated risks among current and long-term users.[259] Although oral contraceptive use has been linked to an increased risk of gallbladder disease, the epidemiologic evidence has been inconsistent. Indeed an Italian case-control study and a report from the Oxford Family Planning Association cohort found no increase in the risk of gallbladder disease in association with oral contraceptive use and no interaction with increasing age or body weight.[260, 261] Keep in mind that even though some studies found a statistically significant modest increase in the relative risk of gallbladder disease, even if the effect were real, it is of little clinical importance because the actual incidence of this problem is very low.

Liver Adenomas

Hepatocellular adenomas can be produced by steroids of both the estrogen and androgen families. Actually, there are several different lesions: peliosis, focal nodular hyperplasia, and adenomas. Peliosis is characterized by dilated vascular spaces without endothelial lining, and may occur in the absence of adenomatous changes. The adenomas are not malignant; their significance lies in the potential for hemorrhage. The most common presentation is acute right upper quadrant or epigastric pain. The tumors may be asymptomatic, or they may present suddenly with hematoperitoneum. There is some evidence that the tumors and focal nodular hyperplasia regress when oral contraception is stopped.[262, 263] Epidemiologic data have not supported the contention that mestranol increased the risk more than ethinyl estradiol.

The risk appears to be related to duration of oral contraceptive use and to the steroid dose in the pills. This is reinforced by the rarity of the condition ever since low-dose oral contraception became available. The ongoing prospective studies have accumulated many woman-years of use and have not identified an increased incidence of such tumors.[256] In a collaborative study of 15 German liver centers, no increase in risk for liver adenomas in contemporary oral contraceptive users could be detected.[264] An Italian case-control study found an increase in risk for focal nodular hyperplasia associated with low-dose oral contraceptives, a risk that reached statistical significance only with 3 or more years of use (with a very wide confidence interval because of only 13 cases).[265] In our view it is not even worth mentioning liver problems during the informed consent (choice) process.

No reliable screening test or procedure is currently available for the detection of liver adenomas. Routine liver function tests are normal. Computed tomography (CT) scanning or magnetic resonance imaging (MRI) is the best means of diagnosis; angiography and ultrasonography are not reliable. Palpation of the liver should be part of the periodic evaluation in oral contraceptive users. If an enlarged liver is found, oral contraception should be stopped, and regression should be evaluated and followed by imaging.

Other Metabolic Effects

Nausea and breast discomfort continue to be disturbing effects, but their incidence is significantly less with low-dose oral contraception. Fortunately, these effects are most intense in the first few months of use and, in most cases, gradually disappear. In placebo-controlled trials with low-dose oral contraceptives, the incidence of "minor" side effects such as headache, nausea, dysmenorrhea, and breast discomfort actually occurred at the same rate in the treated group and the placebo group![266–268] Weight gain usually responds to dietary restriction, but for some patients, the weight gain is an anabolic response to the sex steroids, and discontinuation of oral contraception is the only way that weight loss can be achieved. This must be rare with low-dose oral contraception because data in published studies, especially in placebo-controlled trials, fail to indicate a difference in body weight between users and nonusers.[268–273 274–276]

There is no association between oral contraception and peptic ulcer disease or inflammatory bowel disease.[277, 278] Oral contraception is not recommended for patients with problems of gastrointestinal malabsorption because of the possibility of contraceptive failure, although vaginal administration would be appropriate.

Chloasma, a patchy increase in facial pigment, was, at one time, found to occur in approximately 10% of oral contraceptive users. It is now a rare problem due to the decrease in estrogen dose. Unfortunately, once chloasma appears, it fades only gradually following discontinuation of the pill and may never disappear completely. Skin-blanching medications may be useful.[279]

Hematologic effects include an increased sedimentation rate, increased total iron-binding capacity due to the increase in globulins, and a decrease in prothrombin time. The use of oral contraceptives results in a decrease in iron deficiency anemia because of a reduction in menstrual bleeding.[280, 281] Indeed, in anemic women, an increase in hemoglobin and ferritin levels accompanies the use of oral contraceptives.[282]

The continuous, daily use of oral contraceptives may prevent the appearance of symptoms in porphyria precipitated by menses. Changes in vitamin metabolism have been noted: a small nonharmful increase in vitamin A and decreases in blood levels of pyridoxine (B_6) and the other B vitamins, folic acid, and ascorbic acid. Despite these changes, routine vitamin supplements are not necessary for women eating adequate, normal diets.[283]

Mental depression is very rarely associated with oral contraceptives. In studies with higher dose oral contraceptives, the effect was due to estrogen interference with the synthesis of tryptophan that could be reversed with pyridoxine treatment. It seems wiser, however, to discontinue oral contraception if depression is encountered.

Though infrequent, a reduction in libido is occasionally a problem and may be a cause for seeking an alternative method of contraception.[284] Indeed most women report an increase or no change in libido during oral contraceptive use.[285, 286] Nevertheless, the estrogen-induced increase in sex hormone-binding globulin and the resulting decrease in free testosterone may produce an impact on sexuality in some women. An empiric trial of a nonestrogen-containing contraceptive method is warranted.

Adverse androgenic voice changes were occasionally encountered with the use of the first very high-dose oral contraceptives. Vocal virilization can be a serious and devastating problem for some women, especially when vocal performance is important. Careful study of women on low-dose oral contraceptives indicates that this is no longer a side effect of concern.[287]

The Risk of Cancer

Endometrial Cancer

The use of oral contraception protects against endometrial cancer. Use for at least 12 months reduces the risk of developing endometrial cancer by ***50%***, with the greatest protective effect gained by use for more than 3 years.[288–293] This protection persists for 20 or more years after discontinuation (the actual length of duration of protection is unknown) and is greatest in women at highest risk: nulliparous and low parity women.[293, 294] This protection is equally protective for all three major histologic subtypes of endometrial cancer: adenocarcinoma, adenoacanthoma, and adenosquamous cancers. Finally, protection is seen with all monophasic formulations of oral contraceptives, including pills with less than 50 μg estrogen.[288, 290, 293, 295] There are no data as yet with multiphasic preparations or the new progestin formulations, but because these products are still dominated by their progestational component, there is every reason to believe that they will be protective.

Ovarian Cancer

Protection against ovarian cancer, the most lethal of female reproductive tract cancers, is one of the most important benefits of oral contraception. Because this cancer is detected late

and prognosis is poor, the impact of this protection is very significant. Indeed, a decline in mortality from ovarian cancer has been observed in several countries since the early 1970s, perhaps an effect of oral contraceptive use.[296] Women with increased exposure to oral contraceptives have demonstrated a marked decrease in the incidence of ovarian cancer.[297–300] The risk of developing epithelial ovarian cancer of all histologic subtypes in users of oral contraception is reduced by ***40%*** compared with that of nonusers.[290, 292, 301–307] This protective effect increases with duration of use and continues for at least 20 years after stopping the medication.[308] This protection is seen in women who use oral contraception for as little as 3 to 6 months (although at least 3 years of use are required for a notable impact), reaching an 80% reduction in risk with more than 10 years of use, and is a benefit associated with all monophasic formulations, including the low-dose products.[305–307, 309] The protective effect of oral contraceptives is especially observed in women at high risk of ovarian cancer (nulliparous women and women with a positive family history).[310, 311] Continuous use of oral contraception for 10 years by women with a positive family history for ovarian cancer can reduce the risk of epithelial ovarian cancer to a level equal to or less than that experienced by women with a negative family history.[310] Again, the multiphasic and new progestin products have not been in use long enough to yield any data on this issue, but because ovulation is effectively inhibited by these formulations, protection against ovarian cancer should be exerted. ***The same magnitude of protection has been observed in women with BRCA1 or BRCA2 mutations.***[312–314]

Cancer of the Cervix

Studies have indicated that the risk for dysplasia and carcinoma in situ of the uterine cervix increases with the use of oral contraception for more than 1 year.[315–320] Invasive cervical cancer may be increased after 5 years of use, reaching a 2-fold increase after 10 years. It is well recognized, however, that the number of partners a woman has had and age at first coitus are important risk factors for cervical neoplasia. Other confounding factors include exposure to human papillomavirus, the use of barrier contraception (protective), and smoking. These are difficult factors to control, and, therefore, the conclusions regarding cervical cancer are not definitive. An excellent study from the Centers for Disease Control and Prevention (CDC) concluded there is no increased risk of invasive cervical cancer in users of oral contraception, and an apparent increased risk of carcinoma in situ is due to enhanced detection of disease (because oral contraceptive users have more frequent Pap smears).[318] In the World Health Organization Study of Neoplasia and Steroid Contraceptives, a Pap smear screening bias was identified, nevertheless the evidence still suggested an increased risk of cervical carcinoma in situ with long-term oral contraceptive use.[319]

A case-control study of patients in Panama, Costa Rica, Colombia, and Mexico concluded that there was a significantly increased risk for invasive adenocarcinoma.[321] Similar results were obtained in a case-control study in Los Angeles and in the World Health Organization Collaborative Study.[322, 323] In Los Angeles, the relative risk of adenocarcinoma of the cervix increased from 2.1 with ever use to 4.4 with 12 or more years of oral contraceptive use.[322] Because the incidence of adenocarcinoma of the cervix (10% of all cervical cancers) has increased in young women over the last 20 years, there is concern that this increase reflects the use of oral contraception.[324] Oral contraceptives increase cervical ectopia, but whether this increases the risk of cervical adenocarcinoma is unclear.

A large meta-analysis concluded that the relative risk of cervical cancer increased with increasing duration of use (for in situ and invasive cancer and both squamous cancer and adenocarcinoma); however, the risk was confined to the cases who tested positively for human papillomavirus (HPV).[325] A pooled analysis of case-control studies concluded that

the risk of cervical cancer in women with HPV increases about 3-fold but not until after 5 years of use.[326]

A meta-analysis of 24 epidemiological studies included 16,573 women with cervical cancer and 35,509 women without cervical cancer.[327] The risk of cervical cancer in-situ and invasive cancer increased about 2-fold with increasing duration of use of oral contraceptives. After discontinuation of oral contraceptives, the risk steadily declined. There was no increase in risk with duration of use less than 5 years or after 10 or more years since last use. The same relationships were seen comparing women likely to have been screened and those likely not to have been screened. Although the numbers were smaller, women who were positive for human papillomavirus (HPV) demonstrated the same pattern of risk. A similar but smaller risk was associated with injectable progestin-only methods. Insufficient numbers were available to assess oral progestin-only contraceptives.

It has been well-demonstrated that oral contraceptive users have more sexual partners, more HPV infections, and more Pap smears. The data in the meta-analyses confirm these facts, plus a greater incidence of smoking in oral contraceptive users. Perhaps, the analyses adjusted effectively for some of these factors, but a problem remains. Adequate data are not available to allow case-control and cohort studies to control for Pap smear screening and for differences in condom use.

We currently know three important facts about cervical cancer. (1) Cervical cancer is caused by HPV. (2) Cervical cancer is reduced in prevalence by Pap smear screening. (3) HPV vaccination can potentially eliminate most cases of cervical cancer. Oral contraceptive users have behavior patterns that place them at greater risk of HPV infections. In addition, oral contraceptives could increase the ability of HPV to establish itself in the cervix. Once infected with HPV, oral contraceptives could influence the response of a woman to HPV, for example, suppressing the immune response. Therefore, we don't know for certain whether case-control and cohort data reflect an enhancing effect of oral contraceptives on the risk of cervical cancer or whether the results are affected by the confounding problems of HPV infections and the prevalence of Pap smear screening.

These studies reinforce the HPV and Pap smear screening that we already support, and provide another argument in favor of HPV vaccination. The liquid-based methods along with HPV DNA testing will provide even better identification of at-risk women. Fortunately, steroid contraception does not mask abnormal cervical changes, and the necessity for prescription renewals offers the opportunity for improved screening for cervical disease. ***It is reasonable*** to perform Pap smears every 6 months in women using oral contraception for 5 or more years who are also at higher risk because of their sexual behavior (multiple partners, history of sexually transmitted infections). Oral contraceptive use is appropriate for women with a history of cervical intraepithelial neoplasia (CIN), including those who have been surgically treated.

Liver Cancer

Oral contraception has been linked to the development of hepatocellular carcinoma.[328, 329] However, the very small number of cases, and, thus, the limited statistical power, requires great caution in interpretation. The largest study on this question, the WHO Collaborative Study of Neoplasia and Steroid Contraceptives, found no association between oral contraception and liver cancer.[330] Even case-control analysis of oral contraceptives containing cyproterone acetate (known to be toxic to the liver in high doses) could detect no evidence of an increased risk of liver cancer.[331] In the U.S., Japan, Sweden, England, and

Wales, the death rates from liver cancer did not change despite introduction and use of oral contraception.[332, 333] More recently, there has been an increase in liver cancer incidence and mortality in the U.S., but this is believed to be due to infection with hepatitis C and hepatitis B.[334]

Breast Cancer

Because of breast cancer's prevalence and its long latent phase, concern over the relationship between oral contraception and breast cancer continues to be an issue in the minds of both patients and clinicians. Worth emphasizing is the protective effect of higher dose oral contraception on benign breast disease, an effect that became apparent after 2 years of use.[335–337] After 2 years there was a progressive reduction (about 40%) in the incidence of fibrocystic changes in the breast. Women who used oral contraception were one-fourth as likely to develop benign breast disease as nonusers, but this protection was limited to current and recent users. In the large Oxford Family Planning Association cohort, the incidence of benign breast disease decreased with increasing duration of use.[338] A French case-control study indicated a reduction of nonproliferative benign breast disease associated with low-dose oral contraceptives used before a first full-term pregnancy, but no effect on proliferative disease or with use after a pregnancy.[339] A Canadian cohort study that almost certainly reflected the use of modern low-dose oral contraceptives concluded that oral contraceptives do protect against proliferative benign disease, with an increasing reduction in risk with increasing duration of use.[340]

The Royal College of General Practitioners,[341] Oxford Family Planning Association,[342, 343] the Nurses' Health Study, [344] and Walnut Creek[345] cohort studies indicated no significant differences in breast cancer rates between users and nonusers. However, patients were enrolled in these studies at a time when oral contraception was used primarily by married couples spacing out their children. Beginning in the 1980s, oral contraception was primarily being used by women early in life, for longer durations, and to delay an initial pregnancy (remember, a full-term pregnancy early in life protects against breast cancer).

Case-control studies have focused on the use of oral contraception early in life, for long duration, and to delay a first, full-term pregnancy. Because the women who have used oral contraception in this fashion are just now beginning to reach the ages of postmenopausal breast cancer, many studies have had to focus on the risk of breast cancer diagnosed before age 45 (only 13% of all breast cancer). The results of these studies have not been clear-cut. Some studies have indicated an overall increased relative risk of early, premenopausal breast cancer,[346–354] while others indicated no increase in overall risk.[355–357] The most impressive finding indicates a link in most studies,[358–363] but not all,[364–368] of early breast cancer before age 40 with women who used oral contraception for long durations of time.

A collaborative group re-analyzed data from 54 studies in 26 countries, a total of 53,297 women with breast cancer and 100,239 without breast cancer, in order to assess the relationship between the risk of breast cancer and the use of oral contraceptives.[369, 370] Oral contraceptives were grouped into three categories: low, medium, and high dose (which correlated with less than 50, 50 μg, and more than 50 μg of estrogen). At the time of diagnosis, 9% of the women with breast cancer were under age 35, 25% were 35–44, 33% were 45–54, and 33% were age 55 and older. A similar percentage of women with breast cancer (41%) and women without breast cancer (40%) had used combined oral contraceptives at some time in their lives. Overall, the relative risk (RR) of breast cancer in ever users of oral contraceptives was very slightly elevated and statistically significant: RR=1.07; CI=1.03–1.10.

The relative risk analyzed by duration of use in the collaborative study was barely elevated and not statistically significant (even when long-term use, virtually continuous, was analyzed). Women who had begun use as teenagers had about a 20% statistically significant increased relative risk. In other words, recent users who began use before age 20 had a higher relative risk compared with recent users who began at later ages. The evidence was strong for a relationship with time since last use, an elevated risk being significant for current users and in women who had stopped use 1–4 years before (recent use). No influence on this risk was observed with the following: a family history of breast cancer, age of menarche, country of origin, ethnic groups, body weight, alcohol use, years of education, and the design of the study. There was no variation according to specific type of estrogen or progestin in the various products. Importantly, there was no statistically significant effect of low-, medium-, or high-dose preparations. Ten or more years after stopping use, there was no increased risk of breast cancer. Indeed, the risk of metastatic disease compared with localized tumors was reduced: RR=0.88; CI=0.81–0.95.

Data were limited for progestin-only methods. The reanalysis indicated that the results were similar to those with combined oral contraceptives, but a close look at the numbers reveals that not one relative risk reached statistical significance.

Overall, this massive statistical exercise yielded good news. No major adverse impact of oral contraceptives emerged. ***Even though the data indicated that young women who begin use before age 20 have higher relative risks of breast cancer during current use and in the 5 years after stopping, this is a time period when breast cancer is very rare; and, thus, there would be little impact on the actual number of breast cancers.*** The difference between localized disease and metastatic disease was statistically greater and should be observable. Thus many years after stopping oral contraceptive use, the main effect may be protection against metastatic disease. Breast cancer is more common in older years, and 10 or more years after stopping, the risk was not increased.

The Norwegian-Swedish Women's Lifestyle and Health Cohort Study began in the early 1990s to follow over 100,000 women specifically to address the role of hormonal contraceptives on health.[371] A small increase in the risk of breast cancer was reported among current and recent users of oral contraceptives.

What other explanation could account for an increased risk associated only with current or recent use, no increase with duration of use, and a return to normal 10 years after exposure? The slightly increased risk could be influenced by detection/surveillance bias (more interaction with the health care system by oral contraceptive users). It is also possible that this situation is analogous to that of pregnancy. Studies indicate that pregnancy transiently increases the risk of breast cancer (for a period of several years) after a woman's first childbirth, and this is followed by a lifetime reduction in risk.[372] And some have found that a concurrent or recent pregnancy adversely affects survival.[373, 374] It is argued that breast cells that have already begun malignant transformation are adversely affected by the hormones of pregnancy, while normal stem cells become more resistant because of a pregnancy. It is possible that early and recent use of oral contraceptives also affects the growth of a preexisting malignancy, explaining the limitation of the finding to current and recent use and the increase in localized disease. With the accumulation of greater numbers of older women previously exposed to oral contraceptives, a protective effect may become evident. In a case-control study of women in Toronto, Canada, age 40–69 years, those women who had used oral contraceptives for 5 or more years, 15 or more years previously, had a 50% reduced risk of breast cancer.[375] However, a case-control study from Sweden could detect neither a beneficial nor an adverse effect of previous use of oral contraceptives (mainly 50 μg estrogen products) on the risk of breast cancer in women age 50–74 years.[376]

The largest case-control study included 4,575 American women with breast cancer, and most importantly, the women were 35 to 64 years old.[377] The risk of breast cancer was not increased in current users or past users of oral contraception. There was no adverse effect of increasing duration of use or higher doses of estrogen, with no differences in current or recent users. Initiation at a younger age had no impact, and there was no increase in risk in women with a family history of breast cancer. This large American study had consistently negative results. The next largest study, involving women from California, Canada, and Australia, focused on breast cancer diagnosed before age 40, and could not detect an increase in current or past users of oral contraceptives.[378] A multicenter, large case-control study of women younger than 55 years with breast cancer concluded that the use of oral contraceptives or postmenopausal hormone therapy either before or after diagnosis did not increase the risk of the first breast cancer or recurrent breast cancer.[379] This negative finding was not changed by duration of use or age of use. An American case-control study did not find an increased risk of breast carcinoma in-situ associated with oral contraceptive use.[380] Furthermore, no increase in breast cancer mortality can be detected in women who have used oral contraceptives.[381, 382]

A cohort study from Minnesota concluded that women with a first-degree relative with breast cancer had an increased risk of breast cancer with oral contraception; however, this association was present only with oral contraceptives used prior to 1976 (high dose formulations), and the confidence intervals were wide because of small numbers (13 ever users).[383] In a study of women with *BRCA1* and *BRCA2* mutations, an elevated risk of breast cancer associated with oral contraception was based on only a few cases and did not achieve statistical significance.[384] A larger case-control study concluded that *BRCA1* mutation carriers had small increases in the risk of breast cancer in users for at least 5 years, in users before age 30, and in those who developed breast cancer before age 40.[385] In contrast, another case-control study concluded that oral contraceptive use for at least 5 years doubled the risk of breast cancer before age 50 in *BRCA2* carriers, but not in *BRCA1* carriers.[386] A retrospective analysis of an international cohort of *BRCA* carriers indicated that an increased risk of breast cancer with both *BRCA1* and *BRCA2* carriers was present only with 4 or more years of use before a first full-term pregnancy.[387] A study that focused on low-dose oral contraceptives could detect no association with breast cancer risk in *BRCA* mutation carriers.[388] Another case-control study found no increase in the risk of breast cancer diagnosed before age 40 in either *BRCA1* or *BRCA2* carriers.[378] And finally, a case-control study could detect no significant increase in the risk of contralateral breast cancer among *BRCA1* and *BRCA2* carriers or in noncarriers.[389] The data with oral contraceptives in *BRCA* mutation carriers (also discussed in Chapter 16) are all observational and not robust. ***Until better information is forthcoming, it seems reasonable to inform carriers of BRCA mutations that the use of oral contraceptives is likely to reduce the risk of ovarian cancer, but the effect on breast cancer risk is uncertain.***

A team of epidemiologists from several institutions in the U.S. performed a case-control study of the association between oral contraceptive use and lobular and ductal breast cancer occurring in young women (under age 44).[390] Cases included 100 lobular cancers and 1,164 ductal cancers, and the use of oral contraceptives had no meaningful effects on breast cancer according to histologic subtype. Lobular cancer (15% of all breast cancers) has been increasing in the U.S. in recent years, prompting these investigators to ask whether this reflects exposure to exogenous hormones. According to their data, the answer is no. This is very reassuring because it is well-recognized that lobular cancer is more hormonally sensitive than ductal breast cancer.

Even if there is a small increase in premenopausal breast cancer associated with oral contraceptives, this would be a very small number of cases because most cases of breast cancer occur after age 40. Well-done and large case-control studies of modern low-dose oral contraceptives have been consistently negative and reassuring.

Summary: Oral Contraceptives and the Risk of Breast Cancer

- Current and recent use of oral contraceptives may be associated with about a 20% increased risk of early (under age 35) premenopausal breast cancer, essentially limited to localized disease and a very small increase in the actual number of cases (so small, there would be no major impact on incidence figures). This finding may be due to detection/surveillance bias and accelerated growth of already present malignancies, a situation similar to the effects of pregnancy and postmenopausal hormone therapy on the risk of breast cancer (as reviewed in Chapter 18). Further comfort can be derived from the fact that the increase in breast cancer in American women was greater in older women from 1973 to 1994, those who did not have the opportunity to use oral contraception.[391] In women under 50 years of age, there was only a slight increase during this same time period. The large American case-control study of women age 35–64 years was totally negative and very reassuring.
- There is no effect of past use or duration of oral contraceptive use (up to 15 years of continuous use) on the risk of breast cancer, and there is no evidence indicating that higher dose oral contraceptives increased the risk of breast cancer.
- Previous oral contraceptive use may be associated with a reduced risk of metastatic breast cancer later in life, and possibly with a reduced risk of postmenopausal breast cancer.
- Oral contraceptive use does not further increase the risk of breast cancer in women with positive family histories of breast cancer or in women with proven benign breast disease.
- The clinician should not fail to take every opportunity to direct attention to all factors that affect breast cancer. Breastfeeding and control of alcohol intake are good examples, and are also components of preventive health care. Especially important is this added motivation to encourage breastfeeding. The protective effect of breastfeeding is exerted (although it is probably a small one; see Chapter 16) mainly on premenopausal breast cancer, the cancer of concern to younger women using oral contraception.

Colorectal Cancer

The Nurses' Health Study reported about a 40% reduced risk of colorectal cancer associated with 8 years of previous use of oral contraceptives.[392] A meta-analysis of published studies concluded that there is an 18% reduction in risk of colorectal cancer in ever users of oral contraception, with a stronger effect in recent users.[393] Both case-control and cohort studies have reported reductions in the risk of colorectal cancer associated with the ever use of oral contraceptives.[394–396] ***Steroid contraception should be offered to women with a strong family history of colorectal cancer.***

Other Cancers

The Walnut Creek study suggested that melanoma was linked to oral contraception; however, the major risk factor for melanoma is exposure to sunlight. More accurate evaluation utilizing both the Royal College General Practitioners and Oxford Family Planning

Association prospective cohorts and accounting for exposure to sunlight did not indicate a significant difference in the risk of melanoma comparing users to nonusers.[397–399] There is no evidence linking oral contraceptive use to kidney cancer, gallbladder cancer, or pituitary tumors.[400] Long-term oral contraceptive use may slightly increase the risk of molar pregnancy.[401–403] A case-control study concluded that oral contraceptives reduce the risk of salivary gland cancer.[404]

Endocrine Effects

Adrenal Gland

Estrogen increases the cortisol-binding globulin (CBG). It had been thought that the increase in plasma cortisol while on oral contraception was due to increased binding by this globulin and not an increase in free active cortisol. Now it is apparent that free and active cortisol levels are also elevated, but only slightly.[405] Estrogen decreases the ability of the liver to metabolize cortisol, and in addition, progesterone and related compounds can displace cortisol from transcortin, and thus contribute to the elevation of unbound cortisol. The effects of these elevated levels over prolonged periods of time are unknown, but no obvious impact has become apparent. To put this into perspective, the increase is not as great as that which occurs in pregnancy, and, in fact, it is within the normal range for nonpregnant women.

The adrenal gland responds to adrenocorticotropic hormone (ACTH) normally in women on oral contraceptives; therefore, there is no suppression of the adrenal gland itself. Initial studies indicated that the response to metyrapone (an 11β-hydroxylase blocker) was abnormal, suggesting that the pituitary was suppressed. However, estrogen accelerates the conjugation of metyrapone by the liver; and, therefore, the drug has less effect, thus explaining the subnormal responses initially reported. The pituitary-adrenal reaction to stress is normal in women on oral contraceptive pills.

Thyroid

Estrogen increases the synthesis and circulating levels of thyroxine-binding globulin, Prior to the introduction of methods for measuring free thyroxine levels, evaluation of thyroid function was a problem. Measurement of TSH (thyroid-stimulating hormone) and the free thyroxine level in a woman on oral contraception provide an accurate assessment of a patient's thyroid state. Oral contraception affects the total thyroxine level in the blood as well as the amount of binding globulin, but the free thyroxine level is unchanged.[405]

Oral Contraception and Reproduction

The impact of oral contraceptives on the reproductive system is less than initially thought. Early studies that indicated adverse effects have not stood the test of time and the scrutiny of multiple, careful studies. There are two major areas that warrant review: (1) Inadvertent

use of oral contraceptives during the cycle of conception and during early pregnancy, and (2) Reproduction after discontinuing oral contraception.

Inadvertent Use During the Cycle of Conception and During Early Pregnancy

One of the reasons, if not the major reason, why a lack of withdrawal bleeding while using oral contraceptives is such a problem is the anxiety produced in both patient and clinician. The patient is anxious because of the uncertainty regarding pregnancy, and the clinician is anxious because of the concerns stemming from the retrospective studies that indicated an increased risk of congenital malformations among the offspring of women who were pregnant and using oral contraception. Organogenesis does not occur in the first 2 embryonic weeks (first 4 weeks since last menstrual period); however, teratogenic effects are possible between the third and eighth embryonic weeks (5 to 10 weeks since last menstrual period).

Initial positive reports linking the use of contraceptive steroids to congenital malformations have not been substantiated. Many suspect a strong component of recall bias in the few positive studies due to a tendency of patients with malformed infants to recall details better than those with normal children. Other confounding problems have included a failure to consider the reasons for the administration of hormones (e.g., bleeding in an already abnormal pregnancy), and a failure to delineate the exact timing of the treatment (e.g., treatment was sometimes confined to a period of time during which the heart could not have been affected).

An association with cardiac anomalies was first claimed in the 1970s.[406, 407] This link received considerable support with a report from the U.S. Collaborative Perinatal Project; however, subsequent analysis of these data uncovered several methodologic shortcomings.[408] Simpson, in a very thorough and critical review in 1990, concluded that there was no reliable evidence implicating sex steroids as cardiac teratogens.[409] In fact, in his review, Simpson found no relationship between oral contraception and the following problems: hypospadias, limb reduction anomalies, neural tube defects, and mutagenic effects that would be responsible for chromosomally abnormal fetuses. Even virilization is not a practical consideration because the doses required (e.g., 20–40 mg norethindrone per day) are in excess of anything currently used. These conclusions reflect use of combined oral contraceptives as well as progestins alone.

In the past there was a concern regarding the VACTERL complex. VACTERL refers to a complex of vertebral, anal, cardiac, tracheoesophageal, renal, and limb anomalies. While case-control studies indicated a relationship with oral contraception, prospective studies have failed to observe any connection between sex steroids and the VACTERL complex.[410] Meta-analyses of the studies of the risk of birth defects with oral contraceptive ingestion during pregnancy concluded that there was no increase in risk for major malformations, congenital heart defects, or limb reduction defects.[411, 412]

Women who become pregnant while taking oral contraceptives or women who inadvertently take birth control pills early in pregnancy should be advised that the risk of a significant congenital anomaly is no greater than the general rate of 2–3%. This recommendation can be extended to those pregnant woman who have been exposed to a progestational agent such as medroxyprogesterone acetate or 17-hydroxyprogesterone caproate.[413, 414]

Reproduction After Discontinuing Oral Contraception

Fertility

The early reports from the British prospective studies indicated that former users of oral contraception had a delay in achieving pregnancy. In the Oxford Family Planning Association study, former use had an effect on fertility for up to 42 months in nulligravida women and for up to 30 months in multigravida women.[415] Presumably, the delay was due to lingering suppression of the hypothalamic-pituitary reproductive system.

A later analysis of the Oxford data indicated that the delay was concentrated in women age 30–34 who had never given birth.[416] At 48 months, 82% of these women had given birth compared with 89% of users of other contraceptive methods, not a big difference. No effect was observed in women younger than 30 or in women who had previously given birth. Childless women age 25–29 experienced some delay in return to fertility, but by 48 months, 91% had given birth compared with 92% in users of other methods.

This delay was observed in the U.S. as well. In the Boston area, the interval from cessation of contraception to conception was 13 months or greater for 24.8% of prior oral contraceptive users compared with 10.6% for former users of all other methods (12.4% for intrauterine device, IUD, users, 8.5% for diaphragm uses, and 11.9% for other methods).[417] Oral contraceptive users had a lower monthly percentage of conceptions for the first 3 months, and somewhat lower percentage from 4 to 10 months. It took 24 months for 90% of previous oral contraceptive users to become pregnant, 14 months for IUD users, and 10 months for diaphragm users. Similar findings in Connecticut indicated that this delay lasted at least a year, and the effect was greater with higher dose preparations.[418] Despite the possibility of a delay, there is no evidence that infertility is increased by the use of oral contraception. In fact, in young women, previous oral contraceptive use is associated with a lower risk of primary infertility.[419] Furthermore, the studies indicating a delay in conception are influenced by older, higher dose products. In a prospective study from the U.K. reflecting modern, low-dose oral contraceptives, no delay to conception was found and long-term use was actually associated with greater fertility.[420] The European Active Surveillance Study on Oral Contraceptives was a prospective cohort study of 59,510 users of low-dose oral contraception. The early and 1-year pregnancy rates after discontinuation of oral contraceptives were not negatively affected, regardless of progestin type, duration of use, or parity.[421] After 2 years, the pregnancy rate was 88.3% and the average time to pregnancy was 5.5 cycles. The previous reports indicating a delay in achieving pregnancy may have been influenced not only by higher dose products, but also by a failure to account for declining fertility with aging and the prescribing of oral contraceptives to women with anovulatory, irregular menstrual periods. ***It is unlikely that women discontinuing low-dose steroid contraception experience any significant delay in achieving pregnancy compared with the experience in a general population.***

Spontaneous Miscarriage

There is no increase in the incidence of spontaneous miscarriage in pregnancies after the cessation of oral contraception. Indeed, the rate of spontaneous miscarriages and stillbirths is slightly less in former pill users, about 1% less for spontaneous miscarriages and 0.3% less for

stillbirths.[422] A protective effect of previous oral contraceptive use against spontaneous miscarriage has been observed to be more apparent in women who become pregnant after age 30.[423]

Pregnancy Outcome

There is no evidence that oral contraceptives cause changes in individual germ cells that would yield an abnormal child at a later time.[409] There is no increase in the number of abnormal children born to former oral contraceptive users, and there is no change in the sex ratio (a sign of sex-linked recessive mutations).[422, 424] These observations are not altered when analyzed for duration of use. Initial observations that women who had previously used oral contraception had an increase in chromosomally abnormal fetuses have not been confirmed. Furthermore, as noted above, there is no increase in the miscarriage rate after discontinuation, something one would expect if oral contraceptives induce chromosomal abnormalities because these are the principal cause of spontaneous miscarriage.

In a 3-year follow-up of children whose mothers used oral contraceptives prior to conception, no differences could be detected in weight, anemia, intelligence, or development.[425] Former pill users have no increased risks for the following: perinatal morbidity or mortality, prematurity, and low birth weight.[426, 427] Dizygous twinning has been observed to be nearly 2-fold (1.6% vs. 1.0%) increased in women who conceive soon after cessation of oral contraception.[422] This effect was greater with longer duration of use.

The only reason (and it is a good one) to recommend that women defer attempts to conceive for a month or two after stopping the pill is to improve the accuracy of gestational dating by allowing accurate identification of the last menstrual period.

Initiation of Oral Contraception in the Postpartum Period

The 6-week postpartum visit is an anachronism. It is rooted in old texts and teachings from a time when infection was prevalent and before modern methods of contraception were available. The 6-week visit and pelvic examination were based on the understanding that a 6-week period of time would result in sufficient involution of the changes of pregnancy to allow an effective examination that would establish the return of normal pelvic anatomy. It is time for a change. Many women resume sexual activity before the sixth postpartum week, and because ovulation frequently returns before 6 weeks, the obstetrical tradition of scheduling the postpartum visit at 6 weeks should be changed. A 3-week visit would be more effective in avoiding postpartum surprises, preventing postpartum conception by earlier initiation of effective contraception. There is no reason why a complete physical examination cannot be deferred in an asymptomatic woman until the 3-month follow-up visit that is part of good contraceptive care.

Postpartum Return of Ovulation in Nonbreastfeeding Women

In nonbreastfeeding women, gonadotropin levels remain low during the early puerperium and return to normal concentrations during the third to fifth week when prolactin levels

have returned to baseline. In a careful study of 22 women, not one woman ovulated before 25 days after delivery, but 11 ovulated before the sixth postpartum week.[428, 429] In addition, two-thirds of the women ovulated before their first menses. ***Nonbreastfeeding women, therefore, begin to ovulate after 3 postpartum weeks, and the return of menses cannot be used as an indicator.***

The Contraceptive Effect of Lactation

In primitive human societies, the duration of the birth interval was very important for the survival of the young. Throughout human history, no preliterate society achieved a fertility rate at the maximal level possible. The hunter-gatherer, nomadic !Kung women had a high suckling frequency and gave birth about every 4 years.[430] Lactational amenorrhea, lasting up to 2 years, has been nature's most effective form of contraception.[431] Indeed, lactation is the mechanism that maintains a reasonable interval between pregnancies in all non-seasonally breeding animals. In Africa and Asia, breastfeeding reduces the fertility rate by an average of about 30%.[432] Birth intervals of less than 2 years are associated with a greater incidence of low birth weight, preterm birth, and neonatal death for the new infant and malnutrition, infection, and increased second year mortality for the previous child.[433]

The contraceptive effectiveness of lactation, i.e., the length of the interval between births, depends on the level of nutrition of the mother, the intensity of suckling, and the extent to which supplemental food is added to the infant diet. Well-nourished and undernourished women resume ovulating at the same time postpartum.[434]

Amenorrheic women who exclusively breastfeed (full breastfeeding) at regular intervals, including nighttime, during the first 6 months postpartum have the contraceptive protection equivalent to that provided by oral contraception (98% efficacy); but with menstruation or after 6 months, the chance of ovulation increases.[429, 435] With full or nearly full breastfeeding, approximately 70% of women remain amenorrheic through 6 months and only 37% through 1 year; nevertheless with exclusive breastfeeding, the contraceptive efficacy at 1 year is high, about 92%.[435] Fully breastfeeding women commonly have some vaginal bleeding or spotting in the first 8 postpartum weeks, but this bleeding is not due to ovulation.[436] Half of women studied who are not fully breastfeeding ovulate before the sixth week. Lactation, therefore, provides a contraceptive effect, but it is variable and not reliable for every woman.

Prolactin concentrations are increased in response to the repeated suckling stimulus of breastfeeding. Given sufficient intensity and frequency, prolactin levels will remain elevated. Under these conditions, follicle-stimulating hormone (FSH) concentrations are in the low normal range (having risen from extremely low concentrations at delivery to follicular range in the 3 weeks postpartum) and luteinizing hormone (LH) values are also in the low normal range. These low levels of gonadotropins do not allow the ovary during lactational hyperprolactinemia to display follicular development and secrete estrogen. Therefore, vaginal dryness and dyspareunia are commonly reported by breastfeeding women. *The use of vaginal estrogen preparations is discouraged because absorption of the estrogen can lead to inhibition of milk production. Vaginal lubricants should be used until ovarian function and estrogen production return.*

The Return of Ovulation in Breastfeeding Women

The return of ovulation in breastfeeding women has been documented in women from all parts of the world. In Chile, 14% of women ovulated during full breastfeeding, although

full nursing provided effective contraception up to 3 months postpartum.[437, 438] It has been argued that the threshold for suppression of ovulation is at least five feedings for a total of at least 65 min/day suckling duration.[439] However in the studies from Chile, the frequency of nursing was the same in breastfeeders who ovulated and those who did not.

In Mexico, a study of 29 breastfeeding mothers and 10 nonbreastfeeders observed that in the absence of bleeding and supplementary feedings, 100% of the breastfeeders remained anovulatory for 3 months postpartum, and 96% up to 6 months.[440] The median time from delivery to first ovulation was 259 days for breastfeeders compared to 119 days for nonbreastfeeders. However, by the third postpartum month, 18% of the breastfeeders had ovulated.

In a well-nourished population in Australia, less than 20% of breastfeeding women ovulated by the sixth postpartum month, and less than 25% menstruated.[441] Neither time of first supplement nor the amount of supplement predicted the return of ovulation or menstruation. In other words, even in women giving their infants supplemental feedings, there is effective inhibition of ovulation during the first 6 months of breastfeeding.

What do Couples Really Do?—The Risk of Pregnancy

All too often, writers have emphasized that many couples do not resume sexual intercourse before the sixth postpartum week. What should be emphasized is that longitudinal surveys indicate that impressive numbers do. In England, about one-third of primiparous women resumed intercourse by 6 weeks, and nearly everyone by 3 months.[442] A random survey of all women who delivered in the Grampian region of Scotland in 1990–1991 revealed that 71% had experienced intercourse by 8 weeks postpartum, most by 5 weeks.[443] In another longitudinal follow-up study, nearly 60% of women had resumed intercourse *within* 6 weeks after birth.[444] In North Carolina, 57% of women resumed intercourse by the sixth postpartum week.[445] In Thailand, 35% of women reported resumption of sexual activity before the sixth postpartum week, and no differences were noted comparing vaginal or casarean deliveries or with and without episiotomies.[446] Breast-feeding in a Canadian study was found to be associated with earlier, not later, resumption of intercourse.[447] In an English study, 166 of 328 women had resumed sexual intercourse, most before the sixth postpartum week, and only 55 of the 166 sexually active women recalled medical and nursing advice regarding resumption of sexual activity.[448] In Nigeria, 32% of breast-feeding mothers resumed sexual activity by 6 weeks postpartum, but only 5% began a method of contraception.[449]

Roger Short has documented in Australia that among women who have unprotected intercourse during lactation amenorrhea and use contraception when menses resume, 1.7% become pregnant in the first 6 months of breastfeeding, 7% after 12 months, and 13% after 24 months.[450] In a study of 422 middle-class women in Santiago, Chile, there was only one pregnancy (in month 6) when lactational amenorrhea was consciously relied upon for contraception.[451] This was equal to a cumulative 6-month life-table pregnancy rate of 0.45%. However, this accomplishment required an extensive program of education and support. In this study, 9% of exclusively breastfeeding women had resumption of menses by the end of 3 months and 19% by the end of 6 months. This increased suppression of fertility undoubtedly reflected the intensity of the breastfeeding program and the motivation of the participants. In Chile, the probability of pregnancy in breastfeeding women who are amenorrheic was reported as 0.9% at 6 months and 17% at 112 months, in menstruating women, 36% at 6 months and 55% at 12 months.[452] In Pakistan, women who deliberately chose lactational amenorrhea as a method of contraception experienced a pregnancy rate of only 1.1% at 12 months if they remained amenorrheic.[453]

An international group of researchers in the area of lactational infertility reached the following consensus in 1989, called the Bellagio Consensus (after the site of the conference at Bellagio, Italy)[454]:

"The maximum birth spacing effect of breastfeeding is achieved when a mother 'fully' or nearly fully breastfeeds and remains amenorrheic. When these two conditions are fulfilled, breastfeeding provides more than 98% protection from pregnancy in the first 6 months."

Full breastfeeding means that the infant's total suckling stimulus is directed to the mother. There is no diminution of suckling by supplementation or the use of a pacifier. The Bellagio degree of protection in the first 6 months of full or nearly full breastfeeding has been confirmed in clinical studies.[435, 455]

The World Health Organization conducted a large prospective study examining the relationship between infant feeding and amenorrhea, as well as the rate of pregnancy during lactational amenorrhea.[456–459] Women who were still breastfeeding and remained amenorrheic had pregnancy rates of 0.8% at 6 months and 4.4% at 12 months, again confirming the Bellagio Consensus.

The duration of postpartum lochia is variable and can make it difficult to detect the onset of menstrual bleeding. In the World Health Organization study, postpartum lochia was present from a minimum of 2 days to a maximum of 90 days, with an average duration of 27 days.[459] Most women with lactational amenorrhea will not experience true menstrual bleeding before postpartum day 56 (8 weeks). Breastfeeding frequency has no effect on the duration of postpartum lochia.

Therefore, only amenorrheic women who exclusively breastfeed at regular intervals, including nighttime, during the first 6 months have the contraceptive protection equivalent to that provided by oral contraception; with menstruation or after 6 months postpartum, the risk of ovulation increases.[429, 435, 460] Supplemental feeding increases the risk of ovulation (and pregnancy) even in amenorrheic women.[452] Total protection against pregnancy is achieved by the exclusively breastfeeding woman for a duration of only 10 weeks.[436]

Women skilled in the cervical mucus method can detect evidence of fertile type mucus prior to the first menses in the postpartum period. However, there are many false-positive and false-negative interpretations.[461] This method cannot be used with a great deal of confidence until regular menses are resumed.

Postpartum Contraception

A recent pregnancy and a new infant provide strong motivation for the mother to consider contraception. In nonbreastfeeding women, ovulation can occur during the fourth postpartum week. We urge clinicians and patients to start a new tradition: schedule the first postpartum visit during the ***third week after delivery.*** Even breastfeeding women should be evaluated at this time, to consider whether breastfeeding is full and exclusive, or whether an additional contraceptive method is necessary.

Additional contraception is necessary during lactation for most women. That is not to say that full breastfeeding should not be encouraged and that the protection obtained in

the first 6 months of breastfeeding shouldn't be emphasized. But after 3 months, the first ovulation can precede the first menstrual bleed. Half of women studied who are not fully breastfeeding ovulate before the sixth week, the time of the traditional postpartum visit; a visit during the third postpartum week for both breastfeeding and nonbreastfeeding mothers is strongly recommended for contraceptive counseling.

The Rule of 3's.

In the presence of FULL breastfeeding, a contraceptive method should be used beginning in the *third postpartum month.*

With PARTIAL breastfeeding or NO breastfeeding, a contraceptive method should begin during the *third postpartum week.*

After the spontaneous or elective termination of a pregnancy of less than 12 weeks, steroid contraception can be started immediately. After a pregnancy of 12 or more weeks, the third postpartum week rule should be followed if the pregnancy is term or near term to avoid the postpartum risk of venous thromboembolism (discussed later).

The suppression of prolactin secretion with a dopamine agonist (e.g., bromocriptine), not surprisingly, is associated with the return of gonadotropin secretion in the second postpartum week, and an earlier return to ovulation and menses.[462] In women who receive dopamine agonist treatment at or immediately after delivery, contraception is required a week earlier, in the second week postpartum.[463]

Oral contraception even in low-dose formulations has been demonstrated to diminish the quantity and quality of lactation in postpartum women. Although there has been concern regarding the potential hazard of transfer of contraceptive steroids to the infant (a significant amount of the progestational component is secreted into breast milk),[464, 465] no adverse effects have thus far been identified. Because iron is an important factor in the bacteriostatic activity of breast milk, it is good to know that iron and copper concentrations in breast milk are not affected by the use of oral contraceptives.[466]

Women who use oral contraception have a lower incidence of breastfeeding after the 6th postpartum month, regardless of whether oral contraception is started at the first, second, or third postpartum month.[467–469] In adequately nourished women, no impairment of infant growth and development can be detected; presumably compensation is achieved either through supplementary feedings or increased suckling.[470–472] In an 8-year follow-up study of children breastfed by mothers using oral contraceptives, no effect could be detected on diseases, intelligence, or psychological behavior.[473] This study also found that mothers on oral contraceptives lactated a significantly shorter period of time than controls, a mean of 3.7 months vs. 4.6 months in controls.

Because of the concern regarding the impact of oral contraceptives on breastfeeding, a useful alternative is to combine the contraceptive effect of lactation with the progestin-only minipill; there is no evidence for any adverse effect on breastfeeding as measured by milk volume and infant growth and development.[470–472]

In contrast to the combined oral contraceptive, the progestin-only minipill even provides a modest boost to milk production, and women using the minipill breastfeed longer and add supplementary feeding at a later time.[470, 474, 475] The combination of lactation and the progestin-only minipill is associated with near total contraceptive efficacy. In addition, the minipill can protect against the bone loss associated with lactation, a potential advantage in undernourished women.[476] Because of the slight positive impact on lactation, the minipill can be started immediately after delivery.[477] Progestins alone do not increase the

risk of venous thrombosis.[478, 479] In breastfeeding, overweight, Latina women with prior gestational diabetes, the progestin-only minipill was associated with a 3-fold increased risk of non-insulin dependent diabetes mellitus.[251] It is not known whether this might be a risk in all women who have experienced gestational diabetes; a prudent course would be to advise other methods for this special group of women.

The Risk of Postpartum Venous Thromboembolism

Although venous thromboembolism is a rare event, the potential for a catastrophic consequence makes this an important consideration. In the past, the risk of venous thromboembolism was believed to be concentrated in the postpartum period. It is now apparent that the postpartum incidence of this problem has decreased, and antepartum diagnosis is now more common. Undoubtedly a contributing factor to this change is the now common practice of early ambulation after delivery. Over a 30-year period in Minnesota, the incidence of antepartum venous thromboembolism remained constant, but postpartum cases decreased more than 2-fold; nevertheless the number of cases was still 5 times higher among postpartum women compared with pregnant women.[480] In a large American surveillance study, half of the events occurred during pregnancy and half postpartum.[481] In Spain, however, cases were slightly more frequent during pregnancy compared with the postpartum period.[482] In a large multicenter American registry, there were almost twice as many antepartum cases than postpartum cases.[483]

In Sweden, over the 10-year period of the 1990s, the venous thromboembolism mortality rate was similar comparing pregnancy with oral contraceptive users, about 8 per million women per year; the incidence rate was about 100 cases per 100,000 women per year.[484, 485] Most cases of venous thromboembolism occur in individuals with well-recognized risk factors, including older age, obesity, severe varicose veins, cancer, fractures, surgery, and inflammatory bowel disease.[486] Pregnancy and combined estrogen-progestin contraception must be added to this list.

The increase of venous thromboembolism begins shortly after conception, is maintained throughout pregnancy and the first week postpartum, and then gradually declines, reaching baseline levels about 4 to 6 weeks postpartum.[480] Venous thromboembolism more than 6 weeks postpartum is very uncommon. To minimize the risk of postpartum venous thromboembolism, good contraceptive practice has for decades emphasized the avoidance of exposure to pharmacologic levels of estrogen immediately after delivery. ***Therefore, we recommend that nonbreastfeeding mothers use a progestin-only contraceptive method beginning in the third postpartum week; a change to a combination estrogen-progestin method can be initiated in the seventh postpartum week. This recommendation also applies to the vaginal and transdermal methods of estrogen-progestin contraception.***

Other Considerations

Prolactin-Secreting Adenomas

Because estrogen is known to stimulate prolactin secretion and to cause hypertrophy of the pituitary lactotrophs, it is appropriate to be concerned over a possible relationship between oral contraception and prolactin-secreting adenomas. Case-control studies have uniformly concluded that no such relationship exists.[487, 488] Data from both the Royal College of

General Practitioners and the Oxford Family Planning Association studies indicated no increase in the incidence of pituitary adenomas.[400, 489] Previous use of oral contraceptives is not related to the size of prolactinomas at presentation and diagnosis.[489, 490] Oral contraception can be prescribed to patients with pituitary microadenomas without fear of subsequent tumor growth.[491, 492] ***We have routinely prescribed oral contraception to patients with pituitary microadenomas and have never observed evidence of tumor growth.***

Postpill Amenorrhea

The approximate incidence of "postpill amenorrhea" is 0.7–0.8%, which is equal to the incidence of spontaneous secondary amenorrhea,[427, 493, 494] and there is no evidence to support the idea that oral contraception causes secondary amenorrhea. If a cause-and-effect relationship exists between oral contraception and subsequent amenorrhea, one would expect the incidence of infertility to be increased after a given population discontinues use of oral contraception. In those women who discontinue oral contraception in order to get pregnant, 50% conceive by 3 months, and after 2 years, a maximum of 15% of nulliparous women and 7% of parous women fail to conceive.[427] rates comparable with those quoted for the prevalence of spontaneous infertility. Attempts to document a cause-and-effect relationship between oral contraceptive use and secondary amenorrhea have failed.[495] Although patients with this problem come more quickly to our attention because of previous oral contraceptive use and follow-up, there is no cause-and-effect relationship. Women who have not resumed menstrual function within 12 months should be evaluated as any other patient with secondary amenorrhea.

Use During Puberty

Should oral contraception be advised for a young woman with irregular menses and oligo-ovulation or anovulation? The fear of subsequent infertility should not be a deterrent to providing appropriate contraception. Women who have irregular menstrual periods are more likely to develop secondary amenorrhea whether they use oral contraception or not. The possibility of subsequent secondary amenorrhea is less of a risk and a less urgent problem for a young woman than leaving her unprotected. The need for contraception takes precedence.

There is no evidence that the use of oral contraceptives in the pubertal, sexually active girl impairs growth and development of the reproductive system.[419] Again, the most important concern is and should be the prevention of an unwanted pregnancy. For most teenagers, oral contraception, dispensed in a 28-day package for better compliance, is the contraceptive method of choice. However, even better continuation can be achieved with the vaginal and transdermal methods of estrogen-progestin contraception (Chapter 23).

Eye and Ear Diseases

In the 1960s and 1970s, there were numerous anecdotal reports of eye disorders in women using oral contraception. An analysis of the two large British cohort studies (the Royal College of General Practitioners' Study and the Oxford Family Planning Association Study) could find no increase in risk for the following conditions: conjunctivitis, keratitis, iritis, lacrimal disease, strabismus, cataract, glaucoma, and retinal detachment.[496] Retinal vascular

lesions were slightly more common in recent users of oral contraception, but this finding did not reach statistical significance. Contact lens may be less well tolerated, requiring more frequent use of wetting solutions.

The Oxford Family Planning Association Study could detect no evidence of any adverse effects of oral contraception on ear disorders.[497]

Multiple Sclerosis

There is no evidence in third cohort studies (the Royal College of General Practitioners' Study, the Oxford Family Planning Association Study, and the Nurses' Health Study) that there is any effect of oral contraceptive use on the risk or course of multiple sclerosis.[498–500] A case-control study suggested that recent users of oral contraceptives had a substantial reduction in the incidence of multiple sclerosis, perhaps delaying the onset of the disease.[501] A small (7 women on oral contraceptives) prospective study indicated that symptoms increased during the pill-free interval.[502] Patients with multiple sclerosis should consider using the regimen of daily, continuous administration.

Infections and Oral Contraception

Viral STIs

The viral sexually transmitted infections (STIs) include human immunodeficiency virus (HIV), human papillomavirus (HPV), herpes simplex virus (HSV), and hepatitis B (HBV). At the present time, no known associations exist between oral contraception and the viral STIs. Of course, significant prevention includes barrier methods of contraception.

Thus far, most studies have found no association between oral contraceptive use and HIV seropositivity, and some have indicated a protective effect.[503–505] Steroid contraception used by HIV-positive women does not affect the progress of the disease.[506, 507]

Antiretroviral drugs may decrease oral contraceptive efficacy by affecting drug metabolism or causing diarrhea and vomiting. Treatment usually consists of a combination of three or four drugs, and the available studies often reflect drugs and doses not currently used. Therefore, the degree of clinical impact on hormonal contraception, if any, is not well studied.[508] Nevertheless, the studies suggest that some of the antiretroviral drugs are associated with a decrease in steroid levels during oral contraceptive use.

A dual approach is recommended, combining the contraceptive efficacy and protection against PID offered by steroid contraception with the use of a barrier method for prevention of viral STIs and added contraception.

Bacterial STIs

Sexually transmitted infections (STIs) are one of the most common public health problems in the United States. Pelvic inflammatory disease (PID) is usually a consequence of STIs. The best estimate of subsequent tubal infertility is derived from an excellent Swedish report; approximately 12% after one episode of PID, 23% after two episodes, and 54%

after three episodes.[509] Because pelvic infection is the single greatest threat to the reproductive future of a young woman, the now recognized protection offered by oral contraception against pelvic inflammatory disease is highly important.[510–512] ***The risk of hospitalization for PID is reduced by approximately 50–60%, but at least 12 months of use are necessary, and the protection is limited to current users.***[510, 513] Furthermore, if a patient does get a pelvic infection, the severity of the salpingitis found at laparoscopy is decreased.[514, 515] The mechanism of this protection remains unknown. Speculation includes thickening of the cervical mucus to prevent movement of pathogens and bacteria-laden sperm into the uterus and tubes, and decreased menstrual bleeding, reducing movement of pathogens into the tubes as well as a reduction in "culture medium." This protection probably accounts for the greater fertility rate observed in previous users of oral contraception.[419, 420]

The argument has been made that this protection is limited to gonococcal disease, and chlamydial infections may even be enhanced. Most published studies have reported a positive association of oral contraceptives with lower genital tract chlamydial cervicitis.[516, 517] Because lower genital tract infections caused by chlamydia are on the rise (now the most prevalent bacterial STI in the U.S.) and the rate of hospitalization for PID is also increased, it is worthwhile for both patients and clinicians to be alert for symptoms of cervicitis or salpingitis in women on oral contraception who are at high risk of sexually transmitted infections (multiple sexual partners, a history of STI, or cervical discharge). The mechanism for the association between chlamydial cervicitis and oral contraceptives may be the well recognized extension of the columnar epithelium from the endocervix out over the cervix (ectopia) that occurs with oral contraceptive use.[518] This ectropion may allow a more effective collection of cervical specimens for culture, thus introducing detection bias into the epidemiologic studies.

Despite this potential relationship between oral contraception and chlamydial infections, it should be emphasized that there is no evidence for an impact of oral contraceptives increasing the incidence of tubal infertility.[519] In fact, case-control studies indicated that oral contraceptive users with chlamydia infection are protected against symptomatic PID.[520, 521] A case-control study suggested that oral contraceptive users are more likely to harbor unrecognized endometritis, and that this would explain the discrepancy between the observed rates between lower and upper tract infection.[522] However, this would not explain the lack of an association between oral contraceptive use and tubal infertility. Thus, the influence of oral contraception on the upper reproductive tract may be different than on the lower tract. These observations on fertility are derived mostly, if not totally, from women using oral contraceptives containing 50 μg of estrogen. The continued progestin dominance of the lower dose formulations, however, should produce the same protective impact. Early evidence indicated protection with low-dose oral contraceptives, but a later study failed to find a reduction in upper genital tract disease associated with oral contraceptives or barrier methods.[513, 523]

Other Infections

In the British prospective studies of high-dose oral contraceptives, urinary tract infections were increased in users of oral contraception by 20%, and a correlation was noted with estrogen dose. An increased incidence of cervicitis was also reported, an effect related to the progestin dose. The incidence of cervicitis increased with the length of time the pill was used, from no higher after 6 months to 3 times higher by the sixth year of use. A significant increase in a variety of viral diseases, e.g., chickenpox, was observed, suggesting steroid effects on the immune system. The prevalence of these effects with low-dose oral contraception is unknown.

Oral contraception appears to protect against bacterial vaginosis and infections with *Trichomonas*.[524–527] Evidence is lacking to convincingly implicate oral contraception with vaginal infections with *Candida* species[524]; however, clinical experience is sometimes impressive when recurrence and cure repeatedly follow use and discontinuation of oral contraception.

Patient Management

Absolute Contraindications to the Use of Oral Contraception

1. Thrombophlebitis, thromboembolic disorders (including a close family history, parent or sibling, suggestive of an inherited susceptibility for venous thrombosis), cerebral vascular disease, coronary occlusion, or a past history of these conditions, or conditions predisposing to these problems.
2. Markedly impaired liver function. Steroid hormones are contraindicated in patients with hepatitis until liver function tests return to normal.
3. History of coronary heart disease or cerebrovascular disease.
4. Migraine headaches with aura.
5. Diabetes mellitus with vascular disease.
6. Known or suspected breast cancer.
7. Undiagnosed abnormal vaginal bleeding.
8. Known or suspected pregnancy.
9. Smokers over the age of 35.
10. Severe hypercholesterolemia or hypertriglyceridemia.
11. Uncontrolled hypertension.

Relative Contraindications Requiring Clinical Judgment and Informed Consent

1. Migraine headaches without aura.
2. Controlled hypertension.
3. Uterine leiomyoma.
4. Gestational diabetes.
5. Diabetes mellitus.
6. Elective surgery.
7. Seizure disorders.
8. Obstructive jaundice in pregnancy.

9. Sickle cell disease or sickle C disease.
10. Gallbladder disease.
11. Mitral valve prolapse.
12. Systemic lupus erythematosus.
13. Hyperlipidemia.
14. Smoking.
15. Hepatic disease

Clinical Decisions

Surveillance

Many women can be prescribed hormonal contraception without a clinical breast and pelvic examination.[528] Problems requiring further evaluation can be identified with a careful medical history and measurement of blood pressure. Subsequently, in view of the increased safety of low-dose preparations for healthy young women with no risk factors, patients need be seen only every 12 months for exclusion of problems by history, measurement of the blood pressure, urinalysis, breast examination, palpation of the liver, and pelvic examination with Pap smear. Women with risk factors should be seen every 6 months by appropriately trained personnel for screening of problems by history and blood pressure measurement. Breast and pelvic examinations are necessary only yearly. It is worth emphasizing that better continuation is achieved by reassessing new users within 1–2 months. It is at this time that subtle fears and unvoiced concerns need to be confronted and resolved.

Oral contraception is safer than most people think it is, and the low-dose preparations are extremely safe. Health care providers should make a significant effort to get this message to our patients (and our colleagues). We must make sure our patients receive adequate counseling, either from ourselves or our professional staff. The major reason why patients discontinue oral contraception is fear of side effects.[529] Let's take time to put the risks into proper perspective, and to emphasize the benefits as well as the risks.

Laboratory surveillance should be used only when indicated. Routine biochemical measurements fail to yield sufficient information to warrant the expense. Assessing the cholesterol-lipoprotein profile and carbohydrate metabolism should follow the same guidelines applied to all patients, users and nonusers of contraception. The following is a useful guide as to who should be monitored with blood screening tests for glucose, lipids, and lipoproteins:

Young women, at least once.
Women 35 years or older.
Women with a strong family history of heart disease, diabetes mellitus, or hypertension.
Women with gestational diabetes mellitus.
Women with xanthomatosis.
Obese women.
Diabetic women.

Choice of Pill

The therapeutic principle remains: utilize the formulations that give effective contraception and the greatest margin of safety. You and your patients are urged to choose a low-dose preparation containing less than 50 μg of estrogen, combined with low doses of new or old progestins. Current data support the view that there is greater safety with preparations containing less than 50 μg of estrogen. The arguments in this chapter indicate that all patients should begin oral contraception with low-dose products, and that patients on higher dose oral contraception should be changed to the low-dose preparations. Stepping down to a lower dose can be accomplished immediately with no adverse reactions such as increased bleeding or failure of contraception.

An extended regimen or continuous dosing deserves consideration. Breakthrough and withdrawal bleeding is reduced with these schedules, and because of greater suppression of FSH and follicular growth, it is very likely that the failure rates with typical use will be less.

The multiphasic preparations do have a reduced progestin dosage compared with some of the existing monophasic products; however, based on currently available information there is little difference between the low-dose monophasics and the multiphasics.

The pharmacologic effects in animals of various formulations have been used as a basis for therapeutic recommendations in selecting the optimal oral contraceptive pill. *These recommendations (tailor-making the pill to the patient) have not been supported by appropriately controlled clinical trials. All too often this leads to the prescribing of a pill of excessive dosage with its attendant increased risk of serious side effects*. It is worth repeating our earlier comments on potency. Oral contraceptive potency (specifically progestin potency) is no longer a consideration when it comes to prescribing birth control pills. The potency of the various progestins has been accounted for by appropriate adjustments of dose. Clinical advice based on potency is an artificial exercise that has not stood the test of time. The biologic effect of the various progestational components in current low-dose oral contraceptives is approximately the same. Our progress in lowering the doses of the steroids contained in oral contraceptives has yielded products with little serious differences.

Pill Taking

Effective contraception is present during the first cycle of pill use, provided the pills are started no later than the fifth day of the cycle, and no pills are missed. Thus, starting oral contraception on the first day of menses ensures immediate protection. The Sunday start packages, beginning on the first Sunday following the onset of menstruation, can be easier to remember, and it usually avoids menstrual bleeding on weekends. It is probable, *but not totally certain*, that even if a dominant follicle should emerge in occasional patients after a Sunday start, an LH surge and ovulation would still be prevented.[530] ***We prefer the first-day start and no longer recommend the Sunday start. If the Sunday start is used, it is prudent to advise patients to use added protection in the first week of use.***

The conventional approach to starting oral contraceptives, either with menses or on Sunday, carries with it a delay in achieving contraception for many women. Many clinicians advocate the Quick-Start method, an *immediate or same-day start* on the day the patient receives her prescription, regardless of the patient's day in her cycle.[531] Combined with a backup method for the first week, preferably condoms, an immediate start may avoid unwanted pregnancies occurring during the delay before initiating oral contraception

with the conventional methods. In some instances of appropriate coital history, a sensitive pregnancy test would be a wise precaution before starting treatment. Women who use the immediate start method have better continuation rates and do not experience an increase in breakthrough bleeding.[532, 533]

Starting Choices for Steroid Contraception

1. **Start on first day of menses**
   Advantage: assured immediate protection

2. **Start on first Sunday after onset of menses**
   Advantage: avoids weekend menses
   Disadvantage: backup protection advisable during first week

3. **Quick-Start (Immediate, same-day start)**
   Advantage: better compliance
   Disadvantage: backup protection necessary during first week; rule out pregnancy before beginning

There is no rationale for recommending a pill-free interval "to rest." The serious side effects are not eliminated by pill-free intervals. This practice all too often results in unwanted pregnancies.

How important is it to take the oral contraceptive at the same time every day? Although not well studied, there is reason to believe precise pill taking minimizes breakthrough bleeding. In addition, compliance is improved by a fixed schedule that is habit-forming.

Avoiding Menstrual Bleeding

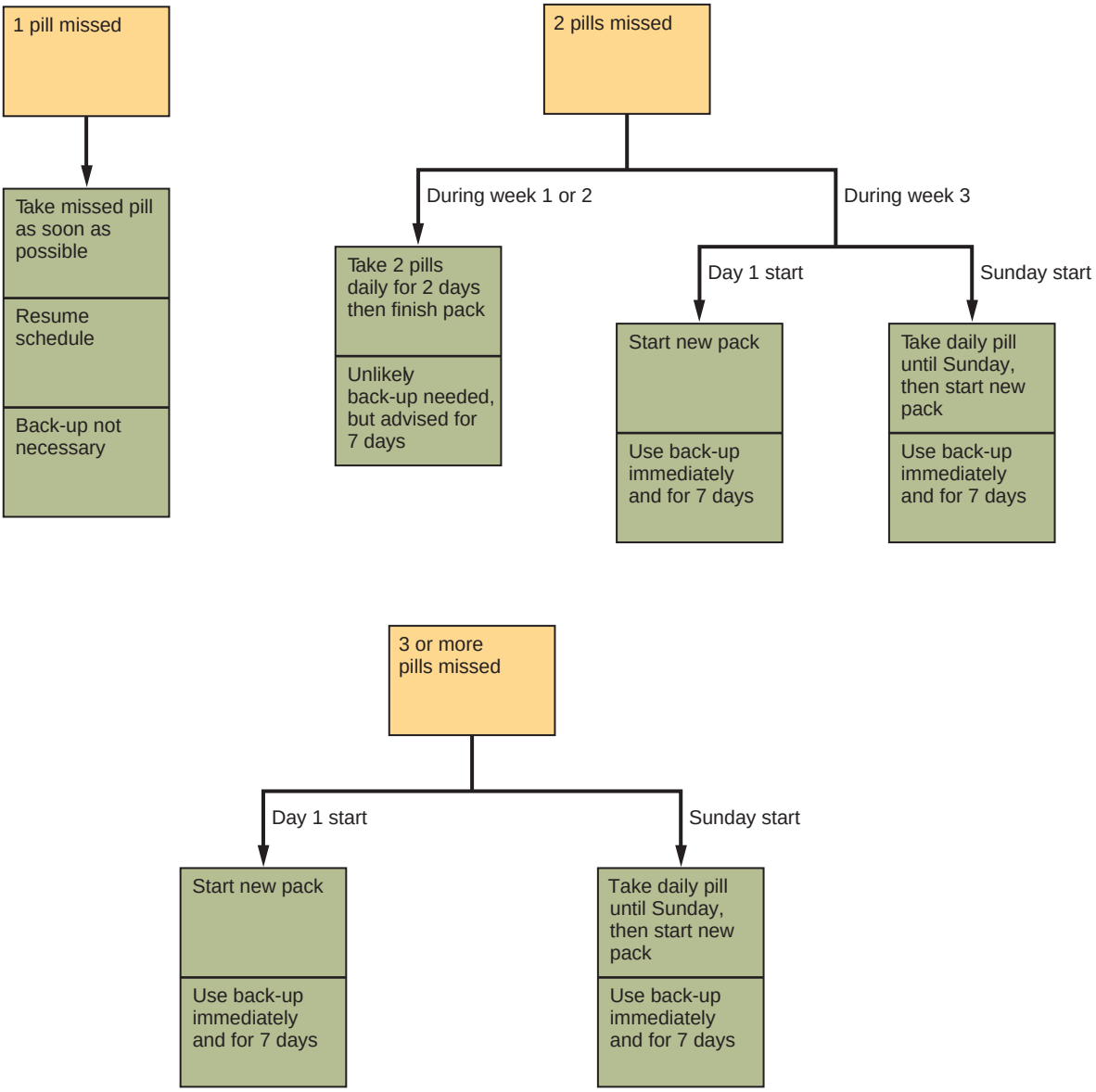
Clinicians for years have prescribed unlimited daily oral contraceptives to treat conditions such as endometriosis, bleeding disorders, menstrual seizures, and menstrual migraine headaches, even to avoid bleeding in athletes and busy individuals. Withdrawal bleeding is not a desired experience for many women, and today, monthly bleeding, periodic bleeding, or no bleeding are choices available for women. Any combination oral contraceptive can be used on a daily continuous dosing basis; even the lowest estrogen dose formulations provide excellent bleeding and side effect profiles in a continuous regimen.[82, 91] A further benefit of continuous dosing is simplification of the pill-taking schedule with the potential of better compliance and a lower failure rate.

Occasionally patients would like to postpone a menstrual period; e.g., for a wedding, holiday, or vacation. This can be easily achieved by omitting the hormone-free interval. Simply start a new package of pills the next day after finishing the series of *active* pills in the previous package.

What to Do When Pills are Missed

Irregular pill taking is a common occurrence. Using an electronic monitoring device to measure compliance, it was apparent that consistency of pill taking is even worse than what patients report; only 33% of women were documented to have missed no pills in cycle 1,

and by cycle 3, about one-third of the women missed 3 or more pills with many episodes of consecutive days of missed pills.[534] These data indicate that women become less careful over time, emphasizing the importance of repeatedly reviewing with patients what to do when pills are missed.



If a woman misses 1 pill, she should take that pill as soon as she remembers and take the next pill as usual. No backup method is needed.

If she misses 2 pills in the first 2 weeks, she should take two pills on each of the next 2 days; it is unlikely that a backup method is needed, but the official consensus is to recommend backup for the next 7 days, especially if the missed pills occurred in the first week. ***If the patient is using an oral contraceptive with 20 µg or less of estrogen and the missed pills occur in the first week, consider the use of emergency contraception.***

If 2 pills are missed in the third week, or if more than 2 active pills are missed at any time, another form of contraception should be used as backup immediately and for 7 days; if a Sunday starter, keep taking a pill every day until Sunday, and on Sunday start a new package; if a non-Sunday starter, start a new package the same day.

Studies have questioned whether missing pills has an impact on contraception. One study demonstrated that skipping 4 consecutive pills at varying times in the cycle did not result in ovulation.[530] Studies in which women deliberately lengthen their pill-fee interval up to 11 days have failed to show signs of ovulation.[84, 535] So far there is no evidence that moving to lower doses has had an impact on the margin of error. Despite greater follicular activity with the lowest-dose oral contraceptives, ovulation is still effectively prevented.[58] However, the studies have involved small numbers of women, and given the large individual variation, it is possible that some women might be at risk with a small increase in the pill-free interval. Although the progestational effects on endometrium and cervical mucus serve to ensure good contraceptive efficacy,[100] conservative advice regarding missing pills is the safest message to convey.

The most prevalent problems that can be identified associated with apparent oral contraceptive failures are vomiting and diarrhea.[101, 102] ***Even if no pills have been missed, patients should be instructed to use a backup method for at least 7 days after an episode of gastroenteritis.***

Clinical Problems

Breakthrough Bleeding

A major continuation problem is breakthrough bleeding. Breakthrough bleeding gives rise to fears and concerns; it is aggravating, and even embarrassing. Therefore, on starting oral contraception, patients need to be fully informed about breakthrough bleeding.

There are two characteristic breakthrough bleeding problems: irregular bleeding in the first few months after starting oral contraception, and unexpected bleeding after many months of use. Effort should be made to manage the bleeding problem in a way that allows the patient to remain on low-dose oral contraception***. There is no evidence that the onset of bleeding is associated with decreased efficacy, no matter what oral contraceptive formulation is used, even the lowest dose products.*** Indeed, in a careful study, breakthrough bleeding did not correlate with changes in the blood levels of the contraceptive steroids.[64]

The most frequently encountered breakthrough bleeding occurs in the first few months of use. The incidence is greatest in the first 3 months, ranging from 10–30% in the first month to less than 10% in the third. Breakthrough bleeding rates are higher with the lowest dose oral contraceptives, but not dramatically.[59, 60] Breakthrough bleeding is higher in women who smoke and in smokers who use formulations with 20 μg ethinyl estradiol.[62] However, the differences among the various formulations currently available are of minimal clinical significance. The basic pattern is the same, highest in the first month and a greater prevalence in smokers, especially in later cycles.

Breakthrough bleeding is best managed by good educational anticipatory preparation of the patient at initiation of treatment, with encouragement and reassurance when bleeding occurs. This bleeding usually disappears by the third cycle in the majority of women. Women should be strongly advised to call the clinician before discontinuing contraception. If necessary, even this early pattern of breakthrough bleeding can be treated as outlined below. It is helpful to explain to the patient that this bleeding represents tissue breakdown as the endometrium adjusts from its usual thick state to the relatively thin state allowed by the hormones in oral contraceptives.

Breakthrough bleeding that occurs after many months of oral contraceptive use is a consequence of the progestin-induced decidualization. This endometrium and the blood vessels within the endometrium tend to be fragile and prone to breakdown and asynchronous bleeding.

There are two recognized factors (both preventable) that are associated with a greater incidence of breakthrough bleeding. Consistency of use and smoking increase spotting and bleeding, but inconsistency of pill taking is more important and has a greater effect in later cycles, whereas smoking exerts a general effect at any time.[536] Reinforcement of consistent pill taking can help minimize breakthrough bleeding. Young women who smoke are more likely to discontinue oral contraception,[537] and this may be partly due to irregular bleeding.

Cervical infection can be another cause of breakthrough bleeding; the prevalence of cervical chlamydial infections is higher among oral contraceptive users who report breakthrough bleeding.[538]

If bleeding occurs just before the end of the pill cycle, it can be managed by having the patient stop the pills, wait 7 days, and start a new cycle. If breakthrough bleeding is prolonged or if it is aggravating for the patient, regardless of the point in the pill cycle, control of the bleeding can be achieved with a short course of exogenous estrogen. Conjugated estrogen, 1.25 mg, or estradiol, 2 mg, is administered daily for 7 days when the bleeding is present, no matter where the patient is in her pill cycle. The patient continues to adhere to the schedule of pill taking. Usually, one course of estrogen solves the problem, and recurrence of bleeding is unusual (but if it does recur, another 7-day course of estrogen is effective).

Remember that there is a significant reduction in the number of bleeding and spotting days, as well as the amount of withdrawal bleeding, in women using a 24-day regimen. Studies comparing 24- and 21-day regimens have indicated that some of the breakthrough bleeding experienced by women on 21-day regimens is due to follicular growth, a rise in endogenous estrogen levels, followed by demise of the follicle and estrogen withdrawal bleeding. There are two operative mechanisms, therefore, in breakthrough bleeding: progestin-induced atrophic endometrium with vascular fragility and withdrawal bleeding in response to the rise and fall of endogenous estrogen levels associated with follicular growth and demise. Greater suppression of follicular growth with 24-day regimens results in less bleeding.

Breakthrough bleeding in women using steroid contraception without a break, daily continuous dosing of oral contraceptives, is best treated by discontinuing medication for 3 or 4 days (no more than once every 3 weeks), allowing a withdrawal menstrual sloughing.[539, 540]

Responding to irregular bleeding by having the patient take 2 or 3 pills is not effective. The progestin component of the pill will always dominate; hence, doubling the number of pills will also double the progestational impact and its decidualizing, atrophic effect on the endometrium and its destabilizing effect on endometrial blood vessels. The addition of extra estrogen while keeping the progestin dose unchanged is logical and effective. This allows the patient to remain on the low-dose formulation with its advantage of greater safety. Breakthrough bleeding, in our view, is not sufficient reason to expose patients to the increased risks associated with higher dose oral contraceptives. Any bleeding that is not handled by this routine requires investigation for the presence of pathology.

There is no evidence that any oral contraceptive formulations that are approximately equivalent in estrogen and progestin dosage are significantly different in the rates of breakthrough bleeding. Clinicians often become impressed that switching to another product effectively stops the breakthrough bleeding. It is more likely that the passage of time is the responsible factor, and bleeding would have stopped regardless of switching and regardless of product.

Amenorrhea

With low-dose pills, the estrogen content is not sufficient in some women to stimulate endometrial growth. The progestational effect dominates to such a degree that a shallow atrophic endometrium is produced, lacking sufficient tissue to yield withdrawal bleeding. It should be emphasized that permanent atrophy of the endometrium does not occur, and resumption of normal ovarian function will restore endometrial growth and development. Indeed, there is no harmful, permanent consequence of amenorrhea while on oral contraception.

The major problem with amenorrhea while on oral contraception is the anxiety produced in both patient and clinician because the lack of bleeding may be a sign of pregnancy. The patient is anxious because of the uncertainty regarding pregnancy, and the clinician is anxious because of the medicolegal concerns stemming from the old studies, which indicated an increased risk of congenital abnormalities among the offspring of women who inadvertently used oral contraception in early pregnancy. We reviewed this problem earlier, and emphatically stated that there is no association between oral contraception and an increased risk of congenital malformation, and there is no increased risk of having abnormal children.

The incidence of amenorrhea in the first year of use with low-dose oral contraception is less than 2%. This incidence increases with duration, reaching perhaps 5% after several years of use. It is important to alert patients upon starting oral contraception that diminished bleeding and possibly no bleeding may ensue.

Amenorrhea is a difficult management problem. A pregnancy test will allow reliable assessment for the presence of pregnancy even at this early stage. However, routine, repeated use of such testing is expensive and annoying, and may lead to discontinuation of oral contraception. ***A simple test for pregnancy is to assess the basal body temperature during the END of the pill-free week or the last active pill-free day in a 24-day regimen; a basal body temperature less than 98 degrees (36.7°C) is not consistent with pregnancy, and oral contraception can be continued.***

Many women are reassured with an understanding of why there is no bleeding and are able to continue on the pill despite the amenorrhea. Some women cannot reconcile themselves to a lack of bleeding, and this is an indication for trying other formulations (a practice unsupported by any clinical trials, and, therefore, the expectations are uncertain). But again, this problem does not warrant exposing patients to the greater risks of major side effects associated with higher dose products.

Some clinicians have observed that the addition of extra estrogen for 1 month (1.25 mg conjugated estrogens or 2 mg estradiol daily throughout the 21 days while taking the oral contraceptive) will rejuvenate the endometrium, and withdrawal bleeding will resume, persisting for many months.

Weight Gain

The complaint of weight gain is frequently cited as a major problem with compliance. Yet, studies of the low-dose preparations fail to demonstrate a significant weight gain with oral contraception, and no major differences among the various products.[268–271, 274, 276] This is obviously a problem of perception, a conclusion supported by finding the weight gain identical in treated and placebo groups. The clinician has to carefully reinforce the lack of association between low-dose oral contraceptives and weight gain and focus the patient on the real culprit: diet and level of exercise. Most women gain a moderate amount of weight as they age, whether they take oral contraceptives or not.

Acne

Low-dose oral contraceptives improve acne regardless of which product is used.[242, 266, 267, 275, 541–546] The low progestin doses (including levonorgestrel formulations) currently used are insufficient to stimulate an androgenic response, and they provide effective treatment for acne and hirsutism. In addition to the inhibitory action of the progestational component on LH and the subsequent reduction in ovarian androgen production, estrogen-progestin contraceptives provide a further benefit because of the increase in SHBG levels induced by the estrogen component. The increase in SHBG results in a greater androgen-binding capacity with a decrease in free testosterone levels. The progestins in estrogen-progestin contraceptives also inhibit 5α-reductase activity in skin, further contributing to the clinical impact of oral contraceptives on hirsutism.[547]

Ovarian Cysts

Anecdotal reports suggested that functional ovarian cysts are encountered more frequently and suppress less easily with multiphasic formulations. This observation failed to withstand careful scrutiny.[548, 549] Functional ovarian cysts occurred less frequently in women on higher dose oral contraception.[550] This protection is reduced with the current lower dose products to the point where little effect can be measured.[549, 551–554] Thus, the risk of such cysts is not eliminated; and, therefore, clinicians can encounter such cysts in patients taking any of the oral contraceptive formulations.

Drugs that Affect Efficacy

There are many anecdotal reports of patients who conceived on oral contraceptives while taking antibiotics. There is little evidence, however, that antibiotics such as ampicillin, metronidazole, quinolone, and tetracycline, which reduce the bacterial flora of the gastrointestinal tract, affect oral contraceptive efficacy. Studies indicate that while antibiotics can alter the excretion of contraceptive steroids, plasma levels are unchanged, and there is no evidence of ovulation.[555–558] A review of a large number of patients derived from dermatology practices failed to find an increased rate of pregnancy in women on oral contraceptives and being treated with antibiotics (tetracyclines, penicillins, cephalosporins).[559] The anti-HIV drug, etravirine, does not produce important changes in the pharmacodynamics of ethinyl estradiol and norethindrone.[560]

There is good reason to believe that drugs that stimulate the liver's metabolic capacity can affect oral contraceptive efficacy. St. John's wort must be added to this list.[561] On the other hand, a search of a large database failed to discover any evidence that lower dose oral contraceptives are more likely to fail or to have more drug interaction problems when other drugs are used.[562]

To be cautious, patients on medications that affect liver metabolism should choose an alternative contraceptive. A list, which may not be complete, includes the following:

Carbamazepine (Tegretol)
Felbamate
Lamotrigine
Nevirapine

Oxcarbazepine
Phenobarbital
Phenytoin (Dilantin)
Primidone (Mysoline)
Rifabutin
Rifampicin (Rifampin)
St. John's wort
Topiramate
Vigabatrin
Possibly valproic acid, ethosuximide, griseofulvin, and troglitazone

Other Drug Interactions

Although not extensively documented, there is reason to believe that oral contraceptives potentiate the action of diazepam (Valium), chlordiazepoxide (Librium), tricyclic antidepressants, and theophylline.[563] Thus, lower doses of these agents may be effective in oral contraceptive users. Because of an influence on clearance rates, oral contraceptive users may require larger doses of acetaminophen and aspirin.[564]

Migraine Headaches

True migraine headaches are more common in women, while tension headaches (90% of all headaches) occur equally in men and women.[565] There have been no well-done studies to determine the impact of oral contraception on the severity of migraine headaches. Patients may report that their headaches are worse or better.

There are two categories of migraine headaches: common migraine which is migraine without aura and classic migraine which is migraine with aura (essentially migraine headaches with visual aura or other neurologic symptoms, occurring in 30% of migraine sufferers). These symptoms begin *before* the headache and resolve with the onset of the headache. Symptoms that occur during headaches, especially photophobia, are not signs of aura. Symptoms that indicate a premonition of a headache, such as light or sound sensitivity, poor concentration, and fatigue occurring 1 to 2 days before a headache are also not considered signs of aura.

Clues to Migraine with Aura:

- Scotomata (blind spots) or blurred vision.
- Bright zig-zag lines.
- Episodes of blindness.
- Numbness, paresthesias.
- Speech difficulties.
- Unilateral symptoms, such as weakness.

Because of the seriousness of this potential complication, the onset of visual symptoms or severe headaches requires a response. If the patient is at a higher dose, a move to a low-dose formulation may relieve the headaches; however, this practice has not be studied. Switching to a different brand is worthwhile, if only to evoke a placebo response. True vascular headaches (migraine with aura) are an indication to avoid or discontinue oral contraception. Oral contraceptives should be avoided in women who have migraine with

complex or prolonged aura, or if additional stroke factors are present (older age, smoking, hypertension, diabetes mellitus, obesity, family history of arterial disease at a young age).[566] Oral contraceptives can be considered in women under age 35, who have migraine *without* aura, and who are otherwise healthy and not smokers.

Migraine headaches, especially with aura, are a risk factor for stroke.[567] The risk is greater in women with hypertension, in smokers, with a family history of migraine, and in women with a long history of migraine or with more than 12 attacks per year of migraine with aura.[568, 569] Studies with high-dose pills indicated that migraine headaches were linked to a risk of stroke. More recent studies reflecting the use of low-dose formulations yield mixed results. One failed to find a further increase in stroke in patients with migraine who use oral contraception, another concluded that the use of oral contraception by migraineurs was associated with a 4-fold increase of the already increased risk of ischemic stroke.[570, 571] The World Health Organization case-control study indicated an increased risk in migraine oral contraceptive users who smoked.[568] Because 20–30% of women experience migraine headaches, one would expect the study populations in the most recent studies of thrombosis to have included substantial numbers of migraineurs. An adverse effect of low-dose oral contraceptives on stroke risk in migraineurs should have manifested itself in the data. The lack of an increased risk of stroke in these studies is reassuring. Nevertheless, it is believed that migraineurs on oral contraceptives have an increased risk of stroke; the absolute risk in a 20-year-old woman is estimated to be 10 per 100,000 and for a 40-year-old woman, 100 per 100,000.[566] Because of the small numbers of young women who have strokes, most studies could not differentiate between migraine with and without aura. However in the American Stroke Prevention in Young Women Study, oral contraceptive use in smokers was associated with an increased risk of stroke in migraineurs with aura.[572]

In some women, a relationship exists between their fluctuating hormone levels during a menstrual cycle and migraine headaches, with the onset of headaches characteristically coinciding with menses (also seen during the pill-free week of oral contraception). We have had personal success (anecdotal to be sure) alleviating headaches by eliminating the menstrual cycle, either with the use of *daily* continuous dosing oral contraceptives or the daily administration of a progestational agent (such as 10 mg medroxyprogesterone acetate) or the use of depot-medroxyprogesterone acetate. An extended regimen, rather than continuous dosing, may also improve menstrual migraine headaches. Some women with migraine headaches have extremely gratifying responses. Women who experience an exacerbation of their headaches with estrogen-progestin contraception should consider one of the progestin-only methods or intrauterine contraception.

In summary, estrogen-progestin contraception should not be used by women who have migraine with aura. Women who have migraine without aura and who are less than 35 years of age (the risk of stroke increases with age), healthy, and nonsmoking can use combined steroid contraception. Migraineurs without aura over the age of 35 should avoid the pharmacologic use of estrogen.

Summary: Oral Contraceptive Use and Medical Problems

Hypertension

Low-dose oral contraception can be used in women less than age 35 years old with hypertension well controlled by medication, and who are otherwise healthy and do not smoke. We recommend the lowest estrogen dose formulations.

Pregnancy-Induced Hypertension

Women with pregnancy-induced hypertension can use oral contraception as soon as the blood pressure is normal in the postpartum period.

Uterine Leiomyoma

This is not a contraindication for low-dose oral contraceptives. There is evidence that the risk of leiomyomas was decreased by 31% in women who used higher dose oral contraception for 10 years.[573] However, case-control studies with lower-dose oral contraceptives have found neither a decrease nor an increase in risk, although the Nurses' Health Study reported a slightly increased risk when oral contraceptives were first used in early teenage years.[574–576] One case-control study indicated a decreasing risk of uterine fibroids with increasing duration of oral contraceptive use.[577] The administration of low-dose oral contraceptives to women with leiomyomas does not stimulate fibroid growth, and is associated with a reduction in menstrual bleeding.[578]

Gestational Diabetes

Low-dose formulations do not produce a diabetic glucose tolerance response in women with previous gestational diabetes, and there is no evidence that combined oral contraceptives increase the incidence of overt diabetes mellitus.[250, 251] We believe that women with previous gestational diabetes can use oral contraception with annual assessment of the fasting glucose level. There is a concern with breastfeeding women using the progestin-only minipill (discussed later in this chapter).

Diabetes Mellitus

Oral contraception can be used by diabetic women less than 35 years old who do not smoke and are otherwise healthy (especially an absence of diabetic vascular complications). A case-control study could find no evidence that oral contraceptive use by young women with insulin-dependent diabetes mellitus increased the development of retinopathy or nephropathy.[253] In a 1-year study of women with insulin-dependent diabetes mellitus who were using a low-dose oral contraceptive, no deterioration could be documented in lipoprotein or hemostatic biochemical markers for cardiovascular risk.[254] And finally, no effect of oral contraceptives on cardiovascular mortality could be detected in a group of women with diabetes mellitus.[255] Women with diabetes and vascular disease or major cardiovascular risk factors should avoid pharmacologic doses of exogenous estrogen.

Elective Surgery

The recommendation that oral contraception should be discontinued 4 weeks before elective major surgery to avoid an increased risk of postoperative thrombosis is based on data derived from high-dose pills. If possible, it is safer to follow this recommendation when

a period of immobilization is to be expected. With major surgery and immobilization, including arthroscopy, especially if the patient is overweight, prophylactic treatment should be considered for a current or recent user of oral contraceptives. It is prudent to maintain contraception right up to the performance of a sterilization procedure, and this short, outpatient operation carries very minimal, if any, risk.

Seizure Disorders

Oral contraceptives do not exacerbate epilepsy, and in some women, improvement in seizure control has occurred.[579, 580] A potential problem is repeatedly raised regarding antiepileptic drugs that affect liver metabolism, and that is that they may decrease the effectiveness of oral contraception. Some clinicians advocate the use of higher dose (50 μg estrogen) products; however, no studies have been performed to demonstrate that this higher dose is necessary. Another concern is that moving to a higher dose product increases the estrogen dose and the risk of side effects. Most importantly, even though studies indicate that some antiepileptic drugs lower steroid levels in oral contraceptive users, ovulation or pregnancies could not be detected in these studies.[581–584] The efficacy of most antiepileptic drugs is not affected by steroid contraception; however, there is evidence that lamotrigine and valproic acid may need dosage adjustments because of estrogen-induced metabolism.[584–587] Consideration should be given to methods that neither affect antiepileptic drug metabolism nor the method affected by the drugs. These include intrauterine contraception with a copper IUD or the levonorgestrel-releasing IUS,[588] long-acting progestin-only methods, barrier methods, and sterilization.

Obstructive Jaundice in Pregnancy

Not all patients with this history will develop jaundice on oral contraception, especially with the low-dose formulations.

Sickle Cell Disease

Patients with sickle cell trait can use oral contraception. The risk of thrombosis in women with sickle cell disease or sickle C diseases is theoretical (and a medicolegal concern). We believe effective protection against pregnancy in these patients warrants the use of low-dose oral contraception. In the only long-term (10 years) follow-up report of women with sickle cell disease and using oral contraceptives, no apparent adverse effects were observed (at a time when higher dose products were prevalent).[589] A study of erythrocyte deformability in women with sickle cell anemia could detect no adverse effects of contraceptive steroids.[590] Keep in mind that depot-medroxyprogesterone acetate used for contraception is associated with inhibition of sickling and improvement in anemia in patients with sickle cell disease.[591]

Gallbladder Disease

Oral contraception use may precipitate a symptomatic attack in women known to have stones or a positive history for gallbladder disease and, therefore, should either be used very cautiously or not at all.

Mitral Valve Prolapse

Oral contraception use is limited to nonsmoking patients who are asymptomatic (no clinical evidence of regurgitation). There is a small subset of patients with mitral valve prolapse who are at increased risk of thromboembolism. Patients with atrial fibrillation, migraine headaches, or clotting factor abnormalities should consider progestin-only methods or the IUD (prophylactic antibiotics should cover IUD insertion if mitral regurgitation is present).

Systemic Lupus Erythematosus

Oral contraceptive use can exacerbate systemic lupus erythematous, and the vascular disease associated with lupus, when present, represents a contraindication to estrogen-containing oral contraceptives.[592] A large case-control study using the General Practice Research Database in the U.K. indicated an increased risk of this autoimmune disease in current and recent users of oral contraceptives; however, the only significant increase was in users of 50 μg high dose formulations.[593] The progestin-only methods are a good choice for women with systemic lupus erythematosus. However, in patients with stable or inactive disease, without renal involvement and high antiphospholipid antibodies, low-dose oral contraception can be considered.[594]

In a clinical trial conducted in Mexico City, 162 women with systemic lupus erythematosus were randomized to treatment with one of three contraceptive methods: estrogen-progestin oral contraceptives, oral progestin-only, or a copper IUD.[595] Disease activity remained equally mild and stable over 1 year in all three groups. There were no differences in the use of anti-inflammatory drugs. There were four cases of lower-limb thromboses, two in the group receiving oral contraceptives and two with progestin-only pills. All four had low titers of antiphospholipid antibodies. Estrogen-containing contraceptives did not exacerbate systemic lupus erythematosus.

The NIH-supported OC-SELENA (Safety of Estrogens in Lupus Erythematosus National Assessment) trial was a double-blind, randomized trial of 183 women with stable systemic lupus erythematosus treated either with a 35 μg ethinyl estradiol oral contraceptive or placebo and followed for 1 year.[596] Lupus flares, the primary endpoint, occurred equally in the two groups. Venous thrombosis did not occur more frequently in the oral contraceptive group. The results indicated that low-dose, estrogen-progestin oral contraceptives can be used by patients with inactive or stable, moderate systemic lupus erythematosus who are at low risk for thrombosis. Patients with high-titer anticardiolipin antibodies, lupus anticoagulant, or previous thrombosis were excluded from the SELENA study. If hormone therapy is to be provided to these patients, some form of chronic anticoagulation should be considered (such as low-dose aspirin).

These studies are important for at least two good reasons. First, there has been a general clinical impression that exogenously administered estrogens would increase lupus disease activity. Second, there are important effects of oral contraceptives that would be beneficial for patients with lupus. These beneficial effects include: (1) Contraception is a chief component of care for lupus patients because pregnancy outcome is adversely affected by unstable, active disease; (2) Patients with lupus experience major bone loss and an increase in fractures as an unwanted side effect of their medical treatment; and (3) Estrogen-progestin contraceptives may moderate the intensity of lupus.

Hyperlipidemia

Because low-dose oral contraceptives have negligible impact on the lipoprotein profile, hyperlipidemia is not an absolute contraindication, with the exception of very high levels of triglycerides (which can be made worse by estrogen). In women with triglyceride levels greater than 250 mg/dL, estrogen should be provided with great caution. If vascular disease is already present, oral contraception should be avoided. If other risk factors are present, especially smoking, oral contraception is not recommended. Dyslipidemic patients who begin oral contraception should have their lipoprotein profiles monitored monthly for a few visits to ensure no adverse impact. If the lipid abnormality cannot be held in control, an alternative method of contraception should be used.[597] Oral contraceptives containing desogestrel, noregestimate, or gestodene can increase HDL levels, but it is not known if this change is clinically significant. If hypertriglyceridemia is the only concern, keep in mind that the triglyceride response to estrogen is rapid. A repeat level should be obtained in 2–4 weeks. ***A level greater than 750 mg/dL represents an absolute contraindication to estrogen treatment because of the risk of pancreatitis.***

Smoking

Oral contraception is absolutely contraindicated in smokers over the age of 35. In patients 35 years old and younger, heavy smoking (15 or more cigarettes per day) is a relative contraindication. The relative risk of cardiovascular events is increased for women of all ages who smoke and use oral contraceptives; however, because the actual incidence of cardiovascular events is so low at a young age, the real risk is very low for young women, although it increases with age. An ex-smoker (for at least 1 year) should be regarded as a nonsmoker. Risk is only linked to active smoking. Is there room for judgment? Given the right circumstances, low-dose oral contraceptives might be appropriate for a light smoker or the user of a nicotine patch. A 20-μg estrogen formulation is a better choice for smoking women, regardless of age (because this dose of estrogen has no impact on clotting factors and platelet activation).[121, 122]

Hepatic Disease

Oral contraception can be utilized when liver function tests return to normal. Follow-up liver function tests should be obtained after 2–3 months of use.

Hemorrhagic Disorders

Women with hemorrhagic disorders and women taking anticoagulants can use oral contraception, although the safety of estrogen-progestin methods has not been documented in women who are anticoagulated because of a past history of venous thrombosis. Inhibition of ovulation can avoid the real problem of a hemorrhagic corpus luteum in these patients. A reduction in menstrual blood loss is another benefit of importance. Progestin-only methods do not increase the risk of venous thrombosis. Depot medroxyprogesterone acetate would suppress ovulation and the risk of ovarian hemorrhage, whereas the levonorgestrel-releasing IUD is especially effective for reducing menstrual bleeding.

Obesity

An obese woman who is otherwise healthy can use low-dose oral contraception. However, there are special considerations associated with obesity:

- Obesity and aging are independent risk factors for venous thrombosis, and case-control studies have indicated this risk adds to that associated with oral contraceptives.[142, 150, 598, 599] Progestin-only methods should receive serious consideration for use by obese women, especially the levonorgestrel intrauterine system. However, the additional risk with oral contraceptives is less than the risk associated with pregnancy and the postpartum period.[600] ***Oral contraceptives with the lowest doses of estrogen should be used for overweight and older women, but progestin-only methods are an even better choice. The same considerations apply to the other methods of estrogen-progestin contraception, the vaginal and transdermal methods.***
- There is modest evidence that hormonal contraceptive failure is increased in overweight women (over 155 lb).[601–606] On the other hand, no effect of body weight on oral contraceptive failure was detected in a large, prospective European cohort.[607] Older clinical trials excluded women with high body weight, and for this reason, the effect of body weight on contraception was not well studied. Selecting a 50 μg estrogen product for overweight women might overcome a possible failure rate, but this would add the risks associated with a higher dose of estrogen to those already linked with obesity. Keep in mind that the positive conclusions regarding failure rates and weight were based on differences of only 2 to 4 pregnancies per 100 women per year. Efficacy in overweight women would still be greater than that with barrier methods. But most importantly, recent clinical trials, especially those with extended regimens, detected no increase in failure rates associated with heavier body weights.[73, 87, 607, 608]
- The epidemiologic data are not absolutely convincing, but a pharmacokinetic study provides a good reason to believe that excessive body weight is associated with a reduction in steroid contraceptive efficacy. Obese women on oral contraceptives with a traditional 7-day pill-free period took twice as long (10 days) to achieve a steady state concentration of the administered progestin, and this was associated with more frequent ovulation.[609] Outliers can take even longer! ***We strongly encourage the use of an extended regimen or continuous dosing in these patients.***
- Obesity is increasingly being treated with a method of bariatric surgery. Postoperatively, it is currently recommended that pregnancy be avoided for 12 to 18 months. Studies indicate that steroid levels after ingestion of oral contraceptives are lower in these patients for at least several years.[610, 611] Alternative routes of administration are best for these patients, including the vaginal insertion of oral contraceptives (discussed later), injections, implants, and intrauterine contraception.

Benign Breast Disease

Benign breast disease is not a contraindication for oral contraception; with 2 years of use, the condition may improve.

Congenital Heart Disease or Valvular Heart Disease

Oral contraception is contraindicated only if there is marginal cardiac reserve or a condition that predisposes to thrombosis.

Depression

Low-dose oral contraceptives have minimal, if any, impact on mood.

Polycystic Ovaries and Insulin Resistance

Because older, high-dose oral contraceptives increased insulin resistance, it has been suggested that this treatment should be avoided in anovulatory, overweight women. However, low-dose oral contraceptives have minimal effects on carbohydrate metabolism, and the majority of hyperinsulinemic, hyperandrogenic women can be expected to respond favorably to treatment with oral contraceptives.[612] Insulin and glucose changes with low-dose (less than 50 μg ethinyl estradiol) oral contraceptives are so minimal, that it is now believed that they are of no clinical significance.[244] Long-term follow-up studies have failed to detect any increase in the incidence of diabetes mellitus or impaired glucose tolerance (even in past and current users of high-dose pills).[246, 248] Furthermore, there is no evidence of an increase in risk of cardiovascular disease among past users of oral contraceptives.[126, 127] In addition, low-dose oral contraceptives have been administered to women with recent gestational diabetes without an adverse impact, and in women with insulin-dependent diabetes mellitus, low-dose oral contraceptives have not produced deterioration of lipid and biochemical markers for cardiovascular disease or increased the development of retinopathy or nephropathy.[250, 251, 253, 254] The administration of a low-dose oral contraceptive to women with extreme obesity and very severe insulin resistance resulted in only a mild deterioration of glucose tolerance.[613] Impressively, in a follow-up study (about 10 years) of women with polycystic ovaries and hyperinsulinism, comparing oral contraceptive users with nonusers, the metabolic parameters not only did not worsen in the users, but they actually improved, including body weight, glucose tolerance, insulin levels, and HDL-cholesterol levels, which was in striking contrast to the metabolic worsening observed in the nonusers.[614] This experience supports the safety of oral contraceptive treatment for anovulatory, hyperandrogenic, hyperinsulinemic women.

Eating Disorders

In patients with eating disorders, bone density correlates with body weight. The bone response to hormone therapy will be impaired as long as an abnormal weight is maintained and low caloric intake pesists.[615] The failure to respond to estrogen treatment with an increase in bone density may be due to the adverse bone effects of the hypercortisolism associated with stress disorders. Furthermore, because the pubertal gain in bone density is so significant, individuals who fail to experience this adolescent increase may continue to have a deficit in bone mass despite hormone treatment. Reduced menstrual function for any reason early in life (even beyond adolescence) may leave a residual deficit in bone density that cannot be totally retrieved with resumption of menses or with hormone treatment.[616, 617] Steroid contraception can be used in patients with eating disorders, and a failure to gain bone density is a clue that the eating disorder is not resolved.

Pituitary Prolactin-Secreting Adenomas

Low-dose oral contraception can be used in the presence of microadenomas.

Infectious Mononucleosis

Oral contraception can be used as long as liver function tests are normal.

Inflammatory Bowel Disease

There is no association between *low-dose* oral contraception and ulcerative colitis,[278] although an increase in risk was reported with older high-dose estrogen products.[618–620] Oral contraceptives are absorbed mainly in the small bowel; women who have an ileostomy after lower bowel surgery have normal steroid absorption with the use of oral contraceptives.[621, 622]

The risk of regional enteritis (Crohn's disease) was reported to be increased in oral contraceptive users, but only in current smokers.[623] Other reports indicted that smoking and oral contraceptive use were independently associated with an increased risk of both incidence and relapse.[618–620, 624, 625] However these associations were not strong and mostly reflected the use of older, higher-dose formulations. At least one case-control study could detect no link between oral contraceptive use and the incidence of Crohn's disease.[626] In a prospective cohort of women with regional enteritis, no adverse impact of oral contraceptives could be detected on the clinical course, specifically on flare-ups.[627]

The impact of low-dose steroid contraception on the risk of inflammatory bowel disease is negligible. In addition, there is no firm reason to withhold these methods from women with these conditions.

Organ Transplantation

Organ transplantation is followed by immunosuppressive therapy to prevent rejection. Although barrier methods of contraception are usually recommended, there is no reason that postoperative women with good hepatic function and normal blood pressures cannot use steroid contraception.[628, 629] The use of steroid contraception in transplant patients will not only provide wanted, effective contraception, but also prevent the irregular and heavy menses that are commonly experienced.

An Alternative Route of Administration

Occasionally, a situation may be encountered when an alternative to oral administration of contraceptive pills is required. For example, patients receiving chemotherapy can either have significant nausea and vomiting, or mucositis, both of which would prevent oral drug administration. Pregnancy should be avoided in the 18 months following bariatric surgery, but this time period is associated with gastrointestinal malabsorption. The low-dose oral contraceptives can be administered vaginally. Initially, it was claimed that two pills must be placed high in the vagina daily in order to produce contraceptive steroid blood levels comparable with the oral administration of one pill.[630] However, a large clinical trial demonstrated typical contraceptive efficacy with one pill administered vaginally per day.[631] In a comparative study, a major reduction in side effects was associated with vaginal administration.[632]

Athletes and Oral Contraception

Because athletes are often amenorrheic and hypoestrogenic, oral contraceptives provide not only confidence against the risk of an unwanted pregnancy, but also estrogen support against bone loss. This is a situation where bone density measurements are worthwhile. A low bone density can help motivate an athlete to take hormone therapy, and a subsequent bone density measurement that reveals a failure of response to estrogen can indicate the presence of a hidden eating disorder.

Competing athletes are often concerned that oral contraceptives could reduce exercise performance. A rationale for the concern can be traced to the physiologic increase in ventilation during pregnancy, mediated by progesterone. Thus, progestin enhancement of ventilatory response could consume energy otherwise available for athletic performance. Indeed, reports have generated conflicting data as measured by laboratory testing. However, experimental studies that simulate athletic events can find no adverse effects on oxygen uptake or respiratory rate.[633] In fact, one study of strenuous exercise in a laboratory setting documented improved aerobic power with increases in time to exhaustion and total work performed in oral contraceptive users.[634] Another study reported decreased soreness, both perceived and with palpation, after exercise in women using oral contraceptives.[635] Oral contraceptive use has no effect on prevalence or severity of low back pain, a common problem among female athletes.[636]

Oral contraceptives have a lot to offer with no serious drawbacks for athletes. In athletes who wish to avoid menstrual bleeding, oral contraceptives can be administered on a daily basis, with no breaks, preventing withdrawal bleeding. Continuous dosing is also a good choice for women in the military. The vaginal and transdermal methods (Chapter 23) can be used in a similar fashion.

The Noncontraceptive Benefits of Oral Contraception

The noncontraceptive benefits of low-dose oral contraception can be grouped into two main categories: benefits that incidentally accrue when oral contraception is specifically utilized for contraceptive purposes and benefits that result from the use of oral contraceptives to treat problems and disorders.

The noncontraceptive incidental benefits can be listed as follows:

Effective Contraception.
-Less need for induced abortion.
-Less need for surgical sterilization.
Less Endometrial Cancer.
Less Ovarian Cancer.
Less Colorectal Cancer.
Fewer Ectopic Pregnancies.
More Regular Menses.
-Less flow.
-Less dysmenorrhea.
-Less anemia.

Less Salpingitis.
Less Benign Breast Disease.
Increased Bone Density.
Probably Less Endometriosis.
Possibly Less Rheumatoid Arthritis.
Possibly Protection against Atherosclerosis.
Possibly Fewer Fibroids.
Possibly Fewer Ovarian Cysts.

Many of these benefits have been previously discussed. Protection against pelvic inflammatory disease is especially noteworthy and a major contribution to not only preservation of fertility but to lower health care costs. Also important is the prevention of ectopic pregnancies. Ectopic pregnancies have increased in incidence (partly due to an increase in STIs) and represent a major cost for our society and a threat to both fertility and life for individual patients. Of course, prevention of benign and malignant neoplasia is an outstanding feature of oral contraception. A 40% reduction in ovarian cancer and a 50% reduction in endometrial cancer represent substantial protection. In the Oxford Family Planning Association cohort, the use of low-dose oral contraceptives was associated with a declining incidence of benign breast disease with increasing duration of use.[338]

Studies with higher-dose formulations documented in long-term users a 31% reduction in uterine leiomyomas and, in current users, a 78% reduction in corpus luteum cysts and a 49% reduction in functional ovarian cysts.[550] Two case-control studies with low-dose oral contraceptives found no impact on the risk of uterine fibroids, neither increased nor decreased,[574, 575] and one indicated a decreasing risk with increasing duration of use, reaching a 50% reduction after 7 or more years of use (the effect was limited to current users).[577] Epidemiologic studies have indicated that a progressive decline in the incidence of ovarian cysts is proportional to the steroid doses in oral contraceptives.[551, 552] Current low-dose monophasic and multiphasic formulations provide no protection against functional ovarian cysts. [551–554] This apparent weaker protection afforded by the current low-dose formulations makes it very likely that clinicians will encounter such cysts in their patients on oral contraceptives.

The low-dose contraceptives are as effective as higher dose preparations in reducing menstrual flow and the prevalence and severity of dysmenorrhea.[637–639] The use of oral contraception is associated with a lower incidence of endometriosis, although the protective effect is probably limited to current or recent use.[640–642] These benefits involving two common gynecologic problems have an important, positive impact on compliance.

Progestational drugs have been used to treat the pain of endometriosis for a long time. In fact, norethindrone and norethynodrel were approved by the FDA for this purpose in 1957, 3 years before approval of the first oral contraceptive. By 1960, 500,000 women were using these agents, although it is unlikely that all had endometriosis or even dysmenorrhea. Therefore there is an enormous clinical history supporting the use of oral contraceptives for the treatment of endometriosis.

A Japanese double-blind, placebo-controlled, randomized multicenter trial evaluated the use of a low-dose oral contraceptive for the treatment of dysmenorrhea associated with endometriosis.[643] The oral contraceptive, given for 3 out of 4 weeks for 4 cycles, consisted of 35 μg ethinyl estradiol and 1 mg norethindrone. Dysmenorrhea pain interestingly decreased in the placebo group, but the decrease in the oral contraceptive group was about twice as great. The treatment group demonstrated a decrease in pelvic induration that did not achieve statistical significance. Only the treated group demonstrated a reduction in the size of ovarian endometriomas that were larger than 3 cm diameter at baseline. This Japanese study provides clinical trial data for a low-dose oral contraceptive that confirms years

of experience. It is worth noting that a randomized, placebo-controlled trial using a 20 μg estrogen oral contraceptive has documented effective treatment of primary dysmenorrhea in adolescents.[644]

For years, many clinicians have believed that the daily use of an oral contraceptive without a break is more effective for the treatment of endometriosis. Clinicians have also believed that monophasic products are superior to multiphasic products. Unfortunately, these two beliefs are derived from historical experience and reported in the literature as uncontrolled studies. In an Italian prospective study (but not a randomized trial), women experiencing recurrent dysmenorrhea associated with endometriosis while being treated with a cyclic oral contraceptive regimen improved when switched to a daily, continuous dosing regimen with a 20 μg oral contraceptive.[645] However, a randomized trial with 239 women found equally effective reductions in the recurrence of endometriomas comparing cyclic and continuous oral contraceptive regimens over 2 years.[646]

Low-dose estrogen-progestin contraception is a good choice to treat pain associated with endometriosis. It is a less expensive option than GnRH analogues, side effects are not a major problem, and treatment can be maintained for long durations. This is also a good option to maintain suppression of endometriosis after surgical or GnRH analogue treatment; remember that treatments only suppress and don't cure or eliminate endometriosis. Another advantage of oral contraceptive treatment is that endometriosis may be associated with a slight increase in ovarian cancer (as well as adenocarcinoma in endometriosis tissue), and the profound reduction in the risks of ovarian and endometrial cancer well-demonstrated in women without endometriosis is observed equally in women with endometriosis.[647]

An Austrian study concluded that osteoporosis occurs later and is less frequent in women who have used long-term oral contraception.[648] Most studies indicate that prior use of oral contraception is associated with higher levels of bone density and that the degree of protection is related to duration of exposure.[649–655] However, other studies reflecting modern use of low-dose products indicate little impact of oral contraceptive use on bone.[656–658] These measurements of bone density are not as important as the clinical outcome: fractures. The available evidence fails to provide a clear-cut picture. Retrospective studies indicated a reduction in fractures in postmenopausal women who had previously used oral contraceptives.[659–662] In the Royal College of General Practitioners Study, the overall risk of fractures in ever users of oral contraceptives was actually slightly increased.[663] Similar results have been observed in the Oxford Family Planning Association Study.[664] It is likely that the increased risk reflects lifestyle effects among oral contraceptive users, but there was no evidence of a protective effect against fractures. In contrast, a case-control study from Sweden found a reduction in the risk of postmenopausal hip fractures when oral contraceptives (mostly older high dose products) were used after age 40 by women who were not overweight, with an increasing benefit with increasing duration of use.[665] Previous oral contraceptive users are just now becoming elderly and reaching the age of greatest fracture prevalence. Future studies of postmenopausal women should eventually reveal the accurate relationship between oral contraceptive use and osteoporotic fractures.

The literature on rheumatoid arthritis has been controversial, with studies in Europe finding evidence of protection and studies in North America failing to demonstrate such an effect. An excellent Danish case-control study was designed to answer criticisms of shortcomings in the previous literature.[666] Ever use of oral contraception reduced the relative risk of rheumatoid arthritis by 60%, and the strongest protection was present in women with a positive family history. One meta-analysis concluded that the evidence consistently indicated a protective effect, but that rather than preventing the development of rheumatoid arthritis, oral contraception may modify the course of disease, inhibiting the progression from mild to severe disease, whereas a later meta-analysis concluded there was no evidence of a protective effect.[667, 668] More recent studies suggest a reduction in severity with long-term use.[669]

Oral contraceptives are frequently utilized to manage the following problems and disorders:

Definitely Beneficial:
- **-Dysfunctional uterine bleeding.**
- **-Dysmenorrhea.**
- **-Mittelschmerz.**
- **-Endometriosis prophylaxis.**
- **-Acne and hirsutism.**
- **-Hormone therapy for hypothalamic amenorrhea.**
- **-Prevention of menstrual porphyria.**
- **-Control of bleeding (dyscrasias, anovulation).**

Possibly Beneficial:
- **-Functional ovarian cysts.**
- **-Premenstrual syndrome.**

Oral contraceptives have been a cornerstone for the treatment of anovulatory, dysfunctional uterine bleeding; the only randomized, placebo-controlled trial documented the beneficial impact long recognized by clinicians.[639] For patients who need effective contraception, oral contraceptives are a good choice to provide hormone therapy for amenorrheic patients, as well as to treat dysmenorrhea. Oral contraceptives are also a good choice to provide prophylaxis against the recurrence of endometriosis in a woman who has already undergone more vigorous treatment with surgery or the gonadotropin-releasing hormone (GnRH) analogues.

The low-dose oral contraceptives are effective in treating acne and hirsutism. Suppression of free testosterone levels is comparable with that achieved with higher dosage.[541, 670] The beneficial clinical effect is the same with low-dose preparations containing levonorgestrel, previously recognized to cause acne at high dosage.[541, 671] Formulations with desogestrel, gestodene, and norgestimate are associated with greater increases in sex hormone-binding globulin and significant decreases in free testosterone levels. Comparison studies with oral contraceptives containing these progestins can detect no differences in effects on various androgen measurements among the various products or when compared with older products.[39, 543, 672] Theoretically, these products would be more effective in the treatment of acne and hirsutism; however, this has not been documented by clinical studies. It is likely that all low-dose formulations, through the combined effects of an increase in sex hormone-binding globulin and a decrease in testosterone production, produce an overall similar clinical response, especially over time (a year or more).

Oral contraceptives have long been used to speed the resolution of ovarian cysts, but the efficacy of this treatment has not been established. Randomized trials have been performed with women who develop ovarian cysts after induction of ovulation.[673–675] No advantage for the contraceptive treatment could be demonstrated. The cysts resolved completely and equally fast in both treated and nontreated groups. Of course, these were functional cysts secondary to ovulation induction, and this experience may not apply to spontaneously appearing cysts. Two short-term (5 and 6 weeks) randomized studies could document no greater effect of oral contraceptive treatment on resolution of spontaneous ovarian cysts when compared with expectant management.[676, 677] Clinical experience (untested by studies) leads us to believe that oral contraception does provide protection in women against the recurrent formation of ovarian cysts.

Continuation: Failure or Success?

Despite the fact that oral contraception is highly effective, hundreds of thousands of unintended pregnancies occur each year in the U.S. because of the failure of oral contraception.

Worldwide, millions of unintended pregnancies result from poor compliance. In general, the most important determinants of pill failure are age, intention toward a future birth, parity, and marital status. Interestingly, once these factors are accounted for, duration of use, race, ethnicity, and poverty status no longer affected the risk of pill failure.[105] Overall, the failure rate with actual use is as high as 8%. This difference between the theoretical efficacy and actual use largely reflects noncompliance. Noncompliance includes a wide variety of behavior: failure to fill the initial prescription, failure to continue on the medication, and incorrectly taking oral contraception. Compliance (continuation) is an area in which personal behavior, biology, and pharmacology come together. Oral contraceptive continuation reflects the interaction of these influences. Unfortunately, women who discontinue oral contraception often utilize a less effective method or, worse, fail to substitute another method.

There are three major factors that affect continuation:

1. The experience of side effects, such as breakthrough bleeding and amenorrhea, and perceived experience of "minor" problems, such as headaches, nausea, breast tenderness, and weight gain. Multiple side effects dramatically and progressively increase the likelihood of discontinuation.[63, 678] Because these complaints respond well even to placebo treatment,[679] it is reasonable to expect a favorable response to sensitive and attentive counseling, as well as changing to a different product.

2. Fears and concerns regarding cancer, cardiovascular disease, and the impact of oral contraception on future fertility.

3. Nonmedical issues, such as inadequate instructions on pill taking, complicated pill packaging, difficulties arising from the patient package insert, and most importantly, contraceptive access and expense.

The information in this chapter is the foundation for good continuation, but the clinician must go beyond the presentation of information and develop an effective means of communicating that information. We recommend the following approach to the clinician–patient encounter as one way to improve continuation with oral contraception.

1. Explain how oral contraception works.

2. Review briefly the risks and benefits of oral contraception, but be careful to put the risks in proper perspective, and to emphasize the safety and noncontraceptive benefits of low-dose oral contraceptives.

3. Show and demonstrate to the patient the package of pills she will use.

4. Explain how to take the pills:
   -When to start.
   -The importance of developing a daily routine to avoid missing pills.
   -What to do if pills are missed (identify a backup method).
   -Consider the use of an extended regimen (with the potential of greater efficacy).
   -Dispense as many packages as possible; evidence indicates that the more packages dispensed, the higher the continuation rate.[99]

5. Review the side effects that can affect continuation: amenorrhea, breakthrough bleeding, headaches, weight gain, nausea, etc., and what to do if one or more occurs. The key is to provide anticipatory guidance.

6. Explain the warning signs of potential problems: abdominal or chest pain, trouble breathing, severe headaches, visual problems, leg pain or swelling.

7. Ask the patient to be sure to call if another clinician prescribes other medications.

8. Ask the patient to repeat critical information to make sure she understands what has been said. Ask if the patient has any questions.

9. Schedule a return appointment in 1–2 months to review understanding and address fears and concerns; a visit at 3 months is too late because most questions and side effects occur early.[63] Inconsistent use of oral contraceptives is more common in women who are new starters.[680]

10. Make sure a line of communication is open to clinician or office personnel. Ask the patient to call for any problem or concern before she stops taking the oral contraceptives.

11. Inform the patient about the availability and the proper use of emergency contraception.

The Progestin-Only Minipill

The minipill contains a small dose of a progestational agent and must be taken daily, in a continuous fashion.[681, 682] There is no evidence for any major differences in clinical behavior among the available minipill products.

Minipills available worldwide:

1. Micronor, Nor-QD, Noriday, Norod ———— 0.350 mg norethindrone.
2. Microval, Noregeston, Microlut ———— 0.030 mg levonorgestrel.
3. Ovrette, Neogest ———— 0.075 mg norgestrel. (equivalent to 0.0375 mg levonorgestrel)
4. Exluton ———— 0.500 mg lynestrenol.
5. Femulen ———— 0.500 mg ethynodial diacetate.
6. Cerazette ———— 0.075 mg desogestrel.

Mechanism of Action

After taking a progestin-only minipill, the small amount of progestin in the circulation (about 25% of that in combined oral contraceptives) will have a significant impact only on those tissues very sensitive to the female sex steroids, estrogen and progesterone. The contraceptive effect is more dependent on endometrial and cervical mucus effects, because gonadotropins are not consistently suppressed. The endometrium involutes and becomes hostile to implantation, and the cervical mucus becomes thick and impermeable. Approximately 40% of patients will ovulate normally.[683, 684] Tubal physiology may also be affected,

but this is speculative. The progestin-only minipill containing 0.075 mg desogestrel appears to be slightly more effective, probably because it exerts a greater inhibition of ovulation.[685]

Because of the low dose, the minipill must be taken every day at the same time of day. The change in the cervical mucus requires 2–4 h to take effect, and, most importantly, the impermeability diminishes 22 h after administration, and by 24 hours some sperm penetration occurs. Midday administration is recommended.

Ectopic pregnancy is not prevented as effectively as intrauterine pregnancy. Although the overall incidence of ectopic pregnancy is not increased (it is still much lower than the incidence in women not using a contraceptive method), when pregnancy occurs, the clinician must suspect that it is more likely to be ectopic. A previous ectopic pregnancy should not be regarded as a contraindication to the minipill.

There are no significant metabolic effects (lipid levels, carbohydrate metabolism, and coagulation factors remain unchanged),[180, 241, 686, 687] and there is an immediate return to fertility on discontinuation. An increase in the risk of venous thrombosis has not been observed in users of progestin-only minipills containing levonorgestrel, norethindrone, or desogestrel.[150, 167] Only one disturbing observation has been reported; progestin-only oral contraception was associated with about a 3-fold increased risk of diabetes mellitus in lactating women with recent gestational diabetes, an observation that is difficult to explain.[251] Because this increased risk is not observed with the use of combined oral contraceptives, it is speculated that the low levels of estrogen associated with breastfeeding allow an unimpeded progestin effect on insulin resistance.

Efficacy

Failure rates have been documented to range from 1.1 to 9.6 per 100 women in the first year of use.[688] The failure rate is higher in younger women (3.1 per 100 woman-years) compared with women over age 40 (0.3 per 100 woman-years).[689] In motivated women, the failure rate is comparable to the rate (less than 1 per 100 woman-years) with combination oral contraception.[690, 691]

Pill Taking

The minipill should be started on the first day of menses, and although a backup method for the first 7 days has been the standard recommendation, this extra precaution should not be necessary. A backup method for 7 days is necessary with either a Quick-Start or a Sunday start. The minipill can be started immediately postpartum or after a miscarriage or induced abortion.

Pill taking should be keyed to a daily event to ensure regular administration at the same time of the day. If pills are forgotten or gastrointestinal illness impairs absorption, the minipill should be resumed as soon as possible, and a backup method should be used immediately and until the pills have been resumed for at least 2 days. If 2 or more pills are missed in a row and there is no menstrual bleeding in 4–6 weeks, a pregnancy test should be obtained. ***If more than 3 h late in taking a pill, a backup method should be used for 48 h. Cerazette is more forgiving, allowing a 12-h late time period.***

Problems

In view of the unpredictable effect on ovulation, it is not surprising that irregular menstrual bleeding is the major clinical problem. The daily progestational impact on the endometrium also contributes to this problem. Patients can expect to have normal, ovulatory cycles (40–50%), short, irregular cycles (40%), or a total lack of cycles ranging from irregular bleeding to spotting and amenorrhea (10%). This is the major reason why women discontinue the minipill method of contraception.[691] Doxycycline, 100 mg b.i.d. for 5 days, is effective in decreasing bleeding associated with progestin-only treatment, apparently by suppressing matrix metalloproteases in endometrial cells.[692, 693]

Women on progestin-only contraception develop more functional, ovarian follicular cysts.[550, 694] Nearly all, if not all, regress. This is not a clinical problem of any significance. Women who have experienced frequent ovarian cysts would be happier with methods that effectively suppress ovulation (combined oral contraceptives and depot-medroxyprogesterone acetate).

The levonorgestrel minipill may be associated with acne. The mechanism is similar to that seen with Norplant. The androgenic activity of levonorgestrel decreases the circulating levels of sex hormone-binding globulin (SHBG).[695] Therefore free steroid levels (levonorgestrel and testosterone) will be increased despite the low dose. This is in contrast to the action of combined oral contraception where the effect of the progestin is countered by the estrogen-induced increase in SHBG.

The incidence of the other minor side effects is very low, probably at the same rate that would be encountered with a placebo.

Clinical Decisions

There are two situations in which excellent efficacy, probably near-total effectiveness, is achieved: lactating women and women over age 40. In lactating women, the contribution of the minipill is combined with prolactin-induced suppression of ovulation, adding up to very effective protection.[696] In breastfeeding, overweight, Latina women with prior gestational diabetes, the progestin-only minipill was associated with a 3-fold increased risk of non-insulin dependent diabetes mellitus.[251] It is not known whether this might be a risk in all women who have experienced gestational diabetes; a prudent course would be to advise other methods for this special group of women. In women over age 40, reduced fecundity adds to the minipill's effects.

There is another reason why the minipill is a good choice for the breastfeeding woman. There is no evidence for any adverse effect on breastfeeding as measured by milk volume and infant growth and development.[470–472] In fact, there is a modest positive impact; women using the minipill breastfeed longer and add supplementary feeding at a later time.[474] Because of the slight positive impact on lactation, the minipill can be started immediately after delivery. A study investigating the impact of early initiation found no adverse effects on breastfeeding.[477]

The minipill is a good choice in situations where estrogen is contraindicated, such as women smokers over the age of 35 or patients with serious medical conditions (diabetes with vascular disease, severe systemic lupus erythematosus,[697] cardiovascular disease). It should be noted that the freedom from estrogen effects, although likely, is presumptive. Substantial data, for example on associations with vascular disease, blood pressure, and cancer, are not available because relatively small numbers have chosen to use this method

of contraception. On the other hand, it is logical to conclude that any of the progestin effects associated with the combination oral contraceptives can be related to the minipill according to a dose-response curve; all effects should be reduced. Both the World Health Organization case-control study and the Transnational case-control study could find no indication for increased risks of stroke, myocardial infarction, or venous thromboembolism with oral progestin-only contraceptives.[478, 479] No impact can be measured on the coagulation system.[686, 698] The minipill can probably be used in women with previous episodes of thrombosis, and the package insert in the United States was revised, eliminating vascular disease as a contraindication.

The minipill is a good alternative for the occasional woman who reports diminished libido on combination oral contraceptives, presumably due to decreased androgen levels. The minipill should also be considered for the few patients who report minor side effects (gastrointestinal upset, breast tenderness, headaches) of such a degree that the combination oral contraceptive is not acceptable.

Because of the relatively low doses of progestin administered, patients using medications that increase liver metabolism should avoid this method of contraception. These drugs include the following:

Carbamazepine (Tegretol)
Felbamate
Lamotrigine
Nevirapine
Oxcarbazepine
Phenobarbital
Phenytoin (Dilantin)
Primidone (Mysoline)
Rifabutin
Rifampicin (Rifampin)
St. John's wort
Topiramate
Vigabatrin
Possibly valproic acid, ethosuximide, griseofulvin, and troglitazone

Do the noncontraceptive benefits associated with combination oral contraception apply to the minipill? Studies are unable to help us with this issue, again because of the relatively small numbers of users. However, the progestin impact on cervical mucus, endometrium, and ovulation leads one to think the benefits will be present (reduced risks of pelvic infection, endometrial cancer, and ovarian cancer). Although limited by small numbers, one case-control study indicated that protection against endometrial cancer was even greater with progestin-only pills than with combination oral contraceptives.[699]

Good efficacy with the minipill requires regularity, taking the pill at the same time each day. There is less room for forgetting, and, therefore, the minipill is probably not a good choice for a disorganized adult or for the average adolescent.

Emergency Postcoital Contraception

The use of a method after intercourse to prevent pregnancy is commonly called postcoital contraception, or the "morning after" treatment. "Emergency contraception" is a more accurate and appropriate name, indicating the intention to be one-time protection. It is an

important option for patients, and should be considered when condoms break, sexual assault occurs, if diaphragms or cervical caps dislodge, or with the lapsed use of any method. In studies at abortion units, 50–60% of the patients would have been suitable for emergency contraception and would have used it if readily available.[700, 701] In the U.S., it is estimated that emergency contraception could annually prevent 1.7 million unintended pregnancies and the number of induced abortions would decrease by about 40%.[702] Nevertheless, increasing access to emergency contraception has not had an impact on pregnancy or abortion rates in clinical trials, apparently because the actual use is insufficient to have an impact on the general population.[703–705] Availability must be accompanied by education and motivation. Clinicians should be aware that younger adolescents can safely use emergency contraception.[706]

Many women do not know of emergency contraception, and it has been difficult to obtain.[701, 707] Even if women are aware of this method, accurate and detailed knowledge is lacking.[708] A favorable attitude toward this method requires both knowledge and availability. Availability should be favorably influenced in the U.S. by the recent Food and Drug Administration approval making levonorgestrel emergency contraception available without a prescription to women older than age 16.

Women who have used emergency contraception are very satisfied with the method, and most importantly, do not express an intention to substitute this method for regular contraception.[709] The use of emergency contraception by adolescents is not associated with an increase in STIs.[710]

Information for patients and clinicians, including the latest available products, can be obtained from the web site and hot line maintained by the Office of Population Research at Princeton University:

http://ec.princeton.edu
Telephone hotline: 1-888-NOT-2-LATE (1-888-668-2528)

The use of large doses of estrogen to provide emergency contraception was pioneered by Morris and van Wagenen at Yale in the 1960s. The initial work in monkeys led to the use of high doses of diethylstilbestrol (25–50 mg/day) and ethinyl estradiol in women.[711] It was quickly appreciated that these extremely large doses of estrogen were associated with a high rate of gastrointestinal side effects. Albert Yuzpe developed a method utilizing a combination oral contraceptive, resulting in an important reduction in dosage.[712] The following treatment regimens, or their generic equivalents, have been documented to be effective:

Ovral: 2 tablets followed by 2 tablets 12 h later.

Alesse: 5 tablets followed by 5 tablets 12 h later.

Lo Ovral, Nordette, Levlen, Triphasil, Trilevlen: 4 tablets followed by 4 tablets 12 h later.

THE METHOD OF CHOICE FOR EMERGENCY CONTRACEPTION IS LEVONORGESTREL ALONE:

Levonorgestrel in a dose of 0.75 mg given twice, 12 h apart, is more successful and better tolerated than the combination oral contraceptive method.[713, 714] In many countries, special packages of 0.75 mg levonorgestrel (Plan B, NorLevo, Vikela) are available for emergency contraception. Greater efficacy and fewer side effects make low-dose levonorgestrel the treatment of choice.

In the U.S., Plan B containing only levonorgestrel (two 0.75 mg tablets) was approved by the Food and Drug Administration for nonprescription sales to women aged 17 and older, and by prescription for those younger, one tablet taken within 120 h of intercourse and the

second 12 h later. ***The two tablets can be combined into a single, one-time dose of 1.5 mg levonorgestrel with no loss of efficacy or increase in side effects.***[715, 716] ***The single-dose product, Plan B One-Step, is approved and available in the U.S., also without a prescription for women 17 and older.*** The two-dose product is now available in the U.S. in a generic version known as Next Choice.

Clinicians should consider providing advance information and a prescription or an emergency contraceptive kit to patients (a kit can be a simple envelope containing instructions and the appropriate number of oral contraceptives) to be taken when needed. It would be a major contribution to our efforts to avoid unwanted pregnancies for all patients without contraindications to oral contraceptives to have emergency contraception available for use when needed. In our view, this would be much more effective in reducing the need for abortion than waiting for patients to call. In studies of self-administration, adult women in Scotland and younger women in California increased the use of emergency contraception without adverse effects such as increasing unprotected sex.[717–719] Women are able to use this nonprescription access effectively and do not develop a reliance on emergency contraception as a regular method.[720]

Mechanism and Efficacy

There is strong evidence that treatment with emergency contraception acts primarily by preventing or delaying ovulation and by preventing fertilization.[721–725] Studies have indicated that emergency contraception does not prevent implantation.[726–728] Experiments in monkeys and rats could detect no effect of a high dose of levonorgestrel administrated postcoitally once fertilization had occurred.[729, 730] ***The evidence indicates that a postfertilization effect does not contribute to the efficacy of emergency contraception***.[724, 729–732] ***Clinicians, pharmacists, and patients can be reassured that treatment with emergency contraception is not an abortifacient.***

Efficacy has been confirmed in large clinical trials and summarized in complete reviews of the literature.[714, 733–735] Treatment with high doses of estrogen or with levonorgestrel yields a failure rate of approximately 1%, with the combination oral contraceptive, about 2–3%. The failure rate is lowest with high doses of ethinyl estradiol given within 72 h (0.1%), but the side effects make this a poor choice. In general clinical use, the method with oral contraceptives can reduce the risk of pregnancy by about 75%; this degree of reduction in probability of conception (given the relatively low chance, about 8%, for pregnancy associated with one act of coitus).[736] yields the 2% failure rate measured in clinical studies.[737–739]

Results with levonorgestrel are better, about an 85% reduction in the risk of pregnancy; in the worldwide World Health Organization study, the risk of pregnancy was 60% lower with the levonorgestrel-only method compared with the oral contraceptive method, with less than half as much nausea and vomiting.[714]

Treatment Method

Treatment should be initiated as soon after exposure as possible, and the standard recommendation is that it be no later than 120 h. Careful assessment of the reported experience with emergency contraception indicated that the method is equally effective when started on the first, second, or third day after intercourse (which would allow user-friendly scheduling), and that efficacy extends beyond 72 h.[740, 741] Data from the World Health Organization randomized, clinical trial, however, support the importance of timing, finding a reduction in efficacy after 72 h, and the greatest protection occurring when the medication is taken within 24 h of intercourse.[742] Postponing the dose by 12 h raises the chance of pregnancy by

almost 50%. For this reason, the treatment should be initiated as soon as possible after sexual exposure, an important argument in favor of advance provision.

We should emphasize, in case the patient is already pregnant, that there is no evidence that exposure to the amounts of estrogen and progestin in oral contraceptives is an abortifacient or teratogenic.[409, 411, 412] Thus, emergency contraception is ineffective in the presence of an established pregnancy. A delay in menses after treatment warrants testing for pregnancy and consideration for the possibility of an ectopic pregnancy.

When using oral contraceptives for emergency contraception, it is worth adding an antiemetic, oral or suppository, to the treatment; a long-acting nonprescription agent, 25 or 50 mg meclizine (Bonine, Dramamine II, Antivert) is recommended, to be taken one hour before the emergency contraception treatment. Side effects reflect the high doses used: nausea and vomiting, 50% and 20% with estrogen-progestin oral contraceptives, but only 18% and 4% with levonorgestrel.[714–716] Other reactions include breast tenderness, headache, dizziness, and bleeding or spotting in the month after treatment. If a patient vomits within an hour after taking pills, additional pills must be administered as soon as possible. Nausea and vomiting are experienced at such a lower rate with the levonorgestrel-only method that an antiemetic is not necessary.

An analysis of the U.K. General Practice Research Database could find no evidence for an increased risk of venous thromboembolism with the short-term use of oral contraceptives for emergency contraception (indeed, no cases were found for as long as 60 days after use in more than 100,000 episodes of use).[743] Although short-term treatment with combined oral contraceptives has been documented to have no effect on clotting factors, in our view the usual contraindications for oral contraception apply to this use.[744] ***Because of the high dose of estrogen, emergency contraception with combined oral contraceptives should not be provided to women with either a personal or close family history (parent or sibling) of idiopathic thrombotic disease.*** For women with a contraindication to exogenous estrogen, the progestin-only method with levonorgestrel should be used for emergency contraception. The levonorgestrel-only method is the treatment of choice anyway because of greater efficacy and fewer side effects.

A 3-week follow-up visit should be scheduled to assess the result, and to counsel for routine contraception. ***Even better, a method of contraception should be initiated immediately after the use of emergency contraception to avoid an unwanted pregnancy.***

Could other combination oral contraceptive products be used? A norethindrone-ethinyl estradiol combination was found to be equally effective to the levonorgestrel-ethinyl estradiol formulation, and it is likely that any combination oral contraceptive would be successful.[745] However, this is a moot point because the levonorgestrel-only method is now the treatment of choice. The sustained release of estrogen or progestins from intrauterine (the levonorgestrel-releasing IUS), subdermal, or vaginal delivery methods cannot be used as emergency contraception because systemic hormone concentrations are too low and too slowly achieved.

The Use of Progesterone Receptor Modulators for Emergency Contraception

Mifepristone

Mifepristone in a single oral dose of 600 mg is associated with markedly less nausea and vomiting than with oral contraceptives and an efficacy rate of nearly 100%.[746, 747]

Mifepristone is marked for emergency contraception in China. In randomized trials, 10 mg mifepristone was as effective as 25, 50, or 600 mg, preventing about 80–85% of expected pregnancies (the same efficacy and side effects as with the levonorgestrel method), with a slight decrease in efficacy when treatment was delayed to 5 days after intercourse.[715, 748–750] Because the next menstrual cycle is delayed after mifepristone, contraception should be initiated immediately after treatment. Ironically, mifepristone, around which swirls the abortion controversy, can make an effective contribution to preventing unwanted pregnancies and induced abortions.

Ulipristal Acetate

Ulipristal acetate (ellaOne) has similar biologic effects as mifepristone and is approved for emergency contraception in Europe and is expected to become available in the U.S. in a single oral dose of 30 mg. ***Randomized trials demonstrated that ulipristal acetate is slightly more effective than the single 1.5 mg dose of levonorgestrel when used within 72 h after sexual intercourse and even between 72 h and 120 h.***[751, 752] Side effects are not severe and similar to those with levonorgestrel. Progesterone receptor modulators like ulipristal acetate and mifepristone suppress ovarian follicular growth and also delay endometrial maturation, manifested by a delay in menstruation after treatment. Ovulation can be temporarily postponed. ***Thus, immediate initiation of a contraceptive method after treatment is even more important with the use of these drugs.***

Other Methods

The three major problems with the available methods of emergency contraception are the high rate of side effects, the need to start treatment within 120 h after intercourse, and the small, but important, failure rate.

Another method of emergency contraception is the insertion of a copper IUD, anytime during the preovulatory phase of the menstrual cycle and up to 5 days after ovulation. The failure rate (in a small number of studies) is very low, 0.1%.[733, 734] This method definitely prevents implantation, but it is not suitable for women who are not candidates for intrauterine contraception, e.g., multiple sexual partners or a rape victim. The use of a copper IUD for emergency contraception is expensive, but not if it is retained as an ongoing method of contraception.

The use of danazol for emergency contraception is not effective.[746]

Steroid Contraception for Older Women

The years from age 35 to menopause can be referred to as the transition years. Preventive health care for women is especially important during the transition years. The issues of preventive health care are familiar ones. They include contraception, cessation of smoking, prevention of heart disease and osteoporosis, maintenance of mental well-being (including sexuality), and cancer screening. Management of the transition years should be significantly oriented to preventive health care, and the use of low-dose steroid contraception (including oral, vaginal, and transdermal methods of administration) can now legitimately

be viewed as a component of preventive health care. ***A discussion of the noncontraceptive health benefits of low-dose oral contraception is especially important with patients in their transition years. This group of women appreciates and understands decisions made with the risk/benefit ratio in mind.***

During this period of time, there are several medical needs that must be addressed: the need for contraception, the management of persistent anovulation, and finally, menopausal and postmenopausal hormone therapy.

At approximately 40 years of age, the frequency of ovulation decreases. This initiates a period of waning ovarian function called the climacteric, which lasts several years, carrying a woman through decreased fertility and menopause to the postmenopausal years. Prior to menopause, the remaining follicles perform less well. As cycles become irregular, vaginal bleeding occurs at the end of an inadequate luteal phase or after a peak of estradiol without subsequent ovulation and corpus luteum formation. Eventually, many women live through a period of anovulation. Occasionally, corpus luteum formation and function occur, and therefore the older woman is not totally safe from the threat of an unplanned and unexpected pregnancy.

Because of a high rate of unintended pregnancy and less than adequate use of contraception, American women older than age 40 have had for the last two decades a high ratio of abortions per live births, a ratio very similar to that of teenagers.[753] Fortunately, clinicians and patients have recognized that low-dose steroid contraception is safe for healthy, nonsmoking, normotensive older women. Oral contraception fulfills a need, and we would argue that this population of women has a series of benefits to be derived from oral contraception that tilts the risk/benefit ratio to the positive side. The benefits of oral contraceptives reviewed in this chapter are especially pertinent for older women. A case-control study could find no evidence for an increased risk of breast cancer in women who used oral contraceptives *after age 40*.[754]

Despite the widespread teaching and publicity that smoking is a contraindication to oral contraceptive use over the age of 35, more older women who use oral contraceptives smoke and smoke heavily, compared with young women.[208] This strongly implies that older smokers are less than honest with clinicians when requesting oral contraception. ***A former smoker must have stopped smoking for at least 12 consecutive months to be regarded as a nonsmoker. Women who have nicotine in their bloodstream obtained from patches or gum should be regarded as smokers.*** Smokers over age 35 should continue to be advised that combined estrogen-progestin contraceptives are not a good choice, regardless of the number of cigarettes smoked. In view of the unreported high rate of smoking in older women who use oral contraceptives, clinicians should use 20-μg estrogen products for women over age 35.

A product containing 20 μg ethinyl estradiol and 150 μg desogestrel has been demonstrated in multicenter studies of women over age 30 to have the same efficacy and side effects as pills containing 30 and 35 μg of estrogen.[755–757] In a randomized study of women over age 30, this formulation was associated with the virtual elimination of any effects on coagulation factors.[758] Indeed, formulations with 20 μg ethinyl estradiol have no significant impact on the measurements of clotting factors, even in smokers.[121, 122, 758, 759]

Although it is true that the implied safety of the lowest estrogen dose remains to be documented by epidemiologic studies, it seems clinically prudent to maximize the safety margin in this older age group of women. Although there may be some increase in breakthrough bleeding, we believe that older women who understand the increased safety implicit in the lowest estrogen dose are more willing to endure breakthrough bleeding and maintain continuation. With avoidance of risk factors and use of lowest dose pills, health risks are negligible for healthy nonsmoking, normotensive women. For healthy nonsmoking women, no specific laboratory screening is necessary beyond that which is usually incorporated in a program of preventive health care.

We should also mention the progestin-only minipill. Because of reduced fecundity, the minipill achieves near total efficacy in women over age 40. Therefore, the progestin-only minipill is a good choice for older woman, and especially for those women in whom estrogen is contraindicated. Older women are more accepting of irregular menstrual bleeding when they understand its mechanism, and, thus, are more accepting of the progestin-only minipill. Older women who are smokers or women with hypertension, or migraine headaches with aura, or obesity, or vascular disease should be encouraged to use progestin-only methods or an intrauterine device.

Anovulation and Bleeding

Throughout the transitional period of life there is a significant incidence of dysfunctional uterine bleeding due to anovulation. While the clinician is usually alerted to this problem because of irregular bleeding, clinician and patient often fail to diagnose anovulation when bleeding is not abnormal in schedule, flow, or duration. As a woman approaches menopause, a more aggressive attempt to document ovulation is warranted. A serum progesterone level measured approximately 1 week before menses is simple enough to obtain and worth the cost. The prompt diagnosis of anovulation (serum progesterone less than 3 ng/mL) will lead to appropriate therapeutic management that will have a significant impact on the risk of endometrial cancer.

In an anovulatory woman with proliferative or hyperplastic endometrium (unaccompanied by atypia), periodic oral progestin therapy is mandatory, such as 5–10 mg medroxyprogesterone acetate given daily the first 2 weeks of each month. If hyperplasia is already present, follow-up aspiration office curettage after 3–4 months is required. The follow-up biopsy should be performed 1–2 months after the progestin treatment to allow any progestin-induced masking of atypia to recede. If progestin treatment is ineffective and histologic regression is not observed, more aggressive treatment is warranted. Monthly progestin treatment should be continued until withdrawal bleeding ceases or menopausal symptoms are experienced. These are reliable signs (in effect, a bioassay) indicating the onset of estrogen deprivation and the need for the addition of estrogen in a postmenopausal hormone program.

Two case-control studies, one using data from the WHO Collaborative Study and one using the data from the U.K. general practice research database, assessed the risk of idiopathic venous thrombosis in users of progestins alone for therapeutic purposes and concluded that therapeutic progestins alone may be associated with an increased risk of venous thromboembolism.[760, 761] These epidemiologic conclusions were based on extremely small numbers and had very wide confidence intervals. Patients who receive progestin-only for therapeutic reasons are probably older and are more likely to have family histories of cardiovascular disease. In addition, a problem of preferential prescribing is probably present in that clinicians are more likely to promote the use of progestin-only for women they perceive to be at greater risk of venous thromboembolism. Thus, it is likely that the case groups represented a higher risk group than the control groups in these reports. For these reasons, we do not believe progestins are associated with an increased risk of venous thromboembolism.

If contraception is desired, the clinician and patient should seriously consider the use of steroid contraception. The anovulatory woman cannot be guaranteed that spontaneous ovulation and pregnancy will not occur. The use of a low-dose oral contraceptive will at the same time provide contraception and prophylaxis against irregular, heavy anovulatory bleeding and the risk of endometrial hyperplasia and neoplasia. In some patients, oral contraceptive treatment achieves better regulation of menses than monthly progestin administration.

Clinicians often prescribe a traditional postmenopausal hormone regimen to treat a woman with the kind of irregular cycles usually experienced in the perimenopausal years. This addition of exogenous estrogen without a contraceptive dose of progestin when a woman is not amenorrheic or experiencing menopausal symptoms is inappropriate and even risky (exposing the endometrium to excessively high levels of estrogen). **And most importantly, a postmenopausal hormonal regimen does not inhibit ovulation and provide contraception**.[762] The appropriate response is to regulate anovulatory cycles with monthly progestational treatment along with an appropriate contraceptive method or to utilize low-dose steroid contraception. The oral contraceptive that contains 20 μg estrogen provides effective contraception, improves menstrual cycle regularity, diminishes bleeding, and relieves menopausal symptoms.[763]

When to Change from Oral Contraception to Postmenopausal Hormone Therapy

A common clinical dilemma is when to change from steroid contraception to postmenopausal hormone therapy. It is important to change because even with the lowest estrogen dose oral contraceptive available, the estrogen dose is 4-fold greater than the standard postmenopausal dose, and with increasing age, the dose-related risks with estrogen become significant. One approach to establish the onset of the postmenopausal years is to measure the FSH level, beginning at age 50, on an annual basis, being careful to obtain the blood sample on day 6 or 7 of the pill-free week in a standard regimen (when steroid levels have declined sufficiently to allow FSH to rise). Friday afternoon works well for patients who start new packages on Sunday. When FSH is greater than 20 IU/L, it is time to change to a postmenopausal hormone program. Because of the variability in FSH levels experienced by women around the menopause, this method is not always accurate.[764, 765] Indeed, in some women, FSH will not rise until 2 weeks after the last pill. To wait 2 weeks is not very practical and places the patient at risk for an unwanted pregnancy. However, the pill-free week method is practical and works for most women. Of course, this approach is not usable for women on extended regimens. An empiric approach allows patients to enter their mid-50s on low-dose steroid contraception, and then switches to a postmenopausal hormone regimen.

Concluding Thoughts

In the 1970s, as epidemiologic data first became available, we emphasized in our teaching and in our communication with patients the risks and dangers associated with oral contraceptives. In the 1990s, with better patient screening and epidemiologic data documenting the effects of low-dose products, we appropriately emphasized the benefits and safety of modern oral contraceptives. In the new millennium, we can with confidence promote the idea that the use of oral contraceptives yields an overall improvement in individual health, and from a public health point of view, the collection of effects associated with oral contraceptives leads to a decrease in the cost of health care.

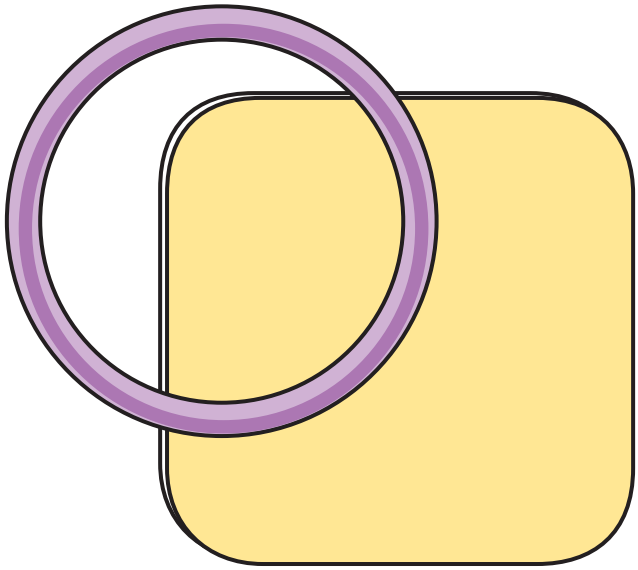
Contraceptive advice is a component of good preventive health care, and the clinician's approach is a key factor. This is an era of informed choice by the patient. Patients deserve to know the facts and need help in dealing with the state of the art and those issues clouded by uncertainty. But there is no doubt that patients are influenced in their choices by their clinician's advice and attitude. Although the role of a clinician is to provide the education

necessary for the patient to make proper choices, one should not lose sight of the powerful influence exerted by the clinician in the choices ultimately made. Emphasizing the safety and benefits of oral contraceptives, and the contribution of oral contraceptives to individual and public health, allows a clinician to present oral contraception with a very positive attitude, an approach that makes an important contribution to a patient's ability to make appropriate health choices.

All references are available online at: http://www.clinicalgynendoandinfertility.com

23

Vaginal and Transdermal Estrogen-Progestin Contraception



The more options available for contraception the more effective family planning is within a society. Vaginal and transdermal estrogen-progestin contraception have considerable appeal for some women unsatisfied with other methods. The use of vaginal contraception in the U.S. has steadily increased since its introduction, and it is now one of the more popular methods. An important advantage of vaginal and transdermal steroid contraception is an improvement in compliance achieved by the elimination of a daily regimen of treatment. Although not yet documented by epidemiologic studies, it is expected that vaginal and transdermal steroid contraception will be associated with the same characteristics, benefits, and problems seen with oral contraceptives.

Vaginal Estrogen-Progestin Contraception

The vagina provides an ideal site for contraceptive steroid absorption. The stratified squamous epithelium, unlike the skin, is not cornified, permitting easier steroid penetration to the underlying lamina propria, made of collagen and elastin, which is richly vascularized. A muscular layer under the epithelium has smooth muscle fibers running in both

circular and longitudinal directions. The final layer of areolar connective tissue contains a second vascular plexus. There are no fat cells, glands (secretions are transudates, not glandular), or hair follicles to interfere with drug absorption.[1]

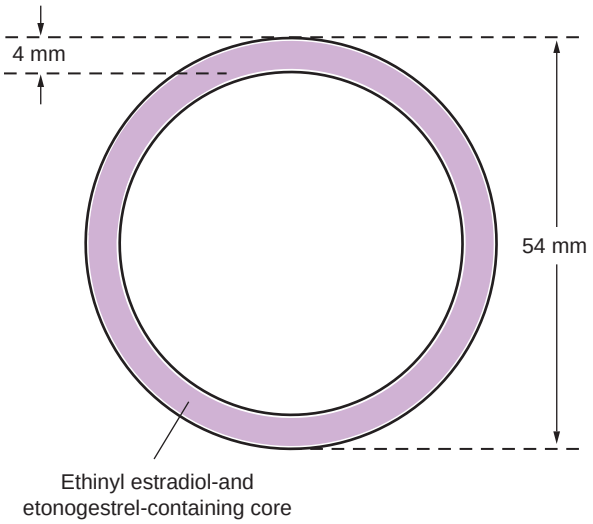
Like subcutaneous, intramuscular, transdermal, and intrauterine contraceptive methods, the vaginal route avoids gastro-intestinal absorption and the first-pass liver effect. Absorption from the gastrointestinal tract can be unpredictable and may be compromised by vomiting, drug-drug interference, or decreased intestinal absorption capacity. The gastrointestinal lumen and the liver are sites of elimination for many compounds, and avoidance of the first-pass effect is particularly advantageous for compounds that undergo a high degree of hepatic metabolism; orally-administered natural estrogens, for example, are 95% metabolized by the liver. Oral administration results in marked fluctuations of contraceptive steroid serum concentrations that may lead to side effects like irregular bleeding and nausea. These daily changes are lower with vaginal than with oral or transdermal administration and lowest with implant and intrauterine methods.[1, 2]

Vaginal contraceptive rings have been studied for 35 years. Six progestin-only and seven different estrogen-progestin combined vaginal contraceptive rings have been designed to provide one week to one year of contraception with weaker progestins like progesterone and medroxyprogesterone in short-acting rings (1 week) and more potent levonorgestrel and nestorone in long-acting ones (up to a year). Only the "NuvaRing" vaginal combined steroid contraceptive is available in the U.S. It is a flexible, soft, transparent ring made of ethylene vinyl acetate copolymer in which are contained crystals of etonogestrel (the biologically active metabolite of desogestrel, previously known as 3-ketodesogestrel) and ethinyl estradiol. This ring is covered with a 2 micron thick membrane of ethylene vinyl acetate. The ring is available in only one size, 4 mm in thickness and 54 mm in diameter (smaller than a diaphragm), that fits all women.

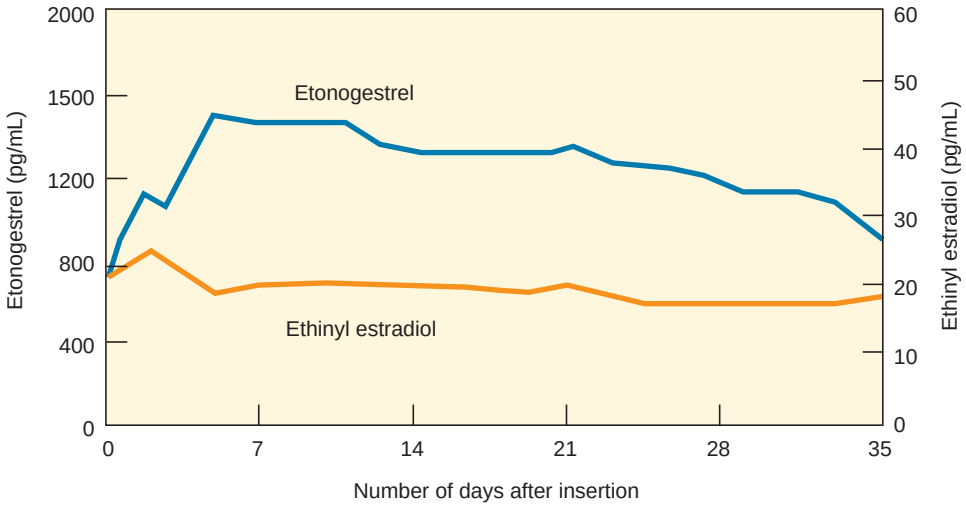
A vaginal ring that is not yet marketed delivers 15 µg ethinyl estradiol and 150 µg nestorone daily and is intended to be effective for a year with periodic removals to induce withdrawal bleeding.[3, 4] Rings that contain only progesterone or nestorone are being developed for use in breastfeeding women.

The Vaginal Ring Method

The NuvaRing releases 15 µg ethinyl estradiol and 120 µg etonogestrel/day.[5] Because the progestin and estrogen are mixed in the ethylene vinyl acetate core, in the unlikely event of damage to the ring, leakage or higher release of the hormones does not occur. Circulating estrogen and progestin levels reach target concentrations within 24 h. The circulating estrogen levels reach a maximum level after 2–3 days, etonogestrel reaches maximum level after 7 days, and remains stable for 35 days.[2] ***The ring is inserted by the patient and worn for 3 weeks. The vaginal ring can be initiated in the same fashion as oral contraceptives, on the first day of menses, a Sunday start, or an immediate, same-day start. Routine use requires the insertion of a new ring every 4 weeks to allow for withdrawal bleeding, but an acceptable and easier method is to insert a new ring on the first of every month. Continuous use is obviously an appropriate option.[6] Because the ring contains enough steroid hormone to inhibit ovulation for at least a total of 5 weeks,[7] when used continuously, a new ring can be inserted every 5 weeks. Breakthrough bleeding with continuous use is effectively managed by a 4-day hormone-free interval.[8]***



The ring produces circulating progestin and estrogen levels that are only 40% and 30%, respectively, of the *peak* levels associated with an oral contraceptive containing 150 μg desogestrel and 30 μg ethinyl estradiol.[2] These levels effectively inhibit ovulation, providing pregnancy rates of less than 1% in clinical trials.[5, 7, 9, 10] ***If the vaginal ring is removed and not replaced within 3 h, the manufacturer recommends backup contraception until the ring has been in place for 7 days.***



Clinical Responses

Taking into account bioavailability as influenced by protein binding, systemic exposure to etonogestrel is similar comparing the vaginal ring to an oral contraceptive containing 150 μg desogestrel; however, the systemic exposure to ethinyl estradiol is about 50% of

that of an oral contraceptive containing 30 μg ethinyl estradiol.[2] Vaginal administration is associated with the lowest estrogen exposure as measured by "area under the curve," compared with oral and transdermal methods.[11] This may explain the low incidence of estrogen-related side effects such as dysmenorrhea, nausea and breast tenderness.[5, 9, 12] In a randomized comparison of the vaginal ring and the transdermal patch, more women preferred the vaginal ring because of a lower rate of side effects.[13]

Breakthrough bleeding and spotting rates are lower (around 6%) when compared with an oral contraceptive containing 30 μg of ethinyl estradiol and much lower than with 15 or 20 μg pills.[9, 14, 15] Bleeding and spotting occur more frequently with extended or continuous regimens.[6] The low serum steroid concentrations do not produce significant changes in LDL or HDL levels, but do increase triglycerides substantially and sex hormone-binding globulin levels markedly.[16] There are no clinically significant changes in clotting parameters. The vaginal ring and oral contraceptives have similar metabolic effects.[17] The vaginal ring has no effect on insulin sensitivity, and no clinically significant adverse changes in carbohydrate and lipid metabolism were observed in vaginal ring users with type 1 diabetes mellitus.[18, 19]

It is not necessary to place the vaginal ring in a specific position; it need only be in contact with vaginal mucosa and need not surround the cervix. The vaginal ring is intended to be placed in a normal vagina; infections and anatomic abnormalities are reasons for clinicians and patients to consider other methods. The most common reasons for discontinuation (about 2–4% in the clinical trials) have been vaginal discomfort, unwanted awareness of the ring's presence, coital problems, or expulsion (during a year of use about 2–3% of women experience spontaneous expulsion). Women report that the ring is easy to insert and remove, and, although about 15% of women and 30% of partners report feeling the ring during intercourse, this is not a common reason for discontinuation.[9, 20] ***Removal for sexual intercourse is not recommended, but efficacy is maintained if the ring is replaced within 3 h.*** Post-marketing surveillance studies documented a high rate of satisfaction with good cycle control.[21, 22]

Cervical cytology and the vaginal flora are not affected by the presence of the ring.[9, 23, 24] Despite no change in inflammatory vaginal flora, one well-done study reported that ring users have slightly more vaginal wetness (not sufficient to cause discontinuation), perhaps due to an increase in lactobacilli.[25]

Vaginally-applied antifungal agents (miconazole) have no effect on absorption of contraceptive steroids released by the vaginal ring, nor does the use of tampons.[9, 26, 27] Amoxicillin and doxycycline do not affect serum estrogen or progestin levels or the steroid pharmacokinetics associated with the vaginal ring.[28]

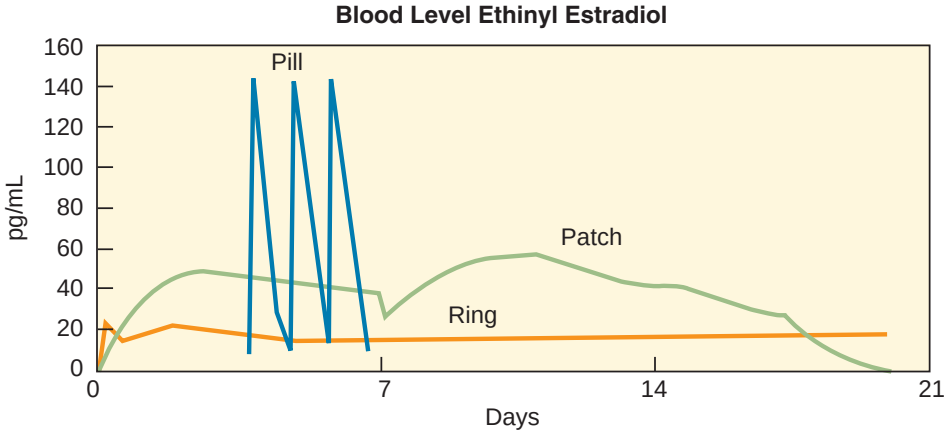
The spermicide, nonoxynol-9, has no effect on the release and absorption of the hormones in NuvaRing, as assessed by the measurement of serum levels of ethinyl estradiol and etonogestrel.[26] Combining a barrier method that contains nonoxynol-9 should not affect the contraceptive efficacy of the ring. However, spermicides do not provide protection against sexually transmitted infections, and there is no good clinical reason to combine the ring with a spermicide.

Summary of Vaginal Ring Advantages

1. Elimination of a daily or coital regimen, resulting in better compliance.
2. Avoidance of gastrointestinal absorption problems.
3. Avoidance of first-pass liver effects.
4. Forgiving of delays; contraceptive efficacy for 5 weeks.
5. Lower systemic estrogen exposure; lower incidence of estrogen side effects.
6. Less frequent breakthrough bleeding and spotting.

Transdermal Estrogen-Progestin Contraception

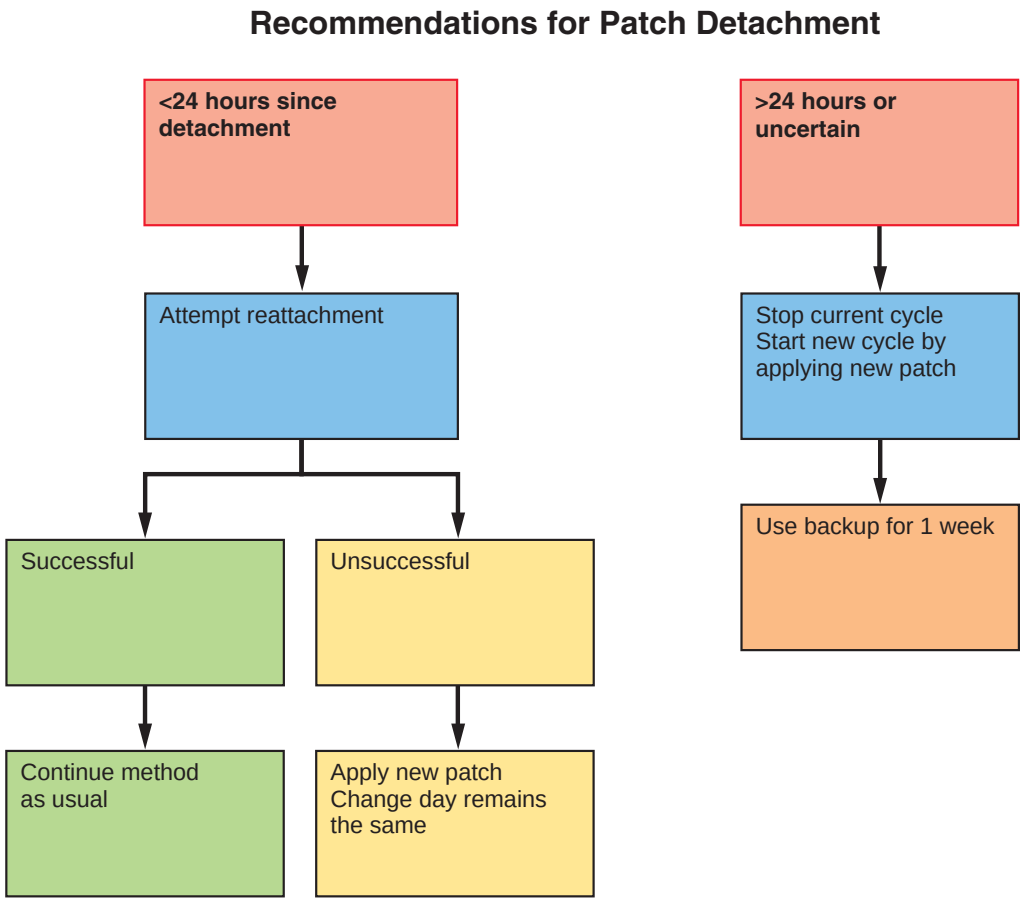
The transdermal contraceptive patch (Ortho-Evra) has an area of 20 cm^2 (4.5 cm × 4.5 cm) and three layers in a matrix-type arrangement. The backing outer polyester layer provides support for the middle layer that contains the adhesive and the hormones, and the inner layer is a polyester liner that is removed from the adhesive layer just before application. The size is that required to deliver an effective dose of the steroid hormones. The patch contains 750 µg ethinyl estradiol and 6 mg of norelgestromin and delivers 20 µg ethinyl estradiol and 150 µg norelgestromin each day when applied to discrete locations on the lower abdomen, upper outer arm, the buttock, or the upper torso (excluding the breast). Norelgestromin is the primary active metabolite of orally administered norgestimate and was previously known as 17-deacetylnorgestimate. Norelgestromin still undergoes liver metabolism with transdermal application; however, the resulting metabolite, levonorgestrel, is highly bound to sex hormone-binding globulin, limiting its biologic impact. About 97% of norelgestromin is bound to albumin and 3% is unbound.[29]



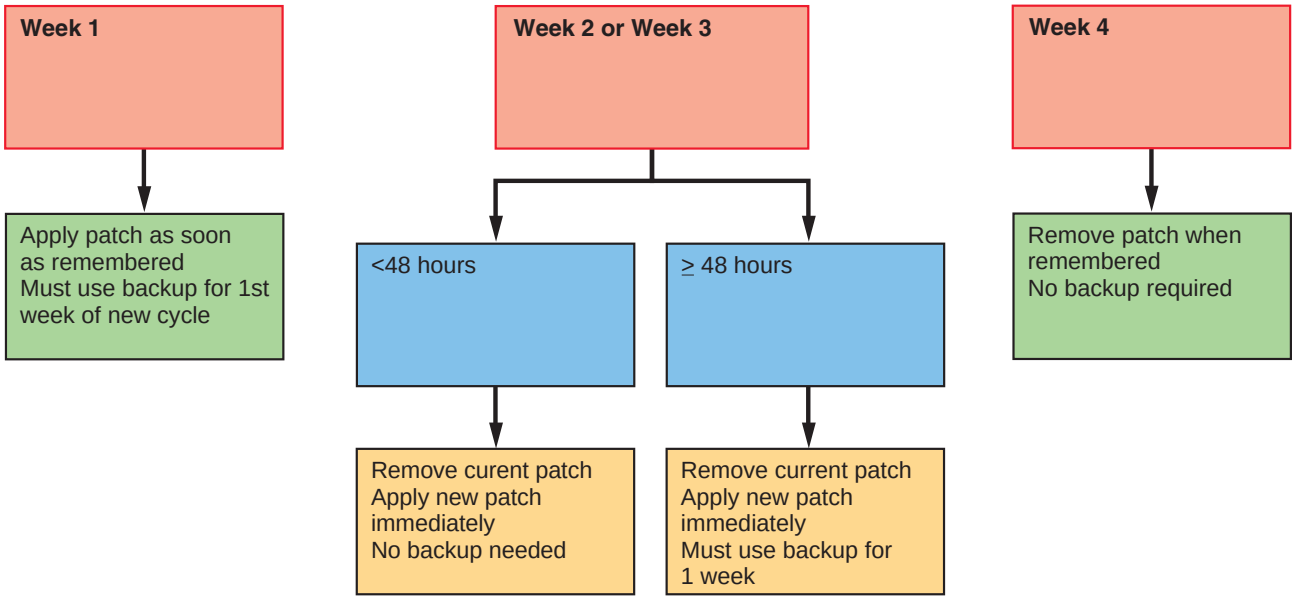
The Transdermal Method

The patch is applied on the same day, but not on the exact same site, once each week for 3 weeks, followed by a week without use of the patch. ***As with oral contraceptives, patient and clinician may choose to use the contraceptive patch continuously, eliminating withdrawal bleeding.*** Timing on the day of application need not be precise; the patch is very forgiving. Instructions for first-day starts, Sunday starts, or immediate same-day starts of oral contraception are also recommended for the patch (backup contraception for 7 days unless the starting day is also day 1 of the menstrual period). Application should be chosen with care to avoid contact with tight clothing, and pressure should be applied for at least 10 s, making sure that the edges stick. The skin should be clear, clean, and dry, and free of irritation or creams and lotions. The patches are stored in their protective pouches at room temperature. ***Many women are bothered by the lint ring that forms around the edges of transdermal patches. In our experience, this problem can be eliminated by lightly dusting talcum powder ("baby" powder) around the edges after application. If a ring does form, it is easily removed with mineral oil.***

Detachment occurs with about 5% of patches, and about half occur in cycle 1 with inexperienced patients.[30, 31] In studies of at least a year's duration, about 2 to 5% of patches were replaced.[32, 33] Use of the patch requires common sense; anything other than a secure adhesion deserves a conservative approach. If the patch seems loose or has been partially or totally off for less than 24 h, the same patch can be reapplied or replaced with a new patch (the patch change day remains the same). Single extra patches are provided for this



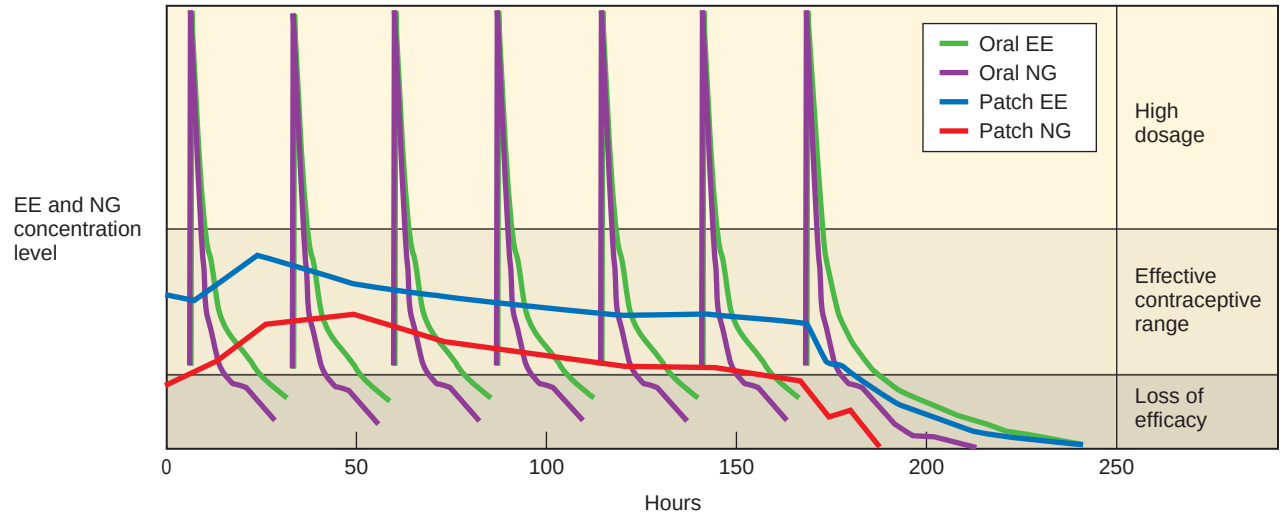
Recommendations for When a New Patch Is Forgotten



purpose by pharmacies and clinics if the clinician writes a separate prescription for an extra patch along with the regular prescription. If the patch has been detached for more than 24 h, a new patch is applied initiating a new cycle and new change day (backup contraception for 7 days is recommended). Delay of a new patch cycle requires a new start with the usual 7-day backup. ***A delay within the patch cycle of no more than 2 days has no risk and does not change the cycle, but a delay more than 2 days also requires the initiation of a new cycle and change day with backup.*** However, in a study that compared three treatment-free days in oral contraceptive and patch users, ovulation occurred significantly less with the patch compared with oral contraception.[34]

Transdermal contraception produces the same spectrum of actions associated with oral contraceptives, achieving the same high level of efficacy in clinical trials. Therefore, the same considerations reviewed in the chapter on oral contraception apply to transdermal contraception (as well as the vaginal ring), including the same contraindications and noncontraceptive benefits. The avoidance of the liver first-pass effect offers the potential for less interaction with other drugs, but this is not known, and patients on medications that affect liver metabolism should choose an alternative contraceptive. Tetracycline administration does not affect the blood concentrations of the steroid hormones with transdermal contraception, a neutral impact as seen with oral contraceptives.[35]

Serum hormonal concentrations are achieved rapidly after application: an average of about 0.7 ng/mL, with a range of 0.6 to 1.2 ng/mL for norelgestromin, and an average of about 50 pg/mL, with a range of 25 to 75 pg/mL for ethinyl estradiol. These are ranges that are maintained by an oral formulation containing 250 μg norelgestromin and 35 μg ethinyl estradiol.[36,37] However, the kinetics are not identical to orally administered hormones; daily fluctuations are avoided. These blood levels allow maintenance of contraceptive efficacy if patch replacement is delayed up to 2 days.[38] Gonadotropin levels return to baseline values by 6 weeks after discontinuation.[34] Daily use has been well studied, and activities such as exercise, bathing, swimming, and the use of a sauna or hot tub do not cause detachment or changes in the blood levels of the hormones.[39]



Transdermal administration has effects on clotting proteins and lipoproteins like those seen with low-dose oral contraceptives.[40] Triglycerides increase modestly, and the LDL/HDL ratio declines slightly.[41] As with oral contraceptives, women who have very high triglycerides or who are at high risk of venous thrombosis due to genetic defects like Factor V-Leiden or protein C or S deficiencies should consider hormonal contraceptives that do not contain estrogen.

Clinical Responses

Ovulation suppression is comparable to that achieved with oral contraception, and failure rates in clinical studies are less than 1.0%.[30, 32, 33, 42] Breakthrough bleeding and spotting rates with the transdermal method in randomized trials were comparable to one monophasic and two triphasic formulations, except for a slightly higher incidence of spotting in the first two cycles.[30, 32]

Modern steroid contraception does not cause weight gain. The transdermal method is not an exception; body weight changes were identical in a randomized trial comparing the contraceptive patch with an identical placebo patch.[41] There were 15 pregnancies in the contraceptive patch clinical trials, and five of these were among women with body weights greater than 198 lb (90 kg).[42] Remember that there is modest evidence that hormonal contraceptive failure is increased in overweight women (over 155 lb).[42–47] On the other hand, no effect of body weight on oral contraceptive failure was detected in a large, prospective European cohort.[48] The positive conclusions regarding failure rates and weight were based on differences of only 2 to 4 pregnancies per 100 women per year. Efficacy in overweight women would still be greater than that with barrier methods. But most importantly, recent clinical trials, especially those with extended regimens, detected no increase in failure rates associated with heavier body weights.[48–51] ***It is not certain that excessive body weight is associated with a reduction in steroid contraceptive efficacy, nevertheless we encourage the use of an extended regimen or continuous dosing in these patients.***

Poor compliance is a major contributor to the typical failure rate associated with oral contraception. The once-a-week schedule with transdermal contraception is simpler and less susceptible to delays and omissions. In randomized trials ranging from 4 cycles to 13 cycles, about 10 to 20% more of the participants demonstrated good compliance with the transdermal method compared with oral contraception.[30, 32] In the clinical trials with transdermal contraception, lower overall pregnancy rates with the patch compared with oral

contraception have been attributed to better compliance. Most importantly, young women, especially those younger than age 20, demonstrated greater compliance with transdermal contraception compared with oral contraceptives than did older women.[52] Indeed, compliance is better at all ages comparing the patch to the pill.[53]

About 20% of patients experience some degree of skin reaction at the application site, and about 2% discontinue the method for this reason.[30, 32, 33, 41] Breast discomfort is experienced during the first few months by 20% of users, more often with transdermal contraception compared with oral contraceptives, but it is usually not severe and has led to discontinuation in only 1% of users.[41]

Summary of Transdermal Patch Advantages

1. Better continuation rates with weekly regimen.
2. Avoidance of gastro-intestinal malabsorption and first-pass liver effects.
3. Contraceptive efficacy maintained with 1- or 2-day delays.
4. Noncontraceptive benefits associated with oral contraceptives are expected.

The Venous Thromboembolism Controversy

The U.S. Food and Drug Administration (FDA) issued a press release on November 10, 2005 calling attention to the fact that women using the transdermal contraceptive patch are exposed over time to a greater amount of estrogen. Subsequently, the patch labeling was updated to include a warning about this higher exposure. Here are the facts:

1. The contraceptive patch delivers daily 20 μg ethinyl estradiol and 150 μg norelgestromin (the primary active metabolite of orally administered norgestimate).
2. The peak estrogen blood levels with the contraceptive patch are about 25% to 35% lower compared with oral products containing 30 or 35 μg ethinyl estradiol.[11]
3. Over time, without the peaks and nadirs experienced by oral contraceptive users, patch users are exposed to about 60% more estrogen compared with an oral product containing 35 μg ethinyl estradiol.

Which is more important, a higher peak level or greater exposure over time? Or maybe it doesn't make a difference. The first concern that there might be an increased risk of venous thrombosis with the patch was a consequence of stories in the media based on anecdotal reports provided to the FDA (a numerator without a denominator). In response, Johnson & Johnson, the parent company for the contraceptive patch, provided research funds for 3 epidemiologic studies, comparing transdermal and oral contraception.

The first case-control study of nonfatal venous thrombosis used information derived from a very large database that records prescriptions and diagnoses longitudinally in managed health care plans.[54] The study over a 3-year time period compared new users of the contraceptive patch with new users of an oral contraceptive containing 35 μg ethinyl estradiol and norgestimate. 68 cases of venous thrombosis and 266 controls were identified and matched

for year of birth and for the date of the thrombotic episode (thus providing comparable dates for exposure). An extension of this study added 56 more cases of venous thrombosis.[55] A comparison of the patch to the oral contraceptive indicated no difference in the risk of venous thrombosis. A specific search for the risk of cerebral venous sinus thrombosis found no increase.[56] This same study also detected no difference in the rates of myocardial infarction or stroke.[57]

The second case-control study indicated that the use of the transdermal contraceptive system was associated with a 2.4-fold increase in the risk of venous thromboembolism compared with users of a 35 μg ethinyl estradiol and norgestimate oral contraceptive.[58] The study used insurance claims data from a major national provider of health insurance to identify women using transdermal or oral contraception. A limitation of the study was the inability to confirm diagnoses in 100% of the cases through review of medical records (completed for 83% of the cases). A strength of the study was the thorough attempt to control for factors that influence venous thrombosis. A strong point of both studies was a focus on new users, eliminating the problem known as attrition of susceptibles. Comparing new users, who are more likely to experience venous thrombosis, to old users would be comparing two different groups of subjects.

A postmarketing study compared the rate of venous thromboembolism in patch users and users of oral contraceptives containing levonorgestrel and 30 μg ethinyl estradiol.[59] The difference between the two methods was *not* statistically significant.

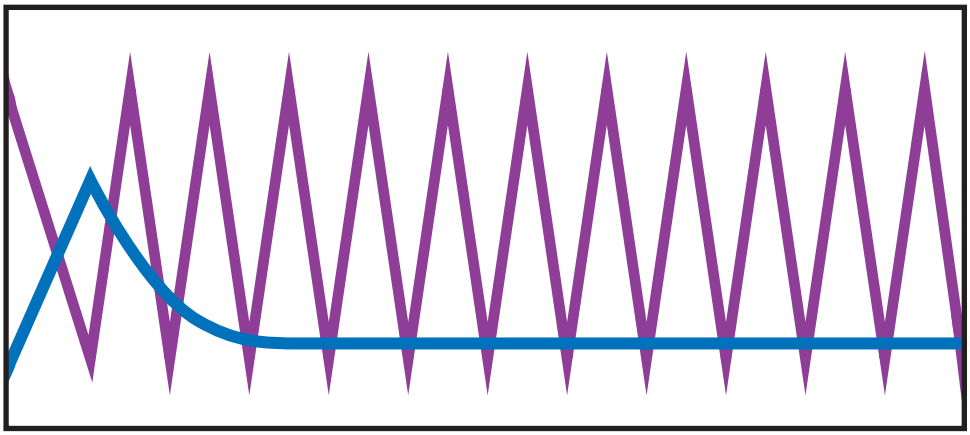
Comparing the two case-control studies, it is worth noting that the negative report had about 6 times as many cases as the positive report, giving it more statistical power. Indeed, in the overall case-control analysis in the positive report, there were only 20 cases among transdermal users and 37 cases among oral contraceptive users, and the odds ratio of 2.0 was close, but it did not reach statistical significance. After excluding cases and controls with high-risk factors, the odds ratio with 16 transdermal cases and 26 oral contraceptive cases achieved 2.4 with statistical significance, but with a relatively very wide confidence interval, 1.1–5.5. Similarly the cases in the postmarketing survey amounted to less than half of the number in the negative case-control study.

These are the only epidemiologic data on this important issue. One case-control study is reassuring; one case-control study is disturbing. Because the confidence intervals of the two studies are within the same range (i.e., they overlap), it is not certain that the different results do not reflect a chance finding. At this point in time, it seems that if there is a difference in the risk of venous thrombosis comparing transdermal and oral contraception, the difference has to be very small. Certainly the available evidence does not indicate a major difference.

All references are available online at: http://www.clinicalgynendoandinfertility.com

24

Long-Acting Methods of Contraception



The high rate of unintended pregnancies and the relatively high failure rates with the typical use of reversible methods of contraception are strong indications of a need for long-acting contraceptive methods that simplify compliance. Two effective and popular methods are available, contraceptive implants (systems with the sustained release of a progestin) and depot-medroxyprogesterone acetate (Depo-Provera).

Injectable depot-medroxyprogesterone acetate is a long-acting (3 months) agent that has been part of the contraceptive programs of many countries for more than 25 years. This experience has demonstrated it to be safe, effective, and acceptable. It is not a "sustained-release" system, but its action is the same.

There are three major implant systems, Implanon, Norplant and Jadelle (formerly called Norplant-2). A Chinese version is called Sinoplant II. Both Jadelle and Sinoplant II contain 150 mg levonorgestrel. Unfortunately, Norplant was withdrawn from the U.S. market in a business decision dictated by profit and liability despite the fact that Norplant provided an excellent option for contraception. Jadelle is approved by the U.S. Food and Drug Administration, but it has not been marketed. In many parts of the world, Jadelle has replaced the use of Norplant; however, Norplant is still used worldwide.

Implanon differs from Norplant and Jadelle in many pivotal aspects, chiefly one rod instead of Norplant's 6 capsules and Jadelle's two rods, and a less androgenic progestin.[1, 2] Like Norplant, Implanon has been extensively marketed throughout the world with a good track record and high continuation rates. Contraceptive implants are approved in more than 60 countries and used by approximately 11 million women.[2]

The long-acting progestin methods are as effective as sterilization and IUDs, and more effective than oral and barrier contraception.[3] An important reason for this high efficacy

in actual use is the nature of the delivery systems themselves, which require little effort on the part of the user. Because compliance does not require frequent resupply or instruction in use, as with oral contraception, the actual or typical use effectiveness is very close to the theoretical (lowest expected) effectiveness.

Sustained-release methods require less of the user, but they demand more of the clinician. Implants involve minor operative procedures for placement and for discontinuation. Clinicians have a special responsibility to become skillful in the operations required to remove implants and to be available to women when those skills are required to terminate use. Disturbances of menstrual patterns and other side effects prompt many more questions from patients about these methods than about use of the familiar oral, intrauterine, and barrier contraceptives.[4]

Implant Systems

Norplant was developed by the Population Council and first approved in 1983 in Finland, where it was manufactured. It was approved in the U.S. in 1990, marketed in 1991, and withdrawn from the market in 2002.

Norplant is a "sustained-release" system using silastic tubing permeable to steroid molecules to provide stable circulating levels of synthetic progestin over years of use. The Norplant system consists of 6 capsules, each measuring 34 mm in length with a 2.4 mm outer diameter and containing 36 mg crystalline levonorgestrel. The capsules are made of flexible, medical-grade silastic (polydimethylsiloxane and methylvinyl siloxane copolymer) tubing that is sealed shut with silastic medical adhesive (polydimethylsiloxane). The 6 capsules contain a total of 216 mg levonorgestrel, which is very stable and remained unchanged in capsules examined after more than 9 years of use.

Jadelle was also developed by the Population Council and manufactured in Finland. It was approved in the U.S. in 1996, but never marketed. The thin, flexible Jadelle rods are wrapped in silastic tubing (the same material used by Norplant), 43 mm in length and 2.5 mm in diameter, thus slightly longer and thicker than Norplant.[5] Each rod contains 75 mg levonorgestrel for a total of 150, 66 mg less than that in the 6 Norplant capsules. Whereas the levonorgestrel in Norplant is packed into the capsules in crystal form, the core of the Jadelle rod is a mixture of levonorgestrel and an elastic polymer (dimethylsiloxane/methylvinylsiloxane). Long-term clinical trials indicate that the performance and side effects are similar to Norplant, but removal is faster.[6, 7]

Norplant average release rate:

First month	**—**	**85 µg levonorgestrel daily**
After 1 year	**—**	**35 µg**
After 2 years	**—**	**30 µg**

Jadelle average release rate:

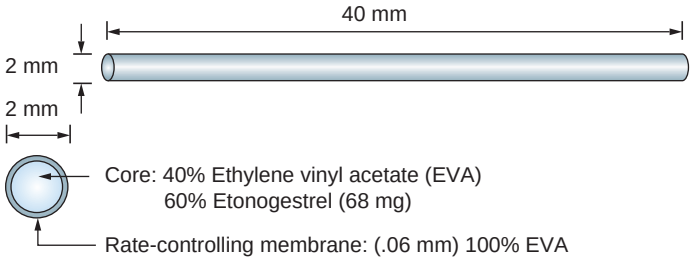
First month	**—**	**100 µg levonorgestrel daily**
After 1 year	**—**	**40 µg**
After 2 years	**—**	**30 µg**

Because the release rates with the two levonorgestrel systems are comparable, it is reasonable to conclude that clinical studies with Norplant and Jadelle should yield similar results.

In the discussion that follows, the more-studied product, Norplant, is often cited, but clinicians can assume that the findings apply as well to Jadelle.

Implanon is a single flexible rod 4 cm long and 2 mm in diameter, that contains 68 mg of 3-keto desogestrel (etonogestrel, the active metabolite of desogestrel) dispersed in a core of ethylene vinyl acetate wrapped with a 0.6 mm thick membrane of the same material. There is no evidence that either ethylene vinyl acetate or silastic have toxic effects when implanted.[8] The hormone is released at an initial rate of about 67 μg/day decreasing to 30 μg after 2 years; concentrations that inhibit ovulation are achieved within 8 h of insertion.[9] A steady state is achieved after 4 months; after which there is a gradual decline.[9] Implanon, placed subdermally with a disposable inserter, suppresses ovulation for 2.5 years, and provides effective contraception for at least 3 years. Side effects are similar to those with Norplant or Jadelle, except for less bleeding and a higher rate of amenorrhea with Implanon.[10–13]

The Implanon Single Rod



Indications

Contraceptive implants are a good choice for women of reproductive age who are sexually active and desire long-term, continuous contraception. Implants should be considered for women who:

1. Want to delay the next pregnancy for at least 2–3 years.
2. Desire a highly effective, long-term method of contraception.
3. Experience serious or minor estrogen-related side effects with estrogen-progestin contraception.
4. Have difficulty remembering to take pills every day, have contraindications or difficulty using IUDs, or desire a non-coitus-related method of contraception.
5. Have completed their childbearing but are not yet ready to undergo permanent sterilization.
6. Have a history of anemia with heavy menstrual bleeding.
7. Intend to breastfeed for a year or two.
8. Have chronic illnesses, in which health will be threatened by pregnancy.

Absolute Contraindications

Implant use is contraindicated in women who have:

1. *ACTIVE* thrombophlebitis or thromboembolic disease.
2. Undiagnosed genital bleeding.
3. *ACUTE* liver disease.
4. Benign or malignant liver tumors.
5. Known or suspected breast cancer.

Relative Contraindications

Based on clinical judgment and appropriate medical management, Implants *MAY BE USED* by women with a history of or current diagnosis of the following conditions:

1. Heavy cigarette smoking (15 or more daily) in women older than 35 years.
2. History of ectopic pregnancy.
3. Diabetes mellitus. Because multiple studies have failed to observe a significant impact on carbohydrate metabolism, Implants, in our view, are particularly well suited for diabetic women.
4. Hypercholesterolemia.
5. Hypertension.
6. History of cardiovascular disease, including myocardial infarction, cerebral vascular accident, coronary artery disease, angina, or a previous thromboembolic event. Patients with artificial heart valves.
7. Gallbladder disease.
8. Chronic disease, such as immunocompromised patients.

Implants are not contraindicated in the following situations, but other methods are preferable:

1. Severe acne.
2. Severe vascular or migraine headaches.
3. Severe depression.
4. Concomitant use of medications that induce microsomal liver enzymes:

Carbamazepine (Tegretol)
Felbamate
Lamotrigine

Nevirapine
Oxcarbazepine
Phenobarbital
Phenytoin (Dilantin)
Primidone (Mysoline)
Rifabutin
Rifampicin (Rifampin)
St. John's wort
Topiramate
Vigabatrin
Possibly valproic acid, ethosuximide, griseofulvin, and troglitazone

We do not recommend the use of implants with any of the previously listed drugs because of a likely increased risk of pregnancy due to lower blood levels of the progestin.[14, 15]

Mechanism of Action

The release rate of the contraceptive implants is determined by total surface area and the density of the implant in which the progestin is contained. The progestin diffuses from the implant into the surrounding tissues where it is absorbed by the circulatory system and distributed systemically, avoiding an initial high level in the circulation as with oral or injected steroids. Within 8 h after insertion of Implanon, plasma concentrations of etonogestrel are about 300 ng/mL, high enough to prevent ovulation.[16] A study of cervical mucus changes with Norplant indicates that a backup method should be used for 3 days after insertion of Norplant or Jadelle; this is not necessary when Implanon is inserted as directed.[17, 18] Progestin concentrations are much more variable with Norplant and Jadelle than with Implanon.[16]

The Implanon rod releases 60 μg of etonogestrel per 24 h at 3 months of use. This rate declines gradually to 40–50 μg daily by 12 months and 30 μg/day by 2 years of use. The 85 μg of hormone released by Norplant or the 100 μg released by Jadelle during the first few months of use is about equivalent to the daily dose of levonorgestrel delivered by the progestin-only, minipill oral contraceptive, and 25–50% of the dose delivered by low-dose combined oral contraceptives. After 6 months of use, daily levonorgestrel concentrations are about 0.35 ng/mL; at 2.5 years, the levels decrease to 0.25–0.35 ng/mL. Until the 8-year mark, mean levels remain above 0.25 ng/mL.[19] Mean plasma concentrations below 0.2 ng/mL are associated with increased pregnancy rates for Norplant (lower levels are more likely in heavier women).

Body weight affects the circulating levels of levonorgestrel; the greater the weight of the user, the lower the levonorgestrel concentrations at any time during Norplant or Jadelle use. The greatest decrease over time occurs in women weighing more than 70 kg (154 lb), but even for heavy women, the release rate is high enough to prevent pregnancy at least as reliably as oral contraceptives. In Implanon users, etonogestrel concentrations are affected very little by body weight, and failure rates did not increase with increasing body weight in the small numbers of overweight women in the clinical trials.[20] Although the data with overweight women are limited, it is likely that Implanon is a good contraceptive choice for obese women.

Levonorgestrel levels can also be affected by the circulating levels of sex hormone-binding globulin (SHBG). Levonorgestrel has a higher affinity for SHBG than does etonogestrel. In the week after Norplant or Jadelle insertion, SHBG levels decline rapidly and then return to approximately half of preinsertion levels by 1 year of use. This effect on SHBG is not uniform and may account for some of the individual variations in circulating progestin concentrations.[21]

Implants are highly effective contraceptives. There are three probable modes of action, which are similar to those attributed to the contraceptive effect of the progestin-only minipills, but because daily dosing is not required, implants are more effective than oral methods.

1. The progestin suppresses, at both the hypothalamus and the pituitary, the luteinizing hormone (LH) surge necessary for ovulation. As determined by progesterone levels in many users over several years, approximately one-third of all cycles in Norplant users are ovulatory.[19, 22] During the first 2 years of use, only about 10% of women are ovulatory, but by 5 years of use, more than 50% are. In those cycles that are ovulatory, there is a high incidence of luteal insufficiency. Implanon inhibits ovulation throughout a 3-year period, accounting for almost all of the contraceptive effect.[2] However, follicular development does occur, avoiding the problem of clinically significant hypoestrogenemia, and in the last 6 months of the 3-year period with Implanon, there is an occasional ovulation.[16, 23]

2. The steady release of progestin has a prolonged effect on the cervical mucus. The mucus thickens and decreases in amount, forming a barrier to sperm penetration.[17, 23–25]

3. The progestin suppresses the estradiol-induced cyclic maturation of the endometrium and eventually causes atrophy. These changes could prevent implantation should fertilization occur; however, no evidence of fertilization can be detected in Norplant users.[26]

Advantages

Implants are a safe, highly effective, continuous method of contraception that requires little user effort and, unlike long-acting injectable contraception, is rapidly reversible. Because this is a progestin-only method, it can be used by women who have contraindications for the use of estrogen-containing contraceptives. The sustained release of low doses of progestin avoids the high initial dose delivered by injectables and the daily hormone surge associated with oral contraceptives. Implants are an excellent choice for a breastfeeding woman and can be inserted immediately postpartum. There are no effects on breast milk quality or quantity, and infants grow normally.[27–30] Another advantage of the implant method is that it allows women to plan their pregnancies precisely; return of fertility occurs within a few weeks, in contrast to the 6- to 18-month delay in ovulation that can follow depot-medroxyprogesterone acetate injections.[16, 31–33]

One of the major advantages of sustained-release methods is the high degree of efficacy, nearly equivalent to the theoretical effectiveness. In couples for whom elective abortion is unacceptable in the event of an unplanned pregnancy, the high efficacy rate is especially important. There are no forgotten pills, broken condoms, lost diaphragms, or missed injections. For women who are at high risk of medical complications should they become pregnant, sustained-release implants present a significant safety advantage. Users should be reassured that implant use has not been associated with changes in carbohydrate or lipid metabolism, coagulation, liver or kidney function, or immunoglobulin levels. Because many women wanting implants will have had negative experiences with other contraceptives, it is important that the differences between this method and previous methods be explained.

Exposure of endometriosis to progestin-only contraceptive methods is an effective method to manage the pain associated with this condition. Implanon, depot-medroxyprogesterone acetate, and the levonorgestrel-containing intrauterine device have all been reported to reduce endometriosis pain.[34–38]

Disadvantages

There are some disadvantages associated with the use of the implant systems. Implants cause disruption of bleeding patterns, especially during the first year of use, and some women or their partners find these changes unacceptable.[4] Endogenous estrogen is nearly normal, and unlike the estrogen-progestin contraceptives, progestin is not regularly withdrawn to allow endometrial sloughing. Consequently, the endometrium sheds at unpredictable intervals.

The implants must be inserted and removed in a surgical procedure performed by trained personnel. Women cannot initiate or discontinue the method without the assistance of a clinician. The incidence of complicated removals is approximately 5% for Norplant or Jadelle and lower for Implanon, an incidence that can be best minimized by good training and careful insertion.[39, 40] The implants can be visible under the skin. This sign of the use of contraception may be unacceptable for some women and for some partners.[4]

Implants do not provide protection against sexually transmitted infections (STIs) such as herpes, human papillomavirus, HIV, gonorrhea, or chlamydia. Although users may be less likely to use a second method because of the high contraceptive efficacy,[41] users at risk for STIs must use condoms as a second method to prevent infection.

Because the insertion and removal of implants require minor surgical procedures, initiation and discontinuation costs are higher than with oral contraceptives or barrier methods. The cost of implants plus fees for insertion total an amount that may seem high to patients unless they compare it with the total cost of using other methods for up to 5 years.[42] Nevertheless, short-term use is expensive compared with the relatively low initial costs of other reversible methods, and most women cannot be expected to use long-acting methods for their full duration of action.

Cultural factors can influence the acceptability of menstrual changes. Some cultures restrict a woman from participating in religious activity, household activities, or sexual intercourse while menstruating. All users must be aware of the possible menstrual changes. It is important to stress that all of the menstrual changes are expected, that they do not cause or represent illness, and that most women revert back to a more normal pattern with increasing duration of use.

Insertion and removal of implants will be a new experience for most women. As with any new experience, women will approach it with varying degrees of apprehension and anxiety. In reality, most patients are able to watch in comfort as implants are inserted or removed. Women should be told that the incisions used for the procedures are very small and heal quickly, leaving small scars that are usually difficult to see because of their location and size.

We encourage prospective users to see and touch implants. Women can be reassured that the implants will not be damaged or move if the skin above them is accidentally injured. Normal activity cannot damage or displace the implants. Most women become unaware of their presence. A few women report sensing the implants if they have been touched or manipulated for a prolonged period of time, or after vigorous exercise. The implants are more visible in slender women with good muscle tone. Darker-skinned users may notice further darkening of the skin directly over the implants; this resolves after removal.

Efficacy

Contraceptive implants provide highly effective birth control. In 2-year or 3-year studies in 11 international clinical trials of 942 women using Implanon, no pregnancies occurred.[20]

In studies of Norplant conducted in 11 countries, totaling 12,133 woman-years of use, the pregnancy rate was 0.2 pregnancies per 100 woman-years of use.[15, 31] All but one of the pregnancies that occurred during the Norplant evaluation were present at the time of implant insertion. If these luteal phase insertions are excluded from analysis, the first-year pregnancy rate was 0.01 per 100 woman-years. In adolescents, Norplant implants provide better protection against unwanted pregnancy, compared with oral contraceptives, and an important factor is the better continuation rate with Norplant.[41, 43, 44]

Failure Rates During the First Year of Use, United States[3, 45, 46]

	Percent of Women with Pregnancy	
Method	*Lowest Expected*	*Typical*
No method	85%	85%
Combination Pill	0.3%	8.7%
Progestin only	0.5%	3.0%
IUDs:		
Levonorgestrel IUS	0.1%	0.1%
Copper T 380A	0.6%	1.0%
Implants:		
Six levonorgestrel capsules (Norplant)	0.05%	0.2%
Two levonorgestrel rods (Jadelle)	0.06%	0.06
One etonogestrel rod (Implanon)	0.01%	0.01
Injectable		
3-month	0.3	0.3%
1-month	0.05	3.0%
Patch	0.3	8.0%
Vaginal ring	0.3	8.0%
Female sterilization	0.5%	0.7%
Male sterilization	0.1%	0.2%

There are no weight restrictions for Norplant or Jadelle users, but heavier women (more than 70 kg) may experience slightly higher pregnancy rates in the later years of use compared with lighter women. Even in the later years, however, pregnancy rates for heavier women using Norplant are lower than with oral contraception. The differences in pregnancy rates by weight are probably due to the dilutional effect of larger body size on the low, sustained serum levels of levonorgestrel. Heavier women should not rely on Norplant or Jadelle beyond the 5-year limit. For slender women the duration of efficacy extends well past the fifth year of use. In some extended trials, no pregnancies have occurred into the seventh year. Data are not available regarding the effect of body weight on the efficacy of Implanon, but unlike Norplant and Jadelle, progestin levels are not significantly lower in heavier women.

The contraceptive efficacy of Implanon surpasses that of Norplant and sterilization.[2] Only a rare pregnancy occurs, resulting in a Pearl Index of about 0.01.[23, 47] In over 70,000 cycles, no pregnancies were recorded because of total inhibition of ovulation until ovulations were observed in the last 6 months of the 3-year period.[23, 48] Post-marketing surveillance of pregnancies in Australia, where nearly one-quarter of contraceptors relied on Implanon in 2004, revealed that of 218 pregnancies, only 13 could *possibly* have been failures of the method.[49] In Australia and the Netherlands, pregnancies commonly were the consequence of poor insertion technique, especially allowing the implant to fall unnoticed to the floor.

Implants have an immediate contraceptive effect when inserted within the first 7 days of a menstrual cycle, but when insertion is after day 7, a backup method of contraception is necessary for at least 3 days.[50]

Ectopic Pregnancy

The ectopic pregnancy rate during Norplant use is 0.28 per 1,000 woman-years.[15] *Although the risk of developing an ectopic pregnancy during use of Norplant is low, when pregnancy does occur, ectopic pregnancy should be suspected because approximately 30% of Norplant pregnancies are ectopic*. Because Implanon is more effective in inhibiting ovulation, we would expect the risk of ectopic pregnancy to be considerably less than that associated with Norplant.

Ectopic Pregnancy Rates Per 1,000 Woman-Years[15, 51, 52]	
Non-Contraceptive Users, all ages	3.0–4.5
Copper T-380 IUD	0.20
Norplant	0.28

Menstrual Effects

Menstrual bleeding patterns are highly variable among users of implant contraception. With levonorgestrel implants, some alteration of menstrual patterns will occur during the first year of use in approximately 80% of users, later decreasing to about 40%, and by the fifth year, to about 33%.[53, 54] The changes include alterations in the interval between bleeding, the duration and volume of menstrual flow, and spotting. Oligomenorrhea and amenorrhea also occur but are less common, less than 10% after the first year and diminishing thereafter. Irregular and prolonged bleeding usually occurs during the first year. Although bleeding problems occur much less frequently after the second year, they can occur at any time.[54,55] Studies of the endometrium in Norplant users experiencing abnormal bleeding indicate the presence of enlarged venous sinusoids (fragile vessels) and a reduction in the expression of a protein factor (perivascular stromal cell tissue factor) involved in the initiation of hemostasis.[56] Within weeks after insertion, the density of endometrial small blood vessels increases and the endometrium regresses to an atrophic state.[57] It is believed that bleeding is a consequence of rapid endometrial regression and that the apparent increase in the number of blood vessels may reflect increased tortuosity accompanying the atrophic regression.

Implanon alters menstrual patterns, but amenorrhea occurs more often (21% of users in the first year, 30–40% after 1 year) than with Norplant.[11, 13] A single Implanon rod completely suppresses ovulation for 2.5 years, and, therefore, menses do not become more regular after the first 2 years as with Norplant. After 2 years, ovulation occurs in about half of the menstrual cycles. Bleeding is lighter and less frequent among Implanon users because more profound ovarian suppression results in less follicular estrogen production and less endometrial stimulation, nevertheless irregular bleeding continues to be a a major reason for discontinuation.[13, 58]

Despite an increase in the number of spotting and bleeding days over preinsertion menstrual patterns, hemoglobin concentrations rise in Norplant users because of a decrease in the average amount of menstrual blood loss.[59–62] Implanon likewise does not cause anemia.[11]

Implant users who can no longer tolerate prolonged bleeding will benefit from a short course of oral estrogen: conjugated estrogens, 1.25 mg, or estradiol, 2 mg, administered daily for 7 days.[63] A therapeutic dose of one of the prostaglandin inhibitors given during the

bleeding will help to diminish flow, but estrogen is the most effective treatment.[64, 65] Another approach is to administer an estrogen-progestin oral contraceptive for 1–3 months.[66]

Although implants are very effective, pregnancy must be considered in women reporting amenorrhea who had been ovulating previously, as evidenced by regular menses prior to an episode of amenorrhea. A sensitive urine pregnancy test should be obtained. Women who remain amenorrheic throughout their use of implants are unlikely to become pregnant.[54] It is important to explain to patients the mechanism of the amenorrhea: the local progestational effect causing decidualization and atrophy.

Metabolic Effects

Exposure to the sustained, low doses of progestin delivered by the implants is not associated with significant metabolic changes. Studies of liver function,[10, 67, 68] blood coagulation,[10,69–71] immunoglobulin levels,[72, 73] serum cortisol levels,[74] and blood chemistries[68, 72] have failed to detect changes outside of normal ranges in Norplant users.

No major impact on the lipoprotein profile can be demonstrated with Norplant.[67, 75, 76] Minor changes are transient, and, with prolonged duration of use, lipoproteins return to preinsertion levels. Long-term exposure to the low dose of levonorgestrel released by Norplant is unlikely to affect users' risk of atherogenesis, just as prolonged exposure to combined oral contraception has not. There are no clinically important effects on carbohydrate metabolism.[72, 77 77, 78] No effect on insulin sensitivity can be detected.[79] In a cohort study of 5 years' duration, no increase was observed in diabetes mellitus, depression, lupus erythematosus, cardiovascular diseases—in fact there was no increase in serious morbidity.[80]

***There are no significant metabolic differences comparing Implanon and Norplant*.[81]** Neither implant system has important clinical effects on the lipoprotein profile, carbohydrate metabolism, thyroid and adrenal function, liver function tests, or the clotting mechanism.[10, 58, 82] Implant contraception is a good choice for a woman at risk for estrogen-associated thromboembolism. Because of the lower androgenic characteristic of etonogestrel, Implanon does not cause a decrease in the levels of sex hormone-binding globulin.[82]

Measurements of bone density in young women reveal that Implanon and Norplant do not affect the teenage gain in bone; similar gains in bone were recorded in implant users and control subjects.[83, 84] In older women, an increase in forearm, spine, and femur bone density has been documented after 6 and 12 months of Norplant use.[85, 86] An international cross-section study reported a small loss in bone density with Norplant that was rapidly regained after discontinuation.[87]

A slight increase in gallbladder disease has been noted in Norplant users.[2, 88] This is at best just a word of caution because the association is weak and may reflect preexisting disease, and there is no apparent biologic mechanism.

Effects on Future Fertility

Circulating levels of progestin become too low to measure within 48 h after removal of implants. Most women resume normal ovulatory cycles during the first month after removal. The pregnancy rates during the first year after removal are comparable with those of women not using contraceptive methods and trying to become pregnant. There are no long-term effects on future fertility nor are there any effects on sex ratios, rates of ectopic pregnancy, spontaneous miscarriage, stillbirth, or congenital malformations.[15, 31]

The return of fertility after implant removal is prompt, and pregnancy outcomes are within normal limits. The rate and outcome of subsequent pregnancies are not influenced by duration of use.

For women who are spacing their pregnancies, the difference between implants and depot-medroxyprogesterone acetate in the timing of the return to fertility can be critical. Implants allow precise timing of pregnancy because the return of ovulation after removal is prompt. Etonogestrel serum levels are undetectable within one week after removal of Implanon, and ovulation can be expected in the first month after discontinuation.[9] Depot-medroxyprogesterone acetate, on the other hand, can cause up to 18 months' delay in return to fertility. By that time, 90% of users of either method will have ovulated, but in the first several months, the difference is dramatic. By 3 months after removal, half of implant users will have ovulated, but 10 months must elapse before half of depot-medroxyprogesterone acetate users are ovulatory.

Side Effects

The occurrence of serious side effects is very rare, no different in incidence than that observed in the general population. In addition to the menstrual changes, levonorgestrel implant users have reported the following side effects: headache, acne, weight change, mastalgia, hyperpigmentation over the implants, hirsutism, depression, mood changes, anxiety, nervousness, ovarian cyst formation, and galactorrhea.[15, 31, 53, 55, 89]

It is difficult, of course, to be certain which of these effects were actually caused by the levonorgestrel. For example, careful study fails to reveal a relationship between Norplant use and depressive symptoms.[90] Although most of these side effects are minor in nature, they can cause patients to discontinue the method. Patients often find common side effects tolerable after assurance that they do not represent a health hazard.[4] Many complaints respond to reassurance; others can be treated with simple therapies. The most common side effect experienced by users is headache (16% of Implanon users); approximately 20% of women who discontinue use do so because of headache.[4, 20, 89]

Stroke, thrombotic thrombocytopenic purpura, thrombocytopenia, and pseudotumor cerebri have been reported with Norplant.[91] However, it is by no means established that the incidence of these problems is increased, and there is little reason to suspect a cause-and-effect relationship. In the follow-up study conducted by the World Health Organization in eight countries, no significant excess of cardiovascular events or malignant disease was observed.[92]

Weight Change

Women using levonorgestrel implants more frequently complain of weight gain than of weight loss, but findings are variable.[88] In the Dominican Republic, 75% of those who changed weight lost weight, whereas in San Francisco, two-thirds gained weight. Assessment of weight change in Norplant users is confounded by changes in exercise, diet, and aging. Although an increase in appetite can be attributed to the androgenic activity of levonorgestrel, it is unlikely that the low levels with Norplant have any clinical impact. Counseling for weight changes focuses best on dietary review and dietary changes. Indeed, a 5-year follow-up of 75 women with Norplant implants could document no increase in the body mass index (nor was there a correlation between irregular bleeding and body weight).[93] A similar experience has been documented with Implanon.[12]

Mastalgia

Bilateral mastalgia, often occurring premenstrually, is usually associated with complaints of fluid retention. After pregnancy has been ruled out, reassurance and therapy aimed at symptomatic relief are indicated. This symptom decreases with increasing duration of implant use, and occurs at a lower rate comparing Implanon (10% of users) with Norplant.[12, 20] The most effective treatments for mastalgia are the following: danazol (200 mg/day), vitamin E (600 units/day), bromocriptine (2.5 mg/day), or tamoxifen (20 mg/day), but there are no studies of these treatments in implant users.

Galactorrhea

Galactorrhea is more common among women who have had insertion of the implants on discontinuation of lactation. Pregnancy and other possible causes should be ruled out by performing a pregnancy test and a thorough breast examination. Patients can be reassured that this is a common occurrence among implant and oral contraceptive users. Decreasing the amount of breast and nipple stimulation during sexual relations might alleviate the symptom, but if amenorrhea accompanies persistent galactorrhea, a prolactin level should be obtained.

Acne

Acne, with or without an increase in oil production, is the most common skin complaint among levonorgestrel implant users. The acne is caused by the androgenic activity of the levonorgestrel that produces a direct impact and also causes a decrease in sex hormone-binding globulin (SHBG) levels leading to an increase in free steroid levels (both levonorgestrel and testosterone).[21] This is in contrast to combined oral contraceptives that contain levonorgestrel, in which the estrogen effect on SHBG (an increase) produces a decrease in unbound, free androgens. Etonogestrel implants are less commonly associated with acne because this progestin is less androgenic than levonorgestrel.[12] Common therapies for complaints of acne include dietary change, practice of good skin hygiene with the use of soaps or skin cleansers, and application of topical antibiotics (e.g., 1% clindamycin solution or gel or topical erythromycin).

Ovarian Cysts

Unlike oral contraception, the low serum progestin levels maintained by implants do not suppress follicle-stimulating hormone (FSH), which continues to stimulate ovarian follicle growth in most users. The LH peak during the first 2 years of use, on the other hand, is usually abolished so that these follicles do not ovulate.[22] However, some continue to grow and cause pain or they are palpated at the time of pelvic examination.[94] Adnexal masses are approximately 8 times more frequent in Norplant users compared with normally cycling women. Because these are simple cysts (and most regress spontaneously within 1 month of detection), they need not be sonographically or laparoscopically evaluated.[40] Further evaluation is indicated if they became large and painful or fail to regress. Regular ovulators are less likely to form cysts so the situation is likely to improve after 2 years of implant use. Etonogestrel implants suppress follicular development more profoundly; thus, ovarian cysts are less likely than with levonorgestrel implants.

Cancer

We can speculate on possible effects of implants based on our experience with oral contraceptives and depot-medroxyprogesterone acetate. The risk of endometrial cancer ought to be reduced. A study of the endometrial effects of Norplant failed to find any evidence of hyperplasia, even when levonorgestrel levels were low and endogenous estradiol production was normal.[95] The risk of ovarian cancer is also probably reduced, and we would expect a greater effect with Implanon because it more effectively suppresses ovulation. Breast and cervical cancer effects will be difficult to assess because of confounding variables as they are with oral contraception and depot-medroxyprogesterone acetate. The low progestin dose of implants, however, would be unlikely to have effects different from other hormonal contraceptives. In a very large case-control study, neither depot medroxyprogesterone acetate nor implants were associated with an increase in the risk of breast cancer.[96]

Post-marketing Surveillance Study

A large 5-year follow-up study in developing countries confirmed the low pregnancy rates associated with Norplant, 0.23 per 100 woman-years for intrauterine pregnancy and 0.03 per 100 woman-years for ectopic pregnancy.[92] When the women using Norplant were compared with women using nonhormonal methods of contraception and to the expected population rates, there was no excess of cancers, connective tissue diseases, or cardiovascular events. Importantly, the complaints of headache and mood disturbances (including anxiety and depression) were similar to those reported by women using oral contraceptives, although higher than for women using IUDs.

Insertion and Removal

The usual personal and family medical history and physical examination should concentrate on factors that might contraindicate use of the various contraceptive options. If a patient elects to use contraceptive implants, a detailed description of the method, including effectiveness, side effects, risks, benefits, as well as insertion and removal procedures, should be provided. Before insertion, the patient is asked to read and sign a written consent for the surgical placement of the implants. The consent reviews the potential complications of the procedure that include reaction to the local anesthetic, infection, expulsion of the implants, superficial phlebitis, bruising, and the possibility of a subsequent difficult removal.

Insertion can be performed at any time during the menstrual cycle as long as pregnancy can be ruled out. If the patient's last menstrual period was abnormal, if she has recently had sexual intercourse without contraception, or if there are reasons to suspect pregnancy, a sensitive urine pregnancy test is a wise precaution. Based on cervical mucus changes, a backup method need be used no more than 3 days after insertion.[18] Implants can be inserted immediately postpartum but certainly should be initiated no later than the third postpartum week in non-breastfeeding women and the third postpartum month in breastfeeding women. Acne and headache are less common in women who receive Norplant immediately postpartum, and there is no difference in postpregnancy weight loss compared with women who receive it 4–6 weeks later.[97]

Timing Summary

- Insert anytime during the first 5 days of the menstrual cycle if hormonal contraception is not being used.
- If hormonal contraception is being used, insert anytime during the hormone-free interval. If steroid contraception is used continuously, insert at anytime.
- If progestin-only contraception is being used, insert on the same day the next progestin injection is due or an implant or intrauterine device is removed. With progestin-only oral contraception, insertion can be performed anytime.
- Insert anytime within the first 5 days after an abortion or before the fourth week postpartum in nonbreastfeeding women.
- Insert before the fourth postpartum month in breast-feeding women; however if access to contraception is limited, it is appropriate to insert an implant immediately postpartum.
- No backup method is necessary if timing of insertion follows the above suggestions. If insertion occurs at other times, backup contraception is necessary for at least 4 days after insertion.

Patients should be questioned about allergies to local anesthetics, antiseptic solutions, and tape. A discussion about the technique of insertion and anticipated sensations is an important part of preparing the patient for the experience. All patients approach insertion with some degree of apprehension that can be decreased by detailed explanations and preparation.[98, 99]

Selection of the site for placement of implants is based on both functional and aesthetic factors. Various sites (the upper leg, forearm, and upper arm) have been used in clinical trials. The nondominant, upper, inner arm is the best site. This area is easily accessible to the clinician with minimal exposure of the patient. It is well protected during most normal activities. It is not highly visible, and migration of the implants from this site has not been documented. The site of placement does not affect circulating progestin levels. Careful implant insertion is the key to trouble-free removal.

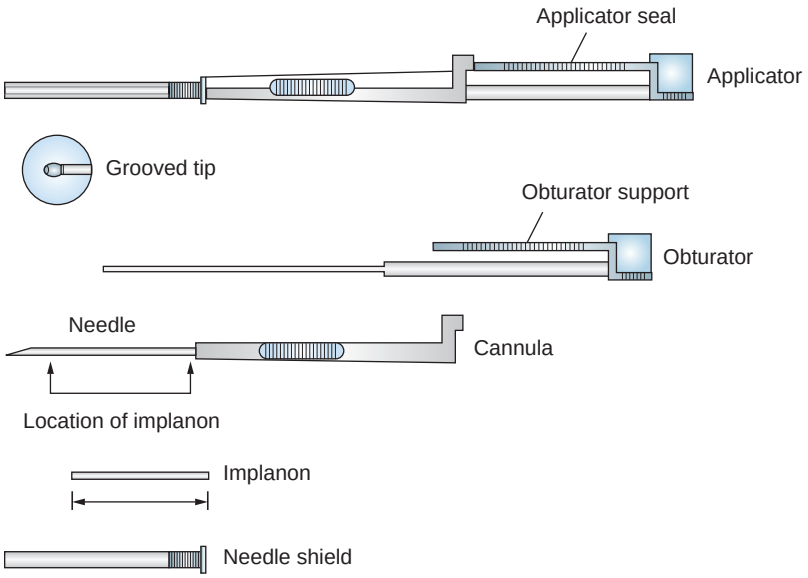
Implanon offers important insertion advantages compared with Norplant.[100] Of course, only one rod simplifies and shortens both insertion and removal. In addition, a pre-loaded applicator is provided that facilitates placement. If necessary, Implanon can usually be visualized by ultrasonography.[101] However, if a nonpalpable rod is not visible on ultrasonography, definitive localization is best achieved with magnetic resonance imaging (MRI).[102] Removal of the Implanon rod uses the “fingers alone” technique with a 2 mm incision.[103] Insertion complications (mainly deep insertions) are rare, and removal complications (difficulty finding the implant or a broken implant) occur at a far greater lower frequency compared with Norplant.[1, 100]

Insertion Technique

Insertion is carried out under local anesthesia in the office or clinic by someone, usually a physician or nurse practitioner, trained in the technique described here.[104] The procedure takes 5–10 min for a six implant system, and 2–3 min for a single implant.[105]

Required Equipment for Implanon Insertion

-2.5 mL syringe.
-0.5 inch, 25-gauge needle for injecting the anesthetic.
-1% chloroprocaine or lidocaine without epinephrine.
-Antiseptic solution.
-Adhesive strip for puncture closures.
-Elastic pressure bandage.



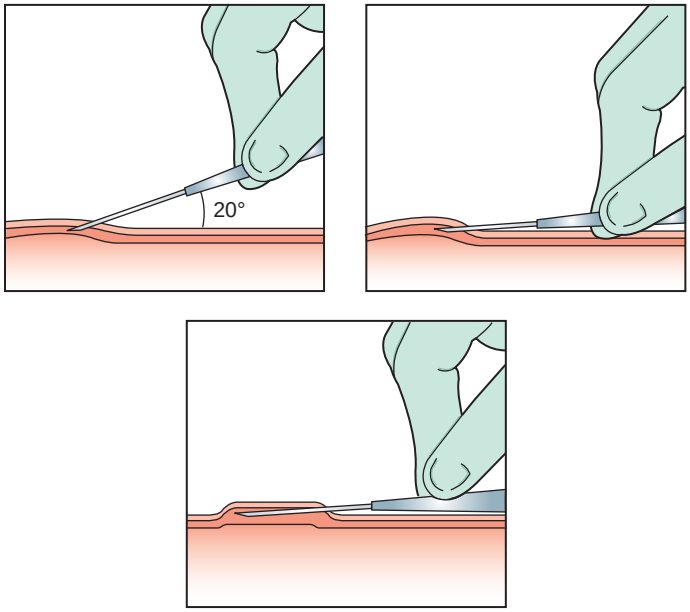
Positioning the Patient

The patient is placed in a supine position with the full length of her arm exposed. The upper inner arm is positioned by bending the elbow to 90 degrees and rotating the arm out, allowing full exposure of the insertion site at the crease between the biceps and triceps muscles. Adequate support under the arm should be provided to ensure comfort. To minimize the risk of infection, strict aseptic technique should be maintained throughout the procedure. An insertion site approximately 3–4 fingerbreadths (6–8 cm) superior and lateral to the medial epicondyle of the humerus is identified. A sterile drape is placed under the arm, and the insertion site on the arm is cleaned with an antiseptic such as povidone-iodine.

Anesthesia

Local anesthesia for the incision is obtained by raising a wheal of 1% chloroprocaine or lidocaine using a 25-gauge needle and injecting 1–2 mL under the skin along the track of the implant insertion needle.

Verify the presence of Implanon by looking carefully at the tip of the needle. If the implant (a white rod) is not visible, turn the applicator needle down and gently tap on a surface with the needle cover in place until the Implanon is seen, then tap the base of the applicator with the needle pointed up until the implant is no longer visible. Keep the applicator sterile.



Incision and Placement

The insertion needle and its obturator can be pushed directly through the skin at no greater than a 20 degree angle without making an incision. The needle is advanced as superficially as possible under the skin by maintaining a slightly upward angle on the trocar. To minimize the chance of too deep an insertion, lift or tent the skin with the tip of the needle. If the skin dimples, the needle is too superficial. Pull the needle back and redirect. Advance the needle to its full length while applying counter traction to the skin at the insertion site.

Once the needle has been fully advanced, break the seal on the applicator by pressing the obturator support. Turn the obturator 90 degrees in either direction with respect to the cannula and fix the obturator with one hand. With the other hand, slowly pull the needle out of the arm, leaving the implant behind under the skin.

Immediately after insertion, palpate the implant to verify correct insertion (both ends should be palpable). Look for the grooved tip of the obturator visible inside the needle. If the implant is not palpable and not within the needle, it must be located before contraception can be assured. If placement is in doubt, another contraceptive method must be used. Ultrasonography is the easiest way to identify Implanon, which is not radio-opaque. If necessary, MRI can localize the implant.[102]

After the insertion, show the patient how to palpate the implant. Place an adhesive closure or bandage over the insertion site. Complete the patient chart label and the "User Card" that must be given to the patient.

Most women experience little pain during the insertion,[98] but if it occurs the discomfort can be relieved with aspirin, acetaminophen, or a nonsteroidal anti-inflammatory agent. Infection or expulsion of the implants is rare (less than 1% with the Norplant system) and usually occurs when an implant is left pressing against the wound.[106]

The most commonly reported discomfort is a burning sensation during the injection of the local anesthetic. This effect of local anesthetic can be eliminated for most patients by adding

1 meq of sodium bicarbonate to each 10 mL of anesthetic (this shortens shelf life to 24 h).[107] After the onset of anesthesia in 2–3 min, most women feel no more than a pressure sensation.

Complications of Insertion

Potential complications include infection, hematoma formation, local irritation or rash over the implants, expulsion of the implant, and allergic reactions to adhesives of the dressing. Implanon can migrate a short distance (less than 2 cm) over time.[108] The incidence of complications is minimized by clinician training and experience, and the use of strict aseptic technique. *Post-insertion pregnancies in Australia and the Netherlands were commonly due to a failure to insert the implant (allowing the implant to fall out prior to insertion). The clinician must make sure the implant is visualized in the trocar prior to insertion and after insertion is palpable under the skin.*

Infection

The rate of infection varies among clinics and countries. The overall risk of infection after Norplant insertion is 0.8%.[106] Infections usually occur within the first week after insertion, but can present as long as 4–5 months later. Infection can be treated either by the removal of the implant or the administration of oral antibiotics while the implant remains in place. One-third of insertion site infections treated with antibiotics are unresponsive to therapy and require removal.[106] There have been no reports of infections leading to serious injury. Rarely, a superficial phlebitis develops. If it resolves over 1–2 weeks with heat and elevation of the arm, the implants need not be removed.

Expulsion

Expulsion of one or more of the implants occurs in 0.4% of Norplant users, usually within the first few months.[106] The majority of expulsions are associated with concurrent infection at the insertion site. Another cause of expulsion is failure to advance the implants far enough from the incision, causing pressure on the incision by the distal tip of the implant.

Local Reactions

Although not common, hematomas can form. The use of a pressure dressing for 72 h will prevent enlargment. Application of an ice pack for 30 min immediately after insertion also helps. Local irritation, rash, pruritus, and pain occur in 4.7% of Norplant users, usually during the first month of use.[106] Allergies to skin closure strip adhesives or to latex gloves account for some reactions.[109] These problems resolve spontaneously, but itching can be relieved by topical corticoid steroids.

Removal Techniques

Although implant removal is an office procedure requiring only a small amount of local anesthesia and a few simple instruments, instruction and practice are necessary.[107]

Practicing on a model arm after viewing an instructional video makes first removals faster and less uncomfortable for both clinician and patient. A removal kit with a model arm, a manual, and compact disc illustrating the technique is available at no charge from Schering-Plough at 877-467-5266. As for insertion, the patient should read and sign an informed consent to be filed in her medical record. We recommend that the patient be given a copy.

Proper positioning of the implant at the time of insertion is the most important factor influencing ease of removal. If the Norplant implants have been inserted with the distal tips (those away from the axilla) far apart or with implants crossing or touching one another, or too deeply, a larger incision and more time are required. Removal is easiest when the implants are just under the skin with their distal tips close together in a fan shape. The fibrous sheaths that form around implants can also make removal more difficult, especially if they are dense. The one and two implant systems can be removed faster and with less pain than the 6 Norplant capsules.[103]

Most removals are not painful (80% of patients reported pain as "none" or "slight"), and systemic analgesia is not required.[110] Time for removal of 6 capsules ranges from 5 to 40 min, with an average of 20 min. For the single rod (Implanon), the average time is about 4 min.[105] The most common cause of discomfort during the procedure is injection of the local anesthetic. Patients may feel pressure or tugging from manipulation of the fibrous sheaths and the implants, but these sensations are not severe if the clinician waits a few minutes after injection of the local anesthetic. The sheaths that form around Implanon are less dense than those around Norplant capsules.

Removal with Instruments

This approach to removal is the one described in the Norplant package insert and has been used around the world for 15 years. The technique requires three small sterile drapes (one fenestrated), sterile gloves, antiseptic solution such as povidone-iodine, 25-gauge 1.5 inch needle with a 3 mL syringe, local anesthetic (1% lidocaine with 1:100,000 epinephrine, buffered with 1 meq sodium bicarbonate per 10 mL lidocaine), one curved and one straight mosquito clamp, 4 × 4 sterile gauze sponges, and a no. 11 blade scalpel. This method is more appropriate for removal of the 6 capsule Norplant system than for one or two rods, which can usually be removed using fingers alone.[103]

The patient is placed in a supine position with her arm flexed and externally rotated as for insertion. A thick book positioned under the patient's arm can make her more comfortable and provide a better operating field. A sterile towel is placed under the arm. The implants are best seen by stretching the skin above and below the implants. Palpate all 6 of the implants before starting; if some portion of every implant cannot be felt, it may be better to sonographically or radiographically image (see below) the impalpable ones before removal because when the palpable implants are gone, they are lost as landmarks.

The skin is cleaned with the antiseptic solution, preparing a wide area above and below the implants so that the incision won't be contaminated during manipulations for removal. Scrape the antiseptic solution from the skin lying over the implants (the sterile stick of a cotton tipped applicator can be used) and let the arm dry. This will leave an impression of the implants that helps find them for removal. Drape the arm with a fenestrated towel and use a third towel to create a sterile field for instruments on a Mayo stand or table.

Wearing sterile gloves, an incision site is selected by pressing down on the proximal ends of the capsules and palpating their distal tips with a finger. Careful selection of the incision

site is the most critical step for easy removal. The best incision site is right at the distal tips, midway between the most medial and lateral implants. This can be the same as the insertion site, but generally the removal incision is made a few millimeters higher up on the arm to ensure placing it as close as possible to the tips of all the implants.

A local anesthetic containing 1:100,000 epinephrine reduces bleeding and allows better visualization of the implants. The 25-gauge needle is used to raise a 1 cm wheal of local anesthetic just under the tips of the implants. About 2 mL are sufficient, although more may be required later. Injection of too much anesthetic over the implants can obscure the tips and make removal more difficult. A 3–5 mm incision is made with the no.11 scalpel right at the mid point of the cluster of implant tips. A larger incision is not usually required and can cause bleeding that can obscure the implants. Implants can be removed by the clinician either sitting or standing, but if sitting, a wheeled stool allows repositioning as needed.

The implant that is most superficial and closest to the incision is removed first. This implant is pushed gently toward the incision with the fingers until the tip is visible and can be grasped with a curved mosquito clamp. The fibrous sheath covering the implant is dissected away using a finger covered with an opened gauze sponge. If the sheath is too dense for the sponge, it can be cautiously dissected with the straight clamp, a needle tip, or, for really dense sheaths, with the scalpel, taking care not to cut open the implant. If the point of the scalpel blade is used to nick the sheath over the thick silastic plug at the tip of the implant, the implant itself will not be cut, but if the sheath is incised across the thin walls of the implant, the implant can be severed and require removal in two portions. If the sheath must be incised with the scalpel, the incision should be along, not across, the implant.

Once the sheath is opened and the white tip of the first implant is exposed, it is grasped with the straight clamp. The curved clamp is released and the implant is gently pulled out. This procedure is repeated with the remaining implants.

If the implant tips cannot be guided to the incision with digital pressure on the skin above the implants, the jaws of the straight mosquito are inserted into the incision and opened just beneath the skin to separate the tissue layers. The straight clamp is removed, and the curved clamp is inserted with the tips pointing upward toward the skin. The clamp is opened and the implant is guided down between the jaws with a forefinger on the skin above the implant. This downward pressure on the tips of the clamp is often the most painful part of the removal procedure. When the implant is pushed between the jaws of the clamp, the clamp is secured at the first or second ratchet. Too much pressure on the implant can fracture the silastic capsule, making removal more difficult. The implant should not be pulled out with the curved clamp. If the implant cannot be seen, after gentle traction, the clamp handle is flipped 180 degrees until it points in the opposite direction, toward the patient's head. A portion of the sheath is cleared with an opened sponge, or if necessary, the scalpel tip, incising longitudinally, not across the implant. The exposed portion is then grasped with the straight clamp, the curved clamp is released, and the implant is removed with gentle traction. The procedure is repeated until all the implants are removed.

At the completion of the procedure, the implants should be counted to ensure that all have been removed. If any of the implants have been broken, the pieces should be aligned and compared with an intact capsule to determine that all of the implant has been removed. An adhesive strip is used to close the incision while pinching the skin edges together. A pressure dressing is then applied as after insertion, and removed the next day. The fibrous sheaths can remain for months causing the patient to think that implants were left behind. For that reason, it is important to show the implants to the patient at the time of removal.

If removal of some of the implants is difficult, painful, or prolonged, the procedure should be interrupted and the patient should return in a few weeks to complete the removal.

The remaining implants will be easier to remove after bleeding and swelling have subsided. A new incision can be made closer to the implants that were difficult to remove the first time. Even if some of the implants remain, the patient should immediately begin to use another method of contraception.

Removal with Fingers Alone

Implants can be removed with less pain and bleeding, and through a smaller incision if the use of instruments is avoided. The amount of trauma and bruising in the surrounding tissues is decreased, the scar is less visible, and the risk of breaking the implants is reduced. The disadvantage of this approach is that it may not be successful for implants that were poorly aligned or too deeply inserted.This technique is especially appropriate for the one and two rod systems.[103]

After preparation of the patient, the distal tip of the implant is palpated. If the implant cannot be felt, removal should be postponed until it has been localized with ultrasonography or radiography. No more than 0.5 mL of buffered lidocaine with epinephrine is injected into the dermis immediately under the implant tip, raising a wheal of about 1 cm in diameter. Too much anesthetic makes it difficult to locate the implant tip under the skin. The area of the injection should be massaged to disperse the anesthetic. Pressure is applied with fingers on the proximal (axillary) end of the implant so that the distal tips presses up against the skin. A 2–3 mm longitudinal incision is made through the skin onto the tip of the implant until the rubbery sensation of the implant can be felt against the point of the scalpel blade. The fibrous sheath is incised by nicking the sheath with the tip of the scalpel blade against the implant tip. It may take several passes across the tip with the scalpel held in different directions to fully open the sheath.

As the sheath is opened, the end of the implant will come into view. With finger pressure on its other end, the implant can be pushed through the incision until it can be grasped with mosquito forceps or fingers and pulled out. The incision is closed with an adhesive strip and covered with a sterile gauze and a pressure bandage.

Holding the implant up against the incision with finger pressure is critical for success with this “Pop Out” technique. If pressure is released, the implant will slip back to the position defined by the fibrous sheath around it. As the implant is manipulated using the fingers of both hands, the scalpel must be held so that it is immediately available to incise the sheath without releasing the implant. It is best to keep the scalpel in one hand with thumb and index finger while manipulating the implant, holding the implant with the rest of the fingers of both hands.

If the implant will not move toward the incision with finger pressure, it can be grasped with a hemostatic or vasectomy clamp, but the incision will usually have to be lengthened 2–3 mm in order to admit the clamp. The procedure followed is then as described for instrumental removals above. It may be necessary to inject more local anesthetic, but not more than 1 mL at a time where the clamp will be applied to the implant.

Difficult Removals

The incidence of difficult removals in the large post-marketing surveillance study of Norplant was 10.1 per 1,000 removals.[92] Removal is more difficult if the implants are broken during attempts to extract them. Once an implant is damaged, it can fracture repeatedly

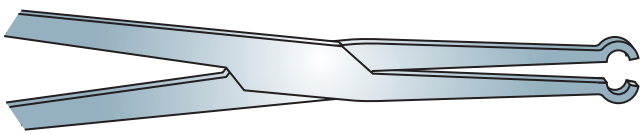
with further attempts to grasp it with clamps. To decrease this risk, the implants should be grasped at their ends whenever possible and as little traction as possible should be used for exposure and removal. If the scalpel is required to open the fibrous sheath around the implant, care should be taken to avoid slicing the capsule. If it has not been possible to grasp the end of implant, in order to open the fibrous sheath, incise along the length of the implant; cut longitudinally, not across, the implant. Rarely, removal of cut or broken implants will require an additional incision at the proximal end of the implant so that the remaining piece can be removed. Even more rarely, an implant can neither be palpated under the skin nor found through an incision. Such "lost" implants are most easily located with a high frequency (7–10 mHz), short focus ultrasound just prior to the removal procedure to help place the incision directly over the implant.[111] Use a transverse orientation to identify the shadow (the implant itself is more difficult to see), measure the depth, and draw lines representing the location on the surface of the skin using a paper clip as a marker.

When an implant is not palpable, imaging techniques for localization are required. Three techniques are particularly useful: mammography, sonography, and digital subtraction fluoroscopy. Compression film screen mammography is superior to standard plain film radiography. Ultrasonography requires a linear array transducer (preferably 7 MHz) applied to the upper arm with its long axis oriented perpendicular to the long axis of the humerus. The transducer is slowly moved until the characteristic acoustic shadowing of the implant is visualized. To measure the depth of each capsule, the transducer is repositioned along the axis of the implant to identify the length and both ends. Depth is determined using electronic calipers. Real-time sonography guidance can be useful during the removal procedure. Digital subtraction fluoroscopy relies on ghost images caused by motion during acquisition. Because it is readily available, and can be employed during removal, sonography is a first choice.

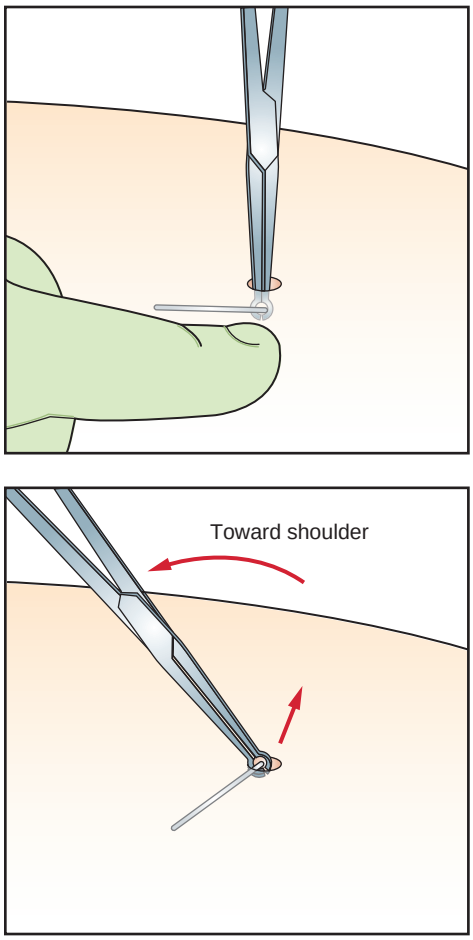
Another instrumental technique employs a modified vasectomy forceps and is very useful for removing deeply or asymmetrically placed Norplant implants. It requires a larger incision made in the center of the field of implants. The vasectomy forceps is advanced under the skin toward the mid portion of the implants. Those in the center are grasped first (in the middle of each implant), pulled into the incision, and cleaned free of their fibrous sheath as in the standard technique. The implant is then extracted, bending it in the middle in a "U" shape. The implants furthest away from the incision are removed last by advancing the forceps under the skin.[112]

We have found this approach to be especially useful for deeply placed, single capsules that are otherwise difficult to remove. The incision is made directly above the mid portion of the implant as determined by sonography or compression radiography. The scalpel blade (or a 25-gauge needle) is advanced to the depth of the implant as determined by imaging to feel for the capsule. The vasectomy forcep is advanced along the same track until the capsule can be grasped and elevated into the incision, freed from its fibrous sheath, and extracted.

Experienced clinicians agree that about half of difficult removals are due to improper placement.[39, 113] With one and two implant systems, removals are easier, but careful insertion will remain the real secret to trouble-free removal.



Vasectomy forceps



Reinsertion

A new implant can be inserted immediately through the same incision used to remove the old implant, or a new implant can be placed in the other arm.

Reasons for Termination

Although implants are long-term methods (2–7 years), only approximately 30% of women continue Norplant for 5 years (although in some cultures 5-year continuation rates reach 65–70%). Discontinuation occurs at a rate of 10–15% yearly, about the same as for intrauterine contraception, but lower than for barrier or oral contraception.[15, 55, 98, 114] Bothersome side effects, such as menstrual changes, headache, or weight change, are the primary reasons for termination of implant use.[4, 89, 115, 116] Menstrual changes are the most common cause for discontinuation of implants in the first year of use. An unspoken concern for many patients and their partners is the fact that bleeding irregularity interferes with sexual interactions. Users who cannot tolerate these symptoms request removal in the first 2 years of use whereas women who want another pregnancy, the most common personal reason for removal, are more likely to terminate use in the third or fourth year. Headache has been observed at a lower rate with Implanon compared with Norplant.[12]

User Acceptance of Contraceptive Implants

Overall, interview surveys throughout the world have indicated that women perceive sustained-release methods as highly acceptable methods of contraception.[99, 116–119] The most popular feature of implants is their ease of use. Approximately 20% of U.S. patients reported that friends and relatives notice their implants. This may be a greater problem in warmer climates with less encompassing clothing. Only 25% of the women who report that the implants were noticed were bothered by this attention.[4]

In the U.S., the primary motivations for implant use have been problems with previous contraceptive methods and ease of implant use. Although fear of pain during implant insertion is a prominent source of anxiety for many women, the actual pain experienced does not match the expectations. The level of satisfaction has been high in self-motivated and well-informed users.[120] Teenagers provide an example of well-documented success. Their 1-year pregnancy rates are much lower, and continuation rates much higher than that with oral contraceptives.[41, 121–124] However, teenage discontinuation of the method due to side effects (especially irregular bleeding and weight gain) is more common with Norplant.[43] The lower rate of irregular bleeding with Implanon contributes to higher patient acceptability, but irregular bleeding continues to be the major reason for discontinuation.[13, 33, 116]

Counseling Women

Frank information about negative factors such as irregular bleeding will avoid surprise and disappointment and encourage women to continue use long enough to enjoy the positive attributes such as convenience, safety, and efficacy. Open discussion of side effects will lead to public and media awareness of the disadvantages as well as the advantages of these methods. Helping women decide if they are good candidates for use of implants before they invest too much time and money in long-acting contraception is a very important objective of good counseling.

Common patient questions regarding contraceptive implants are as follows:

- Is it effective?
- How is it inserted and removed; how long do these procedures take; does it hurt, and will it leave scars?
- Will the implants be visible under the skin?
- Will the implants be uncomfortable or restrict movement of the arm?
- Will the implants move in the body?
- Will the implants be damaged if they are touched or bumped?
- Will this contraceptive change sexual drive and enjoyment?
- What are the short- and long-term side effects?
- Are there any effects on future fertility?
- What do the implants look and feel like?
- What happens if pregnancy occurs during use?
- How long will it take for the method to be effective after insertion?
- Can a partner tell if this method is being used?

Other Single Rod Systems

Uniplant (also Surplant) is a single implant contraceptive, containing 55 mg nomegestrol acetate in a 4 cm silicone capsule with a 100 μg/day release rate. It provides contraception

for 1 year.[125–127] A single silicone implant containing nestorone (Nestorone) is effective for 2 years; another version (Elcometrine) lasts 6 months.[2]

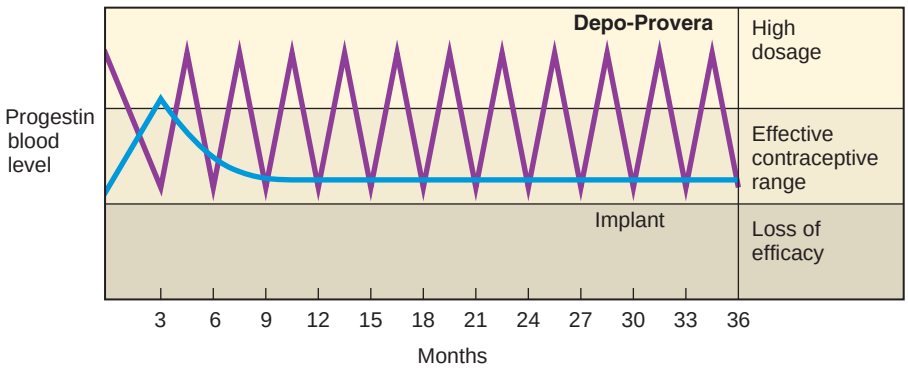
Injectable Contraception

Depot-medroxyprogesterone acetate (Depo-Provera) is the most thoroughly studied progestin-only contraceptive. Although its approval for contraception in the U.S. is relatively recent (1992), it has been available in some countries since the mid-1960s. Much of our knowledge of the safety, efficacy, and acceptability of long-acting hormonal contraception comes from Indonesia, Sri Lanka, Thailand, and Mexico where depot-medroxyprogesterone acetate has been used and studied for decades. The long-delayed approval as a contraceptive in the U.S. was based on political and economic considerations not scientific ones.[128]

Depot-medroxyprogesterone acetate is formulated as microcrystals, suspended in an aqueous solution. The correct dose for contraceptive purposes is 150 mg intramuscularly (gluteal or deltoid) every 3 months. A comparative trial established that the 100-mg dose is significantly less effective.[129] The contraceptive level is maintained for at least 14 weeks, providing a safety margin for one of the most effective contraceptives available, about 1 pregnancy per 100 women after 5 years of consistent use.[129, 130]

A newer formulation allows the self administration into a thigh or the abdomen of a subcutaneous dose of 104 mg every 3 months.[131–133] Pre-filled syringes contain 0.65 mL of an aqueous suspension of medroxyprogesterone acetate.

Depot-medroxyprogesterone acetate is not a “sustained-release” system; it relies on higher peaks of progestin to inhibit ovulation and thicken cervical mucus. The difference between serum levels of progestins in a sustained-release system like Implanon and a depot system like depot-medroxyprogesterone acetate is illustrated in the diagram.



Other widely-used injectables are norethindrone enanthate, 200 mg every 2 months, and the monthly injectables, Lunelle (25 mg medroxyprogesterone acetate and 5 mg estradiol cypionate) and Mesigyna (50 mg norethindrone enanthate and 5 mg estradiol valerate).

Mechanism of Action

The mechanism of action of depot-medroxyprogesterone acetate is different from the other lower dose, progestin-only methods because, in addition to thickening of the cervical mucus

and alteration of the endometrium, the circulating level of the progestin is high enough to effectively block the LH surge, and, therefore, ovulation does not occur.[134] FSH is not suppressed as it is with the combination oral contraceptive; therefore, follicular growth is maintained sufficiently to produce estrogen levels comparable to those in the early follicular phase of a normal menstrual cycle.[135] Symptoms of estrogen deficiency, such as vaginal atrophy or a decrease in breast size, do not occur.

Accidental pregnancies occurring at the time of the initial injection of depot-medroxyprogesterone acetate have been reported to be associated with higher neonatal and infant mortality rates, probably due to an increased risk of intrauterine growth restriction.[136, 137] The timing of the first injection is, therefore, very important. To ensure effective contraception, the first injection should be administered within the first 5 days of the menstrual cycle (before a dominant follicle emerges), or a backup method is necessary for 7 days.[131, 138–140] The Quick Start, same-day start protocol can be used with depot-medroxyprogesterone acetate, with special care to rule out pregnancy and providing a backup method for 7 days.[141] Same-day starts with depot medroxyprogesterone acetate produce better continuation rates with fewer unintended pregnancies.[142]

The duration of action can be shortened if attention is not paid to proper administration. The intramuscular injection must be given deeply by the Z-track technique and not massaged. It is prudent to avoid locations at risk for massage by daily activities.

When given properly, there is an effective 2-week grace period that allows for late reinjections; a study of women arriving late for reinjections concluded that even a 4-week late reinjection provided equivalent protection against pregnancy.[143] Women who are more than 4 weeks late for reinjection should be tested for pregnancy, reinjected if the test is negative, and advised to use backup contraception for 7 days.

Efficacy

The efficacy of this method (in both the intramuscular and subcutaneous formulations) is slightly better than that of sterilization and better than that of all the other temporary methods.[46, 131–133, 144, 145] In a comparison of the intramuscular and subcutaneous methods, the blood levels of medroxyprogesterone acetate are approximately 30% lower with subcutaneous administration of the lower dose, but efficacy is not impaired.[131] Because serum concentrations are relatively high, efficacy is not influenced by weight (making this method a good choice for overweight women) or by the use of medications that stimulate hepatic enzymes.[131, 132]

Indications

1. At least 1 year of birth spacing desired.
2. Highly effective long-acting contraception not linked to coitus.
3. Private, coitally independent method desired.
4. Estrogen-free contraception needed.
5. Breastfeeding.
6. Sickle cell disease.
7. Seizure disorder.

Absolute Contraindications

1. Pregnancy.
2. Unexplained genital bleeding.
3. Severe coagulation disorders.
4. Previous sex steroid-induced liver adenoma.

Relative Contraindications

1. Liver disease.
2. Severe cardiovascular disease.
3. Rapid return to fertility desired.
4. Difficulty with injections.
5. Severe depression.

Advantages

Like sustained-release forms of contraception, depot-medroxyprogesterone acetate is not associated with compliance problems and is not related to the coital event. Continuation rates are better and repeat pregnancy rates are reduced compared with oral contraceptive use in teenagers; however, continuation and repeat pregnancy rates are similar when adolescents begin these methods in the immediate postpartum period.[146, 147] Depot-medroxyprogesterone acetate is useful for women whose ability to remember contraceptive requirements is limited. It should be considered for women who lead disorganized lives or who are mentally retarded.

The freedom from the side effects of estrogen allows depot-medroxyprogesterone acetate to be considered for patients with congenital heart disease, sickle cell anemia, patients with a previous history of thromboembolism, and women over age 30 who smoke or have other risk factors such as hypertension or diabetes mellitus. The absolute safety in regard to thrombosis is mainly theoretical; it has not been proven in a controlled study. However, an increased risk of thrombosis has not been observed in epidemiologic evaluation of depot-medroxyprogesterone acetate users, and a World Health Organization (WHO) case-control study could find no evidence for increased risks of stroke, myocardial infarction, or venous thromboembolism.[130, 148]

Two case-control studies, one using data from the WHO Collaborative Study and one using the data from the U.K. general practice research database, assessed the risk of idiopathic venous thrombosis in users of progestins alone for therapeutic purposes, not for contraception, and concluded that therapeutic progestins alone may be associated with an increased risk of venous thromboembolism.[149, 150] These epidemiologic conclusions were based on extremely small numbers (only three cases in one report and five in the other) and had very wide confidence intervals. Patients who receive progestin-only for therapeutic reasons are

probably older and are more likely to have family histories of cardiovascular disease. In addition, a problem of preferential prescribing is present in that clinicians are more likely to promote the use of progestin-only for women they perceive to be at greater risk of venous thromboembolism. Thus, it is likely that the case groups represented a higher risk group than the control groups in these reports. For these reasons, we do not believe progestins are associated with an increased risk of venous thromboembolism.

An important advantage exists for patients with sickle cell disease because evidence indicates an inhibition of in vivo sickling with hematologic improvement during treatment.[151] Both the frequency and the intensity of painful sickle cell crises are reduced.[152]

Another advantage is the finding that depot-medroxyprogesterone acetate increases the quantity of milk in nursing mothers, a direct contrast to the effect seen with estrogen-progestin oral contraception. The concentration of the drug in the breast milk is negligible, and no effects of the drug on infant growth and development have been observed.[153–155] In a careful study of male infants being breastfed by women treated with depot-medroxyprogesterone acetate, no metabolites of depot-medroxyprogesterone acetate could be detected in the infant's urine and no alterations could be observed in the infant levels of FSH, LH, testosterone, and cortisol.[156] Because of the slight positive impact on lactation, depot-medroxyprogesterone acetate can be administered immediately after delivery. A study to investigate the impact of early initiation found no adverse effects on breastfeeding.[157] In breastfeeding, overweight, Latina women with prior gestational diabetes, the progestin-only oral minipill was associated with a 3-fold increased risk of non-insulin dependent diabetes mellitus.[158] In a similar cohort of Hispanic women, depot medroxyprogesterone acetate was associated with a small increase in subsequent diabetes mellitus that was not statistically significant, a risk that was even lower and less significant when adjusted for the higher body weights and a greater prevalence of family history for diabetes in the users of injected contraception.[159] When compared with oral contraceptive use in Navajo women, depot medroxyprogesterone users were more likely to gain weight and develop diabetes mellitus.[160, 161] ***It is possible that overweight women who already have significant insulin resistance become overtly diabetic by the added effect of progestins in a low-estrogen environment (lactation) or in an induced low-estrogen state (depot medroxyprogesterone acetate). However, it is likely that the independent contribution of excess body weight is the more critical factor.***

Depot-medroxyprogesterone acetate is an excellent contraceptive choice for women taking antiepileptic drugs because the high progestin levels raise the seizure threshold.[162] An improvement in seizure control can be achieved probably because of the sedative properties of progestins.[162]

Women who are anticoagulated or who have bleeding disorders are prone to develop heavy menstrual bleeding and hemorrhagic ovarian cysts. Experience with depot medroxyprogesterone acetate in these patients is limited; however, we would expect a beneficial reduction in bleeding and a reduced risk of ovarian hemorrhage, especially from a corpus luteum.[163]

Other benefits associated with depot-medroxyprogesterone acetate use include a decreased risk of endometrial cancer comparable with that observed with oral contraceptives[164] and probably the same benefits associated with the progestin impact of oral contraceptives: reduced menstrual flow and anemia, less pelvic inflammatory disease (PID), less endometriosis, fewer uterine fibroids,[165] and fewer ectopic pregnancies. A failure to document a reduced risk of ovarian cancer by the World Health Organization probably reflects the study's low statistical power and the high parity in the depot-medroxyprogesterone acetate users.[166]

Depot-medroxyprogesterone acetate, like oral contraception, may reduce the risk of pelvic inflammatory disease; however, the only study was hampered by small numbers.[167] Suppression of ovulation means that ectopic pregnancies are abolished and ovarian cysts are rare.

The greater the number of choices that women have, the more likely they are to find a contraceptive that works well for them. For some women the primary advantages of depot-medroxyprogesterone acetate are privacy and ease of use. No one but the user need know about the injection, and the 3-month schedule can be easy to maintain for women who do not mind injections. In some societies, injections are respected as efficacious, and depot-medroxyprogesterone acetate is the most popular contraceptive despite bleeding changes and other side effects.

Summary of Advantages

1. Easy to use, no daily or coital action required.
2. Safe, no serious health effects.
3. Very effective, as effective as sterilization, intrauterine contraception, and implant contraception.
4. Free from estrogen-related problems.
5. Private, use not detectable.
6. Lactation enhanced.
7. Noncontraceptive benefits.

Problems With Depot-Medroxyprogesterone Acetate

Major problems with depot-medroxyprogesterone acetate are irregular menstrual bleeding, breast tenderness, weight gain, and depression.[129, 130] By far, the most common problem is the change in menstrual bleeding. Up to 25% of patients discontinue in the first year because of irregular bleeding.[121] The bleeding is rarely heavy; in fact, hemoglobin values rise in depot-medroxyprogesterone acetate users. The incidence of irregular bleeding is 70% in the first year, and 10% thereafter. Bleeding and spotting decrease progressively with each reinjection so that after 5 years, 80% of users are amenorrheic (compared with 10% of Norplant users).[168] With the subcutaneous preparation, the bleeding pattern is similar to that with the intramuscular product; 55% achieve amenorrhea at the end of the first year of treatment and 70% after 2 years.[132, 145, 169] Irregular bleeding can be disturbing and annoying, and, for many patients, it inhibits sexuality; therefore, most users prefer the amenorrhea that comes with prolonged use.

If necessary, breakthrough bleeding can be treated with exogenous estrogen, 1.25 mg conjugated estrogens, or 2 mg estradiol, given daily for 7 days. A nonsteroidal anti-inflammatory product given for a week is also effective, and another option is to administer an oral contraceptive for 1–3 months. Giving the depot-medroxyprogesterone acetate injection earlier (more frequently) does not change the bleeding pattern.[170] Most women can wait for amenorrhea without treatment if they know what to expect with time. Trying to regulate breakthrough bleeding with cyclic, repeated estradiol exposure has proved to be ineffective.[171] Chlamydia infection of the endometrial cavity is *not* a cause of the irregular bleeding associated with depot medroxyprogesterone acetate.[172]

In a large international study, the most common medical reasons for discontinuing depot-medroxyprogesterone acetate during the first 2 years of use were the following[130]:

1.	Headaches	2.3%
2.	Weight gain	2.1%
3.	Dizziness	1.2%
4.	Abdominal pain	1.1%
5.	Anxiety	0.7%

In Western societies, depression, fatigue, decreased libido, and hypertension are also encountered. Whether medroxyprogesterone acetate causes these side effects is difficult to know because they are very common complaints in nonusers as well.[173] When studied closely, no increase in depressive symptoms can be observed, even in women with significant complaints of depression prior to treatment.[174, 175]

Attempts to document a greater weight gain specifically associated with depot-medroxyprogesterone acetate have had mixed results, some finding no increase and others a small increase (e.g., about 4 kg over 5 years in one study and 11 kg over 10 years in another).[176–179] In an excellent large cohort study, 3-year users of depot-medroxyprogesterone acetate increased their body weight by 5.1 kg, body fat by 4.1 kg, percent body fat by 3.4% and developed an increase in central, visceral fat, all mostly in the first 18 months and in contrast to no weight gain in oral contraceptive users.[180] With the subcutaneous method, an average weight gain of 1.5 kg was observed after 1 year and 4.5 kg after 3 years, changes that are comparable to the intramuscular formulation.[132, 145, 181] In a cohort comparison study of adolescents, obese girls gained more weight (9.4 kg after 18 months) with depot medroxyprogesterone acetate when compared with oral contraceptives or no hormonal method.[182] A longitudinal 4–5-year follow-up study of adolescents in South Africa documented about a 4 kg greater weight gain with injectable contraception (both depot medroxyprogesterone acetate and norethindrone enanthate) compared with nonusers or oral contraceptive users.[183] Specific individuals and certain ethnic groups may be more susceptible to weight gain; for example, significant weight gain was reported in Navajo women using depot-medroxyprogesterone acetate.[160] In a prospective cohort study, women who experienced a significant weight gain within 6 months gained an average of about 7 kg more over 3 years, thus identifying a susceptible group of women, about 25% of depot medroxyprogesterone acetate users.[184]

Is weight gain a general reaction to depot medroxyprogesterone acetate or does it occur only in vulnerable individuals? An answer to this question is hindered by limitations in the available studies. The evidence is not derived from randomized trials (something that is probably impossible to do). Therefore, results can be influenced by those reasons for which subjects choose a certain method and responses that affect continuation with methods. The individuals who choose to use depot medroxyprogesterone differ in their socioeconomic status, contraceptive practices, and sexual histories; thus the difficulty in matching users and nonusers.

Although it is difficult to separate the hormone effect from the impact of lifestyle and aging, it is best to conclude that depot-medroxyprogesterone acetate injections are associated with a small increase in body fat and body weight, but not in all, probably not in most, women.

Remember that if symptoms are truly due to the progestin, unlike pills and implants, depot-medroxyprogesterone acetate takes 6–8 months to be gone after the last injection.[131] Clearance is slower in heavier women. Approximately half of women who discontinue depot-medroxyprogesterone acetate can expect normal menses to return in 6 months after the last injection, but 25% will wait a year before resumption of a normal pattern.[131, 168]

Several prospective studies have suggested that the use of depot-medroxyprogesterone acetate is associated with an increased risk of cervical infections, especially chlamydia and/or gonorrhea.[185–188] This finding could very well be influenced by the inability to perform a randomized trial and reflect the difficulty in matching users and nonusers. The women who chose to use depot-medroxyprogesterone acetate in these cohort studies were notably different in their socioeconomic status, contraceptive practices, and sexual histories; thus the results could reflect a higher rate of infection in the user group at baseline. Studies have not linked the use of depot-medroxyprogesterone acetate and an increased risk of HIV infection.[189, 190]

Anaphylaxis

As of 2010, there were three case reports of anaphylactic shock within minutes after receiving intramuscular injections of depot-medroxyprogesterone aceate.[191–193] The reactions were most likely to one of the inert substances present in the injections: parabens, polyethylene glycols, and polysorbates. Depot-medroxyprogesterone acetate intramuscular injections are best given by trained personnel in a clinic or office setting with resuscitation equipment and drugs available. No such reactions have been reported with subcutaneous administration.

Breast Cancer

Medroxyprogesterone acetate, in large continuous doses, produced breast tumors in beagle dogs (perhaps because in dogs progestins stimulate growth hormone secretion, known to be a mammotrophic agent in dogs).[194] This is an effect unique with dogs and has not appeared in women after years of use. A very large, hospital-based, case-control WHO study conducted over 9 years in three developing countries indicated that exposure to depot-medroxyprogesterone acetate is associated with a very slightly increased risk in breast cancer in the first 4 years of use, but there was no evidence for an increase in risk with increasing duration of use.[195] The number of cases with recent use was not large, and the confidence intervals reflected this. A possible explanation for this finding is the combination of detection/surveillance bias and accelerated growth of an already present tumor, a situation similar to those described with oral contraceptives (Chapter 22) and postmenopausal hormone therapy (Chapter 18).

Two earlier population-based, case-control studies indicated a possible association between breast cancer and depot-medroxyprogesterone acetate. One, from Costa Rica, was based on only 19 cases.[196] The other, from New Zealand, did not find an increased relative risk in ever users but did find an indication of increased risk shortly after initiating use at an early age, younger than age 25.[197] A pooled analysis of the WHO and New Zealand data indicated an increased risk in recent users and that the highest risk was in women who had received a single injection.[198] The risk, if real, is very slight, and it is equally possible that the suggestions of increased risk based on a small number of cases have not been free of confounding variables. A case-control study from Cape Town, South Africa, and a very large case-control study in the U.S. found no overall increase in the risk of breast cancer and no effect of increasing duration of use or recent use.[96, 199] Because recent use may be the key factor, it is appropriate to emphasize that all of these studies did not find evidence for an overall increased risk of breast cancer, and the risk did not increase with duration of use. However, clinicians should consider informing patients that depot-medroxyprogesterone acetate might accelerate the growth of an already present occult cancer. We would expect such tumors to be detected at an earlier stage and grade of disease and to be associated with a better outcome.

Other Cancers

An increased risk of cervical dysplasia cannot be documented even with long-term use (4 or more years).[200] No increase in adenocarcinoma or adenosquamous carcinoma of the cervix could be detected in the WHO study.[201] The WHO study did not detect an increased risk of invasive squamous cell cancer of the cervix in depot-medroxyprogesterone acetate users; however, the risk of cervical carcinoma *in situ* was slightly elevated in the WHO case-control study, and it is not certain whether this is a real finding or a consequence of unrecognized biases, especially detection bias.[202, 203] In New Zealand, a modest increase in the risk of cervical dysplasia among users of depot-medroxyprogesterone acetate could be attributed to an increased prevalence of known risk factors for dysplasia among women who choose this method of contraception.[200] Nevertheless, it is prudent to insist on annual Pap smear surveillance in all users of contraception, no matter what method. Women at higher risk because of their sexual behavior (multiple partners, history of STIs) should have Pap smears every 6 months. An immature cellular pattern in Pap smears from long-term depot-medroxyprogesterone acetate users can suggest the presence of squamous intraepithelial premalignant lesions, but biopsies in these cases reveal epithelial atrophy.[204]

As noted, depot-medroxyprogesterone acetate is associated with a reduction in the risk of endometrial cancer, and there is probably a modest reduction in the risk of ovarian cancer. There is no evidence that liver cancer risk is changed by the use of depot-medroxyprogesterone acetate.[205]

Metabolic Effects

The impact of depot-medroxyprogesterone acetate on the lipoprotein profile is uncertain. Although some fail to detect an adverse impact and claim that this is due to the avoidance of a first-pass through effect in the liver, others have demonstrated a decrease in HDL-cholesterol and increases in total cholesterol and LDL-cholesterol.[206, 207] In a multicenter clinical trial by the World Health Organization, a transient adverse impact was present only in the few weeks after injection when blood levels were high.[208] The clinical impact of these changes, if any, have yet to be reported. It seems prudent to monitor the lipid profile annually in women using depot-medroxyprogesterone acetate for long durations. The emergence of significant adverse changes in LDL-cholesterol and HDL-cholesterol warrant reconsideration of contraceptive choice.

There are no clinically significant changes in carbohydrate metabolism or in coagulation factors in healthy women.[209, 210] As noted earlier, the use of progestin-only contraception may increase the risk of type 2 diabetes mellitus, especially in breast-feeding women.[158] However, these studies have also indicated that this risk is influenced by baseline body weights and by weight gain.[159] It is possible that overweight women who already have significant insulin resistance become overtly diabetic by the added effect of progestins in a low-estrogen environment (lactation) or in an induced relatively low-estrogen state (depot medroxyprogesterone acetate). But, it is likely that excess body weight is the more critical factor.

Effect on Bone Density

The contraceptive use of depot-medroxyprogesterone acetate is associated with the short-term loss of bone. This is attributed to the fact that blood levels of estrogen with depot-medroxyprogesterone acetate are relatively lower over a period of time compared

with a normal menstrual cycle, an explanation that is supported by the demonstration that estrogen treatment prevents the bone loss.[211, 212] Lumbar and hip bone loss has been documented in both cross-sectional and longitudinal studies.[87, 213–217] This bone loss has also been observed in women receiving a high oral dose of medroxyprogesterone acetate, 50 mg daily, a dose that suppresses LH, resulting in low estrogen levels.[218] Another study documented decreased bone density that was enhanced by lower body weight and duration of amenorrhea.[219] An American cross-sectional study indicated a greater bone loss with increasing duration of use, especially in younger women, 18–21 years old.[220] The bone loss is comparable comparing intramuscular with the subcutaneous administration of depot medroxyprogesterone acetate.[133]

The bone loss associated with the use of depot medroxyprogesterone acetate can be minimized by focusing on risk factors. Women at greater risk for losing bone density with this method of contraception are those who have not experienced childbirth, who smoke, and who have an inadequate intake of calcium.[221] Calcium and vitamin D supplementation should be encouraged in most women, not just depot medroxyprogesterone acetate users, and an emphasis on smoking cessation is always worthwhile.

Bone density increases rapidly and significantly during adolescence. Almost all of the bone mass in the hip and the vertebral bodies will be accumulated in young women by age 18, and the years immediately following menarche are especially important.[222, 223] For this reason any drug that prevents this increase in bone density might increase the risk of osteoporosis later in life. Studies in adolescents have documented bone loss with depot medroxyprogesterone acetate compared with normal controls and young women using oral contraceptives.[83, 224, 225] But, there are reasons to expect this bone loss is regained.

An example of bone loss that is regained is the bone loss associated with lactation. Secretion of calcium into the milk of lactating women approximately doubles the daily loss of calcium.[226] In women who breastfeed for 6 months or more, this is accompanied by significant bone loss even in the presence of a high calcium intake.[227] However, bone density rapidly returns to baseline levels in the 6 months after weaning.[228–230] The bone loss is due to increased bone resorption, probably secondary to the relatively low estrogen levels associated with lactation. Calcium supplementation has no effect on the calcium content of breast milk or on bone loss in lactating women who have normal diets.[231] Most importantly, studies indicate that any loss of calcium and bone associated with lactation is rapidly restored; therefore, there is no impact on the risk of postmenopausal osteoporosis.[232–236]

The degree of bone loss in the previously mentioned depot medroxyprogesterone acetate studies is not as severe as that observed in the early postmenopausal years. Furthermore, this amount of bone loss is not so great that it cannot be regained. Bone density measurements in women who stopped using depot-medroxyprogesterone acetate indicated that the loss was regained in the lumbar spine but not in the femoral neck within 2 years even after long-term use, but in other cohorts of past users, both spinal density and hip density were regained in the 3–4 years after discontinuation.[217, 237, 238] A 7-year prospective cohort study demonstrated unequivocally that bone is regained, almost 100% within 2 years after depot medroxyprogesterone acetate is discontinued in both the lumbar spine and the hip.[239] Cohort studies in adolescents and young women demonstrated a similar gain in bone after discontinuation.[240, 241] Most importantly, cross-sectional studies of postmenopausal women in New Zealand and a large multicenter, worldwide population could not detect a difference in bone density comparing former users of depot-medroxyprogesterone acetate to never users, indicating that any loss of bone during use is regained.[87, 242]

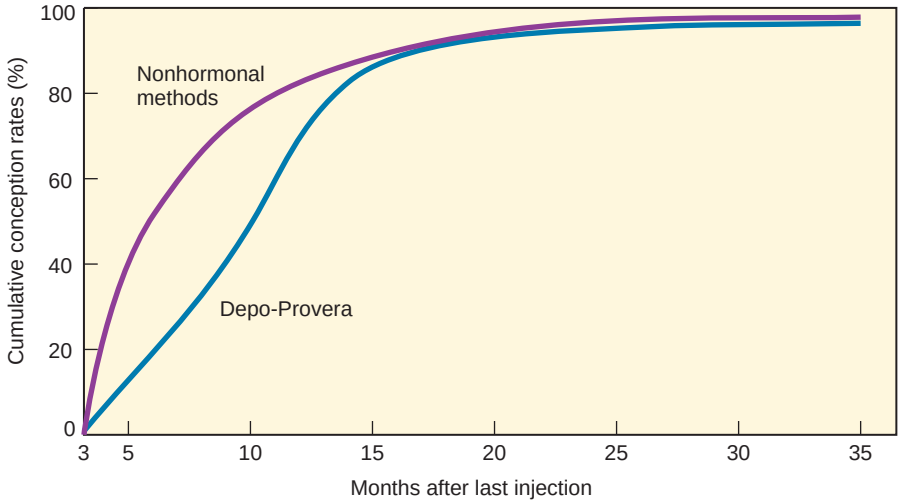
In 2004, the U.S. Food and Drug Administration indicated a concern for the bone loss associated with depot-medroxyprogesterone acetate and warned that this method should not be used longer than 2 years unless it was the only option. We agree that concern is appropriate, but disagree with the warning. The degree of bone loss and the evidence that the bone loss is regained, plus the similarity to the benign bone loss associated with lactation, all argue that the use and duration of use of depot-medroxyprogesterone acetate should not be limited by this concern, and that measurement of bone density or treatment with supplemental estrogen or bisphosphonates is not indicated (and would influence and complicate compliance). At the present time, in our view, the concern over bone loss should not be a reason to avoid this method of contraception, and there is no need to impose a time limit on duration of use. ***It is unlikely that bone loss occurs sufficiently to raise the risk of osteoporosis later in life. However, women who discontinue DMPA at or near their menopause should be encouraged to use hormone therapy in order to regain the lost bone.***

Galactorrhea

Galactorrhea is not associated with the use of depot medroxyprogesterone acetate. Prolactin gene transcription is stimulated by estrogen and mediated by estrogen receptor binding to estrogen responsive elements. The increase in prolactin during pregnancy parallels the increase in estrogen beginning at 7–8 weeks gestation, and the mechanism for increasing prolactin secretion is believed to be estrogen suppression of the hypothalamic prolactin-inhibiting factor, dopamine, and direct stimulation of prolactin gene transcription in the pituitary.[243, 244] Although requiring estrogen for prolactin secretion, prolactin stimulation of breast milk production is prevented by progestational agents and pharmacologic amounts of estrogen. Only colostrum (composed of desquamated epithelial cells and transudate) is produced during gestation. Full lactation is inhibited by progesterone, which interferes with prolactin action at the alveolar cell prolactin receptor level. Both estrogen and progesterone are necessary for the expression of the lactogenic receptor, but progesterone antagonizes the positive action of prolactin on its own receptor while progesterone and pharmacologic amounts of androgens reduce prolactin binding.[245–247] In the mouse, inhibition of milk protein production is due to progesterone suppression of prolactin receptor expression.[248] For these reasons, exposure to high levels of progestational agents, such as depot medroxyprogesterone acetate, is not associated with the clinical problem of galactorrhea.

Effect on Future Fertility

The delay in becoming pregnant after ceasing use of depot-medroxyprogesterone acetate is a problem unique to injectable contraception; all the other temporary methods allow a more prompt return to fertility.[249] However, medroxyprogesterone acetate does not permanently suppress ovarian function, and the concern that infertility with suppressed menstrual function may be caused by depot-medroxyprogesterone acetate has not been supported by epidemiologic data. The pregnancy rate in women discontinuing the injections because of a desire to become pregnant is normal.[250] By 18 months after the last injection, 90% of depot-medroxyprogesterone acetate users have become pregnant, the same proportion as with other methods.[251] The delay to conception is about 9 months after the last injection, and the delay does not increase with increasing duration of use. Because of this delay, women who want to conceive promptly after discontinuing their contraceptive should not use depot-medroxyprogesterone acetate. Suppressed menstrual function persisting beyond 18 months after the last injection is not due to the drug and deserves evaluation.



Determining Menopause in Long-Term Users

Depot medroxyprogesterone acetate will prevent the appearance of the two common markers for the onset of menopause: the loss of menstrual periods and hot flushing. Although the risk is small after age 50, pregnancy is still possible in some women. Knowing a woman is postmenopausal is important in order to minimize, if not eliminate, the risk of pregnancy, and there is reason to be concerned over the relatively low estrogen state associated with depot medroxyprogesterone acetate. Follicle-stimulating hormone (FSH) secretion is regulated by estrogen and inhibin, and therefore pharmacologic amounts of progestins do not restore postmenopausal levels of FSH to premenopausal levels.[252] Greater reliability in measuring FSH is achieved with injectable methods of contraception if the blood sample is obtained prior to the next scheduled injection. We are reluctant to follow the empiric practice that allows women to continue oral contraceptives to age 55 because we believe there is some urgency to expeditiously transfer the patient from the low estrogen state associated with depot medroxyprogesterone acetate to the early benefits provided by a program of postmenopausal hormone therapy. We recommend measuring FSH annually beginning at age 50. An FSH level greater than 20 mIU/mL is a indicator of postmenopause, although a level of 35–40 mIU/mL is considered more reliable.[253]

Short-Term Injectable Contraceptives

Monthly or every-other-month injectable combinations of estrogen and progestin are not new, having been developed over several decades.[254] This method of contraception is popular in China, Latin America, and Eastern Asia. A preparation widely used in China consists of 250 mg 17α-hydroxyprogesterone caproate and 5 mg estradiol valerate, known as Chinese Injectable No.1. A large follow-up study of Chinese women concluded that this monthly injection was associated with a significant reduction in endometrial cancer and no increase in other cancers, including breast cancer.[255]

Lunelle (also called Cyclofem, Cyclo-Provera, Feminena, or Lunella)

Lunelle consists of 25 mg depot-medroxyprogesterone acetate and 5 mg estradiol cypionate and is administered every 28–30 days (not to exceed 33 days) as a deep intramuscular injection. This method is as effective as depot-medroxyprogesterone acetate, but avoids the problems of menstrual irregularity and heavy bleeding, as well as amenorrhea.[256–261] In addition, the method is rapidly reversible; fertility rates after discontinuation are similar to oral contraceptives.[262] Besides the need for a monthly injection, another disadvantage is the likelihood that the combination of estrogen and progestin will inhibit lactation. The requirement for a monthly injection can be made more convenient by the use of an automatic device for self-administration.[263] Approximately 80% of women who are amenorrheic on depot-medroxyprogesterone acetate will develop vaginal bleeding if switched to Lunelle.[264] The same contraindications, concerns, problems, and probably benefits reported with oral contraception should apply to Lunelle.

Norethindrone Ethanthate

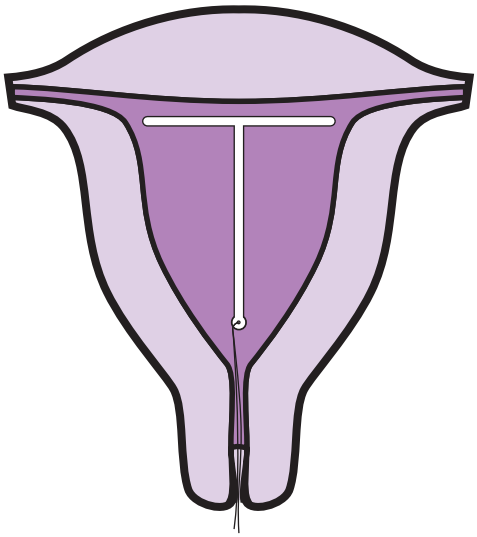
Norethindrone enanthate is given in a dose of 200 mg intramuscularly every 2 months. This progestin acts in the same way as depot-medroxyprogesterone acetate, and has the same problems.[130] A combination (Mesigyna, Norigynon, Noristerat, Norigest, NET-EN) of norethindrone enanthate (50 mg) with estradiol valerate (5 mg) given monthly provides effective contraception with good cycle control.[265] Compared with Lunelle, this combination has less bleeding problems.[266] Fertility returns rapidly (by 1 month) after discontinuation.[262]

Dihydroxyprogesterone Acetophenide and Estradiol Enanthate

The combination of 150 mg dihydroxyprogesterone acetophenide with 10 mg estradiol enanthate (various brand names) is the most widely used injectable contraceptive in Latin America. As with Lunelle and the norethindrone combination, the monthly regimen allows regular, and even reduced, cyclic bleeding.[267] A lower dose (90 mg dihydroxyprogesterone acetophenide and 6 mg estradiol enanthate) provides the same effective contraception as the higher dose with similar bleeding patterns.[268]

All references are available online at: http://www.clinicalgynendoandinfertility.com

25 Intrauterine Contraception



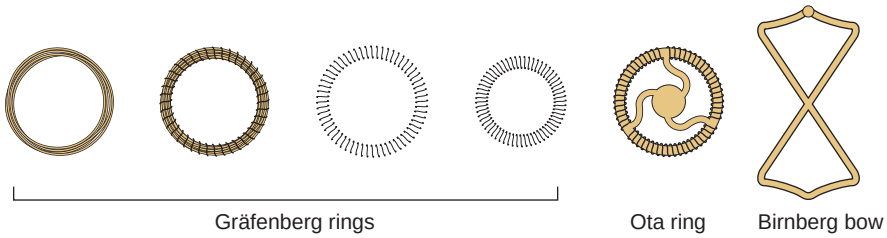
Intrauterine contraceptives are used by over 180 million women worldwide, but only about 3 million of these are American. The growing need for reversible contraception in the U.S. would be well served by increasing utilization of intrauterine contraception with the intrauterine device (the IUD). The efficacy of modern IUDs in actual use is superior to that of oral contraception. Problems with IUD use can be minimized to a very low rate of minor side effects with careful screening and technique. Unfortunately, clinicians in the U.S. still have limited intrauterine contraception knowledge and training.[1] We hope that American clinicians and patients will "rediscover" this excellent method of contraception.

History

A frequently told, but not well-documented story assigns the first use of intrauterine devices to caravan drivers who allegedly used intrauterine stones to prevent pregnancies in their camels during long journeys.

The forerunners of the modern IUD were small stem pessaries used in the 1800s, small button-like structures that covered the opening of the cervix and were attached to stems extending into the cervical canal.[2] It is not certain whether these pessaries were used for contraception, but this seems to have been intended. In 1902, a pessary that extended into the uterus was developed by Hollweg in Germany and used for contraception. This pessary was sold for self-insertion, but the hazard of infection was great, earning the condemnation of the medical community.

In 1909, Richter, in Germany, reported success with a silkworm catgut ring that had a nickel and bronze wire protruding through the cervix.[3] Shortly after, Pust combined Richter's ring with the old button-type pessary and replaced the wire with a catgut thread.[4] This IUD was used during World War I in Germany, although the German literature was quick to report infections with its insertion and use. In the 1920s, Gräfenberg removed the tail and pessary because he believed this was the cause of infection. He reported his experience in 1930, using rings made of coiled silver and gold and then steel.[5]



The Gräfenberg ring was short-lived, falling victim to Nazi political philosophy that was bitterly opposed to contraception. Gräfenberg was jailed, but later he managed to flee Germany, dying in New York City in 1955. He never received the recognition that was his just due.

The Gräfenberg ring was associated with a high rate of expulsion. This was solved by Ota in Japan who added a supportive structure to the center of his gold- or silver-plated ring in 1934.[6] Ota also fell victim to World War II politics (he was sent into exile), but his ring continued to be used.

The Gräfenberg and Ota rings were essentially forgotten by the rest of the world throughout World War II. An awareness of the explosion in population and its impact began to grow in the first two decades after World War II. In 1959, reports from Japan and Israel by Ishihama and Oppenheimer once again stirred interest in the rings.[7,8] The Oppenheimer report was in the *American Journal of Obstetrics and Gynecology*, and several American gynecologists were stimulated to use rings of silver or silk, and others to develop their own devices.

In the 1960s and 1970s, the IUD thrived. Techniques were modified and a plethora of types were introduced. The various devices developed in the 1960s were made of plastic (polyethylene) impregnated with barium sulfate so that they would be visible on an x-ray. The Margulies Coil, developed by Lazer Margulies in 1960 at Mt. Sinai Hospital in New York City, was the first plastic device with a memory, which allowed the use of an inserter and reconfiguration of the shape when it was expelled into the uterus. The Coil was a large device (sure to cause cramping and bleeding), and its hard plastic tail proved risky for the male partner.

In 1962, the Population Council, at the suggestion of Alan Guttmacher, who that year became president of the Planned Parenthood Federation of America, organized the first international conference on IUDs in New York City. It was at this conference that Jack Lippes of Buffalo presented experience with his device, which fortunately as we will see, had a single filament thread as a tail, the first IUD to use a tail to establish position and for easy removal. The Margulies Coil was rapidly replaced by the Lippes Loop. Acquired by the Ortho Pharmaceutical Corporation in 1966, it quickly became the most widely prescribed IUD in the U.S. in the 1970s. The Population Council acquired international rights for the Lippes Loop, and it became used by millions of women throughout the world.

A former World War II test pilot and engineer, Paul H. Bronnenkant, was making plastic parts for jukeboxes in his company, Hallmark Plastics, located next door to the Wurlitzer factory in Buffalo. Lippes' enlistment of Bronnenkant in 1959 to develop his

polyethylene and barium sulfate loop was so successful that Bronnenkant became an energetic advocate of the Lippes Loop; he carried heavy metal molds throughout the Far East to establish local production. A succeeding company, Finishing Enterprises, directed by Bronnenkant's son, Lance Bronnenkant, was the original and is the continuing manufacturer of the TCu-380A since its U.S. approval in 1984. Beginning in 2004, an affiliate of Finishing Enterprises, FEI Women's Health, assumed responsibility for both the manufacturing and the marketing of the ParaGard TCu-380A IUD in the U.S. In 2005, FEI Women's Health was acquired by Duramed Pharmaceuticals, a subsidiary of Barr Pharmaceuticals.

The 1962 conference also led to the organization of a program established by the Population Council, under the direction of Christopher Tietze, to evaluate IUDs, the Cooperative Statistical Program. The Ninth Progress Report in 1970 was a landmark comparison of efficacy and problems with the various IUDs in use.[9]

Many other devices came along, but, with the exception of the four sizes of Lippes Loops and the two Saf-T-Coils, they had limited use. Stainless steel devices incorporating springs were designed to compress for easy insertion, but the movement of these devices allowed them to embed in the uterus, making them too difficult to remove. The Majzlin Spring is a memorable example.

The Dalkon Shield was introduced in 1970. Within 3 years, a high incidence of pelvic infection was recognized. There is no doubt that the problems with the Dalkon Shield were due to defective construction, pointed out as early as 1975 by Tatum.[10] The multifilamented tail (hundreds of fibers enclosed in a plastic sheath) of the Dalkon Shield provided a pathway for bacteria to ascend protected from the barrier of cervical mucus.

Although sales were discontinued in 1975, a call for removal of all Dalkon Shields was not issued until the early 1980s. The large number of women with pelvic infections led to many lawsuits against the pharmaceutical company, ultimately causing its bankruptcy. Unfortunately, the Dalkon Shield problem tainted all IUDs, and for a long time, media and the public in the U.S. inappropriately regarded all IUDs in a single, generic fashion.

About the time of the introduction of the Dalkon Shield, the U.S. Senate conducted hearings on the safety of oral contraception. Young women who were discouraged from using oral contraceptives after these hearings turned to IUDs, principally the Dalkon Shield, which was promoted as suitable for nulliparous women. Changes in sexual behavior in the 1960s and 1970s, and failure to use protective contraception (condoms and oral contraceptives), led to an epidemic of sexually transmitted infections (STIs) and pelvic inflammatory disease (PID) for which IUDs were held partially responsible.[11]

The first epidemiologic studies of the relationship between IUDs and PID used women who depended on oral contraception or barrier methods as controls, and who were, therefore, at reduced risk of PID compared with noncontraceptors and IUD users.[12, 13] In addition, these first studies failed to control for the characteristics of sexual behavior that are now accepted as risk factors for PID (multiple partners, early age at first intercourse, and increased frequency of intercourse).[14] The Dalkon Shield magnified the risk attributed to IUDs because its high failure rate in young women who were already at risk of STIs led to septic spontaneous abortions and, in some cases, death.[15] Reports of these events led the American public to regard all IUDs as dangerous, including those that, unlike the Dalkon Shield, had undergone extensive clinical trials and postmarketing surveillance.

The 1980s saw the decline of IUD use in the U.S. as manufacturers discontinued marketing in response to the burden of litigation. Despite the fact that most of the lawsuits against the copper devices were won by the manufacturer, the cost of the defense combined with declining use affected the financial return. It should be emphasized that this action was the result of corporate business decisions related to concerns for profit and liability, not for medical or scientific reasons. It was not until 1988 that the IUD was returned to the U.S. market.

The reason for the decline in the U.S. was the consumer fear of IUD-related pelvic infection. The final blow to the IUD in the U.S. came in 1985 with the publication of two reports indicating that the use of IUDs was associated with tubal infertility.[16, 17] Later, better controlled studies identified the Dalkon Shield as a high-risk device and failed to demonstrate an association between PID and other IUDs, except during the period shortly after insertion. Efforts to point out that the situation was different for the copper IUDs, and that, in fact, pelvic inflammatory disease was not increased in women with a single sexual partner,[18] failed to prevent the withdrawal of IUDs from the American market and the negative reaction to IUDs by the American public. Ironically, the IUD declined in the country that developed the modern IUD.

The number of reproductive-aged women using the intrauterine device in the U.S. decreased by two-thirds from 1982 to 1988 and further decreased in 1995, from 7.1% to 2% to 0.8%, respectively.[19] Since 1995, the use of the IUD in the U.S. has climbed to 5%, reflecting the popularity of the levonorgestrel-releasing system.[20] In the rest of the world, the IUD is the most widely used method of reversible contraception; currently, more than 180 million women use the IUD, 16.5% of reproductive age women in developing countries and 9.4% in the developed world.[21]

Use of the IUD in the U.S. and the World[19, 21, 22]

	U.S.A.	*China*	*Total World*
1981:	**2.2 million women**	**42 million**	**60 million**
1988:	**0.7 million women**	**59 million**	**83 million**
1995:	**0.3 million women**	**75 million**	**106 million**
2010:	**>3 million women**	**>115 million**	**>180 million**

The Modern IUD

The addition of copper to the IUD was suggested by Jaime Zipper of Chile, whose experiments with metals indicated that copper acted locally on the endometrium.[23] Howard Tatum in the U.S. combined Zipper's suggestion with the development of the T-shape to diminish the uterine reaction to the structural frame and produced the copper-T. The first copper IUD had copper wire wound around the straight shaft of the T, the TCu-200 (200 mm^2 of exposed copper wire), also known as the Tatum-T.[24] Tatum's reasoning was that the T-shape would conform to the shape of the uterus in contrast to the other IUDs that required the uterus to conform to their shape. Furthermore, the copper IUDs could be much smaller than those of simple, inert plastic devices and still provide effective contraception. Studies indicate that copper exerts its effect before implantation of a fertilized ovum; it may be spermicidal, or it may diminish sperm motility or fertilizing capacity. The addition of copper to the IUD and reduction in the size and structure of the frame improved tolerance, resulting in fewer removals for pain and bleeding.

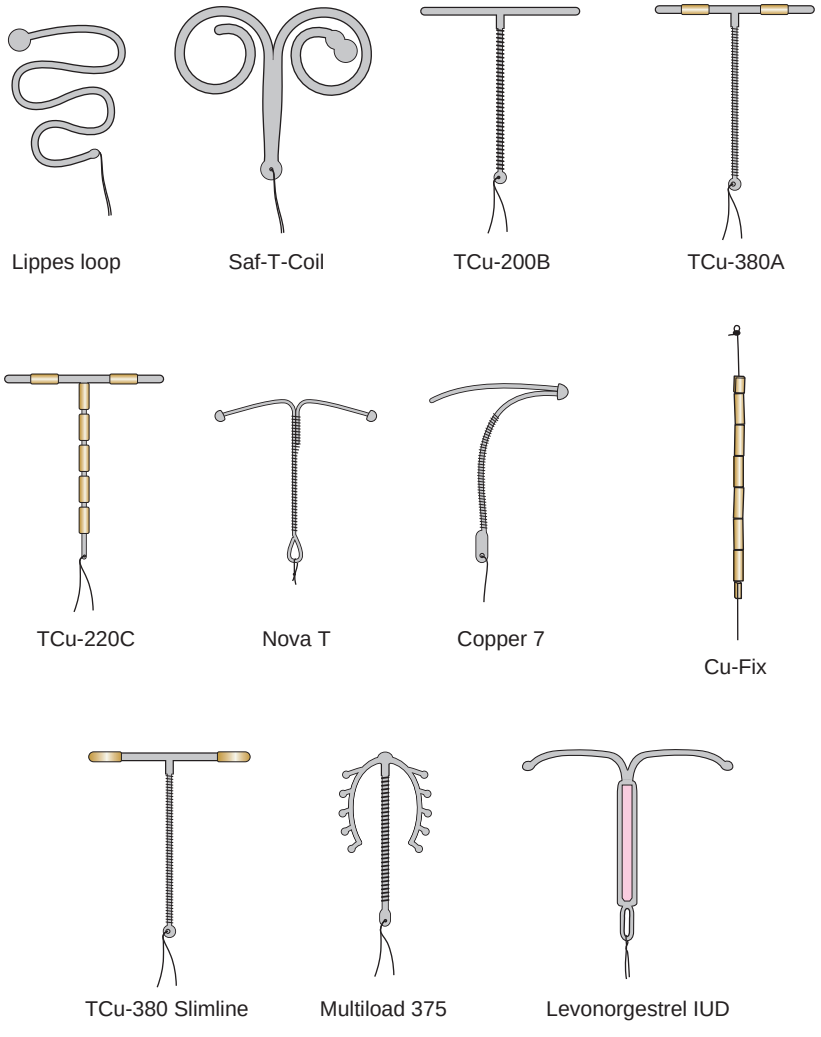
The Cu-7 with a copper wound stem was developed in 1971 and quickly became the most popular device in the U.S. Both the Cu-7 and the Tatum-T were withdrawn from the U.S. market in 1986 by G.D. Searle and Company.

IUD development continued, however. More copper was added by Population Council investigators, leading to the TCu-380A (380 mm^2 of exposed copper surface area) with copper wound around the stem plus a copper sleeve on each horizontal arm.[25] The "A" in TCu-380A is for arms, indicating the importance of the copper sleeves. Making the copper solid and tubular increased effectiveness and the lifespan of the IUD. The TCu-380A has been in use in more than 30 countries since 1982, and, in 1988, it was marketed in the U.S. as the "ParaGard."

Types of IUDs

Unmedicated IUDs

The Lippes Loop, made of plastic (polyethylene) impregnated with barium sulfate, is still used throughout the world (except in the U.S.). Flexible stainless steel rings were widely used in China but not elsewhere until 1994 when copper IUDs were mandated.[26, 27]



Copper IUDs

The first copper IUDs were wound with 200 to 250 mm^2 surface area of wire, and two of these are still available (except in the U.S.): the TCu-200 and the Multiload-250. The more modern copper IUDs contain more copper, and part of the copper is in the form of solid tubular sleeves, rather than wire, increasing efficacy and extending lifespan. This group of IUDs is represented in the U.S. by the TCu-380A (the ParaGard) and in the rest of the world by the TCu-220C, the Nova T, and the Multiload-375. The Sof-T is a copper IUD used only in Switzerland.

The TCu-380A is a T-shaped device with a polyethylene frame holding 380 mm^2 of exposed surface area of copper that provides contraception for at least 10 years. Although the data are sparse, wearing the TCu-380A for 20 years carries only a very small risk of pregnancy.[28] The pure electrolytic copper wire wound around the 36-mm stem weighs 176 mg, and copper sleeves on the horizontal arms weigh 66.5 mg. A polyethylene monofilament is tied through the 3 mm ball on the stem, providing two white threads for detection and removal. The ball at the bottom of the stem helps reduce the risk of cervical perforation. The IUD frame contains barium sulfate, making it radiopaque. The TCu-380Ag is identical to the TCu-380A, but the copper wire on the stem has a silver core to prevent fragmentation and extend the lifespan of the copper. The TCu-380 Slimline has the copper sleeves flush at the ends of the horizontal arms to facilitate easier loading and insertion. The performance of the TCu-380Ag and the TCu-380 Slimline is equal to that of the TCu-380A.[29, 30]

The Multiload-375 has 375 mm^2 of copper wire wound around its stem. The flexible arms were designed to minimize expulsions. This is a popular device in many parts of the world. The Multiload-375 and the TCu-380A are similar in their efficacy and performance.[31]

The Nova T is similar to the TCu-200, containing 200 mm of copper; however, the Nova T has a silver core to the copper wire, flexible arms, and a large, flexible loop at the bottom to avoid injury to cervical tissue. There was some concern that the efficacy of the Nova T decreased after 3 years in World Health Organization (WHO) data; however, results from Finland and Scandinavia indicate low and stable pregnancy rates over 5 years of use.[31]

The CuSAFE-300 IUD has 300 mm^2 of copper in its vertical arm and a transverse arm with sharply bent ends that are adapted to the uterine cavity and help hold this IUD in the fundus. It is made from a more flexible plastic and is smaller than the world's two most popular IUDs, the TCu-380A and the Multiload-375. Pregnancy rates with the CuSAFE-300 are comparable to these two devices, but rates of removal for pain and bleeding are reported to be lower.[32]

The Hormone-Releasing IUD

The LNG-IUS (levonorgestrel-releasing intrauterine system, Mirena), manufactured by Schering-Oy in Finland, releases *in vitro* 20 μg of levonorgestrel per day.[33] This T-shaped device has a collar attached to the vertical arm, which contains 52 mg levonorgestrel dispersed in polydimethylsiloxane and released initially at a rate of 20 μg/day *in vivo*, progressively declining (reaching half of the initial rate after 5 years). The levonorgestrel IUS is approved for 5 years, but lasts for 7 years, and perhaps up to 10 years, and reduces menstrual blood loss and pelvic infection rates.[34–36]

The levonorgestrel IUS is about as effective as endometrial ablation for the treatment of menorrhagia.[37, 38] The local progestin effect directed to the endometrium can be utilized in patients on tamoxifen,[39] patients with dysmenorrhea,[40] and in postmenopausal women

receiving estrogen therapy.[41–46] A slightly smaller T-shaped device that releases 20 μg of levonorgestrel daily is called the Femilis LNG-IUS.[47]

Smaller devices releasing 5 or 10 μg levonorgestrel have been developed in Europe for use for at least 5 years in postmenopausal women.[45]

Other IUDs

The Ombrelle-250 and Ombrelle-380, designed to be more flexible in order to reduce expulsion and side effects, have been marketed in France. A frameless IUD, the FlexiGard (also known as the Cu-Fix or the GyneFIX), invented by Dirk Wildemeersch in 1983 in Belgium, consists of 6 copper sleeves (330 mm^2 of copper) strung on a surgical nylon (polypropylene) thread that is knotted at one end. The knot is pushed into the myometrium during insertion with a notched needle that works like a miniature harpoon. Because it is frameless, it has a low rate of removal for bleeding or pain, but a more difficult insertion may yield a higher expulsion rate. However, when inserted by experienced clinicians, the expulsion rate is very low, and the device is especially suited for nulligravid and nulliparous women.[48–50] This IUD is increasingly popular in Europe. A shorter system combined with a reservoir for the sustained release of 14 μg levonorgestrel per day (FibroPlant) is being tested for perimenopausal and postmenopausal use.[51, 52] FibroPlant effectively treats endometrial hyperplasia and menorrhagia.[53, 54]

Mechanism of Action

The contraceptive action of all IUDs is mainly in the uterine cavity. Ovulation is not affected, and the IUD is not an abortifacient.[55–57] It is currently believed that the mechanism of action for IUDs is the production of an intrauterine environment that is spermicidal.

Nonmedicated IUDs depend for contraception on the general reaction of the uterus to a foreign body. It is believed that this reaction, a sterile inflammatory response, produces tissue injury of a minor degree but sufficient enough to be spermicidal. Very few, if any, sperm reach the ovum in the fallopian tube. Normally cleaving, fertilized ova cannot be obtained by tubal flushing in women with IUDs in contrast to noncontraceptors, indicating the failure of sperm to reach the ovum, and, thus, fertilization does not occur.[58] In women using copper IUDs, sensitive assays for human chorionic gonadotropin (hCG) do not find evidence of fertilization.[59, 60] This is consistent with the fact that the copper IUD protects against both intrauterine and ectopic pregnancies.

The copper IUD releases free copper and copper salts that have both a biochemical and morphologic impact on the endometrium and also produce alterations in cervical mucus and endometrial secretions. There is no measurable increase in the serum copper level. Copper has many specific actions, including the enhancement of prostaglandin production and the inhibition of various endometrial enzymes. The copper IUD is associated with an inflammatory response, marked by production in the endometrium of cytokine peptides known to be cytotoxic.[61] An additional spermicidal effect probably takes place in the cervical mucus.

The progestin-releasing IUD adds the endometrial action of the progestin to the foreign body reaction. The endometrium becomes decidualized with atrophy of the glands.[62] The progestin IUD probably has two mechanisms of action: inhibition of implantation and

inhibition of sperm capacitation, penetration, and survival. The levonorgestrel IUS produces serum concentrations of the progestin about half those of Norplant so that ovarian follicular development and ovulation are also partially inhibited; after the first year, cycles are ovulatory in 50–75% of women, regardless of their bleeding patterns.[63] Finally, the progestin IUD thickens the cervical mucus, creating a barrier to sperm penetration.

Following removal of IUDs, the normal intrauterine environment is rapidly restored. ***In large studies, there is no delay, regardless of duration of use, in achieving pregnancy at normal rates, which belies the assertion that IUD use is associated with infection leading to infertility.***[64–67] There has been no significant difference in cumulative pregnancy rates between parous and nulliparous or nulligravid women.[66, 67]

The levonorgestrel IUS may be associated with a slight increase in the formation of ovarian cysts, but they are asymptomatic and resolve spontaneously.[68] An increase in the risk of venous thrombosis has not been observed in users of the levonorgestrel IUS.[69]

Noncontraceptive Benefits with the Levonorgestrel IUS

Because of the favorable impact of locally released progestin on the endometrium, the levonorgestrel IUS is very effective for the treatment of menorrhagia, more effective than the administration of oral progestins, steroid contraceptives, or inhibitors of prostaglandin synthesis, and compares favorably with surgical treatment (hysterectomy or endometrial ablation).[37, 70–79]

The progestin IUD rapidly decreases dysmenorrhea and menstrual blood loss (about 40–50%); with the levonorgestrel IUS, bleeding over time can be reduced by 90% and about 30–40% of women become amenorrheic 1 year after insertion.[80–82] Average long-term hemoglobin and iron levels increase compared with preinsertion values.[83]

Bleeding is even reduced in the presence of leiomyomas.[84–88] There is some evidence that this IUD reduces the prevalence of myomas and the uterine volume in the presence of myomas.[84, 85, 88–90] Not all women with myomas respond favorably to the levonorgestrel IUS; this is usually because of the presence of submucosal myomas.[91] The levonorgestrel IUS also effectively reduces uterine volume and relieves dysmenorrhea secondary to adenomyosis.[92–94]

Women with hemostatic disorders, such as von Willebrand's disease, and women who are anticoagulated commonly have heavy menstrual bleeding. The insertion of the levonorgestrel IUS effectively reduces the amount of bleeding in many, but not all, of these patients, and there is a suggestion that with time the beneficial effect wanes and earlier replacement is necessary.[95–99]

In a randomized, 5-year trial comparing a copper IUD with the levonorgestrel IUS, pelvic infection was lower than the incidence in a general population with both devices, but the rate with the levonorgestrel IUS was significantly lower compared with the copper IUD.[100]

The levonorgestrel IUS has been used successfully to treat endometriosis, and especially pelvic pain and dysmenorrhea associated with endometriosis.[40, 101–105] Results may even be

better than with GnRH agonist treatment.[106] The advantages of levonorgestrel IUS treatment of endometriosis include many years of efficacy, fewer side effects, and contraception.

The levonorgestrel IUS effectively protects the endometrium against hyperplasia and polyps in women using tamoxifen or postmenopausal estrogen therapy.[39, 41–45, 107–109] In addition, this IUD can be used to treat endometrial hyperplasia.[54, 110–114] Comparison studies indicate that the levonorgestrel IUS is as effective, and probably better than standard treatment with an oral progestin.[111, 115, 116] ***However, the persistence of atypia at biopsy follow-up after 6 months is an indication that regression is unlikely to occur.***

Although the levonorgestrel IUD confidently provides good protection against endometrial hyperplasia, clinicians should maintain a high degree of suspicion of unusual bleeding (bleeding that occurs after a substantial period of amenorrhea) and aggressively assess the endometrium. At least two cases of endometrial adenocarcinoma have been identified in users of the levonorgestrel IUS.[117, 118] As noted however, many studies have documented protection against and even regression of endometrial hyperplasia.

Summary of Noncontraceptive Benefits with the Levonorgestrel IUS

- Reduction of heavy menstrual bleeding and improvement of related anemia.
- Treatment of primary dysmenorrhea.
- Reduction of myoma prevalence as well as uterine volume and bleeding associated with myomas.
- Decrease in uterine volume and pain associated with adenomyosis.
- Reduction of menstrual bleeding in women with hemostatic disorders and in anticoagulated women.
- Protection against pelvic inflammatory disease.
- Treatment of endometriosis, and pain associated with endometriosis.
- Suppression of endometriosis.
- Protection against endometrial hyperplasia and polyps associated with postmenopausal estrogen therapy or tamoxifen treatment.
- Prevention of ectopic pregnancy.
- Reduction of endometrial cancer risk.

Efficacy of IUDs

Intrauterine Pregnancy

The TCu-380A is approved for use in the U.S. for 10 years. However, the TCu-380A has been demonstrated to maintain its efficacy over at least 12 years of use.[119] As previously noted, based on a small number of long-term users, wearing the TCu-380A for 20 years carries a very small risk of pregnancy.[28] The TCu-200 is approved for 4 years and the Nova T for 5 years. The levonorgestrel IUD can be used for at least 7 years and probably for 10 years.[31,120] The levonorgestrel device that releases 15–20 μg levonorgestrel per day is as effective as the new copper IUDs that contain more than 250 mm^2 of copper surface area.[30, 34, 121, 122]

The nonmedicated IUDs never have to be replaced. The deposition of calcium salts on the IUD can produce a structure that is irritating to the endometrium. If bleeding increases after a nonmedicated IUD has been in place for some time, it is worth replacing it. Some

clinicians (as do we) recommend replacing all older IUDs with the new, more effective medicated IUDs.

First Year Clinical Trial Experience in Parous Women[123–125]

Device	Pregnancy Rate (%)	Expulsion Rate (%)	Removal Rate (%)
Lippes Loop	3	12–20	12–15
Cu-7	2–3	6	11
TCu-200	3	8	11
TCu-380A	0.5–0.8	5	14
Levonorgestrel IUS	0.2	6	17

Considering all IUDs together, the actual use failure rate in the first year is approximately 3%, with a 10% expulsion rate and a 15% rate of removal, mainly for bleeding and pain. With increasing duration of use and increasing age, the failure rate decreases, as do removals for pain and bleeding. The performance of the TCu-380A in recent years, however, has proved to be superior to previous IUDs.

Ten-Year Experience with Paragard, TCu-380A
Rate per 100 Users per year

	Year									
	1	2	3	4	5	6	7	8	9	10
Pregnancy	0.7	0.3	0.6	0.2	0.3	0.2	0.0	0.4	0.0	0.0
Expulsion	5.7	2.5	1.6	1.2	0.3	0.0	0.6	1.7	0.2	0.4
Bleeding/Pain Removal	11.9	9.8	7.0	3.5	3.7	2.7	3.0	2.5	2.2	3.7
Medical Removals	2.5	2.1	1.6	1.7	0.1	0.3	1.0	0.4	0.7	0.3
Continuation	76.8	78.3	81.2	86.2	89.0	91.9	87.9	88.1	92.0	91.8
Number Starting Each Year	4932	3149	2018	1121	872	621	563	483	423	325

Data from Population Council (n = 3536) and WHO (n = 1396) trials.

In careful studies, with attention to technique and participation by motivated patients, the failure rate with the TCu-380A and the other newer copper IUDs is less than 1 per 100 women per year.[31, 123, 126] The cumulative net pregnancy rate after 7 years of use is 1.5 per 100 woman-years, and after 12 years, only 1.9 per 100 women (not a single pregnancy was reported after 8 years of use).[119, 127] In developing countries, the failure rate with IUDs is less than that with oral contraception.[128] Failure rates are slightly higher in younger (less than age 25), more fertile women.

Women use IUDs longer than other reversible methods of contraception. The IUD continuation rate is higher than that with oral contraception, condoms, or diaphragms. This may reflect the circumstances surrounding the choice of an IUD (older, parous women).

Expulsion

Approximately 5% of patients spontaneously expel the TCu-380A within the first year. Women younger than age 20 have a higher expulsion rate than older women.[34, 129] In a nulligravid, adolescent population, the cumulative expulsion rate with the levonorgestrel intrauterine system was 8%.[130] This event can be associated with cramping, vaginal

discharge, or uterine bleeding. However, in some cases, the only observable change is lengthening or absence of the IUD strings. Patients should be cautioned to request immediate attention if expulsion is suspected. A partially expelled IUD should be removed. If pregnancy or infection is not present, a new IUD can be inserted immediately (in this instance, antibiotic prophylaxis is recommended).[131]

Ectopic Pregnancy

The previous use of an IUD does not increase the risk of a subsequent ectopic pregnancy.[67,132,133] The current use of an IUD offers some protection against ectopic pregnancy.[132–137] The largest study, a WHO multicenter study, concluded that IUD users were 50% less likely to have an ectopic pregnancy when compared with women using no contraception.[132] This protection is not as great as that achieved by inhibition of ovulation with oral contraception. ***Therefore, when an IUD user becomes pregnant, the pregnancy is more likely to be ectopic. However, the actual occurrence of an ectopic pregnancy in an IUD user is a rare event.***

The lowest ectopic pregnancy rates are seen with the most effective IUDs, like the TCu-380A (90% less likely compared with noncontraceptors).[138] The rate is about one-tenth of the ectopic pregnancy rate associated with the Lippes Loop or with devices with less copper such as the TCu-200.[138] The progesterone-releasing IUD (that is no longer produced) had a higher rate, probably because its action was limited to a local effect on the endometrium,[135] but the reported rate was based on very small numbers and may have been inaccurate. Very few ectopic pregnancies have been reported with the levonorgestrel IUS, presumably because it is associated with a partial suppression of gonadotropins with subsequent disruption of normal follicular growth and development and, in a significant number of cycles, inhibition of ovulation.[30, 36, 122, 138, 139]

The risk of ectopic pregnancy does not increase with increasing duration of use with the TCu-380A or the levonorgestrel IUS.[30, 127] In a 7-year prospective study, not a single ectopic pregnancy was encountered with the levonorgestrel IUS, and in a 5-year study, only one.[30,100] In 8,000 woman-years of experience in randomized multicenter trials, there was only a single ectopic pregnancy reported with the TCu-380A (which is one-tenth the rate with the Lippes Loop or TCu-200).[30] Therefore, the risk of ectopic pregnancy during the use of the copper IUD or the levonorgestrel IUS is much lower compared with non-contraceptive users; however, if pregnancy occurs, the likelihood of an ectopic pregnancy is high.[140]

The protection against ectopic pregnancy provided by the TCu-380A and the levonorgestrel IUS makes these IUDs acceptable choices for contraception in women with previous ectopic pregnancies.

Ectopic Pregnancy Rates per 1,000 Woman-Years[138, 141]	
Non-Contraceptive Users, all ages	3.00–4.50
Levonorgestrel IUS	0.20
TCu-380A IUD	0.20

Side Effects

With effective patient screening and good insertion technique, the copper and levonorgestrel IUDs are not associated with an increased risk of infertility after their removal.[142] Even if

IUDs are removed for problems, subsequent fertility rates are normal.[66, 67, 122]

The symptoms most often responsible for IUD discontinuation are increased uterine bleeding and increased menstrual pain. Within 1 year, 5–15% of women discontinue IUD use because of these problems. Smaller copper and progestin IUDs have reduced the incidence of pain and bleeding considerably, but a careful menstrual history is still important in helping a woman consider an IUD. Women with prolonged, heavy menstrual bleeding or significant dysmenorrhea may not be able to tolerate copper IUDs but may benefit from a progestin IUD.[80] Because bleeding and cramping are most severe in the first few months after IUD insertion, treatment with a nonsteroidal anti-inflammatory drug (NSAID, an inhibitor of prostaglandin synthesis) during the first several menstrual periods can reduce bleeding and cramping and help a patient through this difficult time. Even persistent heavy menses can be effectively treated with NSAIDs.[143] NSAID treatment should begin at the onset of menses and be maintained for 3 days. A copper IUD is available in China that also releases a small amount of indomethacin; this device is associated with markedly less bleeding.[144]

It is not unusual to have a few days of intermenstrual spotting or light bleeding. Although aggravating, this does not cause significant blood loss. Such bleeding deserves the usual evaluation for cervical or endometrial pathology. These changes can be objectionable for women who are prevented from having intercourse while bleeding.

Following insertion of a modern copper IUD, menstrual blood loss increases by about 55%, and this level of bleeding continues for the duration of IUD use.[145] This is associated with a slight (1–2 day) prolongation of menstruation. Over a year's time, this amount of blood loss does not result in changes indicative of iron deficiency (e.g., serum ferritin). With longer use, however, ferritin levels are lower, suggesting a depletion of iron stores.[146] Assessment for iron depletion and anemia should be considered in long-term users and in women susceptible to iron deficiency anemia. In populations with a high prevalence of anemia, these changes occur more rapidly, and iron supplementation is recommended.[147]

Because of a decidualizing, atrophic impact on the endometrium, amenorrhea can develop over time with the progestin-containing IUD. With the levonorgestrel IUS, 70% of patients are oligomenorrheic and 30–40% are amenorrheic within 2 years.[82, 148] In a group of women who used the levonorgestrel IUS for more than 12 years, 60% were amenorrheic; 12% experienced infrequent, scanty bleeding; and 28% had regular but light bleeding.[83] For some women, the lack of periods is so disconcerting that they request removal. On the other hand, this effect on menstruation is manifested by an increase in blood hemoglobin levels.[30, 123]

Sufficient progestin reaches the systemic circulation from the levonorgestrel-containing IUS so that androgenic side effects, such as acne and hirsutism, can occur; however, in one study no change could be detected in the circulating levels of sex hormone-binding globulin, and, therefore, marked clinical effects are unlikely.[149] Occasionally a patient will experience facial, androgenic skin changes or breast tenderness. Spironolactone treatment for androgenic reactions (discussed in Chapter 13) is effective and can be titered down to the lowest successful dose. There are no significant metabolic effects, including changes in glucose, insulin sensitivity, and lipids.[150] In a large postmarketing survey, the incidence of breast cancer in users of the levonorgestrel system was equivalent to that in the general population.[151] The combination of the levonorgestrel system and estrogen therapy in postmenopausal women did not increase breast density.[152]

Some women report an increased vaginal discharge while wearing an IUD. This complaint deserves examination for the presence of vaginal or cervical infection. Treatment can be provided with the IUD remaining in place.

Long-term use of the IUD is associated with impressive safety and lack of side effects. In a 7-year prospective study, the use of either the copper IUD or the levonorgestrel IUS beyond 5 years led to no increase in pelvic infection, no increase in ectopic pregnancy rates, no increase in anemia, and no increase in abnormal Pap smears.[30] Duration of use does not affect pregnancy rates or outcome.

The presence of copper may yield some benefits. There are epidemiologic data indicating that both the copper IUD and the inert IUD reduce the risks of endometrial cancer and invasive cervical cancer.[153–158] Presumably, this protective effect is due to induced biochemical alterations that affect cellular responses.

The copper IUD is not affected by magnetic resonance imaging, and therefore, the copper IUD need not be removed before MRI, and neither patients nor workers need be excluded from MRIs or the MRI environment.[159, 160]

Infections

IUD-related bacterial infection is due to contamination of the endometrial cavity at the time of insertion. Mishell's classic study indicated that the uterus is routinely contaminated by bacteria at insertion.[161] Infections that occur 3–4 months after insertion are believed to be due to acquired STIs not the direct result of the IUD. The early, insertion-related infections, therefore, are polymicrobial, and are derived from the endogenous cervicovaginal flora, with a predominance of anaerobes.

A review of the World Health Organization data base derived from all of the WHO IUD clinical trials concluded that the risk of pelvic inflammatory disease was 6 times higher during the 20 days after the insertion compared with later times during follow-up, but, most importantly, PID was extremely rare beyond the first 20 days after insertion.[162] In nearly 23,000 insertions, however, only 81 cases of PID were diagnosed, and a scarcity of PID was observed in those situations in which STIs are rare. There was no statistically significant difference comparing the copper IUD to the inert Lippes Loop or progestin-containing IUD. However, evidence indicates that the levonorgestrel IUS may reduce pelvic infection rates.[35, 100]

These data confirm earlier studies that the risk of infection with intrauterine contraception is highest immediately after insertion and that PID risk does not increase with long-term use.[15, 18] The problem of infection can be minimized with careful screening and the use of aseptic technique. Even women with insulin-dependent diabetes mellitus do not have an increased risk for infection.[163, 164]

Doxycycline (200 mg) or azithromycin (500 mg) administered orally 1 hour prior to insertion can provide protection against insertion-associated pelvic infection, but prophylactic antibiotics are of little benefit for women at low risk for STIs. Even women with Chlamydia infections at the time of insertion are unlikely to develop PID if the infection is treated with the IUD left in place after a positive culture is obtained.[165]

Compared with oral contraception, barrier methods, and hormonal IUDs, there is no reason to think that nonmedicated or copper IUDs can confer protection against STIs.[166] However, the levonorgestrel-releasing IUS has been reported to be associated with a protective effect against pelvic infection, and the copper IUD is associated with lower titers of antichlamydial antibody.[35, 100, 167] In vitro, copper inhibits chlamydial growth in endometrial cells.[168] Thus, the association between IUD use and pelvic infection (and infertility) is now seriously questioned.[169] Women who use IUDs must be counseled to use condoms

along with the IUD whenever they have intercourse with a partner who could be an STI carrier. Because sexual behavior is the most important modifier of the risk of infection, clinicians should ask prospective IUD users about numbers of partners, their partner's sexual practices, the frequency and age of onset of intercourse, and history of STIs.[170] Women at low risk are unlikely to have pelvic infections while using IUDs.[18] Women at high risk should be encouraged to also use condoms.

It is not certain that the IUD is inappropriate for women who are at increased risk of bacterial endocarditis (previous endocarditis, rheumatic heart disease, or the presence of prosthetic heart valves). The bacteriologic contamination of the uterine cavity at insertion is short-lived.[161] Four studies have attempted to document bacteremia during IUD insertion or removal.[171–174] Only one of the three could find blood culture evidence of bacteremia, and it was present transiently in only a few patients.[173] ***In our view, the IUD is acceptable for patients at risk of bacterial endocarditis, but antibiotic prophylaxis (amoxicillin 2 g) should be provided 1 h before insertion or removal.***

Asymptomatic IUD users whose cervical cultures show gonorrhea or chlamydia infection should be treated with the recommended drugs without removal of the IUD. If, however, there is evidence that an infection has ascended to the endometrium or fallopian tubes, treatment must be instituted and the IUD removed promptly. Vaginal bacteriosis should be treated (metronidazole, 500 mg b.i.d. for 7 days), but the IUD need not be removed unless pelvic inflammation is present. There is no evidence that the prevalence of bacterial vaginosis is influenced by IUD use.[175]

For simple endometritis, in which uterine tenderness is the only physical finding, doxycycline (100 mg b.i.d. for 14 days) is adequate. If tubal infection is present, as evidenced by cervical motion tenderness, abdominal rebound tenderness, adnexal tenderness or masses, or elevated white blood count and sedimentation rate, parenteral treatment is indicated with removal of the IUD ***as soon as antibiotic serum levels are adequate.*** The previous presence of an IUD does not alter the treatment of PID. IUD-associated pelvic infection is more likely to be caused by non-STI organisms.[176]

Appropriate outpatient management of less severe infections:
Cefoxitin (2 g intramuscularly) plus probenecid (1 g orally), or
Ceftriaxone (250 mg IM) plus doxycycline (100 mg b.i.d. orally), for 14 days, or
Levofloxacin (500 mg orally) once daily for 14 days, or
Ofloxacin (400 mg orally) b.i.d. for 14 days.

Severe infections require hospitalization and treatment with:
Cefoxitin (2 g intravenous q.i.d.), or
Cefotetan (2 g intravenous b.i.d.)
Plus doxycycline (100 mg b.i.d. orally or intravenous)
Followed by 14 days of an oral regimen of antibiotics.

The following is an alternative regimen:
Clindamycin (900 mg intravenously t.i.d.), plus
Gentamicin (2 mg/kg intravenously or intramuscularly followed by 1.5 mg/kg t.i.d.), or
Ofloxacin (400 mg intravenously b.i.d.) or,
Levofloxacin (500 mg intravenously t.i.d.).

There was a suggestion (in cross-sectional studies) that IUD use increased the risk of human immunodeficiency virus (HIV) transmission from man to woman, especially when the IUD was inserted or removed during exposure to the infected man.[177, 178] However, this is not a strong suggestion, because the risk with IUD use was ascertained compared with other contraceptive methods (which can protect against transmission), and the many and various influencing factors are difficult to adjust and control. In the only longitudinal study,

no association was observed between IUD use and HIV acquisition by women.[179] In the only study reported, no evidence for female-to-male HIV transmission with IUD use was detected.[180] HIV-infected women who utilize IUDs for contraception *do not* have a greater incidence of complications (including pelvic inflammatory disease), and genital shedding of HIV is not affected.[181–184]

Hepatitis C virus is believed to be transmitted with sexual contact. A cross-sectional study in Turkey indicated that IUD use by women in stable, monogamous relationships did not increase the incidence of hepatitis C seropositivity.[185]

Actinomyces

The significance of actinomycosis infection in IUD users is unclear. There are many reports of IUD users with unilateral pelvic abscesses containing gram-positive, bacilli, *Actinomyces*.[186–188] However, *Actinomyces*, part of the normal flora in the gastrointestinal tract, are found in Pap smears of up to 30% of plastic IUD wearers when cytologists take special care to look for the organisms.[188] The rate is much lower (less than 1%) with copper devices and the levonorgestrel-releasing intrauterine system and increases with duration of use.[186, 187, 189–191] Furthermore, *Actinomyces* are sometimes present in the normal vagina.[192] The clinician must decide whether to remove the IUD and treat the patient, treat with the IUD in place, or simply remove the IUD. These patients are almost always asymptomatic and without clinical signs of infection. If uterine tenderness or a pelvic mass is present, the IUD should always be removed after the initiation of treatment with oral penicillin G, 500 mg q.i.d., that should continue for a month. Alternative antibiotic regimens include tetracycline 500 mg q.i.d.; doxycycline 100 mg b.i.d.; amoxicillin/clavulanate 500 mg b.i.d. If *Actinomyces* are present on the Pap smear of an asymptomatic, well woman, in our view, it is not necessary to administer antibiotic treatment or to remove the IUD. Although it has been recommended that the IUD should be removed in this instance and replaced when a repeat Pap smear is negative, there is no evidence to support this recommendation. If the patient is due for IUD replacement, the IUD can be removed and a new one immediately inserted with no adverse effects.[193] Another anaerobic, gram-positive rod, *Eubacterium nodatum*, resembles *Actinomyces* and has also been reported to be associated with colonization of an IUD.[194] *E. nodatum* can be mistaken for *Actinomyces* on Pap smears. Our recommendations can be applied to both *E. nodatum* and *Actinomyces*.

Pregnancy with an IUD In Situ

Spontaneous Miscarriage

Spontaneous miscarriage occurs more frequently among women who become pregnant with IUDs in place, a rate of approximately 40–50%. Because of this high rate of spontaneous miscarriage, IUDs should always be removed if pregnancy is diagnosed and the string is visible. Use of instruments inside the uterus should be avoided if the pregnancy is desired, unless sonographic guidance can help avoid rupture of the membranes.[195] After removal of an IUD with visible strings, the spontaneous miscarriage rate is approximately 30%.[196, 197] Combining ultrasonography guidance with carbon dioxide hysteroscopy, an IUD with a missing tail can be identified and removed during early pregnancy.[198] ***If the***

IUD is easily removed without trauma or expelled during the first trimester, the risk of spontaneous miscarriage is not increased.[199, 200]

Septic Abortion

In the past, if the IUD could not be easily removed from a pregnant uterus, the patient was offered induced abortion because it was believed that the risk of life-threatening septic, spontaneous miscarriage in the second trimester was increased 20-fold if the pregnancy continued with the IUD in utero. However, this belief was derived from experiences with the Dalkon Shield. There is no evidence that there is an increased risk of septic abortion if pregnancy occurs with an IUD in place other than the Shield.[200, 201] There were no deaths in the United States from 1977 to 1993 among women pregnant with an IUD.[202]

If a patient plans to terminate a pregnancy that has occurred with an IUD in place, the IUD should be removed immediately. If there is no evidence of infection, the IUD can safely be removed in a clinic or office.

If an IUD is in an infected, pregnant uterus, removal of the device should be undertaken only after antibiotic therapy has been initiated, and equipment for cardiovascular support and resuscitation is immediately available. These precautions are necessary because removal of an IUD from an infected, pregnant uterus can lead to septic shock.

Congenital Anomalies

There is no evidence that exposure of a fetus to medicated IUDs is harmful. The risk of congenital anomalies is not increased among infants born to women who become pregnant with an IUD in place.[200, 203] A case-control study did not find an increased incidence of IUD use in pregnancies resulting in limb-reduction deformities.[204]

Preterm Labor and Birth

The incidence of preterm labor and delivery is increased approximately 4-fold when an IUD is left in place during pregnancy.[200, 205–207]

Other Complications

Obstetrical complications at delivery (e.g., hemorrhage, stillbirth, and difficulties with placenta removal) have been reported only with the Dalkon Shield in situ.

IUD Insertion

Patient Selection

Patient selection for successful IUD use requires attention to menstrual history and the risk for STIs. Age and parity are not the critical factors in selection; the risk factors for STIs are the most important consideration. Women who have multiple sexual partners, whose partners have multiple partners, who are drug or alcohol dependent, and who are not in a stable sexual relationship are at greater risk of pelvic infection at the time of IUD insertion and at greater risk of acquiring a sexually transmitted infection after IUD insertion.[16–18] It would be appropriate for these women to use condoms for STI protection and an IUD for effective contraception. Current, recent, or recurrent PID is a contraindication for IUD use. Hormonal and barrier methods are better choices for these women. Nulliparous and nulligravid women can safely use the IUD if both sexual partners are monogamous. In a national U.S. survey, only 13% of adults had more than one sexual partner in the previous year.[208] Most women are good candidates for the IUD. The levonorgestrel-releasing IUS performs as well as oral contraceptives when used by young nulliparous women.[209] Complications, including expulsion and perforation, are no different comparing parous and nulliparous IUD users.[210]

Patients with heavy menstrual periods should be cautioned regarding the increase in menstrual bleeding associated with the copper IUD. Women who are anticoagulated or have a bleeding disorder are obviously not good candidates for the copper IUD, but they might benefit from the progestin IUD.

There are other conditions that can compromise success. Women who have abnormalities of uterine anatomy (bicornuate uterus, submucous myoma, cervical stenosis) may not accommodate an IUD. The IUD is not a good choice when the uterine cavity is distorted by leiomyomas. According to conventional wisdom, the few individuals who have allergies to copper or have Wilson's disease (a prevalence of about 1 in 200,000) should not use copper IUDs; however, no cases of difficulty have ever been recorded, and it is doubtful, considering the low exposure to copper, that there would be a problem. The amount of copper released into the circulation per day is less than that consumed in a normal diet.[211] Nevertheless, barrier methods or long-acting progestin-only methods are recommended for individuals with Wilson's disease.[212]

Immunosuppressed patients should not use IUDs. Patients at risk for endocarditis should be treated with prophylactic antibiotics at insertion and removal. In our view, cervical dysplasia does not preclude IUD insertion or continued use.

Because many older women have diabetes mellitus, it is worth emphasizing that no increase in adverse events has been observed with copper IUD use in women with either insulin-dependent or noninsulin-dependent diabetes.[163, 164, 213] Indeed, the IUD can be an ideal choice for a woman with diabetes, especially if vascular disease is present.

The IUD should not be dismissed just because the patient is an adolescent. Although the clinical performance of the IUD in a study of parous adolescents was not as good as in older women, it was still similar or slightly better than other reversible methods used by adolescents.[214] Given appropriate screening, counseling, and care, the IUD can provide long-term effective contraception for adolescents. The continuation rate with the levonorgestrel intrauterine system in a population of adolescents in New Zealand was 85% after 1 year.[130] There are no major differences in IUD efficacy and side effects comparing nulligravid women with parous women.[215] Nulligravid women who are past users of the IUD did not have an increased risk of tubal infertility in a case-control study, but long-term IUD users in a cohort study were reported to have reduced fertility compared with oral contraceptive users and short-term IUD users.[216, 217] The merits of the two studies can be debated (the cohort study had several confounding problems); by far most evidence indicates that IUD use does not impair fertility.[65–67, 122, 218, 219]

A careful speculum and bimanual examination is essential prior to IUD insertion. It is important to know the position of the uterus; undetected extreme posterior uterine position is the most common reason for perforation at the time of IUD insertion. However, perforation is rare; the incidence is estimated to be less than 1 per 3,000 insertions.[220] A very small or large uterus, determined by examination and sounding, can preclude insertion. For successful IUD use, the uterus should preferably not sound less than 6 cm or more than 9 cm.

Preferably, the absence of cervical or vaginal infection should be established before insertion. If this is not feasible, insertion should definitely be delayed if a mucopurulent discharge of the cervix or a significant vaginitis (including vaginal bacteriosis) is present.

Summary: IUD Use and Medical Conditions

1. A woman with a previous ectopic pregnancy can use a copper IUD or the levonorgestrel IUS.
2. Women with heavy menses and dysmenorrhea, including women who have a bleeding disorder or are anticoagulated, should consider the levonorgestrel-releasing IUS.
3. Women at risk for bacterial endocarditis should receive prophylactic antibiotics at insertion and removal.
4. Current, recent, or recurrent PID is a contraindication for IUD use.
5. Women with diabetes mellitus, either insulin-dependent or noninsulin-dependent, can use IUDs.
6. IUD insertion is relatively easier in breastfeeding women, and the rates of expulsion and uterine perforation are not increased.

Key Points in Patient Counseling

1. Protection against unwanted pregnancy begins immediately after insertion.
2. Menses can be longer and heavier (except with the levonorgestrel IUS); tampons can be used.
3. There is a slightly increased risk of pelvic infection in the first few months after insertion.
4. Protection against infections transmitted through the vaginal mucosa requires the use of condoms.
5. Ectopic pregnancies can still occur.
6. The IUD can be spontaneously expelled; monthly palpation of the IUD strings is important to avoid unwanted pregnancies. If the strings are not felt or something hard is palpable (suggestive of the IUD frame), a clinician should be notified as soon as possible. Backup contraception should be provided until the patient can be examined.

Timing

An IUD can be safely inserted at any time after delivery, spontaneous miscarriage or induced abortion, or during the menstrual cycle. Expulsion rates were higher when the older, large plastic IUDs were inserted sooner than 8 weeks postpartum; however, studies indicate that the copper IUDs can be inserted between 4 and 8 weeks postpartum without an increase in pregnancy rates, expulsion, uterine perforation, or removals for bleeding and/or pain.[221, 222] Insertion can even occur immediately after a vaginal delivery; it is not associated with an increased risk of infection, uterine perforation, postpartum bleeding, or uterine subinvolution.[223, 224] Postvaginal delivery insertion is not recommended if intrauterine infection is present, and a slightly higher expulsion rate is to be expected compared with insertion 4–8 weeks postpartum. The IUD can also be inserted at cesarean section; the expulsion rate is slightly lower than that with insertion immediately after vaginal delivery.[225]

Insertion of an IUD in breastfeeding women is relatively easier and is associated with a lower removal rate for bleeding or pain.[223] Reports disagree whether perforation is more common in lactating women.[223, 226, 227]

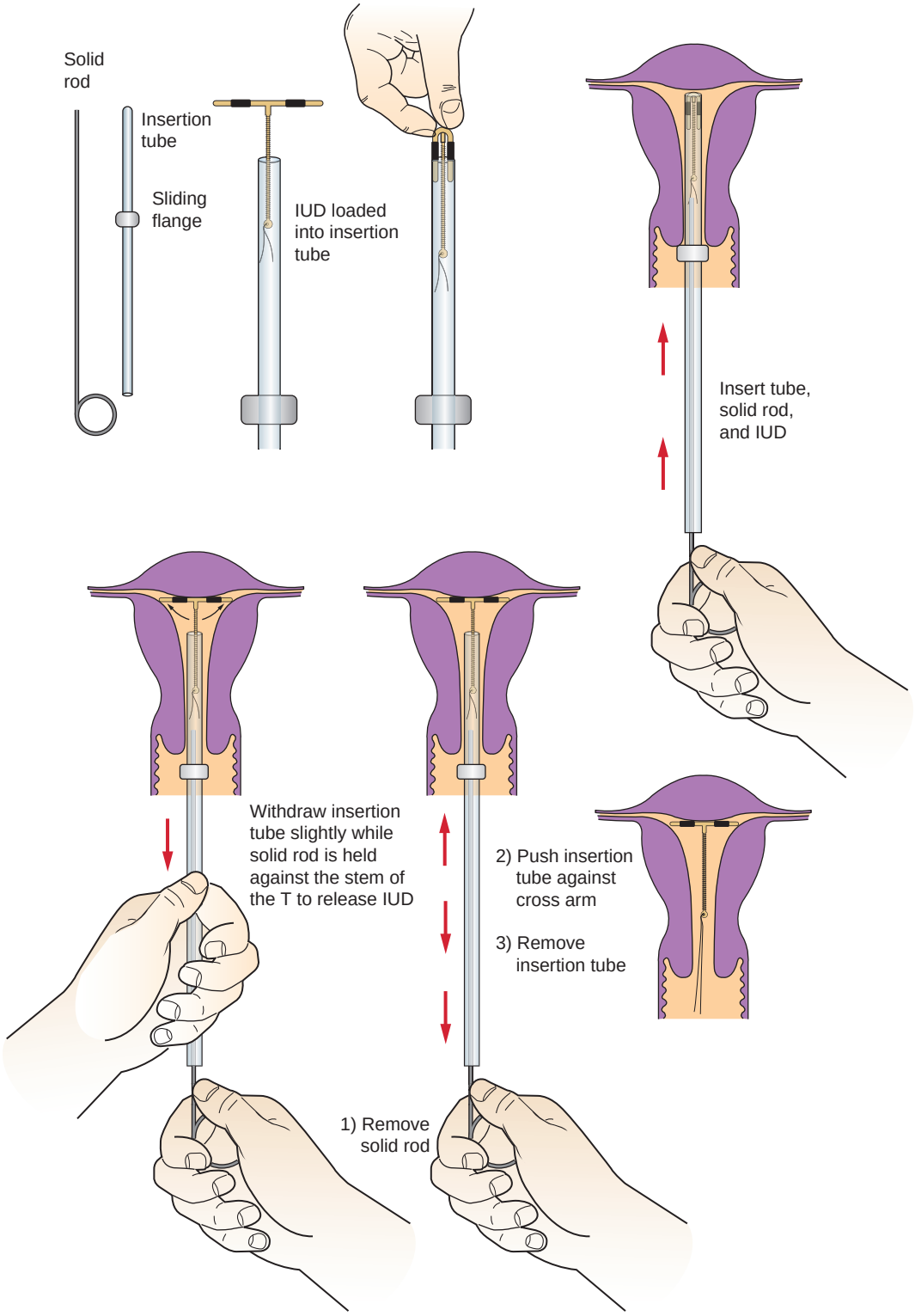
Insertions can be more difficult if the cervix is closed between menses. The advantages of insertion during or shortly after a menstrual period include a more open cervical canal, the masking of insertion-related bleeding, and the knowledge that the patient is not pregnant. These relative advantages may be outweighed by the risk of unintended pregnancy if insertion is delayed to await menstrual bleeding. In addition, there is evidence that the expulsion rate and termination rates for pain, bleeding, and pregnancy are lower if insertions are performed after day 11 of the menstrual cycle, and the infection rate may be lower with insertions after the 17th cycle day.[228]

An IUD can be inserted immediately after a first-trimester abortion, but, after a second-trimester abortion, it is recommended to wait until uterine involution occurs.[229–231] A decision analysis concluded that insertion of an IUD immediately after an abortion procedure would be associated with a major reduction in pregnancies compared with insertion at a follow-up visit, and, therefore, even considering expulsions, post abortion insertions are cost effective.[231]

Technique for the TCu-380A

Inserting an IUD requires only a few minutes, has few complications, and is rarely painful, but preoperative examination, medication, and the right equipment will ensure a good experience for your patient. After introducing a vaginal spectrum, the cervix is cleaned with chlorhexadine or povidone-iodine. Leave the antiseptic-soaked cotton-tipped applicator in the cervical canal during the procedures prior to insertion of the IUD. Place a paracervical block by injecting 1 mL of local anesthetic (1% chloroprocaine) into the cervical lip (anterior if the uterus is anterior in the pelvis and posterior if it lies posteriorly). Inclusion of atropine, 0.4 mg, in the anesthetic will reduce the incidence of vasovagal reactions. After one minute, grasp the cervical lip with the tenaculum ratcheting it only to the first position in a slow, deliberate fashion. Use the tenaculum to move the cervix to the patient's right, revealing the left lateral vaginal fornix. Place the needle tip in the cervical mucosa at 3 o'clock, 1–2 cm lateral to the cervical os, advance it about 1.5 inches (4 cm) under the mucosa and inject about 4 mL of anesthetic, leaving an additional 1 mL behind under the mucosa as the needle is withdrawn. Now deflect the cervix to the patient's left and inject local anesthetic at 9 o'clock in similar fashion.

Wait 2–3 min before proceeding. A very common mistake is to not allow sufficient time for anesthetic action.



Many women can tolerate IUD insertion, especially at the time of menses, without a paracervical block. For some women, however, insertion is less painful with local anesthetic and with administration of a nonsteroidal anti-inflammatory agent 30 min to 1 h

prior to the procedure. If a paracervical block is not used, having the patient cough just as the tenaculum is applied reduces pain and the chance of a vasovagal reaction. An alternative approach for pain relief is to apply benzocaine 20% gel first at the tenaculum site then to leave a gel-soaked cotton-tipped applicator in the cervical canal for about 1 min before inserting the IUD.[232]

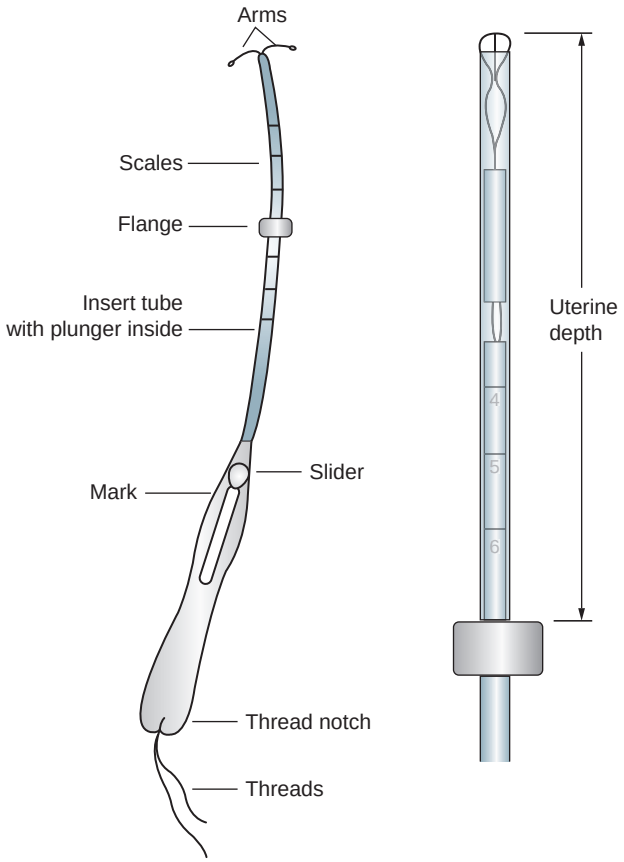
Sound and measure the depth of the uterus (the insertion tube can be used for this purpose). The IUD is loaded into its insertion tube immediately prior to insertion. The arms of the TCu 380A must be folded manually, either with sterile gloves or through the sterile wrapper, and maneuvered into the end of the insertion tube, just enough to hold them in place during insertion. The insertion tube is advanced into the uterus to the correct depth as marked on the tube by a sliding plastic flange. The flange should be twisted to be in the same plane as the horizontal arms. When the insertion tube and IUD reach the fundus, withdraw a few mm. Check to make sure that the transverse arm of the IUD is in the horizontal plane so that the tips of the T will rest in the cornual regions of the endometrial cavity. Placement in the vertical plane increases the risk of expulsion and pregnancy.[233] To release the TCu-380A, advance the solid rod until the resistance of the IUD is felt, fix the rod against the tenaculum which is held in traction, and withdraw the insertion tube while the solid insertion rod is held against the stem of the T, releasing the transverse arms into high fundal position. Remove the solid rod and finally the inserter tube taking care not to pull on the strings. You can ensure that the TCu-380A is in a high fundal position if, after removing the solid rod, you push the inserter tube up against the cross arm of the T prior to withdrawing it completely from the cavity. Trim the strings to about 4 cm from the external os, and record their length in the chart. Shorter strings can cause unpleasant bristle-like sensations.

Patients with newly inserted IUDs should attempt to feel the strings before they leave the examining room. Giving them the cut ends of the strings as a sample of what to feel is helpful. Palpation should be performed monthly by the patient to verify continuing presence of the IUD after each menstrual flow. Caution the patient that the first two menses are typically heavier. As with all office procedures, patients should be provided a 24-h phone number for urgent questions or concerns, and especially to report unusual pain, bleeding, or vaginal discharge.

A 1-month follow-up visit is recommended to identify problems amenable to counseling and treatment. Women experiencing heavier menstrual flow or irregular bleeding in the first month after insertion are at increased risk for removals because of bleeding and pain.[234] Intensified support along with treatment with a nonsteroidal anti-inflammatory agent can maintain continuation of the method.

Technique for the Levonorgestrel IUD

Prepare for insertion as described above for the copper IUD. Open the sterile package, releasing the strings. Make sure that the slider is far forward (toward the patient) as possible and that the arms of the IUD are horizontal. Holding the slider steady, pull on both strings to draw the IUD into the insertion tube until the knobs at the ends of the arms are closing the open end of the inserter. Fix the strings into the notch at the bottom of the inserter. Move the flange so that the forward surface is at the depth of the uterus as measured by the sound. Hold the slider firmly with a thumb or finger; move the inserter into the uterus until the flange is about 1.5–2.0 cm from the cervix (this allows space for the arms to open). Holding the inserter steady, release the arms by pulling the slider back until it reaches the mark (the raised horizontal line); wait 10–30 s for the arms to completely open. Holding the slider in its new position, advance the inserter gently until the flange



touches the cervix, placing the fundus high in the fundus. Release the IUD by pulling the slider down, making sure that the strings release from the notch. Remove the inserter and trim the strings.

Insertion against the anterior fundus is essential for maximal efficacy and low expulsion rates. Correct placement can be assessed with ultrasonography; this is especially useful when inserting IUDs in larger cavities (after a pregnancy or when myomas are present).

Prophylactic Antibiotics

Doxycycline (200 mg) administered orally 1 hour prior to insertion can provide protection against insertion-associated pelvic infection, but 3 double-blind randomized studies, two conducted in Africa and one in Turkey, found no significant advantage in the treated groups.[235–237] Azithromycin in a dose of 500 mg has also been used prophylactically, presumably offering more protection because of a longer half-life.[238] However, a randomized trial in low-risk women could find no effect on the subsequent rate of IUD removal or morbidity when 500 mg azithromycin was administered 1 h before IUD insertion.[239] In women at low risk for STIs, the incidence of infection is so low that there is little benefit to be expected with prophylactic antibiotics.[240]

IUD Removal

Removal of an IUD can usually be accomplished by grasping the string with a ring forcep or uterine dressing forcep and exerting firm traction. If strings cannot be seen, they can

often be extracted from the cervical canal by rotating two cotton-tipped applicators or a Pap smear cytobrush in the endocervical canal. If further maneuvers are required, a paracervical block should be administered. Oral administration of a nonsteroidal anti-inflammatory drug beforehand will reduce uterine cramping.

If IUD strings cannot be identified or extracted from the endocervical canal, a light plastic uterine sound should be passed into the endometrial cavity after administration of a paracervical block. A standard metal sound is too heavy and insensitive for this purpose. The IUD can frequently be felt with the sound and localized against the anterior or posterior wall of the uterus. The device can then be removed using a Facit polyp or alligator-type forcep directed to where the device was felt, taking care to open the forcep widely immediately on passing it through the internal cervical os so that the IUD can be caught between the jaws. If removal is not easily accomplished using this forcep, direct visualization of the IUD with sonography or hysteroscopy can facilitate removal. Sonography is less painful and more convenient and should be tried first.

Fertility returns promptly and pregnancies after removal of an IUD occur at a normal rate, sooner than after oral contraception.[64–67, 121, 122] Pregnancy outcome after IUD removal is assocated with a normal incidence of spontaneous miscarriage and ectopic pregnancy.[122]

If a patient wishes to continue use of an IUD, a new device can be placed immediately after removal of the old one. In this case, antibiotic prophylaxis is advised.

Embedded IUDs

If removal is not easily accomplished, direct visualization of the IUD with sonography or hysteroscopy can be helpful. Sonography is safer and less expensive.[195, 241] Transvaginal ultrasonography provides the best image to confirm the location of the IUD, but there is little room for the removal procedure. A better approach is to fill the bladder and use an abdominal sector transducer to image the uterine cavity as the forceps are introduced. One should open the forceps widely and see if the IUD moves when the forceps close on it. If it moves, one should close the forceps tightly and extract the IUD. If unsuccessful, the forceps is reintroduced in a different plane, keeping one jaw of the open forceps firmly against first the anterior and then the posterior uterine wall. If this approach is not successful, hysteroscopy is indicated.

Finding a Displaced IUD

When an IUD cannot be found, in addition to expulsion, one has to consider perforation of the uterus into the abdominal cavity (a very rare event) or embedment into the myometrium. All IUDs are radiopaque, but localizing them radiographically requires 2–3 views, is time-consuming and expensive, and does not allow intrauterine direction of instruments. A quick, real-time sonographic scan in the office is the best method to locate a lost IUD, whether or not removal is desired. If the IUD cannot be visualized with ultrasonography, abdominal x-rays are necessary because the IUD can be high and hidden.

If the IUD is identified perforating the myometrium or in the abdominal cavity, it should be removed using operative laparoscopy, usually under general anesthesia. If the IUD is in the uterine cavity, but cannot be grasped with a forcep under sonographic guidance, hysteroscopy is the best approach. Both routes may be helpful if an IUD is partially perforated.

Copper in the abdominal cavity can lead to adhesion formation, making laparoscopic removal difficult.[242] Although inert perforated devices without closed loops were previously allowed to remain in the abdominal cavity, current practice is to remove any perforated IUD, including the levonorgestrel-releasing system. Because IUD perforations usually occur at the time of insertion, it is important to check for correct position by identifying the string within a few weeks after insertion. Uterine perforation itself is unlikely to cause more than transient pain and bleeding, and can go undetected at the time of IUD insertion. If you believe perforation has occurred, prompt sonography is indicated so that the device can be removed before adhesion formation can occur.

This problem should be put into perspective. With the new generation of IUDs (copper and levonorgestrel), adhesion formation appears to be an immediate reaction that does not progress, and rarely leads to serious complications.[243] In appropriate situations (in which the risk of surgery is considerable), clinician and patient may elect not to remove the translocated IUD in the absence of concerning symptoms.[244] However, a case has been reported of sigmoid perforation occurring 5 years after insertion, and the general consensus continues to favor removal of a perforated IUD immediately on diagnosis.[245]

The IUD for Older Women

The growing need for reversible contraception in older women would be well served by increased utilization of the IUD. The copper and levonorgestrel IUDs are among the most effective contraceptives, better than some sterilization operations. The decline in IUD use in the U.S. is in direct contrast to the experience in the rest of the world, a complicated response to publicity and litigation. An increased risk of pelvic infection with contemporary IUDs in use is limited to the insertion procedure and the transportation of pathogens to the upper genital tract. This risk is effectively minimized by careful screening for STI risks and the use of good aseptic technique.

The IUD is a good reversible contraceptive choice for older women. An older woman is more likely to be mutually monogamous and less likely to develop PID, and for those women who have already had their children, concern with fertility and problems with cramping and bleeding are both lesser issues. If protection from STIs is not a concern, insertion of a copper IUD can provide very effective contraception until the menopause without the need to do anything other than check the string occasionally. On the other hand, because alterations of bleeding patterns become more common in this age group, it may be necessary to remove an IUD to avoid misinterpreting bleeding that could be due to endometrial disease. Because older women are more likely to have diabetes mellitus, it is worth emphasizing that no increase in adverse events has been observed with copper IUD use in women with either insulin-dependent or non-insulin-dependent diabetes.[164, 213] The IUD is not a good choice when the uterine cavity is distorted by leiomyomas. There are epidemiologic data indicating that both the copper IUD and the inert IUD reduce the risks of endometrial cancer and invasive cervical cancer,[153–157] presumably because of the induced biochemical alterations that affect cellular responses.

The levonorgestrel-releasing IUS (Mirena) is especially worth considering for older women. The levonorgestrel IUS lasts up to 10 years and reduces menstrual blood loss and pelvic infection rates.[34–36] Because of the favorable impact of locally released progestin on the endometrium, the levonorgestrel IUS is very effective for the treatment of menorrhagia, as effective as the administration of oral progestins (with less side effects), and compares favorably with surgical treatment (hysterectomy or endometrial ablation).[72–75, 77] Use of the levonorgestrel-releasing IUS also reduces bleeding in women with uterine leiomyomas,

and there is some evidence that this IUD reduces the prevalence of myomas, although not the volume of existing myomas.[85, 88, 89]

The levonorgestrel IUS effectively protects the endometrium against hyperplasia in women using tamoxifen or postmenopausal estrogen therapy.[39, 41–45, 107] In addition, this IUD can be used to prevent and to treat endometrial hyperplasia.[110, 111, 113] But it is worth emphasizing again that although the levonorgestrel IUD confidently provides good protection against endometrial hyperplasia, clinicians should maintain a high degree of suspicion of unusual bleeding and aggressively assess the endometrium.

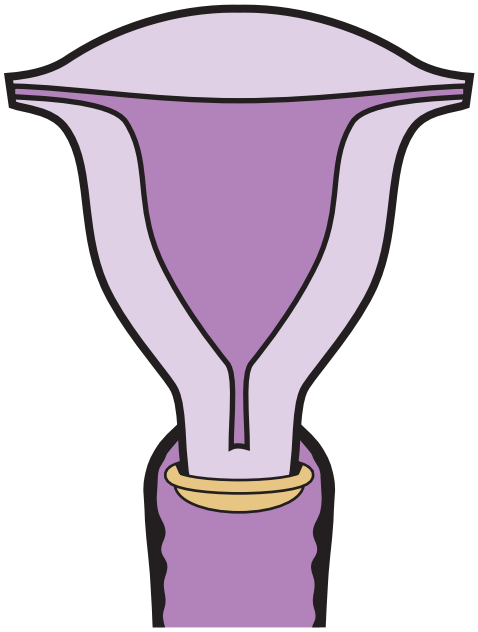
IUD Myths

We hope the information in this chapter lays to rest specific myths associated with IUDs. For emphasis, the following sentences provide the correct responses to what we believe are common misconceptions among clinicians:

1. IUDs are NOT abortifacients.
2. An increased risk of infection with the modern IUD is related ONLY to the insertion.
3. IUD use DOES NOT increase the risk of PID or infertility.
4. IUDs DO NOT increase the risk of ectopic pregnancy and CAN be used by women with a previous ectopic.
5. IUDs CAN be used by nulliparous women.
6. IUDs CAN be inserted immediately postpartum, including after first- and second-trimester abortions.
7. IUDs CAN be inserted in HIV-positive women.
8. The modern IUD HAS NOT exposed clinicians to litigation.

All references are available online at: http://www.clinicalgynendoandinfertility.com

26 Barrier Methods of Contraception and Withdrawal



Barrier methods of contraception have been the most widely used contraceptive techniques throughout recorded history. These methods, the oldest of methods, are now being thrust into the forefront as we respond to the personal and social impact of sexually transmitted infections (STIs). A new need for sexual safety has brought modern respect and new developments to the condom, while the other barrier methods continue to serve well for appropriate couples. Withdrawal is also an ancient method of contraception that deserves modern understanding and appreciation.

History of Barrier Methods

The use of vaginal contraceptives is probably as ancient as *Homo sapiens*. References to sponges and plugs appear in the earliest of writings. However, the diaphragm and the cervical cap were not invented until the late 1800s, the same time period that saw the beginning of investigations with spermicidal agents.

Intravaginal contraception was widespread in isolated cultures throughout the world. The Japanese used balls of bamboo paper, Islamic women used willow leaves, and the women

in the Pacific Islands used seaweed. References can be found throughout ancient writings to sticky plugs, made of gumlike substances, to be placed in the vagina prior to intercourse. In preliterate societies, an effective method had to have been the result of trial and error, with some good luck thrown in.

How was contraceptive knowledge spread? Certainly, until modern times, individuals did not consult clinicians for contraception. Contraceptive knowledge was folklore, undoubtedly perpetuated by the oral tradition. The social and technical circumstances of ancient times conspired to make communication of information very difficult. But even when knowledge was lacking, the desire to prevent conception was not. Hence, the widespread use of potions, body movements, and amulets; all of which can be best described as magic.

Egyptian papyri dating from 1850 B.C. refer to plugs of honey, gum, acacia, and crocodile dung. The descriptions of contraceptive techniques by Soranus are viewed as the best in history until modern times.[1] Soranus of Ephesus lived from 98 to 138 and has often been referred to as the greatest gynecologist of antiquity. He studied in Alexandria, and practiced in Rome. His great text was lost for centuries and was not published until 1838. Soranus gave explicit directions regarding how to make concoctions that probably combined a barrier with spermicidal action. He favored making pulps from nuts and fruits (probably very acidic and spermicidal) and advocated the use of soft wool placed at the cervical os. He described up to 40 different combinations.

The earliest penis protectors were just that, intended to provide prophylaxis against infection. In 1564, Gabriello Fallopius, one of the early authorities on syphilis, described a linen condom that covered the glans penis. The linen condom of Fallopius was followed by full covering with animal skins and intestines, but use for contraception cannot be dated to earlier than the 1700s.

There are many versions accounting for the origin of the word *condom*. Most attribute the word to a Dr. Condom, a physician in England in the 1600s. The most famous story declares that Dr. Condom invented the sheath in response to the annoyance displayed by Charles II at the number of his illegitimate children. All attempts to trace this physician have failed. This origin of the word can neither be proved nor disproved. Condom may be derived from the Latin *condon* that means "receptacle."[1] By 1800, condoms were available at brothels throughout Europe, but nobody wanted to claim responsibility. The French called the condom the English cape; the English called condoms French letters.

The vulcanization of rubber revolutionized transportation and contraception. Vulcanization of rubber dates to 1844, and, by 1850, rubber condoms were available in the U.S. The introduction of liquid latex and automatic machinery ultimately made reliable condoms both plentiful and affordable.

Diaphragms first appeared in publications in Germany in the 1880s. A practicing German gynecologist C. Haase wrote extensively about his diaphragm, using the pseudonym Wilhelm P.J. Mensinga. The Mensinga diaphragm retained its original design with little change until modern times.

The cervical cap was available for use before the diaphragm. A New York gynecologist E.B. Foote wrote a pamphlet describing its use around 1860. By the 1930s, the cervical cap was the most widely prescribed method of contraception in Europe. Why was the cervical cap not accepted in the U.S.? The answer is not clear. Some blame the more prudish attitude toward sexuality as an explanation for why American women had difficulty learning self-insertion techniques.

Scientific experimentation with chemical inhibitors of sperm began in the 1800s. By the 1950s, more than 90 different spermicidal products were being marketed, and some of them were used in the first efforts to control fertility in India.[2] With the availability of

the intrauterine device and the development of oral contraception, interest in spermicidal agents waned, and the number of products declined.

In the last decades of the 1800s, condoms, diaphragms, pessaries, and douching syringes were widely advertised; however, they were not widely used. It is only since 1900 that the knowledge and application of contraception have been democratized, encouraged, and promoted. And it is only since 1960 that contraception teaching and practice became part of the program in academic medicine, but not without difficulty. In the 1960s, Duncan Reid, chair of obstetrics at Harvard Medical School, organized and cared for women in a clandestine clinic for contraception. In "Dr. Reid's Clinic" at the Boston Lying-In Hospital, women were able to receive contraceptives not available elsewhere in the city.

Failure Rates During the First Year of Use, United States[3–5]

	Percent of Women with Pregnancy	
Method	*Lowest Expected*	*Typical*
No method	85%	85%
Combination Pill	0.3%	8.7%
Progestin only	0.5%	3.0%
IUDs:		
Levonorgestrel IUD	0.1%	0.1%
Copper T 380A	0.6%	1.0%
Implant	0.05%	1.0%
Injectable		
3-month	0.3	0.3%
1-month	0.05	3.0%
Patch	0.3	8.0%
Vaginal ring	0.3	8.0%
Female sterilization	0.5%	0.7%
Male sterilization	0.1%	0.2%
Periodic abstinence		25.3%
Calendar	9.0%	
Ovulation method	3.0%	
Symptothermal	2.0%	
Post-ovulation	1.0%	
Spermicides	18.0%	29.0%
Cervical cap		
Parous women	26.0%	32.0%
Nulliparous women	9.0%	16.0%
Sponge:		
Parous women	20.0%	32.0%
Nulliparous women	9.0%	16.0%
Diaphragm and spermicides	6.0%	16.0%
Condom		
Male	2.0%	17.4%
Female	5.0%	27.0%
Withdrawal	4.0%	18.4%

In 1961, C. Lee Buxton, chair of obstetrics and gynecology at Yale Medical School, and Estelle Griswold, the 61-year-old executive director of Connecticut Planned Parenthood, opened four Planned Parenthood clinics in New Haven, in a defiant move against the current Connecticut law. In an obvious test of the Connecticut law, Buxton and Griswold were arrested at the Orange Street clinic, in a prearranged scenario scripted by Buxton and Griswold at the invitation of the district attorney. They were found guilty and fined $100, but imprisonment was deferred because the obvious goal was a decision by the United States Supreme Court. Buxton was forever rankled by the trivial amount of the fine. On June 7, 1965, the Supreme Court voted 7–2 to overturn the Connecticut law on the basis of a constitutional right of privacy. It was not until 1972 and 1973 that the last state laws prohibiting the distribution of contraceptives were overthrown.

Risks and Benefits Common to All Barrier Methods

Barrier methods (condoms and diaphragms) provide protection (about a 50% reduction) against sexually transmitted infections (STIs) and pelvic inflammatory disease (PID).[6–10] This includes infections due to chlamydia, gonorrhea, trichomonas, herpes simplex, cytomegalovirus, and human immunodeficiency virus (HIV); however, only the condom has been proven to prevent HIV infection. STI protection has a beneficial impact on the risk of tubal infertility and ectopic pregnancy.[8, 11] There have been no significant clinical studies on STIs and cervical caps or the female condom, but these methods should be effective. Women who have never used barrier methods of contraception are almost twice as likely to develop cancer of the cervix.[11, 12] The risk of toxic shock syndrome is increased with female barrier methods, but the actual incidence is so rare that this is not a significant clinical consideration.[13] Women who have had toxic shock syndrome, however, should be advised to avoid barrier methods.

Barrier Methods and Preeclampsia

An initial case-control study indicated that methods of contraception that prevented exposure to sperm were associated with an increased risk of preeclampsia.[14] This was not confirmed in a careful analysis of two large prospective pregnancy studies.[15] This latter conclusion was more compelling in that it was derived from a large, prospective, cohort data base.

The Diaphragm

The first effective contraceptive method under a woman's control was the vaginal diaphragm. Distribution of diaphragms led to Margaret Sanger's arrest in New York City in 1918. This was still a contentious issue in 1965 when the Supreme Court's decision in *Griswold vs. Connecticut* ended the ban on contraception in that state. By 1940, one-third of contracepting American couples were using the diaphragm. This decreased to 10% by 1965 after the introduction of oral contraceptives and intrauterine devices and fell to about 1.9% in 1995, and today, diaphragm use has nearly disappeared in the U.S.

Efficacy

Failure rates for diaphragm users vary from as low as 2% per year of use to a high of 23%. The typical use failure rate after 1 year of use is 16%.[3–5] Older, married women with longer use achieve the highest efficacy, but young women can use diaphragms very successfully if they are properly encouraged and counseled. There have been no adequate studies to determine whether efficacy is different with and without spermicides.[16]

Side Effects

The diaphragm is a safe method of contraception that rarely causes even minor side effects. Occasionally, women report vaginal irritation due to the latex rubber or the spermicidal jelly or cream used with the diaphragm. Less than 1% discontinue diaphragm use for these reasons. Urinary tract infections are 2–3-fold more common among diaphragm users than among women using oral contraception.[17, 18] Possibly, the rim of the diaphragm presses against the urethra and causes irritation that is perceived as infectious in origin, or true infection may result from touching the perineal area or from incomplete emptying of the bladder. It is more probable that spermicides used with the diaphragm can increase the risk of bacteriuria with *E. coli*, perhaps due to an alteration in the normal vaginal flora.[19] Clinical experience suggests that voiding after sexual intercourse is helpful, and, if necessary, a single postcoital dose of a prophylactic antibiotic can be recommended. Postcoital prophylaxis is effective, using trimethoprimsulfamethoxazole (1 tablet postcoitus), nitrofurantoin (50 or 100 mg postcoitus), or cephalexin (250 mg postcoitus).

Improper fitting or prolonged retention (beyond 24 h) can cause vaginal abrasion or mucosal irritation. There is no link between the normal use of diaphragms and the toxic shock syndrome.[20] It makes sense, however, to minimize the risk of toxic shock by removing the diaphragm after 24 h and during menses.

Benefits

Diaphragm use reduces the incidence of cervical gonorrhea, trichomonas, and chlamydia,[21] pelvic inflammatory disease,[8, 22] and tubal infertility.[6, 11] There are no data, as of yet, regarding the effect of diaphragm use on the transmission of the acquired immunodeficiency syndrome (AIDS) virus HIV, but because the vagina remains exposed, the diaphragm is unlikely to protect against HIV. A clinical trial demonstrated no added benefit with a diaphragm against HIV when used with condoms.[23] An important advantage of the diaphragm is low cost. Diaphragms are durable and with proper care can last for several years.

Choice and Use of the Diaphragm

There are three major types of latex diaphragms, and most manufacturers produce them in sizes ranging from 50 to 105 mm diameter, in increments of 2.5 to 5 mm. Most women use sizes between 65 and 80 mm. The SILCS diaphragm, is a silicone barrier used with a contraceptive gel.[24] It comes in one size that fits most women, and, therefore, interaction with a clinician and the requirement for fitting can be avoided. The Milex Wide Seal diaphragm is also made of silicone and comes in eight sizes that require fitting.

The latex diaphragm made with a *flat metal spring* or a *coil spring* remains in a straight line when pinched at the edges. This type is suitable for women with good vaginal muscle tone and an adequate recess behind the pubic arch. However, many women find it difficult to place the posterior edge of these flat diaphragms into the posterior cul-de-sac and over the cervix.

Arcing diaphragms are easier to use for most women. They come in two types. The All-Flex type bends into an arc no matter where around the rim the edges are pinched together. The hinged type must be pinched between the hinges to form a symmetrical arc. The hinged type forms a narrower shape when pinched together and, thus, may be easier for some women to insert. The arcing diaphragms allow the posterior edge of the diaphragm to slip more easily past the cervix and into the posterior cul-de-sac. Women with poor vaginal muscle tone, cystocele, rectocele, a long cervix, or an anterior cervix of a retroverted uterus use arcing diaphragms more successfully.

Fitting

Successful use of a diaphragm depends on proper fitting. The clinician must have available aseptic fitting rings or diaphragms themselves in all diameters. These devices should be scrupulously disinfected by soaking in a bleach solution. At the time of the pelvic examination, the middle finger is placed against the vaginal wall and the posterior cul-de-sac, while the hand is lifted anteriorly until the pubic symphysis abuts the index finger. This point is marked with the examiner's thumb to approximate the diameter of the diaphragm. The corresponding fitting ring or diaphragm is inserted, and the fit is assessed by both clinician and patient.

If the diaphragm is too tightly pressed against the pubic symphysis, a smaller size is selected. If the diaphragm is too loose (comes out with a cough or bearing down), the next larger size is selected. After a good fit is obtained, the diaphragm is removed by hooking the index finger under the rim behind the symphysis and pulling. It is important to instruct the patient in these procedures during and after the fitting. The patient should then insert the diaphragm, practice checking for proper placement, and attempt removal.

Timing

Diaphragm users need additional instruction about the timing of diaphragm use in relation to sexual intercourse and the use of spermicide. None of this advice has been rigorously assessed in clinical studies; therefore, these recommendations represent the consensus of clinical experience.

The diaphragm should be inserted no longer than 6 h prior to sexual intercourse. About a tablespoonful of spermicidal cream or jelly should be placed in the dome of the diaphragm prior to insertion, and some of the spermicide should be spread around the rim with a finger. The diaphragm should be left in place for approximately 6 h (but no more than 24 h) after coitus. Additional spermicide should be placed in the vagina before each additional episode of sexual intercourse while the diaphragm is in place.

Reassessment

Weight loss, weight gain, vaginal delivery, and even sexual intercourse can change vaginal caliber. The fit of a diaphragm should be assessed every year at the time of the regular examination.

Care of the Diaphragm

After removal, the diaphragm should be washed with soap and water, rinsed, and dried. Powders of any sort need not and should not be applied to the diaphragm. It is wise to use water to periodically check for leaks. Diaphragms should be stored in a cool and dark location.

The Cervical Cap

The cervical cap was popular in Europe long before its reintroduction into the United States. U.S. trials have demonstrated that the Prentif cervical cap is about as effective as the diaphragm but somewhat harder to fit (it comes in only four sizes) and more difficult to insert (it must be placed precisely over the cervix).[25, 26] Efficacy is significantly reduced in parous women.

The cervical latex Prentif cap has several advantages over the diaphragm. It can be left in place for a longer time (up to 48 h), and it need not be used with a spermicide. However, a tablespoonful of spermicide placed in the cap before application is reported to increase efficacy (to a 6% failure rate in the first year) and to prolong wearing time by decreasing the incidence of foul-smelling discharge (a common complaint after 24 h).[26]

The size of the cervix varies considerably from woman to woman, and the cervix changes in individual women in response to pregnancy or surgery. Proper fitting can be accomplished in about 80% of women. Women with a cervix that is too long or too short, or with a cervix that is far forward in the vagina, may not be suited for cap use. However, women with vaginal wall or pelvic relaxation, who cannot retain a diaphragm, may be able to use the cap.

Those women who can be fitted with one of the four sizes of the Prentif cap must first learn how to identify the cervix and then how to slide the cap into the vagina, up the posterior vaginal wall and onto the cervix. After insertion and after each act of sexual intercourse, the cervix should be checked to make sure that it is covered.

The cervical cap can be left in place for 2 days, but some women experience a foul-smelling discharge by 2 days. It must be left in place for at least 8 h after sexual intercourse in order to ensure that no motile sperm are left in the vagina. To remove the cap (at least 8 h after coitus), pressure must be exerted with a finger tip to break the seal. The finger is hooked over the cap rim to pull it out of the vagina. Bearing down or squatting or both can help to bring the cervix within reach of the finger.

The most common cause of failure is dislodgment of the cap from the cervix during sexual intercourse. There is no evidence that cervical caps cause toxic shock syndrome or dysplastic changes in the cervical mucosa.[27] It seems likely (although not yet documented) that cervical caps would provide the same protection from sexually transmitted infections as the diaphragm.

The FemCap, made of nonallergenic silicone rubber, is shaped like a sailor's hat, a design that allows a better fit over the cervix and in the vaginal fornices and provides a "brim" for easier removal.[28] This cap may be easier to fit and use. There are three sizes, one for nulliparous women and larger sizes for women who have had a vaginal delivery. In a randomized trial, the pregnancy rate with FemCap was nearly 2-fold higher compared with a diaphragm.[29]

Lea's Shield is a vaginal barrier contraceptive composed of silicone.[30, 31] This soft, pliable device comes in one size and fits over the cervix, held in place by the pressure of the vaginal wall around it. There is a collapsible valve that communicates with a 9-mm opening in the bowl that fits over the cervix. This valve allows equalization of air pressure during insertion and drainage of cervical secretions and discharge, permitting a snug fit over the cervix. A thick U-shaped loop attached to the anterior side of the bowl is used to stabilize the device during insertion and for removal. The thicker part of the device is shaped to fill the posterior fornix, thus contributing to its placement and stability over the cervix. The addition of a spermicide, placed in the bowl, is recommended. Lea's Shield is designed to remain in place for 48 h after intercourse. Pregnancy rates are similar to other barrier methods, and no serious adverse effects have been reported.[32]

Ovés is a silastic cervical cap that is available in three sizes, with a loop for insertion and removal. Studies are limited to very small numbers of women, and there are no meaningful data on efficacy.[33]

The Contraceptive Sponge

The vaginal contraceptive sponge is a sustained-release system for a spermicide. The sponge also absorbs semen and blocks the entrance to the cervical canal. The "Today" sponge is a dimpled polyurethane disc impregnated with 1 g of nonoxynol-9. Approximately 20% of the nonoxynol-9 is released over the 24 h that the sponge is left in the vagina. "Protectaid" is a polyurethane sponge available in Canada and Hong Kong (it also can be purchased over the Internet) that contains three spermicides and a dispersing gel.[34] The spermicidal agents are sodium cholate, nonoxynol-9, and benzalkonium chloride. This combination exerts antiviral actions in vitro.[35] The dispersing agent, polydimethysiloxane, forms a protective coating over the entire vagina, providing sustained protection.

To insert, the Today sponge is moistened with water (squeezing out the excess) and placed firmly against the cervix. There should always be a lapse of at least 6 h after sexual intercourse before removal, even if the sponge has been in place for 24 h before intercourse (maximal wear time, therefore, is 30 h). It can be inserted immediately before sexual intercourse or up to 24 h beforehand. It is removed by hooking a finger through the ribbon attached to the back of the sponge. The Protectaid sponge can be inserted up to 12 h before intercourse, and it is easier to remove than the Today sponge. Obviously, the sponge is not a good choice for women with anatomic changes that make proper insertion and placement difficult.

In most studies, the effectiveness of the sponge exceeds that of foam, jellies, and tablets, but it is lower than that associated with diaphragm or condom use.[5, 36–38] Some studies indicated higher failure rates (twice as high) in parous women, suggesting that one size may not fit all users.[39]

Discontinuation rates are generally higher among sponge users, compared with diaphragm and spermicide use. For some women, however, the sponge is preferred because it provides continuous protection for 24 h regardless of the frequency of coitus. In addition, it is easier to use and less messy.

Side effects associated with the sponge include allergic reactions in about 4% of users. Another 8% complain of vaginal dryness, soreness, or itching. Some women find removal difficult. There is no risk of toxic shock syndrome, and, in fact, the nonoxynol-9 retards staphylococcal replication and toxin production. There has been some concern that the

sponge may damage the vaginal mucosa and enhance HIV transmission.[40] Women using the sponge have lower rates of infection with gonorrhea, trichomonas, and chlamydia.[8]

Spermicides

Jellies, creams, foams, melting suppositories, foaming tablets, foaming suppositories, and soluble films are used as vehicles for chemical agents that inactivate sperm in the vagina before they can move into the upper genital tract. Some are used together with diaphragms, caps, and condoms, but even used alone they can provide protection against pregnancy.

Various chemicals and a wide array of vehicles have been used vaginally as contraceptives for centuries. The first commercially available spermicidal pessaries were made in England in 1885 of cocoa butter and quinine sulfite. These or similar materials were used until the 1920s when effervescent tablets that released carbon dioxide and phenyl mercuric acetate were marketed. Modern spermicides, introduced in the 1950s, contain surface active agents that damage the sperm cell membranes. The agents currently used are nonoxynol-9, octoxynol-9, benzalkonium chloride, and menfegol. Most preparations contain 60–100 mg of these agents in each vaginal application, with concentrations ranging from 2–12.5%.

Representative Products

Vaginal Contraceptive Film —	**VCF (70 mg nonoxynol-9)**
Foams —	**Delfen (nonoxynol-9, 12.5%)** **Emko (nonoxynol-9, 8%)** **Koromex (nonoxynol-9, 12.5%)**
Jellies and Creams —	**Conceptrol (nonoxynol-9, 4%)** **Delfen (nonoxynol-9, 12.5%)** **Ortho Gynol (nonoxynol-9, 3%)** **Ramses (nonoxynol-9, 5%)** **Koromex Jelly (nonoxynol-9, 3%)**
Suppositories —	**Encare (nonoxynol-9, 2.27%)** **Koromex Inserts (nonoxynol-9, 125 mg)** **Semicid (nonoxynol-9, 100 mg)**

"Advantage 24" is a contraceptive gel that adheres to the vaginal mucosa and provides longer availability of nonoxynol-9; it is intended to be effective for 24 h. Although available without prescription, adequate clinical trial data are not available. Allendale-N9 is a vaginal contraceptive film that contains more nonoxynol-9 than VCF.[41] An Allendale film has also been developed that contains benzalkonium chloride instead of nonoxynol-9.[42] In addition to spermicidal activity, benzalkonium chloride is microbicidal and demonstrates activity against HIV.[43] Benzalkonium chloride is available for contraceptive use in the form of a suppository, in a sponge, or as a cream in several countries.

Although in vitro studies have demonstrated that spermicides kill or inactivate most STI pathogens, including HIV, it cannot be said that spermicides provide protection against sexually transmitted infections. Spermicides have been reported to prevent HIV seroconversion as well as to have no effect; therefore, spermicides by themselves cannot be counted on for protection against HIV.[44–48] In a controlled, clinical trial in female sex workers, nonoxynol-9 failed to protect against HIV transmission.[49] Clinical studies have indicated

reductions in the risk of gonorrhea,[50–52] pelvic infections,[22] and chlamydial infection.[50, 52] However, these studies probably reflected condom use. In trials with a placebo, nonoxynol-9 provided no protection against gonorrhea or chlamydia.[49, 53–55] Indeed, there is concern that frequent spermicide use may irritate the vagina and enhance HIV transmission.[45] Because of this concern, condom makers discontinued the production of condoms lubricated with nonoxynol-9. There is little difference in the incidence of trichomoniasis, candidiasis, or bacterial vaginosis among spermicide users.[56] ***The best evidence indicates that spermicides do not provide additional protection against STIs over that associated with condoms; therefore, spermicides should not be used without condoms if a primary objective is to prevent infection with HIV, gonorrhea, or chlamydia.***

Efficacy

Only periodic abstinence demonstrates as wide a range of efficacy in different studies as do the studies of spermicides. Efficacy seems to depend more on the population studied than the agent used. Efficacy ranges from less than 1% failure to nearly one-third in the first year of use. Failure rates of approximately 20–25% during a year's use are most typical.[4, 57] A randomized trial comparing VCF vaginal contraceptive film (72 mg nonoxynol-9) with Conceptrol foaming tablets (100 mg nonoxynol-9) recorded similar 6-month pregnancy rates (24.9% with the film and 28.0% with the tablet).[57] A randomized assessment of the various products concluded that a dose of 52.5 mg nonoxynol-9 was less effective (22% in 6 months) than those containing 100 or 150 mg (about 15% in 6 months; intermediate doses were not tested).[58] These are very high rates, amounting to approximately 30–40% for 1 year of use. ***Although better than no method, spermicides alone should not be recommended for contraception unless method failure and pregnancy are acceptable.***

Spermicides require application 10–30 min prior to sexual intercourse. Jellies, creams, and foams remain effective for as long as 8 h, but tablets and suppositories are good for less than 1 h. If ejaculation does not occur within the period of effectiveness, the spermicide should be reapplied. Reapplication should definitely take place for each coital episode.

Vaginal postcoital douches are ineffective contraceptives even if they contain spermicidal agents. Postcoital douching is too late to prevent the rapid ascent of sperm (within seconds) to the fallopian tubes.

Advantages

Spermicides are relatively inexpensive and widely available in many retail outlets without prescription. This makes spermicides popular among adolescents and others who have infrequent or unpredictable sexual intercourse. In addition, spermicides are simple to use.

Side Effects

No serious side effects or safety problems have arisen in all the years that spermicides have been used. The only serious question raised was that of a possible association between spermicide use and congenital abnormalities or spontaneous miscarriages. Epidemiologic analyses, including a meta-analysis, concluded that there is insufficient evidence to support these associations.[59–61] Spermicides are not absorbed through the vaginal mucosa

in concentrations high enough to have systemic effects.[62] Vaginal and cervical mucosal damage (de-epithelialization without inflammation) has been observed with nonoxynol-9, and the overall impact on HIV transmission, although unknown, is of concern.[63, 64]

The principal minor problem is allergy that occurs in 1–5% of users, related to either the vehicle or the spermicidal agent. Using a different product often solves the problem. Spermicide users also have an altered vaginal floral promoting the colonization of *E. coli*, leading to a greater susceptibility to urinary tract infections as noted with diaphragm/spermicide users.[18, 65]

The Search for Contraceptives to Prevent STIs

For the last two decades, extensive research has been devoted to the development of contraceptive microbicides to prevent STIs, especially HIV. The ideal agent would be a topical microbicide that would prevent infection and be spermicidal. Any new agent must match the latex condom, which is nearly 100% effective in blocking bacteria and viruses. The road is long, extending from in vitro work to clinical application. An acceptable agent must avoid damage to vaginal epithelial cells and disruption of vaginal flora, and the delivery system must be user-friendly. Carraguard, a microbicide that contains carrageenan, a substance derived from seaweed, is a good example of the frustration in this area. After extensive development by the Population Council, a large, long-term Phase 3 clinical trial in South Africa concluded that Carraguard did not prevent vaginal transmission of HIV.[66] Acidform and BufferGel, spermicidal and microbicidal acidic gels, have been safe and effective in Phase I clinical trials, when loaded in a diaphragm-like device.[67, 68]

Condoms

Although awareness of condoms as an effective contraceptive method as well as a protector against STIs has increased tremendously in recent years, a great deal remains to be accomplished to reach the appropriate level of condom use. Contraceptive efficacy and STI prevention must be linked together and publicly promoted. The male condom is the only contraceptive proven to prevent HIV infection.

There are three specific goals: correct use; consistent use; and affordable, easy availability. If these goals are met, the 21st century will see the annual manufacture of 20 billion condoms per year.

Various types of condoms are available. Most are made of latex; polyurethane and silicone rubber condoms are also now manufactured. “Natural skin” (lamb’s intestine) condoms are still obtainable (about 1% of sales). Latex condoms are 0.3–0.8 mm thick. Sperm that are 0.003 mm in diameter cannot penetrate condoms. The organisms that cause STIs and AIDS also do not penetrate latex condoms, but they can penetrate condoms made from intestine.[69, 70] Evidence indicates that condom use (latex) also prevents transmission of human papillomavirus (HPV).[71] The use of spermicides or spermicide-coated condoms increases the incidence of *E. coli* bacteriuria and urinary tract infections due to either *E. coli* or *Staphylococcus saprophyticus* because of the spermicide-induced alteration in vaginal flora.[19, 72] Consistent use of condoms when one partner is HIV seropositive is highly effective in preventing HIV transmission; there was no seroconversion in 124 couples who used condoms consistently compared with 12.7% conversion after 24 months in couples with inconsistent use.[73, 74] Women who are partners of condom users are less likely to be HIV-positive.[75]

An evaluation of the world's literature concluded that consistent use of condoms provides protection against HIV to a degree comparable to condom efficacy in preventing pregnancy (reflecting some inconsistent use and other routes of transmission).[74] In addition, condoms protect against transmission of the herpes simplex virus from infected men to women.[76]

Polyurethane condoms are expected to protect against STIs and HIV, based on in vitro efficacy as a barrier to bacteria and viruses. They are odorless, may have greater sensitivity, and are resistant to deterioration from storage and lubricants. Those individuals who have the infrequent problem of latex allergy can use polyurethane condoms. Breakage and slippage have been reported to be comparable with latex condoms.[77] However, in a randomized, well-designed study, the polyurethane condom had a 6-fold higher breakage rate, and another study comparing latex and polyurethane condoms found a higher pregnancy rate with the polyurethane condom.[78, 79]

Condoms can be straight or tapered, smooth or ribbed, colored or clear, lubricated or nonlubricated. These are all marketing ventures aimed at attracting individual notions of pleasure and enjoyment.[80] An often repeated concern is the alleged reduction in penile glans sensitivity that accompanies condom use.[80] This has never been objectively studied, and it is likely that this complaint is perception (or excuse) not based on reality. A clinician can overcome this objection by advocating the use of thinner (and more esoteric) condoms, knowing that any difference is also more of perception than reality.

As is true for most contraceptive methods, older, married couples experienced in using condoms and strongly motivated to avoid another pregnancy are much more effective users than young, unmarried couples with little contraceptive experience. This does not mean that condoms are not useful contraceptives for adolescents, who are likely to have sex unexpectedly or infrequently. The recent decline in the teen pregnancy rate partly reflects wider use of condoms by teens concerned about avoiding HIV infection.

Prospective users need instructions if they are to avoid pregnancy and STIs. A condom must be placed on the penis before it touches a partner. Uncircumcised men must pull the foreskin back. Prior to unrolling the condom to the base of the penis, air should be squeezed out of the reservoir tip with a thumb and forefinger. The tip of the condom should extend beyond the end of the penis to provide a reservoir to collect the ejaculate (a half-inch of pinched tip). If lubricants are used, they must be water based. Oil-based lubricants (such as Vaseline) will weaken the latex. Couples should be concerned that any vaginal medication can compromise condom integrity. After intercourse, the condom should be held at the base as the still erect penis is withdrawn. Semen must not be allowed to spill or leak. The condom should be handled gently because fingernails and rings can penetrate the latex and cause leakage. If there is evidence of spill or leakage, a spermicidal agent should be quickly inserted into the vagina, and treatment should be initiated within 120 h, but preferably within 72 h, with an emergency contraception method.

SUMMARY—Key Steps for Maximal Condom Efficacy

1. Use condoms for every act of coitus.
2. Place the condom before vaginal contact.
3. Create a reservoir at the tip.
4. Withdraw while the penis is still erect.
5. Hold the base of the condom during withdrawal.

These instructions should be provided to new users of condoms who are likely to be reluctant to ask questions. Most condoms are acquired without medical supervision; therefore, clinicians should use every opportunity to inform patients about their proper use.

Inconsistent use explains most condom failures. Incorrect use accounts for additional failures; also, condoms sometimes break. Breakage rates range from 1–8 per 100 episodes of vaginal intercourse (and somewhat higher for anal intercourse), and slippage rates range from 1% to 5%.[81, 82] With experienced couples, condom failure due to breakage and slippage (sufficient to increase the risk of pregnancy or STIs) occurs at a rate of about 1%.[83] In a U.S. survey, one pregnancy resulted for every three condom breakages.[84] Concomitant use of spermicides lowers failure rates in case of breakage. In addition, even when there is slippage or breakage, the condom provides some protection against pregnancy and STIs because there is still a reduction in exposure to semen.[85]

Breakage is a greater problem for couples at risk for STIs. An infected man transmits gonorrhea to a susceptible woman approximately two-thirds of the time.[86] If the woman is infected, transmission to the man occurs one-third of the time.[87] The chances of HIV infection after a single sexual exposure ranges from 1 in 1,000 to 1 in 10.[88, 89]

Condom breakage rates depend on sexual behavior and practices, experience with condom use, the condition of the condoms, and manufacturing quality. Condoms remain in good condition for up to 5 years unless exposed to ultraviolet light, excessive heat or humidity, ozone, or oils. Condom manufacturers regularly check samples of their products to make sure they meet national standards. These procedures limit the proportion of defects to less than 0.1% of all condoms distributed. Contraceptive failure is more likely to be due to nonuse or incorrect use.

When a condom breaks, or if there is reason to believe spillage or leakage occurred, a woman should contact a clinician preferably within 72 h and no later than 120 h. Emergency contraception, as discussed in Chapter 22, should be provided. Couples who rely on condoms for contraception should be educated regarding emergency contraception, and an appropriate method should be kept available for self-medication).

For the immediate future, prevention of STIs and control of the AIDS epidemic will require a great increase in the use of condoms. We must all be involved in the effort to promote condom use. Condom use must be portrayed in the positive light of STI prevention. An important area of concentration is the teaching of the social skills required to ensure use by a reluctant partner. Using scare tactics about STIs in order to encourage condom use is not sufficient. A more positive approach can yield better compliance. It is useful to emphasize that prevention of STIs will preserve future fertility. ***For women not in a stable, monogamous relationship, a dual approach is recommended, combining the contraceptive efficacy and protection against PID offered by estrogen-progestin contraception with the use of a barrier method for prevention of viral STIs.***

The Female Condom

The female condom is a pouch made of polyurethane, which lines the vagina.[90] An internal ring in the closed end of the pouch covers the cervix and an external ring remains outside the vagina, partially covering the perineum. The female condom is prelubricated with silicone, and a spermicide need not be used. The female condom should be an effective barrier to STI infection. The female condom is impervious in vitro to cytomegalovirus and HIV[91]; however, high cost and acceptability are major problems. The integrity of the female condom is maintained with up to 8 multiple uses with washing, drying, and relubricating.[92]

The devices are more cumbersome than condoms, and studies have indicated relatively high rates of problems such as slippage.[93] Women who have successfully used barrier methods and who are strongly motivated to avoid STIs are more likely to choose the female condom. With careful use, the efficacy rate should be similar to that of the diaphragm and the cervical cap.[94–96]

	Diaphragm	Cap	Sponge	Female Condom
Insertion before coitus, no longer than:	6 hrs	6 hrs	24 hrs	8 hrs
After coitus, should be left in place for:	6 hrs	8 hrs	6 hrs	6 hrs
Maximal wear time:	24 hrs	48 hrs	30 hrs	8 hrs

Withdrawal

Withdrawal as a contraceptive method, a method used for centuries and known historically as coitus interruptus, should not be ignored or underrated. The withdrawal method is intuitively learned, has no cost, and efficacy is surprisingly good. If withdrawal before ejaculation occurs with every instance of intercourse, a failure rate over a year of only 4% can be achieved.[5] Clinicians who scoff at this method lose sight of the fact that its typical failure rate of 18.4% during one year is very similar to that achieved with male condoms, which are a lowest expected failure rate of 2% and a 17.4% typical rate in 1 year.

A lack of respect for withdrawal as a contraceptive method can be attributed to two factors: an understandable preference for modern methods and a belief that pre-ejaculate fluid contains sperm. The latter concern is understandable given the difficulty inherent in separating pre-ejaculate fluid from the ejaculate in order to study the question. There is one small study of 5 men with a history of premature ejaculation and 3 men with excessive fluid during foreplay.[97] Not a single sperm was detected in preejaculatory fluid specimens collected at the tip of the urethra. In two other studies, no sperm was found in preejaculatory fluid from 16 men, and in 15 men, a few clumps of sperm were present in 5 men.[98, 99] Scarce data; however, the relatively good failure rates with this method indicate that these studies are correct.

The modern prevalent use of this method (about 4% of contracepting women in the 2002 U.S. National Survey of Family Growth) is difficult to accurately measure because individuals very often combine withdrawal with another method, and the other method is the one reported in family planning surveys.[100] In other words, the use of withdrawal is significantly underreported. Indeed, many individuals do not classify withdrawal as a contraceptive method, saving that designation for modern methods. Nevertheless, it is not surprising that from 25% to 60% of adolescents report the use of withdrawal.[101, 102]

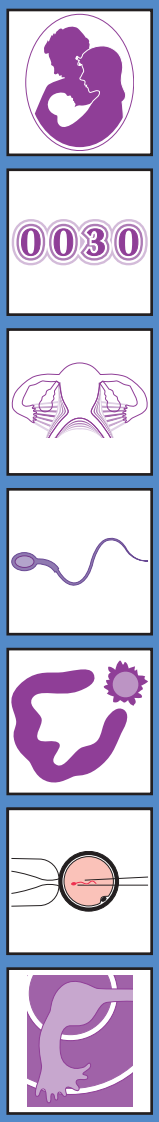
We encourage clinicians to have a greater acceptance for this commonly used method, and it truly is a legitimate contraceptive method as evidenced by the substantial protection against pregnancy that is achievable. Of course, protection against STIs is not to be expected.

All references are available online at: http://www.clinicalgynendoandinfertility.com

SECTION

IV

INFERTILITY



27 Female Infertility

Infertility is generally defined as one year of unprotected intercourse without conception.[1] Some prefer the term ***subfertility*** to describe women or couples who are not sterile but exhibit decreased reproductive efficiency. Approximately 85–90% of healthy young couples conceive within 1 year, most within 6 months.[2, 3] Infertility therefore affects approximately 10–15% of couples and is an important part of the practice of many clinicians. Cycle ***fecundability*** is the probability that a cycle will result in pregnancy and ***fecundity*** is the probability that a cycle will result in a live birth.

Contrary to popular perception, the overall incidence of infertility has remained relatively unchanged over the past 3 decades. However, the evaluation and treatment of infertility have changed dramatically during that time. Three major developments have had the greatest impact. First was the introduction of *in vitro* fertilization (IVF) and other assisted reproductive technologies (ART). ART techniques have provided the means to study reproductive processes in new and more revealing ways and have markedly improved the prognosis for a great many infertile couples, particularly those whose infertility relates to severe tubal damage or male factors. Second, changes in population demographics have resulted in greater numbers of women attempting pregnancy at older ages when they are inherently less biologically fertile. Third, advances in ART and concerns about the age-related decline in fertility have combined to attract greater media attention and to raise public awareness of infertility and modern treatments. Consequently, infertile couples are now more likely to seek medical advice, evaluation, and treatment.

The Epidemiology of Infertility in the U.S.

The first U.S. census was in 1790. At that time, the crude birth rate was 55 per 1,000 total population; in 2007, it was 14.3 per 1,000,[4] representing nearly a 75% decline over the past 200-plus years. The crude birth rate in 2007 was 15% lower than in 1990 (16.7 per 1,000 population), but increased from 2002 (13.9 per 1,000), which was a record low for the nation.[5] The general fertility rate (births per 1,000 women aged 15–44) in 2007 was 69.5, 2% lower than in 1990 (70.9/1,000), 11% lower than in 1970 (87.9/1,000), and 35% lower than in 1950 (106.2/1,000) during the post-war "baby boom."[4, 6, 7] The general fertility rate in 2007 was the highest since 1990.

The overall long-term decline in U.S. birth and fertility rates has been attributed to several factors.

- Greater interest in advanced education and careers among women.
- Later marriage and more frequent divorce.
- Improvements in contraception and access to family planning services.
- Delayed childbearing.
- Decreased family size.

Attitudes towards women and among women in our society have changed dramatically in many ways over the past 30 years. Expanding opportunities have increased interests in advanced education and careers among women. U.S. census data indicate that in 1970, only 8.2% of women age 25 and older had completed 4 or more years of college; by 2001, that proportion had tripled (24.3%).[8] Women have represented the majority of college students in the U.S. since 1979. In recent years, the majority of bachelor's and more advanced degrees have been awarded to women. The proportion of U.S. women with infant children in the work force has steadily increased, from 31% in 1976 to 55% in 2000.[8] In 2006, 85% of all women ages 15 to 44 years were in the labor force.[9]

Greater focus on education and careers among women triggered other trends in modern society. Less frequent and later marriage and more frequent divorce were among the most striking. First marriage rates in the U.S. peaked after World War II, between 1945 and 1947 (143 per 1,000 single women aged 15–44), and declined about 15% every 10 years and approximately 50% overall over the next 5 decades.[10] The median age at first marriage for women has risen steadily since 1960 (20.3 years) and reached an all-time high in 2007 (26.0 years). The probability of future marriage for women decreases as age increases: 84% at age 25, 72% at age 30, 52% at age 35, and 41% at age 40.[11] If and when women do marry, they also are more likely to divorce than in the past.[10–13] Divorce rates among women of reproductive age rose quickly after 1960 to more than double by 1980 (40 per 1,000 married women aged 15–44) and have declined only slightly over the last 30 years. The National Center for Health Statistics estimates that approximately one-third of new marriages among younger people will end in divorce within 10 years and 43% within 15 years. Once-married women also are increasingly less likely to remarry. Remarriage rates peaked in 1968 (166 per 1,000 divorces or widowed women aged 15–44) as divorce rates began to rise, but have since declined steadily by more than one-third, in parallel with first marriage rates.[10–13]

The post-war "baby boom" generation, those born between 1946 and 1964, was the first to be afforded the means to safely and effectively control their fertility. Expanding contraceptive options and access to family planning and legalized abortion services over the past 5 decades definitely have contributed to declining U.S. birth and fertility rates. Their effects have been direct, by reducing the number of unplanned pregnancies and births, and indirect, by helping women to avoid pregnancy until their education and career goals were met, and marriage and family become a priority.

The net result of all of these societal changes was a trend to delayed childbearing among American women. The mean age at first live birth has risen steadily, from 21.4 years in 1970 to an all-time high 25.2 years in 2004 (3.8 years and 18% higher). The percentage of first births occurring to women aged 30 or older increased more than 6-fold between 1970 and 2002.[14] Mean ages for all subsequent live births increased as well; the increase in mean age was greatest (3.6 years) for the second live birth (27.7 years), and lower for the third (2.5 years), fourth (1.6 years), and fifth births (0.4 years).[15] Between 1970 and 2007, birth rates fell for women ages 15–19 (68.3 vs. 42.5/1,000), 20–24 (167.8 vs. 106.4/1,000), and those 25–29 (145.1 vs. 117.5/1,000), and increased for women aged 30–34 (73.3 vs. 99.9/1,000), 35–39 (31.7 vs. 47.5/1,000), and those aged 40–44 (8.1 vs. 9.5/1,000).[4, 7, 16] Predictably, increasing age at first birth and declining fertility rates combined to result in fewer births per woman. At the height of the postwar baby boom, the U.S. total fertility rate (births by age 45) reached a modern high of 3.7 births per woman (1957). Thereafter, the total fertility rate declined to a low of 1.8 in 1976, rose slightly to 2.1 in 2001,[7] and has remained stable since.[4] The total fertility rates in some European countries are significantly lower (Italy, 1.33; Greece, 1.29; Spain, 1.32), and inadequate even to ensure replacement of the population.[17]

The larger number of women born during the postwar baby boom increased markedly the absolute numbers of women with impaired fertility. Over a 20-year interval, a large population of women was attempting pregnancy, often for the first time, when older and less biologically fertile. Whereas in the past many such women might have chosen to adopt, the availability of legal abortion services and society's increasing acceptance of single parenthood greatly reduced the number of infants available for adoption. Women were more likely to seek infertility services, and more likely to pursue the most aggressive forms of treatment, because they offered the greatest probability for success. Now, even the youngest "boomers" are over age 45 and have completed childbearing. In 2000, the median age of the U.S. population was 35.3 years and 16% of people were between the ages of 35 and 44, representing the largest 10-year age segment of the entire population. That same year, 14.2% of the population was 25–34 years of age, 13.9% was 15–24 years, and 14.6% was 5–14 years. ***Even barring any changes in the causes and prevalence of infertility, the absolute numbers of infertile women in the U.S. can be expected to decline in the years ahead.***

Trends in the incidence of infertility among U.S. women have been difficult to define confidently, partly due to confusion over the use of two different measures—impaired fecundity, and infertility, which are defined differently, describe different populations, and can yield conflicting data.[18, 19] However, evidence indicates that the incidence of infertility in the U.S. now is declining.[20] The earliest national estimate of infertility, from the 1965 National Survey of Family Growth (NSFG), was 11.2%. In 1982, 8.5% of married American women were infertile, and in 2002, 7.4% were infertile, representing a 10% decrease over the intervening 20 years.[20] Although the explanation is not entirely clear, the percentage of women ever treated for pelvic inflammatory disease also decreased steadily, from 17.1% in 1982 to 12.0% in 1988, to 8.2% in 1995, to 5.7% in 2002.[20] In a 2007 analysis of data derived from 25 population surveys sampling 172,413 women, the median international prevalence of infertility (12 months) among women ages 20–44 years was 9% (range 3.5% to 16.7%).[21]

The array of infertility services, and their availability, has increased dramatically over the last 25 years. Clinicians have become more aware of infertility and better trained to evaluate and treat its causes. The public too has a greater awareness of infertility and modern treatments, largely due to the increased media attention, good and bad, surrounding the advances and controversies relating to assisted reproductive technologies (ART). As infertility has become more visible, and more socially acceptable, couples have become less reluctant to seek evaluation and treatment.

According to data derived from the National Survey of Family Growth (NSFG) conducted in 1995, 9.3 million women ages 15–44 (15%) had ever received infertility services, an increase from 6.6 million women (12%) in 1982.[22] These data indicated that the demand

for infertility services increased during the 1980s and early 1990s, corresponding to the aging of baby boomers and the time when the availability of ART was rapidly expanding. Compared to the general population, women seeking infertility services were more likely to be older (aged 35–44 years; 43% vs. 36%) nulliparous (36% vs. 16%), married (79% vs. 64%), relatively affluent (61% vs. 51%), and to have health care insurance (83% vs. 74%).[22] Among those who received infertility services, 35% had used ovulation-inducing drugs and 1.6 % had undergone some form of ART. In the 2002 NSFG, 7.3 million women ages 15–44 reported ever having used infertility services, representing a significant decrease of approximately 20% since 1995.[23] Advice (74%) and testing (59%) were the most common types of services received, and nearly half reported receiving drugs to improve ovulation.[24]

Aging and Fertility

The effects of aging on female fertility are perhaps best revealed by the results of studies in "natural" populations wherein couples reproduce without voluntary restrictions;[25] the Hutterites in North America are a classic example. Contraception is condemned in the sect, which emigrated originally from Switzerland in the 16th century and settled ultimately in communal colonies in South Dakota in the late 19th century. Studies of fertility in the Hutterites illustrate how fertility declines with advancing age.[26] Whereas only 2.4% of Hutterite women were infertile, 11% bore no children after age 34, 33% were infertile by age 40, and 87% were infertile at age 45. Although revealing, these and other data derived from studies in natural populations may not reflect true biologic reproductive potential, for several reasons:

- Women who have children when young may be less inclined to conceive again in later life.
- Coital frequency often declines as age increases, reflecting decreasing desire or lack of a partner.
- The incidence of subclinical abortion is unknown.
- The cumulative impact of other diseases or conditions that can adversely affect fertility (e.g, pelvic infections, leiomyomata, endometriosis) is greater in older women.

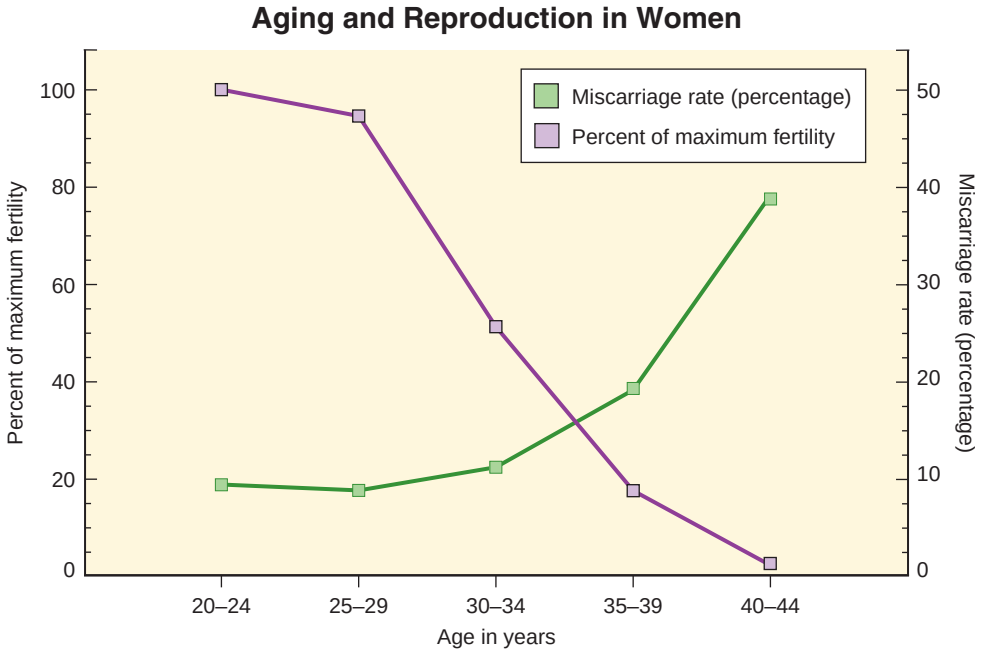
Taken together, data from studies in the Hutterites and other natural populations suggest that fertility in women peaks between the ages of 20 and 24, decreases relatively little until approximately age 30 to 32, and then declines progressively. ***Overall, fertility rates are 4–8% lower in women aged 25–29 years, 15–19% lower in those aged 30–34, 26–46% lower in women aged 35–39, and as much as 95% lower for women aged 40–45 years***.[27, 28] Variations in fertility rates among natural populations could reflect differences in genetic factors or socio-economic conditions at different times and in different places.

Other evidence for the adverse effect of aging on fertility derives from numerous studies of cumulative conception rates among women attempting pregnancy by artificial insemination with donor sperm. Data from donor insemination studies are informative because the women enrolled are less likely to have other infertility factors, and because carefully timed inseminations eliminate the confounding effects of decreasing coital frequency with increasing age. In a French study involving more than 2,000 women across up to 12 insemination cycles, conception rates were highest in those age 25 or younger (73%) and ages 26–30 (74%), 16% lower (62%) in women aged 31–35, and 27% lower in those over age 35 (54%).[29] An American donor insemination study yielded similar results, observing lower overall conception rates and a 2-fold higher number of inseminations per conception in women over age 35.[30] A Dutch study observed that the probability of a healthy live birth decreased by approximately 3.5% per year after age 30.[28] In a large British study of nearly 3,000 donor insemination cycles from a single center, cumulative conception rates

in women over age 30 were 20–35% lower than in younger women after 3 (17% vs. 21%), 6 (26% vs. 40%), and 12 insemination cycles (44% vs. 62%).[31]

Success rates achieved with ART also decline as age increases. The numbers of oocytes retrieved and embryos available are lower, embryo fragmentation rates are higher, and implantation rates are lower in older than in younger women.[32, 33] Although ART pregnancy rates have increased steadily over the past 20 years for women in all age groups, annual reports derived from registry data collected by the Centers for Disease Control and Prevention (CDC) in the U.S. since 1989 demonstrate consistently that age is the single most important factor affecting the probability of success with ART. Pregnancy and live birth rates for ART cycles using fresh, non-donor eggs or embryos vary little for women under age 32, but thereafter decrease steadily in an almost linear fashion as age increases. Regardless whether success rates are calculated per cycle, per oocyte retrieval, or per embryo transfer, the result is the same. In the 2007 U.S. national summary, the live birth rate per embryo transfer was 45.9% for women under age 35, 36.9% for ages 35–37, 27.1% for ages 38–40, 16.0% for ages 41–42, and 8.4% for women aged 43–44 years.[34]

The age-related decline in ART live birth rates reflects not only decreasing fertility, but also increasing pregnancy wastage. Just as fertility decreases with increasing age, the incidence of clinically recognized miscarriage rises as age advances. ***Miscarriage rates in natural conception cycles are generally low before age 30 (7–15%) and rise with age, only slightly for ages 30–34 (8–21%), but to a greater extent for ages 35–39 (17–28%) and ages 40 and older (34–52%).***[27, 35–37] The same pattern is observed in pregnancies resulting from ART. In the 2007 U.S. national summary of IVF outcomes, miscarriage rates were below 15% for women under age 35, almost 30% at age 40, and over 50% for women age 44 and older.[34] Longitudinal studies of healthy young women wherein daily urine samples were monitored for the appearance of human chorionic gonadotropin (hCG) have revealed that true spontaneous miscarriage rates (also including clinically unrecognized "biochemical" pregnancies) are substantially higher.[38–40] ***Up to 60% of all conceptions miscarry within the first 12 weeks of gestation and 20–40% of all early pregnancy losses go unrecognized.*** Whether the incidence of occult early pregnancy loss also is higher in older women than young women has not been determined. If so, the relationship between true spontaneous miscarriage rates and age may be even more dramatic. Even if not, the overall miscarriage risk (recognized and unrecognized) in women over age 40 approaches or exceeds 75%.[39, 41]



Physiology of Reproductive Aging

Established societal trends toward delayed childbearing and the age-related decrease in female fertility have focused a great deal of attention on the physiology of reproductive aging. Consequently, our understanding of the mechanisms that govern the pace of follicular depletion, the endocrinology of reproductive aging, and age-related changes in follicular dynamics and oocyte quality has advanced greatly over the past 20 years. We long ago recognized the changes in menstrual cycle characteristics that accompany advancing age, but now much better understand the mechanisms responsible for those changes. We long ago recognized that fertility declines as age increases, but now have measures of reproductive aging that help to guide our efforts to overcome its limitations. We know that we cannot prevent aging, but now can better help women to set and to realize their reproductive goals.

Follicular Depletion

During fetal life, germ cells rapidly proliferate by mitosis to yield approximately 6–7 million oogonia by 16–20 weeks of pregnancy.[42–44] From that point on, the germ cell population begins an inexorable decline mediated primary by gene-regulated apoptosis.[45] After entering the first meiotic division and becoming oocytes, the number of germ cells falls to between 1 and 2 million at birth,[46] and to about 300,000 by the onset of puberty.[43, 47] Over the next 35–40 years of reproductive life, only about 400 oocytes will ovulate, the rest being lost through atresia. By age 40, the size of the follicular pool declines to approximately 25,000, and at menopause, less than 1,000 follicles remain.[48–51]

Accurate modeling of the pattern of follicle depletion in the human ovary is important because the ability to measure reproductive aging or to predict the number of remaining follicles—to tell time on the biological clock—would help women make informed decisions about their reproductive plans.[52] However, for obvious reasons, accurate measures of the numbers of primordial follicles across a human female reproductive life span are difficult to obtain. The first attempt to define the age-related pattern of follicular depletion was based on an analysis of combined data from older morphometric studies and yielded a bi-exponential model of ovarian aging, describing a biphasic pattern of oocyte depletion, with a distinct increase in the rate of decline beginning at approximately age 37.5 years.[42, 48, 53, 54] The biphasic model was widely accepted, despite the biological implausibility of an abrupt, population-wide, physiologic shift in the rate of follicular depletion.[55, 56] The model still is cited frequently,[57, 58] but subsequent work has demonstrated that a simpler, more biologically plausible, exponential[49, 59] or power function[60] conforms best with available human data and current concepts regarding the mechanisms that govern the rate of follicular depletion.[52, 61] ***The current working model describes a gradually increasing rate of follicular depletion in which the pace of decline increases as the number of follicles remaining decreases, supported by evidence that paracrine factors secreted by primordial follicles inhibit recruitment and regulate the size of the resting follicular pool.***[52, 61–63] The model describes the mean trajectory of follicular depletion, but leaves a great deal of population variation unexplained. Some of the variation among individuals doubtless relates to differences in the size of the initial follicular pool, which could be random but likely is genetically determined, and on lifestyle factors. The current model of reproductive aging still is evolving and does not yet have any real clinical utility because it cannot predict the reproductive lifespan for an individual woman.[52, 60]

Endocrinology of Reproductive Aging

As the pace of follicular depletion increases during the latter reproductive years, but before any discernible change in menstrual regularity, serum follicle-stimulating hormone (FSH) levels begin to rise; luteinizing hormone (LH) concentrations remain unchanged. The subtle "monotropic" rise in circulating FSH concentrations is most apparent during the intercycle transition, when the corpus luteum regresses and menses begins, and could result from age-related changes in the pattern of pulsatile gonadotropin-releasing hormone (GnRH) secretion, or from progressive follicular depletion and lower levels of feedback inhibition from ovarian hormones. The weight of available evidence supports the second explanation.[64, 65]

A variety of studies in animals and women have identified changes in the patterns of hypothalamic-pituitary hormone secretion across the menopausal transition. In rodents, an age-related decrease in pulsatile GnRH and LH secretion and a loss of positive estrogen feedback have been observed, before the follicular pool is exhausted.[66–69] In nonhuman primates, pulsatile GnRH release increases during the perimenopause and the positive feedback response remains intact.[70] Studies in perimenopausal and postmenopausal women have yielded conflicting results. Whereas some have observed changes in sensitivity to estrogen feedback signals[71, 72] or in LH pulse amplitude or frequency,[73–78] others have not.[79–81] The response to exogenous GnRH stimulation also is inconsistent.[77, 82, 83] On balance, these data suggest strongly that age-related changes in pulsatile LH secretion and gonadotropin concentrations merely reflect changes in ovarian feedback signals and do not result from aging of the hypothalamic-pituitary axis.

The bulk of available evidence indicates that the progressive increase in FSH concentrations associated with reproductive aging results from a progressive decrease in the levels of feedback inhibition from the smaller cohorts of follicles recruited from a shrinking follicular pool. Circulating follicular phase inhibin B levels (derived primarily from smaller antral follicles) decrease as or even before FSH concentrations begin to increase.[64, 84–91] Inhibin A levels also decline, but only in the later stages of reproductive aging, after the onset of menstrual irregularity.[88, 92–95] Both inhibins selectively inhibit pituitary FSH secretion. Consequently, FSH levels rise progressively as inhibin production from smaller cohorts of aging follicles decreases, most noticeably in the early follicular phase. Whereas declining inhibin production also could reflect a decrease in the functional capacity of older follicles,[96] the observation that preovulatory follicular fluid inhibin concentrations are similar in young and older cycling women suggests that the number of remaining follicles is more important.[84] Ovarian steroid hormones do not play a major role. The initial rise in FSH levels precedes any measurable decrease in estradiol levels, by several years.[65, 97] Follicular phase estradiol levels in older cycling women generally are similar to those in younger women, and often even higher.[84, 98] Luteal phase estrogen and progesterone levels also do not seem to change consistently with advancing age.[64, 86, 88, 99–102] Moreover, in sporadic ovulatory cycles in aging women, serum concentrations of estradiol and progesterone are comparable to those observed in younger women.[103]

As age and FSH levels increase, the follicular phase becomes shorter;[104–106] LH levels and luteal phase duration remain unchanged. As the follicular phase shortens, estradiol levels rise earlier, suggesting that higher FSH levels stimulate more rapid follicular development.[64] ***However, careful studies have shown that the earlier rise in estradiol levels results not from accelerated follicle growth, but from advanced follicular development at the beginning of the cycle and earlier selection of the dominant follicle***.[99, 105, 107] The earlier increase in follicular phase FSH level also frequently results in more than one dominant follicle,[108–110] explaining the higher prevalence of dizygotic twinning in older cycling women.[99, 108, 111]

Reproductive aging already is quite advanced when the first clinical sign appears. Cycles remain regular, but overall cycle length and variability decrease gradually, reaching a nadir at an average age of 42 years,[104, 112] when fertility is at or near an end. However, women generally take notice only when cycles become irregular, marking the beginning of the menopausal transition.[113] The menopausal transition begins at an average age of 46 years, but can arrive as early as age 34 and as late as age 54 years.[104, 112, 114–116] Thereafter, average cycle length and variability increase steadily as ovulations become less regular and frequent.[112] Regardless of age, the interval from loss of menstrual regularity to menopause is relative fixed, spanning approximately 5–6 years.[47, 117, 118] The age of menopause, recognized only in retrospect, averages 51 years, but ranges widely, between ages 40 and 60 years.[116, 119–124] The variation in menopausal age is very similar across populations and generally follows a normal distribution that is slightly skewed to younger ages.[124–126]

Genetics of Reproductive Aging

Barring any disease that destroys or causes the removal of ovarian tissue and any important environmental insults, the total number of follicles at birth, and the age when the supply is exhausted, are genetically determined.[47, 127–135]

There is good correlation between menopausal age in mothers and daughters and between sisters, suggesting that genetic factors play an important role in determining menopausal age.[136–138] ***Approximately 10% of women become menopausal by the age of 45,[116, 128] probably because they were endowed with a smaller than average ovarian follicular pool that is functionally depleted at an earlier age.*** Pedigree analysis has revealed that the genetic features of early menopause (age 40–45) and premature ovarian failure (POF) are similar, suggesting a dominant pattern of inheritance through maternal or paternal relatives.[139, 140] The same genetic factors that determine the age at menopause also likely determine the age of reproductive milestones preceding the menopause.[141] In natural populations, age at last birth varies as widely as the age at menopause, but occurs on average 10 years earlier.[47] Moreover, women who repeatedly respond poorly to exogenous gonadotropin stimulation also tend to have an earlier menopausal transition,[141–144] suggesting their poor response reflects an advanced stage of follicular depletion, beginning years sooner than would be anticipated normally.[141] Conversely, fertility in women destined for a later than average menopause may not decrease significantly until after age 40.

Genes affecting reproductive hormones (*FSH, FSHR, LH, LHR, CYP17, CYP19*) or involved in the initial growth of primordial follicles (*BMP15, GDF9, GPR3*) impact follicular function; mutations are rare in humans, but polymorphisms could influence the rate of follicular recruitment and depletion and thereby affect the length of reproductive life.[145] Variations in other genes encoding DNA binding proteins and transcription factors (*NOBOX, LHX8*) and RNA binding proteins (*NANOS*) expressed during oogenesis could affect germ cell formation; mutations causing POF have been identified in a few women.[146] Variations in other genes with links to POF also might affect the rate of follicular depletion in normal women (*ADAMts9, FOXL2*).[147, 148] In a Dutch cohort study, common polymorphisms in the gene encoding the receptor for antimüllerian hormone (*AMHR2*) were associated with menopausal age,[149] implicating a decrease in AMH signaling that would weaken its paracrine inhibition of primordial follicle recruitment, leading to more rapid follicular depletion. Careful examinations of these and other candidate genes identified in genome-wide association studies likely will yield new insights and further our understanding of the mechanisms that govern reproductive aging.

The Aging Follicle and Oocyte

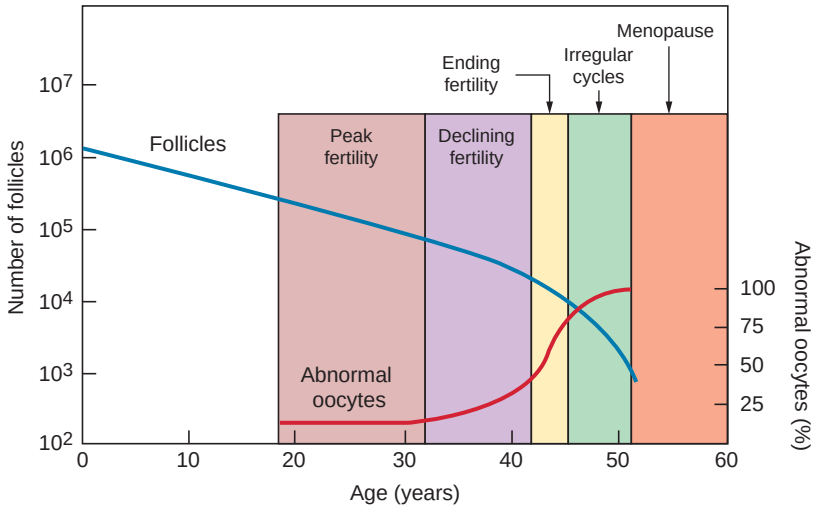
Whereas the number of remaining ovarian follicles steadily declines with increasing age, observations in stimulated cycles suggest that aging follicles also become progressively less sensitive to gonadotropin stimulation. As age increases, the total dose and duration of treatment required to stimulate multiple follicular development increase. The rate of rise and the peak in estradiol levels decrease, reflecting the smaller cohorts of follicles that can be recruited. However, the amount of estradiol secreted by the follicles that do emerge and grow to maturity appears comparable to that in younger women.[150] Although a decrease in exogenous hCG-induced ovarian androgen production can be demonstrated before the age of 30, circulating estradiol levels remain normal throughout and beyond the reproductive years, probably because rising FSH levels are able to compensate.[151] Studies of ovarian follicular development and preovulatory follicular fluid hormones in older and younger cycling women do not suggest any age-related decline in follicular function, once growth and development begins. Preovulatory follicles in older and younger women are similar in size and inhibin content, and follicular fluid progesterone levels and estrogen/androgen ratios are even higher in older than in younger women.[84]

Older cycling women ovulate as regularly and more frequently than younger women. Their rising FSH levels apparently compensate quite effectively for any decrease in follicular sensitivity to gonadotropin stimulation. Preovulatory follicles in older cycling women get an earlier start, but grow at a normal pace and reach a normal size; their follicular fluid characteristics suggest they also are quite healthy. Why then does fertility in women decline progressively with age? ***The available evidence indicates that both the age-related decline in female fertility and the increase in risk of miscarriage can be attributed to an increase in the proportion of abnormal oocytes in an aging and shrinking follicular pool.***

As the number of follicles decreases, oocyte quality also declines (at least by age 31–32 years), primarily because of an increase in meiotic nondisjunction, resulting in an increasing rate of oocyte and embryo aneuploidy in aging women.[50, 152–154] A wide variety of techniques has been used to study the chromosomal composition of human oocytes. The best available evidence, derived from detailed cytogenetic analysis of oocytes retrieved for IVF that failed to fertilize, suggests that the global rate of oocyte aneuploidy increases with advancing maternal age.[155, 156] Oocyte aneuploidy results primarily from premature separation of sister chromatids during meiosis I (resulting in a single chromatid in place of or in addition to one or more whole chromosomes), or from whole chromosome nondisjunction during meiosis II.[156] The prevalence of both types of meiotic segregation errors increases progressively with age, but single chromatid events make the greatest contribution to the age-dependent increase in the prevalence of oocyte aneuploidy.[155–159]

The age-related decrease in the proportion of normal oocytes (23,X) and the corresponding increase in the proportion of aneuploid oocytes bear striking similarity to the age-related decrease in fertility and increase in the incidence of spontaneous miscarriage in women. Fertility and the prevalence of euploid oocytes decrease progressively with age. ***Miscarriage risk and the prevalence of aneuploid oocytes are relatively low and change little until approximately age 35 (about 10%), then increase progressively, reaching nearly 30% at age 40, 50% by age 43, and virtually 100% after age 45.***[155] These observations offer a logical explanation for the age-related increase in the prevalence of aneuploidy in spontaneous abortuses. Whereas at least half of all clinically recognized miscarriages exhibit an abnormal karyotype and the frequency of both euploid (normal) and aneuploid (abnormal) abortuses increases with maternal age, the probability that an abortus will be chromosomally abnormal increases with age, from less than 35% at age 20 to nearly 80% over age 42.[36] Trisomies are by far the most common abnormality observed, followed by polyploidies and monosomy X (45,X).

Studies of meiotic segregation have revealed that factors predisposing to nondisjunction relate to the disruption of chromosomal pairing and recombination.[160, 161] Various mechanisms have been implicated, but all involve an age-dependent deterioration in cellular factors required for proper spindle formation and function.[162] Molecular investigations of chromatid cohesion and separation have implicated cohesins, a specific class of proteins that maintain cohesion between sister chromatids and oppose the splitting forces mediated by the microtubules of the meiotic spindle.[163–166] An age-related premature degradation or deficiency of cohesins may result in unstable bivalent chromatid structures and predispose to premature separation of sister chromatids before they align on the meiotic spindle. The smaller chromosomes appear more prone to premature chromatid separation, possibly because they have fewer of the chiasma that help to prevent such dissociation.[157, 167, 168] Other studies using high-resolution confocal microscopy to examine the meiotic spindle in human oocytes have revealed that abnormalities of the cleavage spindle microtubular matrix or chromosome alignment during meiosis II are four to five times more common in older cycling women (age 40–45) than in younger women (age 20–25).[50] These and other observations of cultured human oocytes collected from unstimulated ovaries further indicate that the meiotic competence of oocytes declines with age.[169] ***In sum, accumulated evidence strongly suggests that the primary cause of the age-dependent decrease in fecundability and increase in the incidence of miscarriage is an increasing prevalence of aneuploidy in aging oocytes resulting from disordered regulatory mechanisms governing meiotic spindle formation and function.***



Adapted from[17, 156]

Aging and the Uterus

Aging does not appear to have any significant adverse effect on the uterus. Although the prevalence of benign uterine pathology (leiomyomata, endometrial polyps, adenomyosis) increases with age,[170–172] little evidence exists to indicate it has much overall impact on fertility in women.[173–176] Age also does not appear to adversely affect endometrial development or function in response to steroid stimulation.[177] The strongest evidence comes from comparing outcomes in nondonor and donor oocyte IVF cycles. Whereas early studies suggested that donor oocyte IVF pregnancy and delivery rates decreased modestly with the age of the recipient,[178–180] the bulk of more recent experience refutes those conclusions.[34, 181, 182]

In the national summary of ART success rates for the year 2007, live birth rates declined progressively with increasing age for nondonor egg cycles, as expected. In contrast, the overall live birth per transfer rate in donor egg IVF cycles was 55% and did not vary significantly with age of the recipient.[34] ***Live birth rates in donor egg IVF cycles relate to the age of the donor, not the age of the recipient.*** In one large series, miscarriage rates increased from 14% in women matched with egg donors ages 20–24 to 44% for women whose donors were over age 35.[183]

Aging and Male Fertility

The relationship between age and fertility in men is discussed in detail in Chapter 30 and summarized here. Modest age-related decreases in semen volume, sperm motility, and morphologically normal sperm, but not sperm density, have been observed.[184] Semen characteristics generally do not accurately predict fertilizing capacity;[185–188] neither do endocrine parameters.[189, 190] In studies of the effect of male partner age on pregnancy rates, female partner age and declining coital frequency with increasing age are obvious and important confounding factors. Among the few studies that have controlled for female age, pregnancy rates for men over 50 have been 23–38% lower than for men under age 30.[184] A British study that examined the effect of men's age on the time to conception (adjusting for the confounding effects of both partner's age and coital frequency) found that increasing men's age was associated with increasing time to conception and declining overall pregnancy rates; time to conception was 5-fold greater for men over age 45 than for men under age 25, and restricting the analysis to men with young partners yielded similar results.[191] Results of two studies that controlled for female partner age have suggested that male fertility may start to decline earlier, beginning in the late 30s.[192, 193]

There are several possible biological mechanisms that might explain an age-related decline in male fertility. Sperm chromosomal abnormalities may increase in frequency with age and adversely affect early embryonic development.[194] There is at least some evidence to suggest that increasing male age may raise the risk of miscarriage in young women.[195] Average FSH levels in men increase during their 30s,[196] suggesting that age-related changes in the hypothalamic-pituitary-gonadal axis may begin during midlife.[197] The testes and prostate also exhibit morphological changes with aging that might adversely affect both sperm production and the biochemical properties of semen.[198] Whatever the mechanism, decreasing fertility with increasing male age in healthy couples suggests that normal sperm overproduction may not fully buffer the effects of increasing age.

On balance, the available evidence indicates that pregnancy rates decrease and time to conception increases as male age increases. However, because there is little or no overall vmeasurable decline in male fertility before age 45–50, male factors generally contribute relatively little to the overall age-related decline in fertility.

Ovarian Reserve Tests

Over the past 20 years, studies of the mechanisms involved in reproductive aging and its clinical consequences have stimulated efforts to measure "ovarian reserve," generally describing the size and quality of the remaining ovarian follicular pool. A number of methods have now been described, all intended to predict fertility or to provide prognostic information regarding the likelihood of successful treatment in infertile women, recognizing that although the number and quality of oocytes decline with age, fertility varies significantly among women of similar age.

Like all screening tests, ovarian reserve tests are aimed at identifying individuals at risk for a disease, in this case a "diminished ovarian reserve" (DOR). ***It is important to emphasize that such tests cannot and do not establish a diagnosis of DOR; they only identify women more likely to exhibit a poor response to gonadotropin stimulation and to have a lower likelihood of achieving pregnancy with treatment.*** The value of a screening test depends on its validity, describing its ability to correctly categorize individuals as affected (sensitivity) or unaffected (specificity). The sensitivity and specificity of a screening test will vary with the chosen threshold value. A choice intended to maximize sensitivity minimizes the number of false-negative results (patients with DOR categorized as normal), but increases the number of false-positive results (patients with a normal ovarian reserve categorized as having DOR). Conversely, a threshold value that maximizes specificity minimizes false-positives, but increases false-negative results. ***For measures of ovarian reserve, test threshold values should have high specificity for DOR, so as to decrease false-positive results (incorrectly categorizing a patient with a normal ovarian reserve as having DOR), thereby avoiding overly aggressive treatment or inappropriate recommendations to abandon treatment or pursue adoption or oocyte donation in women with a normal ovarian reserve. Treating women with unrecognized DOR (the consequence of maximizing specificity) is undesirable, but a less serious error.***

The most important test characteristics of a screening test are its positive predictive value (PPV) and negative predictive value (NPV), which vary with the prevalence of the disease of interest (DOR) in the test population. PPV describes the probability that a woman with a positive test truly has DOR, and NPV is the probability that a woman with a negative test truly has a normal ovarian reserve. If the prevalence of DOR is low, as in young women, the PPV will be low, even if sensitivity and specificity are high. Conversely, if the prevalence of DOR is high, as in older women, the PPV will be high if a highly specific threshold value is chosen. ***If the purpose of ovarian reserve testing is to correctly identify women with DOR, it will be most useful in women at high risk for DOR. When applied in a low prevalence population, many women with a normal ovarian reserve will have a false-positive result and be categorized as having DOR.***

Ovarian reserve tests include both biochemical and ultrasonographic measures of the size and (by inference) the quality of the ovarian follicular pool. Biochemical tests include both basal measurements, such as FSH, estradiol, inhibin B, and antimüllerian hormone (AMH), and provocative tests, such as the clomiphene citrate challenge test. Ultrasonographic measures of ovarian reserve include the antral follicle count and ovarian volume. The clinical utility of any test of ovarian reserve is most easily and efficiently evaluated by examining the relationship between test results and IVF cycle characteristics and outcomes. Considering the costs, logistics, and risks involved with IVF, and the importance of accurate prognostication in counseling candidate couples, correlation with IVF outcome is arguably also the most clinically relevant measure.

Basal FSH and Estradiol Concentrations

Given that rising FSH levels are one of the earliest indications of reproductive aging in women, it was logical to think that the serum FSH concentration might serve as a useful ovarian reserve test. The basal FSH concentration is the simplest and still most widely applied measure of ovarian reserve.

Because serum FSH concentrations vary significantly across the cycle, the serum FSH concentration is best obtained during the early follicular phase (cycle day 2–4). FSH values

vary with the assay method; although values obtained with different assays correlate very well, absolute values can differ significantly. Values also vary with the reference standard, previously an international reference preparation of human menopausal gondotropin (IRP-HMG), and now the World Health Organization Second International Reference Preparation (IRP 78/549).

Numerous studies have investigated the relationship between cycle day 3 FSH concentrations or FSH/LH ratios and IVF cycle outcomes, all observing that these measures correlate with the ovarian response to exogenous gonadotropin stimulation and, to a lesser extent, with the likelihood for success. As values increase, peak estradiol levels, the number of oocytes retrieved, and the probability for pregnancy or live birth steadily decline.[199–205] ***With current assays (using IRP 78/549), FSH levels greater than 10 IU/L (10–20 IU/L) have high specificity (80–100%) for predicting poor response to stimulation, but their sensitivity for identifying such women is generally low (10–30%) and decreases with the threshold value.***[206] Although most women who are tested (including those with DOR) will have a normal result, the test is still useful because those with abnormal results are very likely to have DOR. In a 2008 study, an FSH concentration above 18 IU/L had 100% specificity for failure to achieve a live birth.[207]

Because FSH levels can vary significantly, many clinicians prefer to repeat the test. Not surprisingly, consistently high values are associated with a poor prognosis, but a single elevated FSH concentration (>10 IU/L) does not have high specificity for predicting poor response to stimulation or failure to achieve pregnancy.[208] Serial testing in efforts to select the ideal cycle for treatment does not improve outcomes in women with fluctuating FSH concentrations.[209, 210]

The basal serum estradiol concentration, by itself, has little value as an ovarian reserve test,[211–214] ***but can provide additional information that helps in the interpretation of the basal FSH level.*** An early elevation in serum estradiol reflects advanced follicular development and early selection of a dominant follicle (as classically observed in women with advanced reproductive aging), and will suppress FSH concentrations, thereby possibly masking an otherwise obviously high FSH level indicating DOR. When the basal FSH is normal and the estradiol concentration is elevated (>60–80 pg/mL), the likelihood of poor response to stimulation is increased and the chance for pregnancy is decreased.[215–218] When both FSH and estradiol are elevated, ovarian response to stimulation is likely to be very poor.

Clomiphene Citrate Challenge Test

The clomiphene citrate challenge test (CCCT) is a provocative and possibly more sensitive test of ovarian reserve that probes the endocrine dynamics of the cycle under both basal and stimulated conditions, before (cycle day 3 FSH and estradiol) and after (cycle day 10 FSH) treatment with clomiphene citrate (100 mg/d, cycle days 5–9).[219]

The smaller follicular cohorts in aging women produce less inhibin B and estradiol, resulting in less negative feedback inhibition on clomiphene-induced pituitary FSH release, causing an exaggerated increase in FSH concentrations.[85, 220] Consequently, a frankly elevated cycle day 10 FSH concentration can identify women with DOR who might otherwise go unrecognized if evaluated with basal cycle day 3 FSH and estradiol levels alone.[221, 222]

In studies evaluating CCCT results, stimulated concentrations of FSH, estradiol, and inhibin B have varied widely, limiting the value of the test.[223–225] A 2006 systematic review

of the predictive value of the CCCT over a range of day 10 FSH concentrations (10–22 IU/L) in women at low, average, and high probability of DOR concluded the test had 47–98% specificity and 35–93% sensitivity for predicting poor response to stimulation, and 67–100% specificity and 13–66% sensitivity for predicting treatment failure.[226] ***Overall, stimulated FSH levels have higher sensitivity but lower specificity than the basal FSH concentration.***[226]

Inhibin B

Inhibin B is secreted primarily during the follicular phase by the granulosa cells of smaller antral follicles, and might therefore be expected to have some value as an ovarian reserve test.[227] However, serum inhibin B concentrations increase in response to exogenous GnRH or FSH stimulation and vary widely across and between menstrual cycles.[213, 228] ***Inhibin B is generally not regarded as a reliable measure of ovarian reserve.***

Although inhibin B levels are generally lower in women who respond poorly to exogenous gonadotropin stimulation than in those who respond normally,[229, 230] even low threshold values (40–45 pg/mL) have only 64–90% specificity and 40–80% sensitivity for predicting poor response. Inhibin B has a relatively low PPV (19–22%) but a relatively high NPV for detecting DOR in a general IVF population;[228, 231] in a high prevalence population, the PPV of inhibin B can exceed 80%.[213] In most studies, inhibin B has had poor PPV for failed treatment.[212, 213, 227, 232, 233]

Antimüllerian Hormone

Antimüllerian hormone (AMH) is produced by the granulosa cells of preantal and small antral follicles, beginning when primordial follicles start development and ending when they reach a diameter of 2–6 mm.[234–237] Small antral follicles are likely the primary source because they contain larger numbers of granulosa cells and a more developed microvasculature.[238, 239] Although it functions primarily as an autocrine and paracrine regulator of follicle development, AMH appears in measurable amounts in the serum.[240] The number of small antral follicles correlates with the size of the residual follicular pool and AMH levels decline progressively, becoming undetectable near the menopause.[241–244]

Because AMH derives from preantral and small antral follicles, levels are gonadotropin-independent and exhibit little variation within and between cycles.[245–247] In clinical studies, AMH has been assayed using two different commercial assay kits, and although the results they yield are highly correlated, their standard curves are not parallel and there is no applicable conversion factor; one comparative study observed that concentrations measured with one kit were more than 4-fold lower than those measured with the other.[248] Consequently, when applying results in clinical practice, it is important to know which assay method was used to measure AMH. Commercial assay kits yield consistent results with low interassay variation (<10%).[249]

The performance of AMH as a screening test of ovarian reserve has been examined in the general IVF population and in populations of women at low or high risk for DOR. Overall, lower AMH levels have been associated with poor response to ovarian stimulation and low oocyte yield, embryo quality, and pregnancy rates,[228, 229, 250–252] but studies correlating mean AMH levels with IVF outcomes have not yielded threshold values that can be applied confidently in clinical care.[211, 229, 231, 250] ***In the general IVF population, low AMH threshold values (0.2–0.7 ng/mL) have had 40–97% sensitivity, 78–92% specificity,***

22–88% PPV and 97–100% NPV for predicting poor response to stimulation (<3 follicles, or <2–4 oocytes), but have proven neither sensitive nor specific for predicting pregnancy.[228, 253–255] In women at low risk for DOR, values of 2.5–2.7 ng/mL have had 83% sensitivity, 82% specificity, 67–77% PPV, and 61–87% NPV for clinical pregnancy.[212, 256] The higher threshold values decrease specificity, resulting in lower PPV because the prevalence of DOR was low. A study in women at high risk for DOR (involving older women, those with an elevated FSH, or history of poor response to stimulation) observed that an undetectable AMH had 76% sensitivity, 88% specificity, 68% PPV, and 92% NPV for three or fewer follicles.[229] A higher threshold value (1.25 ng/mL) had 85% sensitivity, 63% specificity, 41% PPV, and 57% NPV for cycle cancellation.[213]

AMH is a very promising screening test for DOR, but is likely to be more useful in a general IVF population or in women at high risk for DOR than in women at low risk for DOR. Low threshold values have good specificity for poor response to ovarian stimulation, but not for predicting pregnancy.

Antral Follicle Count

Reproductive aged women have an estimated 20–150 growing follicles in the ovaries at any one time, although only a few are large enough to be imaged (≥2 mm) by transvaginal ultrasonography.[257–259] Follicles of that size have reached a stage of development where they are responsive to FSH, which stimulates and supports more advanced stages of development. ***Histologic studies have revealed that the number of small antral follicles in the ovaries is proportional to the number of primordial follicles remaining.[260] Therefore, as the supply of primordial follicles decreases, the number of visible small antral follicles also declines.*** The antral follicle count (AFC; total number of antral follicles measuring 2–10 mm in both ovaries) thus provides an indirect but useful measure of ovarian reserve.[258, 261–264]

AFC correlates with onset of the menopausal transition, indicating that it relates to the number of follicles remaining.[242] Some, perhaps as much as half, of the antral follicles that can be imaged are probably in the process of atresia, but there is no way other than observing their response to FSH stimulation to distinguish them from viable growing follicles.[17] However, AFC correlates well with oocyte yield in IVF cycles,[265] suggesting that gonadotroin stimulation can still rescue follicles that may be in the early stages of atresia.[266] Several studies have observed a relationship between the AFC and response to ovarian stimulation in IVF cycles. In the general IVF population, including women at low and high risk for DOR, an AFC threshold value of three to four follicles has high specificity (73–100%) for predicting poor response to ovarian stimulation and failure to conceive (64–100%), but relatively low sensitivity for both endpoints (9–73% for poor response, 8–33% for failure to conceive).[213, 265, 267–272] The PPV and NPV of AFC have varied widely in studies.

A low AFC has high specificity for predicting poor response to ovarian stimulation and treatment failure, making it a useful test, but low sensitivity limits its overall clinical utility.

Ovarian Volume

Not surprisingly, ovarian volume decreases with progressive follicular depletion.[273, 274] However, the measure has high inter-cycle and inter-observer variability,[213, 275–277] and because most studies of ovarian volume have excluded women with ovarian pathology such as endometriomas and polycystic ovary syndrome, results have limited generalizability.[274, 278]

Ovarian volume (length × width × depth × 0.52=volume) generally correlates with the number of oocytes retrieved, but poorly with pregnancy.[267, 272, 279–281] A low ovarian volume (< 3mL) has high specificity (80–90%) and widely ranging sensitivity (11–80%) for predicting poor response to ovarian stimulation.[206] The PPV for poor response can be as low as 17% among women at low risk for DOR, and as high as 53% in women at high risk.[213] ***Overall, ovarian volume has very limited clinical utility as an ovarian reserve test.***

Other Tests of Ovarian Reserve

Numerous other provocative tests of ovarian reserve have been investigated, including exogenous FSH-stimulated estradiol, inhibin B or AMH levels[250, 282–286] and GnRH agonist-stimulated FSH, estradiol, inhibin B, or AMH concentrations.[250, 282, 287–289] In theory, the ovarian and endocrine response to FSH or GnRH agonist stimulation should provide the best estimate of the number of responsive follicles. ***However, a 2006 systematic review found no evidence that these more complex and costly tests predict response to ovarian stimulation or pregnancy any better than basal FSH, AMH, and AFC.***[206]

Combined Tests of Ovarian Reserve

Recognizing that no one test of ovarian reserve has 100% sensitivity and specificity, a number of investigators have examined the performance of varying combinations of ovarian reserve tests. Analysis is difficult, primarily because of differences in chosen threshold values for specific tests. Moreover, because the different tests are highly correlated, using more than one measure in a prediction model does not necessarily improve its performance.[213, 230, 267] Complicated formulas also are generally not useful in clinical practice. One analysis combining AMH, inhbin B, AFC, and ovarian volume found that only AFC and AMH predicted response to stimulation and that the combination predicted outcome no better than the individual tests.[275] A meta-analysis of cohort studies investigating the performance of various combinations of tests concluded that models combining tests do not perform significantly better than individual tests such as the AFC.[290]

SUMMARY

Currently, there is no uniformly accepted definition of diminished ovarian reserve. A number of different measures have been developed, primarily for use in predicting success with IVF. The ideal ovarian reserve test should yield consistent results and be highly specific, to minimize the risk for incorrectly categorizing normal women as having a diminished ovarian reserve. Basal FSH is the most commonly used ovarian reserve test, but antral follicle count and antimüllerian hormone are promising predictors with significant potential advantages.

Ovarian reserve tests predict response to exogenous gonadotropin stimulation reasonably well, but whether the information gained truly affects outcomes is less certain. ***Although the planned amount of gonadotropin stimulation often is increased in predicted poor responders, those adjustments do not improve response predictably, probably because the small cohort of responsive antral follicles is the limiting factor and no amount of stimulation can increase that number appreciably.***[291–293] Even in women

who previously exhibited a poor response to stimulation, changes in treatment regimens generally have not improved response or pregnancy rates in subsequent cycles.[292, 294–296]

None of the ovarian reserve tests currently in use is an accurate predictor of pregnancy in IVF cycles, unless extreme abnormal threshold values are applied, which results in very low sensitivity for identifying women having a poor prognosis.[207] The tests are adequate for predicting poor response, which does have prognostic value, although not as much in young women as in older women.[297–299] Although ovarian reserve tests have become a routine element of pre-treatment evaluation for couples planning IVF, it can be argued that routine testing has limited clinical utility in the large majority of patients and can be misleading, especially in women at low risk for having a diminished ovarian reserve.[17]

Ovarian reserve tests also have become a routine element of the diagnostic evaluation for infertility. Advocates for the liberal application of ovarian reserve tests argue that abnormal tests can help to persuade older women to abandon plans to pursue aggressive, costly, and likely futile treatment, and can help to convince young women to do just the opposite, to take fullest advantage of a rapidly closing window of opportunity. Others more circumspect emphasize correctly that few young women will have an abnormal test, and some of those who do inevitably will be categorized incorrectly, leading to inappropriate counseling and treatment. ***The best overall strategy would seem to limit ovarian reserve testing to women at increased risk for having a diminished ovarian reserve and to apply highly specific threshold values to minimize the risk for a false-positive result.*** In this context, ovarian reserve testing can best be justified for women with any of the following characteristics:[141, 300–303]

- Age over 35.
- Unexplained infertility.
- Family history of early menopause.
- Previous ovarian surgery (ovarian cystectomy or drilling, unilateral oophorectomy), chemotherapy, or radiation.
- Smoking.
- Demonstrated poor response to exogenous gonadotropin stimulation.

Ovarian reserve tests always should be interpreted with caution. Rigid application of test results risks inappropriate recommendations for treatment, or for no treatment, and both must be avoided. An abnormal test result does not preclude the possibility of pregnancy. Except perhaps when grossly abnormal, test results should not be used to deny treatment, but only to obtain prognostic information that may help to guide the choice of treatment and best use of available resources. Although the probability of pregnancy may be low, many with abnormal test results will achieve pregnancy if afforded the chance. Ultimately, regardless of the prognosis, the success rate for any individual woman will be 0% or 100%.

Guiding Principles for Evaluation and Treatment of Infertility

From the beginning, the evaluation of infertility should focus on the *couple* and not on one or the other partner, regardless of past reproductive performance. Both partners should be encouraged to attend each visit during evaluation, whenever possible. Each can provide

information and perspective the other may not have or remember. Joint visits also help to ensure that both partners understand any information, options, and recommendations that may be offered and that each has the opportunity to have their questions addressed directly.

Clinicians caring for infertile couples should keep four basic goals in mind:

- To identify and to correct specific causes of infertility, when possible. With proper evaluation and treatment, the majority of women will achieve pregnancy.
- To provide accurate information and to dispel the misinformation commonly gained from friends and mass media.
- To provide emotional support during a trying time. In many couples, the inability to conceive results in feelings they have lost control over an important and very personal part of their lives, and the process of evaluation adds to that burden. Infertile couples often need the opportunity to express their concerns, frustrations, and fears, and support groups can help to meet that need. Group meetings can help couples to realize that their problem is not unique and to learn how others cope with similar problems. Whereas severe anxieties can have adverse effects on ovulatory function and coital frequency, there is no substantial evidence that the usual anxieties of couples trying to conceive cause or contribute to their infertility.
- To guide couples failing to conceive with other forms of treatment to alternatives, including IVF, the use of donor gametes (oocytes or sperm), and adoption, and to help those who reject or fail treatment to come to closure.

Counseling must be an ongoing process during both evaluation and treatment. Regular visits to review and critique results and to outline recommendations for further evaluation and treatment help to ensure that all of the couple's medical, emotional, and financial needs and concerns are addressed effectively in a timely fashion.

Lifestyle and Environmental Factors

Understandably, all infertile couples are very interested in learning anything they might do to maximize the likelihood of achieving a successful pregnancy. Lifestyle choices and environmental factors influence fertility and deserve consideration and discussion when they are relevant. Over 35% of American women are obese and another 30% are overweight.[304] Obesity is defined as a body mass index (BMI) greater than 30 kg/m^2 and overweight is defined as a BMI between 25 kg/m^2 and 30 kg/m^2. In women, obesity is associated with menstrual dysfunction, decreased fertility, and increased risks of miscarriage and obstetric and neonatal complications. In men, obesity is associated with abnormal semen parameters and can adversely affect fertility.[305]

Substance abuse is one of the few things over which the couple may have specific control, smoking being the most important. Many are not aware of the adverse effects smoking has on fertility and pregnancy outcome.[306] The couple's motivation to maximize their fertility presents a golden opportunity to educate those who smoke and to establish a smoking cessation strategy. Smoking has well-known adverse impact on pregnancy outcome, and evidence strongly suggests that fertility is lower in both men and women who smoke.[307–311] The prevalence of infertility is higher, fecundability is lower, and the time to conception is longer in smoking than in non-smoking women, and the effects of passive smoke exposure are only slightly less than those of active smoking by either partner.[312] The available data suggest that the adverse effects of smoking on fertility are dose-dependent.[308, 313–315] The mechanisms involved ay include accelerated follicular

depletion,[316–318] menstrual cycle abnormalities,[319] or gamete or embryo mutagenesis induced by toxins in cigarette smoke.[320–324] A causal relationship between cigarette smoking and female infertility has not been established. However, based on the results of a meta-analysis including 12 studies (overall OR for risk of infertility in women smokers versus non-smokers 1.60), and assuming a 25% prevalence of smoking in women of reproductive age, up to 13% of female infertility may relate to smoking.[310] Consequently, an active approach to prevention of infertility is justified, discouraging smoking and helping those who smoke to quit.[325]

Other forms of substance abuse also can adversely affect fertility. Marijuana inhibits the secretion of GnRH and can suppress reproductive function in both women and men.[326] In women, marijuana use can interfere with ovulatory function.[327] Cocaine use can impair spermatogenesis in men[326, 328] and has been associated with a greatly increased risk of tubal disease in women.[327] Heavy alcohol consumption in women may decrease fertility;[329–331] in men, it has been associated with decreased semen quality and impotence.[332] Conflicting evidence suggests that moderate alcohol intake can reduce fecundability.[333, 334] In both women and men, even modest amounts of alcohol consumption have been associated with lower pregnancy rates in IVF cycles.[335] Although moderate caffeine ingestion (≤250 mg daily; two standard beverages) appears not to have any adverse effects on fertility, higher levels of consumption may delay conception[311, 336, 337] or increase the risk of pregnancy loss.[338]

Other potentially harmful occupational and environmental exposures, although uncommon, may be identified. Exposures to perchlorethylene in the dry cleaning industry, toluene in the printing business, ethylene oxide, and mixed solvents have been associated with decreased fecundity. Semen abnormalities have been described in men exposed to radiant heat or heavy metals. Environmental exposure to herbicides or fungicides has been associated with decreased fertility in women,[331] and exposure to pesticides and other chlorinated hydrocarbons with an increased risk of miscarriage.[339]

For couples attempting to conceive, there is fair evidence to support recommendations for smoking cessation and efforts to achieve a BMI between 20 and 25 kg/m^2. Recommendations to limit alcohol consumption to four or fewer drinks per week and to limit caffeine intake to less than 250 mg/d also are reasonable and consistent with available evidence. However, there have been no randomized controlled trials demonstrating that such lifestyle modifications improve fertility.

Normal Reproductive Efficiency

As evaluation begins, and again before treatment starts, education on normal human reproductive efficiency can help to provide important perspective for infertile couples. Few realize that, compared to other mammals and even nonhuman primates, humans are not highly fertile. In captive baboons, cycle fecundity ranges as high as 80% when conditions and timing are optimized.[340] ***In normally fertile couples, cycle fecundity averages 20% and does not exceed approximately 35% even when coitus is carefully timed.***[40, 341, 193] That perspective is particularly helpful when discussing and comparing the efficacy of different treatment options, typically viewed in terms of cycle fecundability. When doing so, it is important for couples to realize that the benchmark for comparison is 20–30%, and not 100%.

Given the average 20% cycle fecundability, the cumulative pregnancy rates observed over time in normal fertile couples are easy to understand. The data in the table below have been a standard since 1956, and have been confirmed by more recent studies.[3, 342, 343]

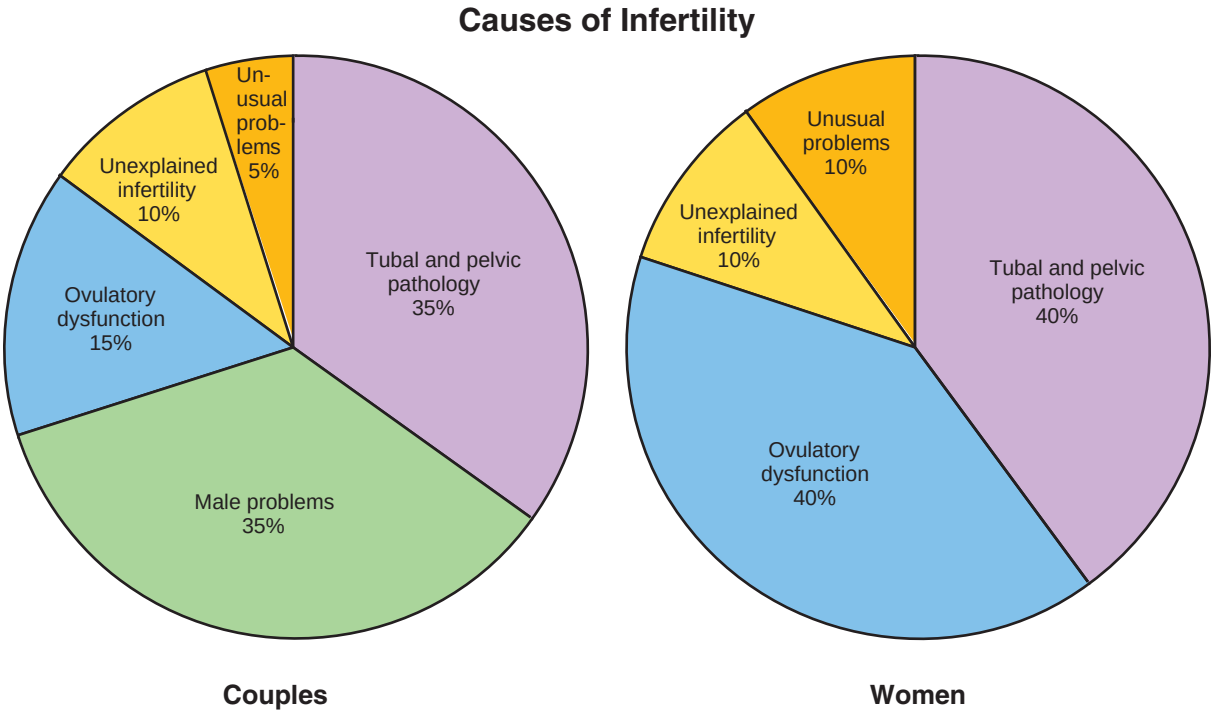
Time Required for Conception Among Couples Who Will Attain Pregnancy[342]	
Months of Exposure	*% Pregnant*
3 months	57%
6 months	72%
1 year	85%
2 years	93%

Normal sperm can survive in the female reproductive tract and retain the ability to fertilize an egg for at least 3 and up to 5 days, but an oocyte can be fertilized successfully for only approximately 12–24 hours after ovulation.[344] Consequently, virtually all pregnancies result from intercourse occurring sometime within the 6-day interval ending on the day of ovulation.[193, 341, 345] Estimates of when fertility peaks vary with the method used to determine the time of ovulation. When ovulation is assumed to occur on the day before the midcycle rise in basal body temperature (BBT), the day of peak fertility falls 2 days prior to ovulation[193]; ovulation generally occurs within 1 day of that predicted.[345] When the time of ovulation is based on daily urine estrogen concentrations, the probability of conception increases steadily as ovulation nears and peaks on the day before and day of ovulation,[341, 345] ranging from about 10% at its low to approximately 33% at its peak. When daily urinary LH excretion is monitored to detect the midcycle surge that triggers ovulation, follicular collapse (as determined by serial transvaginal ultrasonography) and, presumably, ovum release generally follows within 14–26 hours, and almost always within 48 hours.[346, 347] Regardless of the method used, all studies indicate that fertility plummets almost immediately thereafter, declining to near zero within 24 hours after ovulation.

Timed coitus is frequently recommended to infertile couples as a means to increase the likelihood of pregnancy, even though there are few data to support the recommendation. Although BBT and ovulation predictor kits can help define the time of ovulation, they should be used only when necessary. Scheduled intercourse clearly adds to the already significant stress of infertility. Moreover, much of the interval of peak fertility during the menstrual cycle may be inadvertently excluded while awaiting the appropriate "signal." ***For most couples, the simple recommendation for intercourse approximately twice per week can avoid an unnecessary source of stress while also helping to ensure that coitus occurs during the interval of highest fertility.[348] However, timed coitus may be a reasonable recommendation for couples having infrequent intercourse, by preference or because of circumstance.***

Causes of Infertility

Before any formal investigation begins, the major causes of infertility and the basic components of the infertility evaluation should be outlined for the couple. ***The major causes of infertility include ovulatory dysfunction (20–40%), tubal and peritoneal pathology (30–40%), and male factors (30–40%); uterine pathology is relatively uncommon, and the remainder is largely unexplained.*** To some extent, the prevalence of each cause of infertility varies with age. Ovulatory dysfunction is more common in younger than in older couples, tubal and peritoneal factors have a similar prevalence, and male factors and unexplained infertility are observed somewhat more often in older couples.[349, 350] The distribution of causes also varies with the duration of infertility and the level of care.[351–353]



Most couples seeking evaluation have been trying to conceive for 2 or more years, so few will be normally fertile. Those with longer durations of infertility generally have more severe or multiple problems and tend to congregate in tertiary care centers. The average duration of infertility for couples seen in tertiary care centers (42 months)[353] is twice that for couples seen in the primary care setting (21 months).[351] Predictably, the proportion of couples with easily treatable ovulatory dysfunction decreases from primary to tertiary care, and that with more severe tubal/peritoneal or male factors increases.

The human reproductive process is complex, but for purposes of evaluation, it can be dissected into its most important and basic components.

- Sperm must be deposited at or near the cervix at or near the time of ovulation, ascend into the fallopian tubes, and have the capacity to fertilize the oocyte (male factor).
- Ovulation of a mature oocyte must occur, ideally on a regular and predictable basis (ovarian factor).
- The cervix must capture, filter, nurture, and release sperm into the uterus and fallopian tubes (cervical factor).
- The uterus must be receptive to embryo implantation and capable of supporting subsequent normal growth and development (uterine factor).
- The fallopian tubes must capture ovulated ova and effectively transport sperm and embryos (tubal factor).

The infertility evaluation is designed to isolate and test the integrity of each component, insofar as that is possible, and to identify any abnormalities that might impair or prevent conception. The pace and extent of evaluation should be based on the couple's age, duration of infertility, medical history, physical examination, and preferences.

Some infertility problems once considered insurmountable are now amenable to modern treatments. IVF can effectively bypass irreparable tubal occlusive disease, and intracytoplasmic sperm injection (ICSI) can overcome even severe abnormalities of semen quality.

Treatments aimed at increasing gamete density—bringing together more than the usual numbers of oocytes and sperm in the right place at the right time—can increase cycle fecundability for couples with age-related or otherwise unexplained infertility, and include ovarian stimulation with intrauterine insemination (IUI) or IVF. In women with premature ovarian failure, women beyond normal reproductive age, and women without ovaries, IVF using donor oocytes is highly successful.

The advent of evidence-based medical practice has had significant impact on the diagnosis and treatment of infertility. Critical analyses of standard diagnostic tests and common therapies have questioned and, in some cases, proven invalid some of the most time-honored methods of evaluation and treatment.[354] ***The scope and sequence of the modern infertility evaluation have shifted focus, from making a specific diagnosis to using the most efficient and cost-effective tests. The focus of treatment for infertility also has shifted, from systematic correction of each identified factor to applying the most efficient and cost-effective therapy, which often is assisted reproductive technology (ART).***

Indications for Evaluation

When should a formal evaluation for infertility begin? After all, most infertile couples are only subfertile, not truly sterile, and many will conceive, eventually, without treatment. Infertility has a significant spontaneous cure rate that varies with female partner age, duration, past conception history, and the cause(s). ***The probability for achieving a live birth without treatment decreases with increasing age and duration of infertility.***[351–353, 355–357] Overall, the likelihood of pregnancy without treatment declines by about 5% for each additional year of female partner age and by 15–25% for each added year of infertility.[353] ***The largest majority of spontaneous pregnancies occur within 3 years; thereafter, the prognosis for success without treatment is relatively poor.*** Couples that have conceived before generally have a better prognosis than those who have never achieved pregnancy. The cause of infertility also affects the prognosis for success without treatment but, of course, cannot be determined without evaluation. Predictably, the diagnoses of anovulation and unexplained infertility have the best prognosis. The likelihood for success without treatment for couples with male factors, tubal disease, and endometriosis varies widely with the severity of disease; the prognosis is reasonably good for mild oligospermia, tubal adhesions, and mild endometriosis, and quite poor for severe male factors, tubal obstruction, and severe endometriosis.

Evaluation should be offered to all couples who have failed to conceive after a year or more of regular unprotected intercourse, but a year of infertility is not a prerequisite for evaluation. Earlier evaluation is justified for women with irregular or infrequent menses, history of pelvic infection or endometriosis, or having a male partner with known or suspected poor semen quality, and also is warranted after 6 months of unsuccessful effort for women over the age of 35 years.[358]

Education should be offered to any couple who seeks it, regardless whether they have made any active effort to conceive. It is always helpful to explain the reproductive process, to inform couples that normal cycle fecundability is approximately 20% (far lower than most realize), and to discuss the relationship between age and fertility, when it is relevant. In concerned couples who have not yet truly tested their fertility and have no obvious problems, some basic preliminary evaluation is reasonable to perform, if requested. Tests to confirm ovulation and semen quality are easy to perform, relatively inexpensive, minimally

invasive, and quickly can identify some of the most common reproductive problems. In women at high risk for diminished ovarian reserve, an ovarian reserve test is also reasonable, because results may help to determine when and how further evaluation and treatment should be recommended.

Preliminary Evaluation of the Infertile Couple

Any evaluation of infertility must begin with a careful history and physical examination, which often will identify symptoms or signs that suggest a specific cause and help to focus evaluation on the factor(s) most likely responsible. In the female partner, relevant medical history and physical findings include the following[359]:

History

- Gravidity, parity, pregnancy outcomes and associated complications.
- Cycle length and characteristics, and onset and severity of dysmenorrhea.
- Coital frequency and sexual dysfunction.
- Duration of infertility and results of any previous evaluation and treatment.
- Past surgery, its indications and outcome, and past or current medical illnesses, including episodes of pelvic inflammatory disease or exposure to sexually-transmitted infections.
- Previous abnormal pap smears and subsequent treatment.
- Current medications and allergies.
- Occupation and use of tobacco, alcohol, and other drugs.
- Family history of birth defects, mental retardation, early menopause or reproductive failure.
- Symptoms of thyroid disease, pelvic or abdominal pain, galactorrhea, hirsutism, or dyspareunia.

Physical Examination

- Weight and BMI.
- Thyroid enlargement, nodule, or tenderness.
- Breast secretions and their character.
- Signs of androgen excess.
- Pelvic or abdominal tenderness, organ enlargement, or mass.
- Vaginal or cervical abnormality, secretions, or discharge.
- Mass, tenderness, or nodularity in the adnexa or cul-de-sac.

Irregular or infrequent menses indicate ovulatory dysfunction. Previous treatment for cervical intraepithelial neoplasia or observations of a mucopurulent cervicitis or cervical stenosis helps to identify unusual women in whom the cervix may present an obstacle. A history of previous hysteroscopic or reconstructive uterine surgery or recently developing symptoms of menorrhagia suggest an abnormality of the uterine cavity; previous uncomplicated first- and second-trimester pregnancy terminations generally do not adversely affect subsequent fertility.[360, 361] Worsening dysmenorrhea, new onset of dyspareunia, or physical findings of focal tenderness or cul-de-sac nodularity suggest endometriosis. A history of pelvic infection, septic abortion, ruptured appendix, ectopic pregnancy, abdominal myomectomy, or adnexal surgery should raise suspicion for tubal or peritoneal disease.

Screening Tests

Pap smear screening is recommended for all sexually-active women of reproductive age who have a cervix. The date and results of the most recent pap smear should be documented and a pap smear performed, if needed. A blood type, Rh factor, and antibody screening (in Rh-negative women) also are recommended, if not already known.

The American College of Obstetricians and Gynecologists and the American College of Medical Genetics recommend that screening for ***cystic fibrosis*** (CF) be offered to individuals with a family history of CF, reproductive partners of individuals with CF, and couples planning a pregnancy or seeking prenatal care wherein one or both partners are Caucasian or of Ashkenazi Jewish descent, and that the test be made available to all patients on request.[362] Sequential screening (testing one partner, and the second only if the first partner is identified as a carrier) is most cost effective. Interestingly, a 2007 study found that only 22/1,006 (2%) infertile non-Hispanic Caucasian couples offered counseling and screening (carrier frequency 1/25, detection rate 88%) chose to be tested, most citing the cost of screening.[363]

All women attempting pregnancy with undocumented previous ***rubella*** infection or vaccination should be tested for immunity, and vaccinated if seronegative. As there has never been a documented case of congenital rubella syndrome attributed to vaccine, the Centers for Disease Control and Prevention (CDC) has determined that women need not avoid pregnancy for more than 1 month after vaccination.[364] The CDC also recommends that all women without history of previous infection or evidence of immunity or vaccination against ***varicella*** (chicken pox) receive two doses of vaccine and avoid pregnancy for 1 month after each dose.[365]

Screening for ***sexually-transmitted infections*** (STI) is recommended for all women at moderate to high risk for infection. Decisions regarding STI screening should consider that current recommendations from the CDC include screening all pregnant women for chlamydia and gonorrhea (nucleic acid-based tests), syphilis (rapid plasma reagin; RPR), hepatitis B (hepatitis B surface antigen; HBSAg), and voluntary screening for human immunodeficiency virus type 1 (HIV-1) at the first prenatal visit.[366] For women receiving inseminations of donor sperm, the American Society for Reproductive Medicine (ASRM) considers HIV-1 screening mandatory, recommends screening for syphilis, hepatitis B and C, cytomegalovirus (CMV), HIV-2, and human T-cell lymphocyte virus (HTLV) types I and II, and suggests screening for chlamydia and gonorrhea at the discretion of the physician.[367] For male partners of women receiving inseminations of donor sperm, the ASRM strongly recommends HIV-1 and recommends other STI screening. For recipients of donor oocytes or embryos and their male partners, the ASRM recommends screening for syphilis, hepatitis B and C, CMV, and HIV-1.[367] Any additional screening laboratory tests should be directed by the medical history and clinical judgment.

Male Factor: Abnormalities of Semen Quality

The evaluation and treatment of male infertility is the focus of Chapter 30, but must be addressed briefly here because male factors explain or contribute significantly to infertility in up to 35% of couples. Semen analysis is therefore always an appropriate and important initial step in the evaluation of the infertile couple. In the absence of any known genital abnormality, trauma, surgery, or sexual dysfunction, physical examination of the male partner can be deferred pending the results of the initial semen analysis.

When semen analysis yields equivocal results, additional analyses are required to better define a suspected abnormality. A frankly abnormal semen analysis is indication for additional evaluation that may be conducted by a gynecologist having the necessary training and experience, but most often is performed by a urologist or other specialist in male reproduction.[368] Invasive diagnostic procedures in the female partner generally are best deferred until evaluation of the male is completed. The range of effective treatment options for couples with severe male factor infertility is limited, and often will direct or even dictate what additional evaluation may be relevant in the female partner. When semen quality is normal, attention naturally turns to the female partner.

Ovarian Factor: Ovulatory Dysfunction

Overall, disorders of ovulation account for approximately 20% of the problems identified in infertile couples. Ovulatory dysfunction can be severe enough to prevent conception (anovulation), or only a contributing factor (oligoovulation). However, because cycle fecundability averages only approximately 20% even in normally fertile couples, the distinction is moot.

A number of methods can be used to determine if and when ovulation occurs. Directly or indirectly, all are based on one or another of the hormonal events that characterize the normal ovulatory cycle (Chapter 6). Each of the available tests is useful and no one test is necessarily best. Some are simple, noninvasive, and inexpensive, and others are more complicated, invasive, and costly. A few can predict when ovulation is likely, with varying accuracy. However, no test, regardless how sophisticated, can prove that ovulation has actually occurred; the only positive proof of ovulation is pregnancy. The most appropriate test to use varies with the information required. The same tests used to diagnose anovulation can be used to assess the effectiveness of treatment.

Menstrual History

Menstrual history alone often is sufficient to establish a diagnosis of anovulation. Menses in normally ovulating women generally are regular, predictable, consistent in volume and duration, and typically accompanied by a recognizable pattern of premenstrual and menstrual symptoms. Conversely, those in anovulatory women generally are irregular, unpredictable or infrequent, vary in flow characteristics, and exhibit no consistent pattern of molimina. Women with regular menses are almost always ovulatory. ***Women with irregular or infrequent menses may ovulate, but not consistently, and do not require specific diagnostic tests to prove what is already obvious.***

Basal Body Temperature (BBT)

Basal body temperature is body temperature under basal conditions, at rest. For practical purposes, BBT is measured each morning, upon awakening and before arising. Traditionally, BBT is measured with an oral glass-mercury thermometer having an expanded scale, typically ranging from 96.0 to 100.0 degrees Fahrenheit and marked in tenths of one degree; modern electronic thermometers are a suitable alternative, but only if they have the necessary accuracy

and precision. As a test of ovulation, daily BBT recordings are based on the thermogenic properties of progesterone; as levels rise after ovulation, BBT also increases. The effects are more qualitative than quantitative, are subtle but nonetheless distinct, and generally easy to detect when daily BBT recordings are plotted on graph paper.[369] ***Synthetic progestins commonly used to induce menses in amenorrheic women (medroxyprogesterone acetate, norethindrone acetate) have similar thermogenic properties and also raise BBT.***

BBT is typically low and fluctuates between 97.0 and 98.0 degrees during the follicular phase of the cycle, modestly but distinctly higher (0.4–0.8 degrees) during the luteal phase, and falls again to baseline levels just before or after the onset of menses. In ovulatory women, a "biphasic" pattern usually is readily evident. ***The ideal BBT recording is distinctly biphasic and reveals a cycle between 25 and 35 days in length, with menses beginning 12 days or more after the rise in temperature.*** When pregnancy occurs in a monitored cycle, onset of menses is delayed and BBT remains elevated, reflecting the sustained production of progesterone by the corpus luteum stimulated by human chorionic gonadotropin (hCG).

BBT recordings provide objective evidence of ovulation and also reveal the approximate time of ovulation. Unfortunately, the temporal relationship between the thermogenic shift in BBT and ovulation frequently is misunderstood. BBT generally falls to its lowest level on the day before or day of ovulation, but the nadir in BBT cannot be reliably identified until after the temperature rises and remains elevated.[370] The shift in BBT occurs when progesterone concentrations rise above approximately 3–5 ng/mL, 1 to 5 days *after* the midcycle LH surge and up to 4 days *after* ovulation.[371] The temperature rise usually is somewhat abrupt, but may be gradual and difficult to define, and once apparent (2 or more days of temperature elevation), the most fertile interval has passed. ***In cycles monitored with BBT, the interval of highest fertility spans the 7-day interval immediately before the midcycle rise in BBT.*** Much of the uncertainty in predicting the time of ovulation can be avoided by reviewing a series of recordings, noting the earliest and latest days of the cycle on which the temperature shift occurred. ***Coital timing can be optimized by suggesting intercourse on alternate days beginning 7 days before the earliest observed rise in BBT and ending on the latest day it has been observed.***

The principal advantage that BBT has over other tests of ovulation is low cost. BBT recordings also can reveal an abnormally long follicular phase and grossly short luteal phase that otherwise might go unrecognized, for which treatment is warranted. BBT monitoring is easy and non-invasive, but can become tedious over time. For some it also increases stress, serving as a daily reminder of unsuccessful efforts to conceive, each day beginning with thoughts of a family not yet realized. In the few women who menstruate regularly but do not exhibit a biphasic BBT, an alternative method should be used to document ovulation before assuming that treatment is required. Although there are more reliable methods to evaluate ovulatory function, BBT is still useful and may be the best method for couples who are reluctant or unable to pursue more formal and costly evaluations.

Serum Progesterone Concentration

A serum progesterone measurement is the simplest, most common, objective and reliable test of ovulatory function, as long as it is appropriately timed. Progesterone levels generally remain below 1 ng/mL during the follicular phase, rise slightly on the day of the LH surge (1–2 ng/mL) and steadily thereafter, peak 7–8 days after ovulation, and decline again over the days preceding menses. ***A progesterone concentration less than 3 ng/mL implies anovulation, except when drawn immediately after ovulation or just before the onset of menses, when lower levels naturally might be expected.***[372, 373]

When is the best time to measure the serum progesterone concentration to document ovulation? ***Ideally, the serum progesterone level should be drawn approximately one week***

before the expected onset of menses, when the concentration is at or near its peak. Contrary to popular belief and practice, cycle day 21 is not always the best time to measure the serum progesterone concentration. Cycle day 21 is a good choice for women with cycles lasting approximately 28 days, but a poor choice for women with 35 day cycles. The normal ovulatory cycle is 25–35 days long and exhibits a 13–15 day luteal phase. At the extremes of normal, ovulation may occur as early as cycle day 10 (in a 25-day cycle) and as late as day 22 (in a 35-day cycle). If ovulation occurs on cycle day 10, day 21 falls 11 days after ovulation, well after progesterone concentrations peak and when they are again nearing basal levels. If ovulation occurs on cycle day 22, day 21 falls 1 day *before* ovulation, when progesterone levels have not yet started to rise. ***The best time to test will vary with the overall length of the menstrual cycle, aiming for approximately 1 week before the expected menses.***

Serum progesterone levels also have been used to evaluate the quality of luteal function. Whereas the amount and duration of progesterone production certainly does reflect the functional capacity of the corpus luteum, a truly accurate measure requires daily serum progesterone determinations that are costly, and impractical.[374–376] Judgments based on limited sampling, regardless how well timed, have numerous pitfalls and cannot define the quality of luteal function reliably.[374, 377–381] ***There is no consensus minimum serum progesterone concentration that defines normal luteal function.*** A midluteal serum progesterone level greater than 10 ng/mL is a popular standard,[382] but the concentrations observed in normal and abnormal cycles and in conception and non-conception cycles in both fertile and infertile women vary widely and overlap greatly.[383] One reason is that progesterone is secreted by the corpus luteum in distinct pulses, temporally linked to pulsatile luteinizing hormone (LH) secretion[384, 385]; levels ranging from as low as 4 ng/mL to as high as 40 ng/mL can be observed within brief intervals of time.[385] ***A midluteal serum progesterone concentration cannot define the quality of luteal function and has little value beyond documenting ovulation.***

Urinary LH Excretion

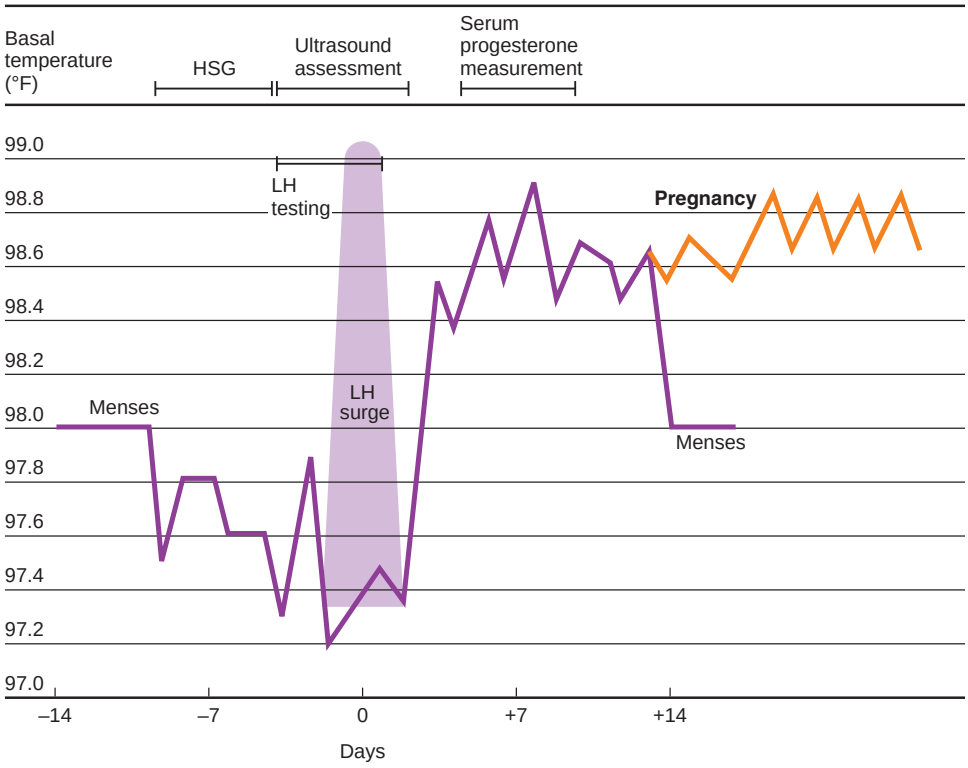
A wide variety of different commercial products allow women to determine not only if they ovulate, but when, in advance of the actual event. Generally known as "ovulation prediction kits" or "LH kits," the products are all designed to detect the midcycle LH surge in urine. Ovulation predictor kits take advantage of advances in hormone measurement technology, reducing what was once a very labor-intensive process in the hospital laboratory to one or two simple steps requiring only a few minutes time in the home.

The midcycle LH surge is a relatively brief event, typically lasting between 48 and 50 hours from start to finish. LH has a short half-life and is rapidly cleared via the urine. Ovulation predictor kits turn positive when the urinary LH concentration exceeds a threshold level normally seen only during the LH surge. In most cycles, the test is positive on a single day, occasionally on 2 consecutive days. To detect the LH surge reliably, testing must be done on a daily basis, generally beginning 2 or 3 days before the surge is expected, based on the overall length of the cycle. The first positive test provides all relevant information; there is no value in continued testing.

Test results are sensitive to both the volume of fluid intake and time of day. There is no need to restrict fluid intake, but patients should be advised to avoid drinking large volumes of fluid a short time before they plan to test. Logically, the first morning void would seem an ideal specimen to test because it is usually the most concentrated. However, results correlate best with the serum LH peak when testing is performed in the late afternoon or early evening hours (4:00–10:00 PM),[371] probably because LH surges often begin in the early morning hours and are not detected in urine for several hours. Twice daily testing decreases

the frequency of false-negative results (failure to detect the LH surge in an ovulatory cycle), but generally is unnecessary. When performed daily and properly timed, testing will detect the LH surge in most ovulatory cycles. True false-positive tests (detection of an LH surge in an anovulatory cycle) occur in approximately 7% of cycles;[386] equivocal or "borderline" results also are common and can be both confusing and frustrating.

The accuracy of ovulation predictor kits varies. All are useful and reasonably reliable, but some are better and easier to use than others.[347, 387] The best products predict ovulation within the subsequent 24 to 48 hours, with greater than 90% probability.[346, 347] ***Ovulation generally occurs 14–26 hours after detection of the LH surge and almost always within 48 hours.[346] Consequently, the interval of greatest fertility includes the day the surge is detected and the following 2 days.*** The day *after* the first positive test generally is the one best day for timed intercourse or insemination.[346, 388, 389] Ovulation predictor kits are non-invasive, widely available, require relatively little time and effort, and invite women to become actively involved in their care. Their greatest advantage over other methods is their ability to predict when ovulation will occur. Accurate identification of the midcycle LH surge also defines the length of the follicular and luteal phases, which may reveal subtle and otherwise unrecognized cycle abnormalities warranting treatment. Urinary LH monitoring is perhaps best reserved for women who ovulate (based on menstrual history, BBT recordings, or an appropriately timed serum progesterone concentration), but have infrequent intercourse or require insemination.



Endometrial Biopsy and Luteal Phase Deficiency

Endometrial biopsy can be used as a test of ovulation, based on the characteristic histologic changes induced by progesterone. During the follicular phase of the cycle, the endometrium exhibits a proliferative pattern, reflecting the growth stimulated by rising levels of estrogen derived primarily from the dominant ovarian follicle. During the luteal phase,

progesterone secreted by the corpus luteum induces the "secretory" transformation of the endometrium. Anovulatory women are always in the follicular phase; their endometrium is always proliferative and even may become hyperplastic with extended exposure to a constant estrogen growth stimulus. ***In the absence of treatment with exogenous progesterone or a synthetic progestin, a secretory endometrium implies recent ovulation.***

Endometrial biopsy is a relatively simple office procedure, usually performed with a disposable plastic aspiration cannula, and complications are few. Pretreatment with a non-steroidal anti-inflammatory drug (NSAID) helps to reduce pain or cramping associated with the procedure. Sedation or anesthetic (paracervical block) is helpful when the biopsy is technically difficult and in women who are very anxious. When properly timed, in the same way and for the same reasons as a serum progesterone concentration, endometrial biopsy is an effective test of ovulation. However, it is also invasive, uncomfortable, costly, and provides little more information than can be obtained from BBT recordings, a serum progesterone concentration, or monitoring urine LH excretion. Therefore, endometrial biopsy has rather limited and specific indications in the evaluation of infertile women. For women with chronic anovulation of long duration, biopsy can identify or exclude endometrial hyperplasia that requires specific treatment. In those few individuals suspected of harboring a chronic endometritis, biopsy is diagnostic. ***Until recently, endometrial biopsy for diagnosis of luteal phase deficiency was considered a basic element of the infertility evaluation, but no longer.***

Inadequate corpus luteum progesterone production or "luteal phase deficiency" (LPD) was long considered an important cause of both infertility and early pregnancy loss.[390, 391] The proposed mechanisms were different but related, representing only different points on a pathophysiologic continuum. In theory, because the human implantation window is relatively narrow (spanning the interval from approximately 6 to 10 days after ovulation)[392–394] low circulating progesterone levels could be expected to result in delayed endometrial maturation, causing a shift in the implantation window and failed or late implantation. A long delay would threaten embryo viability or prevent implantation. A shorter delay would allow implantation but result in a tardy or low amplitude hCG rescue signal that could not stimulate normal amounts of progesterone from an already regressing corpus luteum, or maintain production for the requisite duration,[395–397] with either causing a premature luteal-placental shift and pregnancy loss.[398] In this context, endometrial biopsy was viewed as a bioassay of luteal function because it would reflect both the functional capacity of the corpus luteum and the end organ response.

The classic histologic features of secretory endometrial development were described by Noyes, Hertig, and Rock, in the lead article of the inaugural issue of *Fertility and Sterility*.[399] The pattern was considered sufficiently predictable to allow experienced pathologists to "date" the endometrium, assigning a histologic day that could be compared to the actual day of sampling, estimated by counting backward from onset of the next menstrual period (assuming menses began on the 14th postovulatory day), or defined by the number of days elapsed since detection of the LH surge or observation of follicular collapse by serial ultrasonography.[400] Historically, histologic and sampling dates that agreed, within a 2-day interval, were considered normal, whereas a date more than 2 days "out of phase" was the gold standard criterion for the diagnosis of LPD.[401–403] Traditionally, diagnosis of LPD required abnormal results in two (preferably consecutive) cycles, reasoning that reproductive failure could only be attributed to LPD if it was consistent or recurring, and acknowledging that LPD also could occur in normal fertile women, at least occasionally.[404–410] Endometrial dating was accepted widely by clinicians and pathologists and the practice endured, despite numerous challenges to its validity.

The first and most fundamental criticism of the traditional histologic dating criteria was that the normal standard was based on analysis of tissue specimens obtained from infertile women;[399] the reference population was abnormal, by definition, and also likely heterogeneous because infertility has many different causes. Second, the sampling date was estimated

retrospectively, after the onset of menses, assuming a uniform 14-day luteal phase, despite numerous studies demonstrating that luteal phase duration varied significantly, even in normal women.[104, 109, 112, 411, 412] Moreover, retrospective estimates of the sampling date correlated poorly with the time of ovulation as defined by the LH surge or observations of follicular collapse,[382, 400, 413] and ignored any effect that biopsy might have on the onset of menses, or when it was perceived to start.[400, 409, 414] Third, the traditional histologic dating criteria were inherently subjective, and numerous studies had observed significant intra-observer and inter-observer variations in histologic interpretation that were great enough to affect diagnosis and management in 20–40% of individual women.[403, 415–418]

The standard practice of endometrial biopsy and histologic dating for diagnosis of LPD was proven invalid in 2004, for all intents and purposes. A systematic re-analysis of the histologic features used for endometrial dating confirmed the classically described sequence, but revealed the patterns were much less temporally discrete than originally described, and demonstrated that normal variations among individuals, between cycles in individuals, and among different observers were too great to reliably define any specific luteal day or even a narrow interval of days.[419] Soon thereafter, a large multicenter trial demonstrated conclusively that abnormal histologic dating could not discriminate infertile from proven fertile women.[420] The second study invalidated the practice of endometrial dating, and the first explained why the method failed.

Recent evidence challenges even the basic premise on which the concept of LPD is founded: that abnormally low circulating progesterone concentrations result in delayed endometrial maturation. In normal women treated with a fixed physiologic dose of estrogen after down-regulation with a GnRH agonist, then randomized to receive physiologic (mean progesterone level 19 ng/mL) or grossly low levels of exogenous progesterone treatment (mean progesterone level 5.5 ng/mL), there was no discernible difference in endometrial histology.[421] These observations suggest the histologic features of secretory endometrium relate more to the duration of progesterone exposure than to the concentration. Studies using a similar design have demonstrated that widely ranging concentrations of estradiol also have no discernible impact on secretory endometrial maturation.[422] Altogether, these data indicate that secretory endometrial development can progress normally despite widely varying concentrations of estradiol and progesterone, challenging the traditional paradigm, and served to further invalidate the use of endometrial histologic dating as a diagnostic tool. ***In sum, endometrial dating cannot guide the clinical management of women with reproductive failure and has no place in the diagnostic evaluation of infertility.***

The lack of any valid method for diagnosis of LPD does not refute its existence or its potential importance in the pathophysiology of reproductive failure. The pathogenic mechanisms outlined above are still viable. Evidence supports the notion of a finite implantation window,[392–394] that progesterone is essential for embryo implantation,[423] and that delayed implantation might adversely affect corpus luteum function,[395–397] predisposing to reproductive failure.[398] It is entirely possible, if not likely, that abnormally low levels of progesterone might have important functional consequences with no morphologic correlate. Biochemical or molecular markers of endometrial function provide the means to further explore the possibility. The pattern of endometrial gene expression defines distinct functional phases of the cycle.[424] A number of endometrial proteins exhibit patterns of expression or gene regulation during the putative implantation window, suggesting they might serve as markers of endometrial receptivity, including cytokines (leukemia inhibitory factor, colony-stimulating factor-1, and interleukin-1), cell adhesion molecules (the $\alpha v\beta 3$ integrin), glycodelin, and the polymorphic mucin 1,[425, 426] osteopontin,[427–429] N-acetylglucosamine-6-O-sulfotransferase (important in synthesis of L-selectin ligands),[430] and the L-selectin ligand itself.[431] None has yet been validated as a reliable measure of endometrial function or receptivity, but if and when that occurs, a functional marker may become the basis for diagnosis of LPD and endometrial biopsy again may be viewed as offering valuable information beyond that provided by other tests of ovulation.

Transvaginal Ultrasonography

The last and most complicated test of ovulation involves serial transvaginal ultrasonography (TVUS), which permits direct observation of events in the ovary just before and immediately after ovum release. Although still not providing positive proof that ovulation actually occurred, serial TVUS provides detailed information about the size and number of preovulatory follicles and the most accurate estimate of when ovulation occurs.

In its final stages of development, the preovulatory follicle grows at a predictable pace, approximately 2 mm per day (range: 1–3 mm/day). After ovulation, the follicle collapses, margins become less distinct, the density of internal echoes increases, and the volume of cul-de-sac fluid increases.[432, 433] Abnormal patterns of follicle development also can be observed. The follicle may grow at an abnormal pace, collapse when still relatively small, or continue to grow but fail to rupture and persist as a cyst for days after the LH surge—the luteinized unruptured follicle.[434, 435] Such subtle forms of ovulatory dysfunction cannot be detected otherwise, but also are rare. ***Because treatment with prostaglandin synthase inhibitors (NSAIDs) can disrupt the ovulatory process and predispose to an luteinized unruptured follicle,[436, 437] their use is best limited to the menstrual phase of the cycle in women attempting to conceive.***

Serial TVUS to monitor the size and number of developing follicles is essential to the safety and effectiveness of ovulation induction with exogenous gonadotropins (Chapter 31), but the costs and logistical demands involved are otherwise difficult to justify. Consequently, the method generally should be reserved for the few in whom the safety or effectiveness of treatment truly hinges on the detailed information it offers.

SUMMARY

The evaluation of ovulation is a core component of the evaluation for infertility. All of the different methods are useful and no one method is necessarily best. Whereas some are very simple, noninvasive, and inexpensive, others are more complicated, invasive, and costly. A few provide the means to determine not only if ovulation occurs, but when, with varying accuracy. The best choice among methods varies with the information required. In women with oligomenorrhea or amenorrhea, no formal evaluation is needed to establish a diagnosis of ovulatory dysfunction, but endometrial biopsy to exclude hyperplasia may be prudent, depending on duration. When the only objective is to confirm ovulatory function, as in those with regular monthly menses, a properly timed serum progesterone concentration is the simplest and most reliable method. When circumstances require accurate prediction of ovulation, as in couples having infrequent intercourse or those requiring insemination, urinary LH monitoring generally is the most cost-effective and appropriate choice. In the few who require insemination but consistently fail to detect a midcycle LH surge, serial transvaginal ultrasonography can provide the necessary information. Ultimately, the method chosen should be tailored to the needs of the individual patient.

Infertile women with ovulatory dysfunction are obvious candidates for ovulation induction. In general, only limited additional evaluation is needed to define the initial treatment of choice and most women will respond promptly to one of the simpler treatment strategies (Chapter 31). In the majority of cases, it is reasonable and appropriate to begin treatment immediately, even before other potential causes of infertility have been investigated. If anovulation is the only obstacle to overcome, most couples will conceive promptly without further interventions. Women with amenorrhea or hyperandrogenic anovulation deserve additional preliminary evaluation, applying the principles described in Chapters 11,12, and 13.

Cervical Factor: Abnormalities of Sperm-Mucus Interaction

The cervix participates in the reproductive process in several ways. Cervical mucus accepts or captures sperm from the ejaculate and the vagina, excludes the seminal plasma and morphologically abnormal sperm,[438] nurtures sperm biochemically, and serves as a reservoir, thereby prolonging sperm survival and the fertile interval between intercourse and ovulation. Mucus is a glycoprotein gel with solid and liquid phases and has a mosaic ultrastructure with interstitial channels between mucin strands that expand and contract in response to cyclic changes in the steroid hormone environment across the menstrual cycle to facilitate or inhibit the passage of sperm.[439–443] Estrogen stimulates cervical mucus production, and as levels rise during the follicular phase, mucus becomes more abundant and watery, less cellular, and more easily penetrated by sperm.[444] Progesterone inhibits cervical mucus production and renders it opaque, viscid, and impenetrable. The cyclic changes in cervical mucus characteristics help to explain why the cycle day-specific probability of conception rises steadily as ovulation nears and plummets immediately thereafter.

For most of the past century, the postcoital test for diagnosis of cervical factor infertility was considered a basic element of the infertility evaluation. The test involved collection of cervical mucus (by aspiration or with nasal polyp forceps) shortly before the expected time of ovulation (as determined by BBT or urinary LH monitoring in previous cycles) a few to several hours (typically 2–12 hours) after intercourse.[445] The mucus specimen was evaluated for pH, clarity, cellularity, viscosity (the length to which a column of mucus can be stretched in centimeters, known as "spinnbarkeit"), and salinity (evaluated according to the complexity of the network of crystals that forms when mucus is dried on a glass slide, also known as "fern ing"), and for the number and motility of surviving sperm. The presence of motile sperm confirmed effective coital technique and sperm survival and the number of sperm (per high power field) was used to predict semen quality (sperm density and motility) and cycle fecundability (inverse correlation with time to conception or cumulative conception rates).[446–451] Most considered even a single motile sperm in most fields a "positive" or normal test result.[451–453]

Abnormal or "negative" postcoital test results were common, usually due to improper timing, either too early in the cycle when mucus was relatively scant, or after ovulation when mucus quality was poor.[454] Timing was optimized by performing the test within 2 days before the LH surge or when transvaginal ultrasonography demonstrated a preovulatory follicle.[455] Other explanations for poor quality mucus were cervicitis, previous treatment for cervical intraepithelial neoplasia (e.g., cryotherapy), and treatment with clomiphene citrate. Potential explanations for the absence of motile sperm in good quality mucus included ineffective intercourse, failed ejaculation (frequently resulting from performance anxiety), poor semen quality, and use of spermicidal coital lubricants. Observations of degenerating, immotile, "shaking" or agglutinated sperm were considered reason for antisperm antibody testing.[449] An abnormal result was confirmed by repeat testing to establish the diagnosis of cervical factor infertility,[445, 451, 456] prompting further evaluation with a nucleic acid test for chlamydia and cultures for ureaplasma and myocoplasma (or empirical treatment with antibiotics),[457–459] and semen analysis. Normal semen quality and absence of sperm in good quality mucus was regarded as evidence of "hostile" cervical mucus or a sperm function abnormality, differentiated by comparisons of partner and donor sperm survival and motility in bovine cervical mucus *in vitro* and antisperm antibody testing.[460–462] Strategies for correcting or overcoming cervical factor infertility included treatment with exogenous estrogens (to stimulate mucus production)[463] or mucolytic agents (guanifenesin),[464] precoital douching with a sodium bicarbonate solution,[465] and intrauterine insemination (IUI).[449, 466–468]

Advocates of routine postcoital testing argued that the postcoital test could identify couples who might benefit from a simple treatment and had prognostic value for predicting the probability of pregnancy without treatment.[356, 450] Critics reasoned that results achieved even with IUI suggested only a modest benefit at best,[469] that any prognostic value the test might have was limited to young couples with unexplained infertility of short duration because the test surely had no predictive value in women with infertility due to anovulation or tubal occlusive disease, and that male infertility amenable to treatment with IUI could be more accurately defined by the results of semen analysis. The argument for expectant management in couples with unexplained infertility and a normal postcoital test was dismissed as moot, because few couples seeking evaluation and treatment accepted the recommendation.

The postcoital test for diagnosis of cervical factor is no longer recommended.[359] Abnormalities of cervical mucus production or sperm/mucus interaction are rarely, if ever, the sole or principal cause of infertility. Chronic cervicitis or cervical stenosis resulting from conization or other treatment for cervical disease that might impair sperm-mucus interaction can be identified by speculum examination, and in the absence of such findings, the likelihood that cervical mucus represents an important obstacle is remote. Semen analysis identifies couples with significant male factor infertility. The test has no standard methodology or interpretation,[445, 453] and has poor reproducibility even among trained observers.[470] The only randomized trial comparing outcomes in women with normal and abnormal postcoital tests found the test invalid because neither test results nor treatment for abnormal tests affected outcome.[471–473] Office examination after scheduled intercourse is an inconvenient, embarrassing, and unwelcome intrusion for most couples, adding further to their burden of stress. Finally, postcoital test results seldom change clinical management, because contemporary treatments for unexplained infertility include IUI (usually with ovarian stimulation) or IVF, both of which negate any contributing cervical factor.

Uterine Factor: Anatomic and Functional Abnormalities

Abnormalities of the uterus are a relatively uncommon cause of infertility, but should always be considered. If for no other reason, they may adversely affect the outcome of pregnancies achieved by successful treatment of more common male, ovarian, and tubal factors. The anatomic uterine abnormalities that can adversely affect fertility include congenital malformations, leiomyomas, and intrauterine adhesions; endometrial polyps also have been implicated, but their reproductive implications are less clear. The only functional uterine abnormality of specific interest in the evaluation of infertility is chronic endometritis. Whereas abnormalities of endometrial receptivity (including luteal phase deficiency) might be viewed as another, they can have no practical significance until there is conclusive evidence that infertility can result from intrinsic endometrial dysfunction that impairs or prevents implantation and a method for diagnosis has been validated. In the meantime, luteal phase deficiency is best viewed as a subtle form of ovulatory dysfunction, as discussed in an earlier section of this chapter (see Ovarian Factor).

Anatomic and functional uterine abnormalities that can impair fertility also can adversely affect pregnancy outcome. They are discussed here as a cause of infertility, and elsewhere (Chapter 28) as a cause of recurrent pregnancy loss. The embryology or pathogenesis and obstetric consequences of uterine malformations and of leiomyomas are considered at length in Chapter 4. Discussion here is focused on their diagnosis, their impact on fertility, and how they influence evaluation and treatment.

There are three basic methods for evaluation of the uterine cavity: hysterosalpingography, transvaginal ultrasonography or saline sonohysterography, and hysteroscopy. Each has advantages and disadvantages and the choice among them should be tailored to the needs of the individual patient. HSG is the traditional method and most often still the best initial test because it also evaluates tubal patency. However, in women with no risk factors for tubal disease and those whose tubal status is already known (from earlier surgery for other indications) or is largely irrelevant (as in women who require IVF for severe male factor infertility), ultrasonography offers a simpler and better tolerated alternative that also may reveal unsuspected ovarian pathology (cyst, endometrioma), with no radiation exposure. When symptoms suggest an anatomic lesion of the uterine cavity (menorrhagia, intermenstrual spotting)), sonohysterography is the most sensitive and logical diagnostic test. Hysteroscopy is definitive but has few diagnostic advantages over sonohysterography and generally can be safely reserved for treatment of abnormalities already identified by less invasive and costly methods.

Hysterosalpingography

Hysterosalpingography (HSG) accurately defines the size and shape of the uterine cavity, provides clear images of most uterine developmental anomalies (unicornuate, septate, bicornuate, and didelphys) and, with exceptions, also identifies submucous myomas and intrauterine adhesions that can have important reproductive implications. Although HSG also may reveal endometrial polyps, sonohysterography is a more sensitive method for their detection. A slow injection of contrast medium helps to minimize the risk that a cavitary lesion will be obscured and go undetected.

The normal uterine cavity is symmetrical, roughly triangular in shape, widest at the level of the cornual orifices near the fundus, and relatively smooth in its contours. The various developmental uterine anomalies generally have a fairly characteristic appearance on HSG. A unicornuate uterus is typically somewhat tubular, deviates to the left or right, and has one fallopian tube. Both septate and bicornuate uteri typically exhibit a common lower segment that divides into two distinct horns to yield a Y-shaped configuration with varying distance between the upper arms.[474, 475] The two anomalies cannot be differentiated by HSG alone; additional evaluation is required to establish an accurate diagnosis (standard or three-dimensional ultrasonography, sonohysterography, MRI, or laparoscopy).[476] Either anomaly also can be confused with a unicornuate uterus if only one of the two horns is imaged because they divide near or below the tip of the cannula or catheter inserted into the cervix or uterus. To properly study a uterus didelphys or complete septate uterus, the two hemi-uteri must be imaged via their separate cervical openings, often found on opposite sides of a longitudinal vaginal septum of varying length. Myomas and larger polyps generally produce curvilinear filling defects of various size and shape. HSG in women with intrauterine adhesions usually reveals grossly irregular cavity contours and filling defects, and in many with severe disease, no cavity at all.

The accuracy of HSG for detecting intrauterine pathology in infertile women varies with the nature of the abnormality. A large study involving over 300 women comparing HSG to hysteroscopy (the gold standard) observed that HSG had overall 98% sensitivity, 35% specificity, 70% positive predictive value, and 92% negative predictive value, with a 30% false-positive rate and 8% false-negative rate; misdiagnoses almost entirely related to distinguishing submucous myomas from polyps and were, therefore, relatively unimportant.[477] In another study of similar design, HSG had 75% sensitivity for detection of intrauterine adhesions and only 50% sensitivity for detection of endometrial polyps.[478]

Specific issues concerning the scheduling and preparation for HSG and details regarding technique and interpretation as they relate to the evaluation of tubal factor infertility are addressed in the following section (see Tubal Factor, below).

Transvaginal Ultrasonography and Saline Sonohysterography

Transvaginal ultrasonography (TVUS) is another method for evaluation of uterine factors in infertile women. Saline sonohysterography, involving TVUS during or after introduction of sterile saline through a catheter designed for the purpose, crisply defines cavity contours and readily demonstrates even small, but potentially important, intrauterine lesions.[479]

In all phases of the cycle, the interface between the endometrium and the myometrium is well defined. The interface between the two layers of the endometrium itself (bordering the uterine cavity) can be difficult to identify very early in the cycle and during the secretory phase, but is visible during the latter half of the proliferative phase. Together, the two layers of the endometrium comprise the "endometrial stripe," which changes in appearance and thickness across the cycle. During the proliferative phase, the endometrium is relatively hypoechoic and grows in thickness to yield a prominent "triple line" or trilamminar pattern. During the secretory phase, the endometrium grows little more, or not at all, and increases in echodensity, possibly because the developing network of coiled basilar vessels presents a great many more reflective surfaces. Cycle-dependent changes in uterine artery blood flow parameters (velocity and pulsatility index) measured using color and pulsed Doppler ultrasonography also have been described,[480, 481] but diurnal variations and differences between the two uterine arteries (ipsilateral or contralateral to the dominant ovarian follicle) complicate interpretation. In efforts to define a receptive endometrium, several studies have examined the correlation between endometrial stripe thickness and pattern or uterine artery blood flow parameters with implantation or pregnancy rates in IVF cycles,[482–487] but results are conflicting. Whereas some have found correlations between one or more parameters and treatment outcomes, others have not. The few studies examining the endometrium in unstimulated cycles in infertile women have not demonstrated any important correlation between endometrial thickness, pattern, or blood flow and the cause of infertility or prognosis.[488–490] ***In the diagnostic evaluation of infertile women, transvaginal ultrasonography can identify important uterine pathology but provides no useful measure of endometrial function or receptivity.***

For identification of congenital malformations, standard two-dimensional TVUS complements HSG and improves diagnostic accuracy for differentiating septate and bicornuate uteri by revealing the shape of the fundal contour. The septate uterus presents a single unified fundus that often is somewhat broader than normal and sometimes slightly concave; the bicornuate uterus has two entirely separate fundi divided by a distinct midline cleft of varying depth.[474, 476] The accuracy of saline sonohysterography exceeds that of HSG, by revealing both the double uterine cavity and the shape of the fundal contour. Modern three-dimensional ultrasonography can generate reconstructed images in the coronal plane and offers diagnostic accuracy comparing favorably with magnetic resonance imaging or combined laparoscopy and hysteroscopy (the gold standard).[475, 476, 491]

Results of studies evaluating the accuracy of TVUS for detection of submucous myomas and endometrial polyps have varied, but in general, both two-dimensional and three-dimensional TVUS are more sensitive than HSG, and approach the accuracy of hysteroscopy.[492–495] Whereas an overall or focal increase in endometrial thickness or asymmetry between the two layers suggests a polyp or myoma, saline sonohysterography reveals a polypoid projection into the fluid-filled cavity. For diagnosis of intrauterine adhesions, standard TVUS is reasonably specific, but rather insensitive[478, 496]; a focally narrowed or discontinuous endometrial stripe suggests the diagnosis. Saline sonohysterography compares with HSG, having relatively high sensitivity (75%) and specificity (over 90%), modest positive predictive value (approximately 50%), and excellent negative predictive value (over 95%) for detection of adhesions.[478, 497] Women with mild disease exhibit mobile thin, echogenic bands bridging a normally distensible endometrial cavity. Those with severe disease have more broadly based bands, or no cavity at all.[498]

Hysteroscopy

Hysteroscopy is the gold standard method for both diagnosis and treatment of intrauterine pathology that may adversely affect fertility. Traditionally, hysteroscopy was reserved for treatment of disease identified by other less invasive methods, but modern operative hysteroscopes with an outer diameter measuring 2–3 mm now permit diagnostic and minor operative procedures to be performed safely in the office setting.[499] Major intrauterine pathology generally requires more traditional operative hysteroscopy using instruments having larger caliber and greater capabilities.

Congenital Uterine Anomalies

Developmental uterine anomalies have long been associated with pregnancy loss and obstetric complications, but affected women generally are not infertile. The prevalence of uterine anomalies in infertile women and fertile women with normal reproductive outcomes is similar, approximately 2–4%.[500–505] The prevalence is higher among women with poor pregnancy outcomes, such as recurrent miscarriage (10–13%). Consequently, when discovered during an infertility evaluation, anomalies cannot be regarded as the likely cause or even as an important contributing cause of infertility, but only as another obstacle that must be considered when planning treatment after evaluation is completed. For example, treatments associated with substantial risk for multifetal gestation (ovarian stimulation/IUI, IVF) present even greater risks to women with uterine malformations. In most series, septate uterus is the most common anomaly (35%), followed by bicornate (26%), arcuate (18%), didelphys (8%), and agenesis (3%).[505]

Septate uterus is the anomaly most highly associated with reproductive failure and obstetrical complications, including first- and second-trimester miscarriage, preterm delivery, fetal malpresentation, intrauterine growth restriction, and infertility.[476, 505] The mechanisms responsible are poorly understood, but poor septal blood supply, resulting in poor implantation efficiency and embryo growth, and cervical incompetence are the usual suspects.[506–509]

Although diagnosis of septate uterus is not an automatic indication for metroplasty, the overall reproductive performance of women with a septum *in situ* (at least those who are recognized), is rather poor, with term delivery rates of approximately 40%. Most losses occur in the first trimester (approximately 65%). ***In the select population of women with a septate uterus and recurrent pregnancy loss, live birth rates are approximately 10% before hysteroscopic septum resection and 75–80% after surgery,[476, 505] indicating that hysteroscopic metroplasty restores an almost normal prognosis for term delivery.*** A 2010 systematic review of studies relating to outcomes after hysteroscopic septum resection concluded that the procedure results in fewer pregnancies in infertile patients than in those with recurrent miscarriage (RR=0.7, CI=0.5–0.9).[510] In the past, surgical correction of a septate uterus required abdominal metroplasty, risking post-operative adhesions that might impair fertility, and committed all future successful pregnancies to cesarean birth. Surgical treatment was reserved for women in whom the benefits of surgery more clearly outweighed the risks, but modern hysteroscopic surgery has changed the equation. Hysteroscopic septum resection is usually a relatively straightforward and brief outpatient procedure associated with low morbidity, no risk of adnexal adhesions or obligation to cesarean delivery, and a prompt and uneventful recovery; surgical indications now are appropriately more liberal.

Inevitably, systematic infertility evaluations will identify nulligravid women with a uterine septum who present a management dilemma. Given the high probability of successful hysteroscopic surgery and its low morbidity, we believe it is reasonable and appropriate to

consider preemptive surgical correction of a septate uterus, especially in women over age 35, women with infertility of long duration, women with other indications for surgical treatment, and women who require IVF or other treatments associated with increased risk of multifetal gestation and pregnancy loss.[476, 505, 511] Careful discussion of the relative risks and benefits of surgery is always important, but especially so when the indications for surgery are less clear.

Uterine Myomas

Myomas can be identified in 20–40% of all reproductive aged women and in 5–10% of infertile women[173, 512, 513]; myomas are the only abnormal finding in 1–2% of women with infertility. Although they are an established cause of abnormal bleeding, pain, and symptoms relating to pressure on adjacent organs, the impact of myomas on fertility has been more difficult to define, with the bulk of evidence coming from studies comparing the prevalence of myomas in fertile and infertile women, or the reproductive performance of women with otherwise unexplained infertility before and after myomectomy.[173, 174] Infertility relating to myomas has been attributed to all of the following mechanisms[514]:

- Displacement of the cervix, decreasing exposure to sperm.
- Enlargement or deformity of the uterine cavity, interfering with sperm transport.
- Obstruction of the interstitial segment of the fallopian tubes.
- Distorted adnexal anatomy, inferfering with ovum capture.
- Distortion of the uterine cavity, or increased or abnormal myometrial contractions, inhibiting sperm or embryo transport.
- Impaired uterine blood flow or chronic endometritis, interfering with implantation.

Whereas there is relatively little evidence to support the majority of these mechanisms, a number of observations lend credence to the notion that myomas may impair fertility by interfering with implantation. Glandular atrophy is commonly observed in the endometrium overlying myomas, depending on their proximity, and also can be seen in the opposing endometrium, suggesting it results from mechanical pressure.[515–517] Recent molecular studies indicate that submucous and intramural myomas also induce a local decrease in *HOX* gene expression, which has been implicated in the cascade of molecular events involved in implantation.[518]

The effects of myomas on fertility are best assessed by studies comparing IVF outcomes in infertile women with and without myomas, because IVF effectively controls for the confounding effects of other fertility factors. Numerous studies have examined the effects of myomas of varying size and location.[519–521] Altogether, these observations permit some conclusions regarding the effects of myomas on IVF outcomes, and by inference, on overall fertility.

There is a clear consensus that submucous myomas have significant adverse effect on clinical pregnancy rates (OR=0.3, CI=0.1–0.7) and delivery rates (OR=0.3, CI=0.1–0.8).[174, 514, 522–526] Available data also support the conclusion that submucous myomas increase risk for miscarriage by more than 3-fold.[525, 526] Results of early studies examining the effect of intramural myomas on IVF outcomes were inconsistent, with some finding adverse effects,[520, 521, 527–529] and others not.[519, 525, 530–534] A 2005 systematic review including six studies found that intramural myomas have significant negative impact on implantation rates (OR=0.62, CI=0.48–0.8) and live birth rates (OR=0.69, CI=0.5–0.95), and concluded that myomectomy deserved consideration, particularly in women with previous failed IVF cycles.[523] A 2007 meta-analysis of data from seven relevant studies also found evidence that intramural myomas adversely affect the clinical pregnancy rate (OR=0.8, CI=0.6–0.9) and delivery rate (OR=0.7, CI=0.5–0.8),[524] and a 2009 systematic review including 23 studies concluded that intramural myomas increase risk for miscarriage (RR=1.7, CI=1.2–2.4).[526]

All of the evidence concerning the effects of subserosal myomas is consistent in finding no evidence of adverse effects on IVF outcomes. ***In sum, the accumulated body of evidence indicates that submucous myomas reduce IVF success rates by approximately 70%, intramural myomas by approximately 30%, and subserosal myomas have no adverse impact on outcomes. Submucous myomas increase risk for miscarriage after successful IVF at least 3-fold, and intramural myomas by more than half.***

Logically, decisions regarding the management of infertile women with myomas should be guided by the evidence concerning their likely importance and the outcomes of surgical intervention. It seems clear that submucous myomas (distorting the uterine cavity) have important adverse effects on fertility and pregnancy outcomes and that myomectomy improves both. A 2009 systematic review of studies examining outcomes after submucous myomectomy concluded that clinical pregnancy rates achieved with IVF were 2-fold higher after surgery than in women with submuous myomas *in situ*, and comparable to those observed in women without myomas.[526] A randomized trial comparing the effects of myomecomy and expectant management on fertility in 181 women with a combination of submucous, intramural, and subserosal myomas observed that myomectomy significantly improved pregnancy rates among women with submucous myomas (43% vs. 27%) and those with both submucous and intramural myomas (26% vs. 15%), without other interventions.[535] Younger women having a single small submucous myoma and otherwise unexplained infertility have the best prognosis. Results are less encouraging for older women and those with multiple or large submucous myomas. Although complications of hysteroscopic myomectomy are relatively few, the risk of postoperative intrauterine adhesions increases with the size, number, and extent of intramural extension of submucous myomas.

Evidence for the benefits of myomectomy in women with intramural myomas (not distorting the uterine cavity) is less compelling, probably because their impact on fertility is not as great. Results of a cohort study suggest that myomectomy can improve cumulative clinical pregnancy and live birth rates after up to three IVF cycles in women having at least one intramural myoma larger than 5 cm in diameter.[536] A randomized trial observed a clinically significant trend toward improved fertility in women with intramural myomas after myomectomy (56% vs. 41%).[535] In contrast, the results of two other studies question the therapeutic value of myomectomy in asymptomatic infertile women with intramural myomas;[537, 538] Cumulative conception rates over the first 2 postoperative years related primarily to duration of infertility and the presence or absence of other infertility factors, but *not* to size or site (relationship to the uterine cavity) of the largest myoma removed. Increasing age and posterior myomas (associated with higher risk of postoperative pelvic and adnexal adhesions) were associated with a poorer prognosis, and symptoms (menorrhagia) with a better prognosis.

Decisions regarding the management of infertile women with asymptomatic intramural myomas are among the most difficult clinical judgments. They must consider not only the size, number, and location of myomas and the risks and benefits of the procedure, but also age, duration of infertility, ovarian reserve, other infertility factors, and the treatments they require. In most cases, the benefits of myomectomy are modest or uncertain, and the procedure is not without significant potential risks. Myomectomy commonly results in postoperative pelvic and adnexal adhesions, which can decrease fertility if severe,[538, 539] but are less concerning in women who require IVF for other reasons. Myomectomy generally commits the patient to cesarean delivery to avoid the risk of uterine rupture during labor, which has been reported after myomectomy.[540–543] Whereas excision of large, deep intramural myomas that abut or displace the uterine cavity might reasonably be expected to improve fertility, removal of smaller myomas having no direct anatomical relationship with the cavity probably will not. Whereas excision of anterior and fundal myomas is not likely to result in serious adnexal adhesions, posterior uterine incisions invite the complication. Arguably, excision of any intramural myomas large enough or deep enough to warrant myomectomy also likely warrants recommendation for cesarean delivery. Whereas

myomectomy offers limited, if any, benefits to young women with infertility of short duration and other infertility factors amenable to non-surgical treatments, it is less difficult to justify in older women with unexplained infertility of long duration planning to pursue IVF.

Adherence to basic microsurgical principles—gentle tissue handling, meticulous hemostasis, and minimal exposed suture—help to ensure best surgical results. Adjuvants such as local injection of aqueous pitressin, tourniquets to compress the uterine arteries, and surgical adhesion barriers aim at those goals. Laparoscopic and robotic myomectomy, performed by those having the requisite training and experience, may offer the same benefits as traditional open or minilaparotomy myomectomy for infertile women with intramural myomas, and have the added advantage of lower morbidity (decreased blood loss and shorter recovery time).[544–548] A multicenter, randomized trial comparing reproductive outcomes after laparoscopic and minilaparotomy myomectomy in women with unexplained infertility observed no differences in cumulative pregnancy, live-birth and miscarriage rates between the two procedures.[545] ***The careful selection of patients most likely to benefit from myomecomy is far more important than the choice of surgical technique. If the procedure has little or no likely benefit, the choice of technique is irrelevant.***

Intrauterine Adhesions (Asherman's Syndrome)

Intrauterine adhesions develop as a result of trauma.[549–552] Any insult severe enough to remove or destroy endometrium can cause adhesions. The gravid uterus is particularly susceptible to injury, especially between the second and fourth weeks postpartum.[553] Inflammation or infection also may predispose to adhesions.[554–556] In approximately 90% of cases, intrauterine adhesions relate to curettage for pregnancy complications, such as missed or incomplete abortion or retained products of conception.[557] Adhesions also can develop after abdominal or hysteroscopic myomectomy, septum resection, or other uterine surgery. In the developing world, genital tuberculosis is an important cause of intrauterine adhesions; although rare in the U.S., the possibility must be considered in women emigrated from regions where the disease is prevalent.[558]

Intrauterine adhesions can be asymptomatic or cause menstrual disorders (hypomenorrhea, amenorrhea, dysmenorrhea), pain, recurrent miscarriage, or infertility.[551, 552] The overall incidence of intrauterine adhesions is uncertain, but may be increasing.[557, 559] The risk of intrauterine adhesions associated with elective termination of pregnancy is generally low, but the prevalence and severity of adhesions may increase with the number of procedures.[560] A temporal relationship between symptoms and a predisposing event, the inability to pass a uterine sound, or a negative progestin challenge in amenorrheic women suggest the diagnosis. When suspected, HSG and saline sonohysterography confirm the presence of intrauterine adhesions. Compared to hysteroscopy (the gold standard), HSG has approximately 80% sensitivity and specificity for diagnosis of adhesions.[561] A study comparing HSG and sonohysterography with hysteroscopy concluded the two methods of imaging were equally sensitive for detection of adhesions,[478] but hysteroscopy is required to define the location and extent of disease.

Hysteroscopy can reveal a variety of findings.[549, 556, 562] Central adhesive bands can appear as columns or bridges between the opposing walls of the cavity, dividing it into smaller irregular chambers of varying size and shape. Adhesions at the margins of the cavity often appear as half-drawn curtains that may obscure one or both cornual orifices. Depending on their composition (mucosal, fibromuscular, connective tissue), adhesions may or may not have a surface of endometrium; dense connective tissue adhesions typically do not. Whereas mucosal adhesions generally appear similar to surrounding normal tissue and are easy to lyse, fibromuscular and connective tissue adhesions are thicker, typically pale, and must be

mechanically divided or dissected. Numerous classification systems have been proposed, but no system has gained wide acceptance or has prognostic value validated by prospective studies.[551, 552] Consequently, outcome studies are difficult to interpret and compare.

Hysteroscopy is the method of choice for treatment of intrauterine adhesions and is both safer and more effective than blind curettage. Often, lysis of adhesions can be accomplished using only the tip of the hysteroscope aided by the pressure provided by continuous infusion of distention media. When needed, an assortment of mechanical, electrosurgical, and laser-based instruments allows adhesions to be lysed or cut under direct vision. In general, best results are achieved when central adhesions are lysed first, moving from the lower uterine segment to the fundus and then to the margins of the cavity, gradually restoring normal cavity architecture. When disease is severe and anatomic landmarks are poorly defined, transabdominal ultrasonography or laparoscopy can help to maintain orientation and to limit the risk of uterine perforation.[563]

Various methods have been used to facilitate hysteroscopic surgery or to improve outcomes. In one randomized clinical trial examining the efficacy of vaginally administered misoprostol (200 μg) for cervical softening before operative hysteroscopy, treatment reduced or eliminated the need for mechanical dilation and the incidence of operative complications.[564] Various physical barriers, including both unmedicated IUDs and balloon catheters, are commonly used as a means to maintain separation between the opposing layers of endometrium during the immediate postoperative interval.[556, 557, 565] A study comparing outcomes after insertion of an IUD or a balloon catheter observed more frequent return of normal menses (81% vs. 63%) and higher conception rates (34% vs. 23%) in women receiving a catheter.[566] Postoperative treatment with exogenous estrogens to promote rapid re-epithelialization and reduce risks of recurrent adhesions is frequently used, but its efficacy has not been established;[567] a typical regimen involves treatment with 2.5–5 mg conjugated estrogens daily for 4 weeks, adding a progestin (e.g., medroxyprogesterone acetate 10 mg daily) during the last week.

Complications of hysteroscopic adhesiolysis are the same as with any operative hysteroscopic procedure and are relatively uncommon. Acute complications include uterine perforation, fluid overload and electrolyte imbalance, hemorrhage, and infection; late complications include recurrent adhesions and uterine rupture in a subsequent pregnancy.[568]

Surgical results should be evaluated by HSG or saline sonohysterography after menses.[569] A second operation to lyse persistent or recurrent adhesions may be required when disease is severe. Alternatively, pressure lavage with normal saline under guidance of transvaginal ultrasonography can be used to hydro-dissect recurrent adhesions that are not particularly dense or extensive.[570] Lysis using a balloon catheter under fluoroscopic control and local anesthesia or intravenous sedation also has been described.[571] Normal cyclic menses can be restored in from 70% to 90% of women with intrauterine adhesions, depending on severity.[549] Conception and term delivery rates after successful hysteroscopic lysis of intrauterine adhesions have ranged between 25% and 75%[549, 556, 572–578]; predictably, the prognosis is better for women with mild disease.

Endometrial Polyps

Endometrial polyps are hyperplasic endometrial growths having a vascular center and a sessile or pedunculated shape extending into the uterine cavity. They are generally rare in young women and increase in incidence with age. The overall prevalence of polyps in infertile women ranges between 3% and 10%.[478, 579–584] A number of molecular mechanisms have been implicated in their pathogenesis, including endometrial hyperplasia,[585] overexpresion

of endometrial aromatase,[586, 587] and gene mutations.[588] Saline sonohysterography is the most useful method of imaging for detection of endometrial polyps,[494, 589, 590] although false-positive results due to blood clots, mucus, and shearing of normal endometrium are not uncommon.

Careful, systematic evaluation inevitably will identify polypoid cavitary lesions in some infertile women. Differentiation of small submucous myomas and endometrial polyps can be difficult by any means other than hysteroscopy.[477] Whereas symptomatic women (abnormal bleeding) certainly merit hysteroscopic evaluation and treatment, whether surgery has benefits for asymptomatic infertile women with polyps is less clear. The observation that polyps are resistant to the actions of progesterone suggests they might interfere with implantation[591]; local inflammatory changes or distortion of the uterine cavity also have been implicated.[592]

Evidence from studies examining reproductive performance after hysteroscopic polypectomy is rather weak and conflicting.[175, 176, 592, 593] In a study of infertile women with documented but unresected endometrial polyps (>2 cm), IVF outcomes in treated (preliminary hysteroscopic polypectomy) and untreated women were not different.[176] In two studies examining outcomes in women with polyps (<1.5–2 cm) identified by ultrasonography during ovarian stimulation for IVF, pregnancy rates in women who proceeded to oocyte retrieval and embryo transfer or had hysteroscopic polypectomy after retrieval and later frozen embryo transfer were not different from those in women without polyps having fresh or frozen embryo transfers.[594, 595] The only evidence indicating that polyps adversely affect fertility derives from a study comparing outcomes after up to four cycles of IUI in a group of 215 infertile women with polyps who were randomized to receive preliminary polypectomy or no treatment; among 93 total pregnancies, 64 occurred in women having polypectomy and 29 in those who did not (RR=2.1, CI=1.5–2.9).[593] ***Taken together, the available evidence suggests that polypectomy may improve reproductive performance in infertile women. Treatment must be individualized, depending on the size of a polyp, associated symptoms, and on the circumstances leading to its discovery.*** [584, 596]

Chronic Endometritis

Chronic endometritis has been regarded traditionally as a distinct but uncommon cause of reproductive failure, but its true prevalence in infertile women is unknown.[597] Available evidence suggests that chronic subclinical endometritis is relatively common in women with symptomatic lower genital tract infections, including cervicitis and recurrent bacterial vaginosis[598–601] and may not be altogether rare even in asymptomatic infertile women.[602] Mucopurulent cervicitis is highly associated with chlamydia (*C. trachomatis*) and mycoplasma (*M. genitalium*) infection and both organisms, in turn, are associated with chronic endometritis, which likely plays a role in the pathogenesis of tubal factor infertility.[459, 601, 603–605] Although routine serologic testing for past chlamydia exposure, cervical cultures, and endometrial biopsy may be difficult to justify, further evaluation and treatment are appropriate and prudent in infertile women with clinical cervicitis, chronic or recurrent bacterial vaginosis, or other symptoms that suggest pelvic infection.

Tubal Factor: Tubal Occlusion and Adnexal Adhesions

Tubal and peritoneal pathology is among the most common causes of infertility and the primary diagnosis in approximately 30–35% of both younger and older infertile women.[349]

A history of pelvic inflammatory disease (PID), septic abortion, ruptured appendix, tubal surgery, or ectopic pregnancy strongly suggests the possibility of tubal damage. Unquestionably, PID is the major cause of tubal factor infertility and ectopic pregnancies. Classic studies in women with PID diagnosed by laparoscopy revealed that the risk of subsequent tubal infertility increases with the number and severity of pelvic infections; overall, the incidence is approximately 10–12% after one episode, 23–35% after two, and 54–75% after three episodes of acute PID.[606–610] The risk of ectopic pregnancy is increased 6- to 7-fold after pelvic infection. Although many women with tubal disease or pelvic adhesions have no known history of previous infection, evidence suggests strongly that "silent" ascending infection is the most likely cause.[601, 605] Many such women will have detectable chlamydia antibodies suggesting prior infection (discussed below). Other causes of tubal factor infertility include inflammation related to endometriosis, inflammatory bowel disease, or surgical trauma. Endometriosis is considered at length in Chapter 29; discussion here is focused on intrinsic tubal disease.

The mechanism responsible for tubal factor infertility obviously involves anatomic abnormalities that prevent the union of sperm and ovum. Proximal tubal obstructions prevent sperm from reaching the distal fallopian tube where fertilization normally occurs. Distal tubal occlusions prevent ovum capture from the adjacent ovary. Whereas proximal tubal obstruction is essentially an all-or-none phenomenon, distal tubal occlusive disease exhibits a spectrum ranging from mild (fimbrial agglutination) to moderate (varying degrees of fimbrial phimosis) to severe (complete obstruction). The likelihood or efficiency of ovum capture probably is inversely related to the severity of disease. Inflammatory damage to internal tubal mucosal architecture cannot be detected easily but may nonetheless impair sperm or embryo transport functions.

HSG and laparoscopy are the two classic methods for evaluation of tubal patency in infertile women and are complementary rather than mutually exclusive; each provides information the other does not and each has advantages and disadvantages. HSG images the uterine cavity and reveals the internal architecture of the tubal lumen, neither of which can be evaluated by laparoscopy. Laparoscopy provides detailed information about the pelvic anatomy that HSG cannot, including adhesions, endometriosis and ovarian pathology. HSG is performed in an outpatient setting, is far less costly than laparoscopy, and may have some therapeutic value[611]; it also is often uncomfortable or painful, involves some radiation exposure, and has risk of infectious complications that can further impair fertility.[612] Laparoscopy is more invasive, usually requires general anesthesia, provides no information regarding the uterine cavity (unless hysteroscopy is also performed), and involves the usual risks of surgery. Sonohysterosalpingography is similar to HSG, using ultrasonography and sterile saline instead of fluoroscopy and contrast media, and is another, but less common, method for evaluating tubal factor. Chlamydia antibody tests represent a fourth, albeit indirect, method for evaluating tubal factor that is relatively inexpensive and minimally invasive.[613–616] Chlamydia antibody tests have been used primarily for screening infertile women to identify those at high risk for having tubal disease who merit evaluation with laparoscopy.

Hysterosalpingography (HSG)

HSG is best scheduled during the 2–5 day interval immediately following the end of menses, to minimize risk for infection, avoid interference from intrauterine blood and clot, and to prevent any possibility that the procedure might be performed after conception. Even the most sensitive assays for hCG cannot exclude the possibility when HSG is performed during the early luteal phase of the cycle. HSG does not require any specific preparation, although pretreatment with a NSAID (30–60 minutes before) is helpful to decrease

discomfort associated with the procedure; more potent analgesics and sedatives generally are not required. Infectious complications from HSG are relatively uncommon, even in high risk women (1–3%).[612, 617] Nonetheless, routine prophylactic antibiotic treatment can be justified, considering the potential consequences of a post-procedure infection. ***Treatment with antibiotics (doxycycline 100 mg twice daily for 5 days, beginning 1–2 days before HSG) is prudent when tubal disease is highly suspected, and specifically indicated when HSG reveals distal tubal obstruction, because risk for acute salpingitis is increased (approximately 10%) and treatment can prevent clinical infection.***[612, 618] To minimize the risk of infection, HSG is best avoided altogether for at least several weeks following any episode of acute PID.

The technique for performing an HSG is quite standard. The study should be performed using image intensification fluoroscopy with a limited number of radiographs. The average HSG requires only 20–30 seconds of fluoroscopic time with minimal radiation exposure and has very low risk. Usually, only three basic films are required (a scout, one film to document the uterine contours and tubal patency, and a post-evaluation film to detect any areas of contrast loculation). Additional oblique films may be needed when the uterus obscures the tubes or the uterine cavity appears abnormal. Otherwise, they provide little or no more useful information and increase radiation exposure unnecessarily.[619] Contrast can be introduced using a common metal "acorn" cannula or via a balloon catheter. In general, the latter technique requires less fluoroscopic time, smaller volumes of contrast, produces less pain, and is easier to perform.[620] Slow injection of contrast (typically 3–10 mL) helps to minimize the discomfort associated with HSG.

Debate regarding the relative advantages and disadvantages of oil- and water-soluble contrast media has raged for years. Advocates of water-soluble contrast media emphasize that oil-soluble media is too viscous to reveal internal tubal architecture (having prognostic significance),[621] disperses poorly in the pelvis (and therefore cannot detect adnexal adhesions), and has significant risks (granulomatous reactions, intravasation, and embolism).[622, 623] Those favoring oil-soluble contrast media argue that granulomatous reactions are rare, that intravasation and embolization are uncommon and almost uniformly benign,[624] and cite evidence suggesting that oil-soluble media increases fertility in the months immediately following HSG in women with patent tubes.[611] A 2007 systematic review of 12 studies involving 2,079 patients concluded that tubal perfusion with oil-soluble contrast significantly increased the likelihood of pregnancy, compared to no intervention (OR=3.30, CI=2.0–5.43), but not compared to perfusion with water-soluble contrast (OR=1.21, CI=0.95–1.54). Consequently either choice of media is appropriate.

HSG may reveal bilateral tubal patency (60–75%) or unilateral (15–25%) or bilateral (15–25%) tubal occlusion.[625, 626] Both false-negative (obstructions that are not real) and false-positive results (patency that is not real) occur, the former being much more common than the latter. Injection of contrast may cause "cornual spasm" (uterine contractions that transiently close the interstitial segment and prevent distal perfusion) that can be misinterpreted as proximal tubal occlusion. HSG may reveal unilateral tubal patency and contralateral proximal occlusion. Although the observation may represent a true unilateral proximal obstruction, which is rare, catheter placement allowing contrast to take the path of least resistance is the more common cause; most often, the non-visualizing tube is normal. A false-positive HSG may occur when contrast entering a widely dilated hydrosalpinx is diluted to yield a blush that is misinterpreted as evidence of tubal patency. Peritubular adhesions surrounding an otherwise normal and patent tube can sequester contrast as it escapes from the tube, resulting in a focal loculation that can be misinterpreted as distal obstruction.

Compared to laparoscopy (the gold standard method) as a test of tubal patency, HSG has only moderate sensitivity (ability to detect patency when the tubes are open; 65%), but

relatively high specificity (accuracy when patency is detected; 83%) in a typical infertile population.[627, 628] ***The clinical implications are that when HSG reveals obstruction there is still a relatively high probability (approximately 60%) that the tube is open, but when HSG demonstrates patency there is little chance the tube is actually occluded (approximately 5%).*** However, interpretation of HSG results can vary significantly among different observers.[629, 630] Consequently, when the treating clinician has not performed the HSG, a personal review of the films is prudent before making recommendations for additional evaluation or treatment. The probability of treatment-independent pregnancy is highest when HSG reveals bilateral tubal patency, substantially lower when neither tube is open, and reduced only slightly when one tube is patent.[625, 626] These observations help in deciding whether laparoscopy is needed before starting treatment.

Laparoscopy

Laparoscopy is regarded generally as the definitive test for the evaluation of tubal factors. Issues concerning scheduling, the use of antibiotics, and the risks of infectious complications are the same as for HSG. Diagnostic laparoscopy is usually performed under general anesthesia, but may require only deep sedation and local anesthetic; operative laparoscopy for treatment of disease typically requires general anesthesia. With few exceptions, a systematic and thorough inspection of the pelvis will accurately define the location and extent of any disease. Examination should include the uterus, the anterior and posterior cul-de-sacs, the ovarian surfaces and fossae, and the fallopian tubes. Injection of a dilute blue dye through a cannula attached to the cervix or an intrauterine manipulator permits evaluation of tubal patency ("chromotubation"). Indigo carmine is preferred over methylene blue, which rarely may induce acute methemoglobinemia, particularly in individuals with glucose-6-phosphate dehydrogenase deficiency.[631, 632] As with HSG, slow injection of fluid helps to reduce the incidence of false-negative results.

Laparoscopy provides both a panoramic view of the pelvic reproductive anatomy and a magnified view of the uterine, ovarian, tubal, and peritoneal surfaces. Consequently, it can identify milder degrees of distal tubal occlusive disease (fimbrial agglutination, phimosis), pelvic or adnexal adhesions, and endometriosis that adversely affect fertility but escape detection by HSG. Most importantly, laparoscopy offers the opportunity to treat disease at the time of diagnosis. Lysis of filmy or focal adhesions and excision or ablation of superficial endometriosis are relatively simple procedures well within the capabilities of most surgeons. Excision of ovarian endometriomas, lysis of dense or extensive adhesions involving the cul-de-sac or bowel, excision or ablation of widely disseminated or deeply invasive endometriosis, and fimbrioplasty or salpingoneostomy procedures require greater technical skill and experience.

Although laparoscopy is a better predictor of future fertility than HSG, it is not a perfect test for diagnosis of tubal pathology. Intraoperative chromotubation is subject to the same pitfalls causing false-negative results with HSG. False-positive results with laparoscopy are uncommon but do occur, particularly in cases where the fallopian tubes are obscured by adhesions. Whereas tubal obstructions detected by HSG are frequently not confirmed at laparoscopy, patency almost always is. Laparoscopy also is a better predictor of future treatment-independent pregnancy than HSG because the information gained is more accurate. Again, the prognosis is best when both fallopian tubes are patent, poor when both are blocked, and intermediate when only one tube is open.[626, 633] Because many obstructions detected by HSG are not real and all but a few of those identified by laparoscopy are, the prognoses associated with unilateral and bilateral tubal occlusion diagnosed by laparoscopy are significantly worse than when the same diagnosis is made by HSG.

Sonohysterosalpingography

Sonohysterography is recognized as having greater sensitivity than HSG for detection of intrauterine pathology. A natural extension of that technique, sonohysterosalpingography, has been viewed as a means to evaluate tubal patency at the same time, much like HSG. As originally described, sonohysterosalpingography relied on observations of fluid accumulation in the cul-de-sac as an indication of tubal patency. However, the technique provided no information regarding tubal anatomy and could not determine whether one or both tubes were patent. A new sonographic contrast media consisting of a surfactant that produces microbubbles when stimulated by ultrasound improved sensitivity for detecting tubal patency, but standard two-dimensional imaging in the sagittal and transverse planes was still inadequate to visualize the three-dimensional tubal anatomy.

Technological advances in ultrasonography have expanded the capabilities of sonohysterosalpingography further; three-dimensional transvaginal ultrasonography provides the means to generate coronal images and Doppler techniques have improved visualization of fluid movement through the fallopian tubes. However, even with these improvements, it is unlikely that sonohysterosalpingography will replace traditional HSG anytime soon. Studies directly comparing results of sonohysterosalpingography with those obtained by HSG or laparoscopy have yielded inconsistent results.[634–638] The fallopian tube remains difficult to image with ultrasonography, even with three-dimensional equipment, and sonohysterosalpingography has its own unique pitfalls.[639] A 2006 study comparing results with laparoscopy found that three-dimensional sonohysterosalpingography had excellent sensitivity (100%) and moderate specificity (67%) for detecting tubal patency (100%), but 30% of patients judged the procedure unacceptable.[640] Sonohysterosalpingography may yet become a viable alternative to HSG, but currently is not.

Chlamydia Antibody Tests

A number of studies have suggested that chlamydia antibody tests can be as accurate as HSG or even laparoscopy for detection of tubal pathology, including tubal occlusion, hydrosalpinx, and pelvic adhesions.[613, 614, 641] The performance of the different tests varies widely with the assay method. Commercial assays differ in detection method (immunofluorescence, microimmunofluorescence, ELISA, immunoperoxidase) and in the source of antigen they use (general or genus-specific major outer membrane proteins, an inactivated organism, whole-cell inclusion). Some methods are highly specific for the chlamydia species of interest (*C. trachomatis*) and others do not distinguish antibodies to *C. trachomatis* from those directed against other chlamydia species (*C. pneumoniae*, *C. psittaci*). As expected, tests having the greatest specificity for *C. trachomatis* perform best for detection of tubal pathology.[614, 642, 643] Practical considerations suggest that a rapid, highly sensitive but less specific assay is the most suitable test for screening, using a more specific test to confirm the antibody specificity of sera selected by the screening assay.

The predictive value of any diagnostic test depends on the prevalence of the disease of interest in the population tested. If the prevalence of disease in the population is very low or very high, diagnostic testing has little or no value because the outcome rarely affects management, and false-positive (when the prevalence is very low) or false-negative test results (when the prevalence is very high) are common. Diagnostic tests tend to have greatest utility when the prevalence of disease is somewhere in between the extremes.[628] Some have suggested that chlalmydia antibody tests might be used to select patients likely to

benefit most from laparoscopy, but the predictive value of even some of the more specific chlamydia antibody tests may not be sufficient to justify that approach.[644]

The role for chlamydia antibody tests in the evaluation of infertile women has not been sufficiently defined. Chlamydia antibody tests could prove useful as a pretest to select women who warrant earlier or more detailed evaluation.[645] If applied as a screening tool early in evaluation, a positive test might alert one to the possibility of tubal factors relating to previous chlamydia infection not otherwise suspected. Although selective laparoscopy based on chlamydia antibody tests may be unjustified for all infertile women,[644] it might be effective if limited to women with unexplained infertility (including a normal HSG), identifying those most likely to have undetected tubal factors best addressed before starting aggressive and costly empirical treatments. The utility of chlamydia antibody tests in these or other clinical contexts is uncertain but warrants further investigation. ***In summary, chlamydia antibody tests can provide useful information, but also have pitfalls that limit their clinical utility.***

Tubal Surgery in the Era of ART

For women with tubal factor infertility, treatment options are reconstructive surgery and IVF. Over the last 2 decades, IVF success rates have increased steadily (from approximately 10% to over 40%) and now frequently exceed those achieved with surgery.[34] Consequently, IVF has become the treatment of choice for much or most tubal factor infertility, particularly for couples with other infertility factors or severe tubal disease. However, surgery remains an appropriate option in select circumstances and for couples with ethical or religious objections or financial restrictions that preclude IVF. The indications, preliminary evaluation, techniques, risks, and outcomes for IVF and other forms of ART are the focus of a separate chapter (Chapter 32); discussion here is limited to surgical treatments for tubal factor infertility and the choice between surgery and IVF.

Sterilization Reversal

Approximately 1 million U.S. women have an elective tubal sterilization procedure each year; up to 7% regret the decision and about 1% later request its reversal.[22, 646] The most commonly cited reasons for sterilization reversal requests include new relationships, changes in family planning goals, and death of a child. Regrets are more common in younger women, those who were unaware of the spectrum of contraceptive options, women whose decision for sterilization was influenced by a third-party (partner, other family member, friend, or physician), and those sterilized postpartum or after an abortion.[647, 648] Women 30 years old or younger are twice as likely as older women to express regret, 3.5 to 18 times more likely to request information about reversal of the procedure, and approximately eight times more likely to actually have a sterilization reversal or IVF.[649] For women who want to conceive again, tubal anastomosis is a legitimate option. A preoperative HSG can be useful to assess the proximal segments and to confirm the type of sterilization performed. Laparoscopy may occasionally be necessary to assess the feasibility of surgical repair when the type of procedure is unknown and when destruction or removal of large segments of tube or other pelvic pathology is suspected; otherwise fewer than 5% of women will have irreparable tubes.[650]

The prognosis for achieving a live birth after microsurgical sterilization reversal relates to age, the type and location of procedure, and the final length of the repaired fallopian

tubes. Younger women, those whose sterilization was performed using rings and clips, and women having no other infertility factors have the best prognosis; success rates are lower for older women, those who were sterilized by cautery (particularly multiple-burn techniques), and women with other infertility factors.[651–658] Cumulative pregnancy rates are similar when one or both tubes are repaired, although the time to conception is longer after unilateral anastomosis.[657] ***In properly selected candidates, overall conception rates are generally quite good (45–82%) after microsurgical sterilization reversal.*** Risk for ectopic pregnancy ranges between 1% and 7% and is higher after isthmic-ampullary than after isthmic-isthmic anastomoses.[659, 660] Among all surgical treatments for tubal factor infertility, sterilization reversal has the highest postoperative fecundability. ***Best candidates for the procedure are young women desiring more than one additional pregnancy and having no other infertility factors.*** Compared to IVF, the primary advantages of surgery are the opportunity for natural conception and lower risk for multiple gestation; the disadvantages of surgery include the surgical insult itself, a higher risk for ectopic pregnancy, and the need for future contraception. Laparoscopic tubal anastomosis is an option for highly skilled surgeons experienced in the technique, although success rates may be somewhat lower (25–53%).[661, 662] Early experience with robotic tubal anastomosis indicates that operating time is modestly greater, but hospital stay and recovery time are shorter, compared to open microsurgical procedures[663, 664]; pregnancy rates are comparable, but risk for ectopic pregnancy may be increased.[664]

Distal Tubal Obstruction

Distal tubal occlusive disease exhibits a wide spectrum of severity ranging from adherent fimbrial folds, to varying degrees of phimosis, to complete obstruction with hydrosalpinges. HSG generally will reveal complete distal tubal obstructions but cannot reliably detect or accurately define lesser degrees of disease when the tubes are still patent. Laparoscopy is the definitive method for diagnosis of distal tubal occlusive disease and also provides the means for treatment. Fimbriolysis refers to the separation of adherent fimbria, fimbrioplasty describes the correction of phimotic but patent fimbria, and neosalpingostomy involves the reopening of a completely obstructed tube. Predictably, surgical success inversely relates to the severity of disease. The extent and character of associated tubo-ovarian adhesions, tubal thickness, and the condition of the internal ampullary mucosal architecture are all variables that affect prognosis.[665, 666] For the milder forms of distal tubal disease, postoperative live birth rates can exceed 50%.[667–669] Results achieved with surgery for more severe disease have varied widely but success rates are lower (10–35%) and risk for ectopic pregnancy is higher (5–20%).[666, 670–672] Postoperative tubal patency rates far exceed pregnancy rates; patency is more easily restored than function because mucosal regeneration is slow and often fails altogether.[673, 674]

The majority of pregnancies occur within the first 2 years after surgical treatment of distal tubal obstruction. In general, the results achieved by experienced surgeons using traditional microsurgical techniques or laparoscopic methods have been similar. In a case series of 35 women with distal tubal occlusion treated by laparoscopic fimbrioplasty followed for at least 2 years after surgery, the global conception rate was 74%, the intrauterine pregnancy rate was 51%, the live birth rate was 37% and the ectopic pregnancy rate was 23%.[675] ***In younger women with mild distal tubal occlusive disease, laparoscopic surgery may be viewed as an alternative to IVF, but when disease is severe or pregnancy does not occur during the first postoperative year, IVF is the logical choice. For older women with any significant degree of distal tubal disease, IVF is generally the first and best option because cycle fecundability after distal tubal surgery is low (1–2%), time is limited, and IVF is both more efficient and more effective.***[676]

As success rates with IVF have improved steadily, the indications for reconstructive surgery in women with distal tubal occlusive disease have further declined. However, women with severe distal tubal disease still can benefit from surgery (salpingectomy) because a substantial body of evidence indicates that large hydrosalpinges adversely affect IVF outcomes. Several mechanisms have been implicated to explain the observation, including mechanical interference with implantation and toxic effects on the embryo or endometrium.[677–679] A 2010 systematic review including five randomized controlled trials involving 646 women observed that the odds of achieving an ongoing pregnancy were twice as great after laparoscopic salpingectomy for hydrosalpinges before IVF (OR=2.14, CI=1.23–3.73).[680] Laparoscopic occlusion of the fallopian tubes increased the odds of clinical pregnancy, compared to no intervention (OR=4.66, CI=2.47–10.01), and neither surgical procedure was superior.[680] ***These data demonstrate clearly that laparoscopic salpingectomy or tubal occlusion improve IVF pregnancy rates in women with hydrosalpinges.*** Other treatment strategies, such as ultrasound-guided aspiration of hydrosalpingeal fluid at the time of oocyte retrieval, have been suggested as an alternative treatment,[681] but their effectiveness has not been established and evidence suggests the fluid re-accumulates rapidly.[682]

Proximal Tubal Obstruction

Proximal tubal occlusions represent approximately one-third of all tubal obstructions observed with HSG, many of which are not real (20–40%). ***Efforts to establish a certain diagnosis of true proximal tubal occlusion are justified; otherwise, many women may needlessly undergo major surgery or IVF.*** Repeated HSG can decrease the number of false-negative tests of tubal patency; in a case series including 98 infertile women with a diagnosis of proximal tubal occlusion based on an HSG, repeating the procedure revealed bilateral tubal patency in 14 patients (14%), patency of one tube in 12 others (12%), and confirmed bilateral occlusion in 72 patients (74%).[683] In many, if not most, laparoscopy is required to establish an accurate diagnosis, also providing the opportunity to treat coexisting tubo-ovarian disease that may be observed in up to 20% of women.[684–686] The pathogenesis of proximal tubal occlusive disease is not well understood; most is presumed to result from infection or chronic inflammation. Histologic studies suggest that obliterative luminal fibrosis is most common, followed by salpingitis isthmica nodosa, chronic inflammation, and intratubal endometriosis.[687, 688]

Microsurgical segmental tubal resection and anastomosis is a proven treatment for true proximal tubal obstruction. Experienced surgeons can achieve pregnancy rates ranging between 50% and 60%,[688–691] but the number of surgeons having the necessary expertise is fast declining. Outcomes vary with the cause of the obstruction; reocclusion rates are relatively high with causes other than salpingitis isthmica nodosa. Proximal tubal cannulation using hysteroscopic or fluoroscopic methods is a proven alternative to traditional microsurgical repair. In case series, patency rates between 60% and 80% and pregnancy rates between 20% and 60% have been observed,[635, 683, 684, 692–695] with less morbidity and lower cost. The specialized catheter systems involved require some training and experience but allow selective tubal perfusion for accurate diagnosis (true occlusion or not) and provide the means for treatment when needed.

Bipolar tubal disease involves both proximal and distal tubal obstruction. In general, success rates achieved with surgery have been extremely poor and IVF represents the best treatment option.[690, 696, 697]

SUMMARY

Since only the best surgeons generally publish their results, the best available estimates from surgical series also very likely represent the best possible outcomes. Even so, steady advances in ART have improved IVF outcomes to where they now equal or exceed what can be achieved with tubal reconstructive surgery. Accordingly, surgical treatments for tubal factor infertility are generally in an era of decline; laparoscopic surgery has replaced simple open procedures, and ART has replaced more complicated ones. Tubal surgery remains a legitimate treatment option for women seeking pregnancy after a previous tubal sterilization, for those with mild distal tubal disease (particularly when they are young), and for some women with proximal tubal occlusion. Under virtually all other circumstances, IVF is the best choice. Laparoscopic salpingectomy or proximal tubal occlusion increases IVF success rates by 2-fold and should be recommended to all women with hydrosalpinges planning IVF.

Unexplained Infertility

Unexplained infertility is a diagnosis of exclusion, after systematic evaluation fails to identify a cause. The incidence of unexplained infertility ranges from 10% to as high as 30% among infertile populations, depending on diagnostic criteria.[698–700] ***At a minimum, the diagnosis of unexplained infertility implies evidence of normal semen quality, ovulatory function, a normal uterine cavity, and bilateral tubal patency.*** In the past, the diagnosis also required a "positive" postcoital test (excluding cervical factor infertility) and "in phase" endometrial dating (excluding luteal phase deficiency), but no longer, because the tests have proven invalid. In the past, the diagnosis also required laparoscopy (excluding pelvic adhesions and endometriosis), but laparoscopy is no longer performed routinely, because evidence indicates it has very limited impact on overall outcomes among women with unexplained infertility. Instead, transvaginal ultrasonography is performed to detect unsuspected ovarian pathology, such as endometriomas. Consequently, much of infertility previously attributed to cervical factors, luteal phase deficiency, and mild endometriosis or adhesions is now "unexplained."

Excluding false-negative results of standard diagnostic tests, which do occur but are uncommon, there are two potential explanations for unexplained infertility: 1) there truly is no abnormality and the couple's natural fertility is at the extreme lower end of the normal range, possibly due to female partner age or advanced reproductive aging; and 2) there is a specific cause, but not one that can be identified with existing diagnostic tests.

Undoubtedly, much of unexplained infertility relates to the natural decline in fertility with increasing age. Unexplained infertility is more common in women over age 35; in a study involving over 7,000 infertile women, those over the age of 35 years were nearly twice as likely to have unexplained infertility (OR=1.8, CI=1.4–2.7).[350] ***Logically, the most likely occult causes of infertility relate to abnormalities in gametes or implantation, for which there is no valid diagnostic test.*** Genetic or functional abnormalities in zona pellucida proteins could interfere with sperm penetration and cause fertilization failure.[701] Abnormalities in the centrosome could interfere with normal spindle formation and function, preventing

fertilization or resulting in arrested early embryonic development.[702] Although failed fertilization occurs in less than 5% of IVF cycles and does not always reoccur in subsequent cycles,[703, 704] a marked decrease in fertilization efficiency easily could result in unexplained infertility. A higher incidence of fertilization failure has been observed in several, but not all, studies of IVF outcomes in couples with unexplained infertility.[705–708] Evidence that up to 75% of human pregnancies fail soon after conception implicates early embryopathy and implantation failure as likely causes of unexplained infertility.[39, 709, 710] Although aneuploidy is common in early human embryos,[711, 712] a recurring nonrandom genetic defect in the embryo or trophectoderm could cause early loss. Intrinsic genetic abnormalities in endometrial function and receptivity could interfere with apposition, adhesion, attachment, or invasion of the embryo, causing implantation failure.[713–715] ***It is important to emphasize that all of the potential causes of unexplained infertility could co-exist with known causes for infertility, helping to explain why many couples with identified ovarian, male, uterine, or tubal infertility factors fail to achieve a successful pregnancy despite receiving proven effective treatments.***[716]

Unexplained infertility likely represents either the lower extreme of the normal distribution of reproductive efficiency or abnormalities of sperm or oocyte function, fertilization, implantation, or embryo development that cannot be detected reliably by standard methods of evaluation. Although many couples with unexplained infertility may be expected to conceive without treatment, their already low and steadily declining cycle fecundity provides ample justification for offering treatment to those concerned enough to seek evaluation. The goal of treatment is to increase monthly fecundability to a level more closely approximating that observed in normally fertile couples.

The prognosis for untreated couples with unexplained infertility is similar to that for couples with minor infertility factors, such as mild oligospermia or endometriosis; age of the female partner and duration of infertility are the primary variables that affect pregnancy rates.[353, 717, 718] ***In studies evaluating treatments for unexplained infertility, untreated patients have a cycle fecundability ranging typically between 2% and 4%,[719] or about 80–90% lower than in normal fertile couples (20–25%) The likelihood of pregnancy without treatment decreases progressively with increasing age of the female partner and increasing duration of infertility.***[353, 720] After 3 years of infertility, the likelihood of pregnancy without treatment falls to approximately 40%, and after 5 years to about 20%, of what it was when efforts to conceive first began.[343] Only approximately 14% of couples with unexplained infertility managed expectantly for up to 7 years achieve a pregnancy resulting in a live birth within a year; the prognosis is better when the female partner is under age 30.[353, 718] ***The effect of duration of infertility is important to understand. Because spontaneous pregnancy rates are highest among couples with a relatively short duration of infertility and success rates achieved with all forms of treatment for unexplained infertility other than IVF are similar, treatments can appear more effective in couples with a longer duration of infertility having a lower probability for conceiving without treatment.***

By definition, the cause of unexplained infertility is unknown. Consequently, all treatments for unexplained infertility are empiric. Although methods differ, the basic strategy is the same for all—to bring together more than the usual numbers of oocytes and sperm in the right place at the right time. To this end, the most common treatments include intrauterine insemination (IUI), ovarian stimulation with clomiphene or gonadotropins and IUI, and IVF. It is important to realize that none of the current treatments for unexplained infertility targets the most likely causes, which all involve events occurring during or after fertilization. Empiric treatments for unknown disorders cannot be expected to achieve dramatic results. In small studies, modest effects can be difficult to demonstrate, and large effects can occur by chance.

Intrauterine Insemination (IUI)

Although several studies have examined the effectiveness of intrauterine insemination (IUI) as treatment for unexplained infertility in natural cycles,[449, 467, 719, 721, 722] a 2006 meta-analysis concluded that none provided reliable data because of problems with design, such as cross-over trials that do not include data from the first phase of the study or populations not limited to couples with unexplained infertility.[723] The two most informative studies were published more recently and included only couples with unexplained infertility or an abnormal postcoital test, with expectant management as the control treatment.[724, 725] In the first trial (average age 32 years, average duration of infertility 2.5 years), 43 live births were observed among 191 couples receiving IUI (23%) over 6 months, compared to 32 in 193 couples (17%) managed expectantly.[724] Although the effect difference (6% over 6 months) was not significant (OR=1.46, CI=0.88–2.43), more women randomized to IUI judged their treatment acceptable. In the second trial (average age 30 years, average duration of infertility 1.7 years), 11 ongoing pregnancies were observed among 51 couples receiving IUI (22%), compared to 9 in 48 couples (19%) managed expectantly.[725] ***The best available evidence suggests that treatment with IUI in natural cycles has no clinically important effects.***

Clomiphene Citrate and IUI

Numerous studies have examined the effectiveness of clomiphene therapy without IUI as treatment for unexplained infertility.[726–729] However, only two are truly informative trials, including only patients with unexplained infertility, using placebo or expectant management as the control treatment.[724, 730] In one trial (average age 30 years, average duration of infertility 4.3 years), 10 pregnancies were observed among 76 couples (13%) receiving clomiphene treatment over 290 cycles (3%/cycle), compared to 4 in 72 couples (6%) receiving placebo over 274 cycles (1%/cycle).[730] In the other (average age 32 years, average duration of infertility 2.5 years), 26 pregnancies were observed among 192 couples receiving clomiphene (14%), compared to 32 in 193 couples (17%) managed expectantly.[724] The differences between treatment and control pregnancy rates (per couple or per cycle) were not significant in either trial. ***Although clomiphene is commonly used as a treatment for unexplained infertility, the best available evidence indicates it has no significant benefit.***

Combined treatment with clomiphene and IUI is commonly recommended for couples with unexplained infertility, but evidence for its effectiveness is quite limited. In a review of eight studies involving 932 treatment cycles, the estimated cycle fecundity was 5.6% with clomiphene and 8.3% with clomiphene and IUI.[719] The one trial (average age 33 years, average duration of infertility 3.5 years) including an untreated control group (timed intercourse), included patients with unexplained infertility or treated endometriosis.[731] Limiting analysis to cycles observed before cross-over, eight pregnancies were observed in 23 couples (35%) receiving clomiphene and IUI over 73 treatment cycles (11%/cycle), compared to 4 in 28 couples (14%) over 103 cycles (4%/cycle). The 7.1% absolute difference (CI=-1.0–15.2) in cycle fecundability was not significant, and even if it were, the treatment effect was quite modest; the calculated number needed to treat was 15, implying that one additional pregnancy might be expected for every 15 treatment cycles.

Results of three other cross-over trials involving control groups receiving an active treatment (instead of placebo or no treatment) are difficult to interpret confidently, because no data were provided for the first phase of the study.[732–734] A fourth management trial (the fast track and standard treatment "FASTT" trial) compared outcomes in two groups, one

randomly assigned to receive three cycles of treatment with clomiphene and IUI followed by up to six cycles of IVF, and the other assigned to receive three cycles of clomiphene and IUI, followed by three cycles of treatment with gonadotropins and IUI, followed by up to six cycles of IVF.[735] Notably, 55 pregnancies were observed among 233 couples over 646 treatment cycles (8.5%/cycle) in the first group and 68 in 242 couples over 648 treatment cycles (10.5%/cycle) in the second; overall, 123 pregnancies were observed in 475 couples (26%) over 1,294 cycles (9.5%/cycle). The overall pregnancy rate compares favorably with the expected 2–4% cycle fecundability among couples with unexplained infertility, which supports the use of clomiphene and IUI in the treatment of unexplained infertility. In two large retrospective studies involving a total of more than 8,000 cycles of treatment with clomiphene and IUI, cycle fecundability ranged between 5% and 10% per cycle after four to six cycles for women age 40 years and younger, and were under 5% for those over age 40.[736, 737]

In sum, evidence for the effectiveness of combined treatment with clomiphene and IUI is not compelling. However, considering its relatively modest cost and complexity (compared to the alternatives, discussed below), treatment with clomiphene and IUI seems justified because the cycle fecundability observed in large prospective and retrospective studies is significantly higher than can be expected in couples with unexplained infertility receiving no treatment.

Gonadotropins and IUI

Gonadotropin therapy without IUI for treatment of unexplained infertility has been evaluated in only a few clinical trials. In the largest, pregnancy rates resulting from treatment with gonadotropins and intracervical insemination were higher than was achieved with insemination alone, but the difference was small (3.6%).[738] ***Although treatment with gonadotropins alone can increase cycle fecundability, compared with no treatment, the effect is quite modest and no better than can be achieved by treatment with clomiphene and IUI.***

More commonly, gonadotropin treatment is combined with IUI for the treatment of unexplained infertility. Among four trials comparing gonadotropins and IUI with no treatment, two were cross-over trials providing no results for the first phase of treatment.[739] In a U.S. trial (average age 32 years, average duration of infertility 3.6 years), 77 pregnancies were observed among 231 couples (33%) receiving treatment with gonadotropins and IUI over 618 cycles (12%/cycle), compared to 23 pregnancies in 233 couples (10%) receiving intracervical insemination over 706 cycles (3%/cycle); pregnancy rates per couple were 18% for treatment with insemination alone and 19% for gonadotropins and IUI.[738] A Dutch trial (average age 33 years, average duration of infertility 2 years) observed 29 pregnancies among 127 couples (23%) receiving gonadotropins and IUI over 676 cycles (4%/cycle), compared to 34 in 126 couples (27%) managed expectantly over 737 cycles (5%/cycle).[740]

The differing results of the two trials emphasize again the influence of the duration of infertility on outcomes achieved with treatment for unexplained infertility. In the U.S. trial, involving couples infertile for an average of 3.6 years, fecundability in those receiving treatment with gonadotropins and IUI (12%/cycle) was 9% higher than in couples receiving intracervical insemination (3%/cycle), and only 10% of couples in the latter group conceived. In the Dutch trial, involving couples with an average of 2 years of infertility and a better prognosis for achieving pregnancy without treatment,[718] fecundability of those receiving gonadotropins and IUI (4%/cycle) was no better than in couples managed expectantly (5%/cycle), and 27% of couples receiving no treatment conceived. Together, the results of the two trials indicate that treatment with gonadotropins and IUI has little benefit when the prognosis is reasonably good, and modest benefit when the prognosis is poor (one additional pregnancy for every 11 treatment cycles).

The results of treatment with gonadotropins and IUI for unexplained infertility raise two clinically relevant questions. The first concerns what benefits treatment with gonadotropins and IUI might have in couples first treated with clomiphene and IUI and failing to conceive. The only data addressing the question directly comes from the "FASTT" trial described above, in which 50 pregnancies were observed among 169 couples (30%) receiving treatment with gonadotropins and IUI over 439 cycles (11%/cycle) after failing to conceive over three cycles of treatment with clomiphene and IUI.[735] Although cycle fecundability (11%/cycle) was slightly higher than was achieved with clomiphene and IUI in the same population (9.5%/cycle), the difference is not clinically important, especially when considering the greater costs, complexity, and risks associated with use of gonadotropins. Consistent with that view, a 2002 systematic review of trials comparing outcomes of treatment with clomiphene/IUI and gonadotropins/IUI concluded that evidence is insufficient to suggest that either treatment is superior.[741] The second question relates to whether success with clomiphene and IUI depends on multifollicular development, and there are no reliable data that address the question directly.

A number of studies have examined the efficacy of various adjuvant treatments in couples receiving treatment with gonadotropins and IUI for unexplained infertility. The available evidence indicates that whereas pre-treatment with a GnRH agonist does not improve outcomes,[742] adding a GnRH antagonist to the treatment regimen can (OR=1.6, CI=1.1–2.3).[743]

In summary, treatment with gonadotropins and IUI is modestly effective treatment for couples with longer durations of unexplained infertility (>3years). Treatment with gonadotropins and IUI is reasonable to consider for couples who fail to conceive during treatment with clomiphene and IUI and when clomiphene treatment fails to stimulate multiple follicular development, especially when IVF is not a viable option.

Assisted Reproductive Technology

Observations in ART cycles frequently provide insight into the possible causes of a couple's unexplained infertility because the procedures involved address or eliminate many of the unknown variables. Sperm and oocytes will be combined effectively. Fertilization and early embryonic development can be observed directly, and embryo transfer ensures that embryos will reach the endometrial cavity. Although the chromosomal composition of embryos and endometrial receptivity may seem like the only factors remaining, the list of unknowns is, in truth, much longer.

Although hundreds of studies of ART outcomes have been published, the large majority involve comparisons between two different treatment protocols; few have compared ART with no treatment or a different treatment such as gonadotropins and IUI,[744, 745] and none has been limited to couples with unexplained infertility. Excluding trials comparing IVF with GIFT,[746] which are no longer relevant, and one comparing immediate IVF to IVF after various other treatments, leaves only a single multicenter trial, in which 139 couples were randomly assigned to receive immediate IVF (within 6 weeks) or 3 months of expectant management.[747] In that trial, the average patient age was 33 years and the average duration of infertility was 4.8 years. Among the 51 couples with unexplained infertility (37%), clinical pregnancies were observed in 12/24 (50%) couples receiving immediate IVF and in 3/27 (11%) receiving expectant management, yielding a large difference of 39% per couple or 46% per cycle.[747] In the 2007 U.S. national summary of ART outcomes, the overall live birth rate per cycle start for couples with unexplained infertility (all ages) was 31.8%.[34] Evidence from three relevant trials suggest that intracytoplasmic sperm injection (ICSI) does not significantly improve IVF outcomes, compared to conventional fertilization, although the studies were not limited to couples with unexplained infertility.[748–750]

In summary, IVF is clearly the most effective treatment for couples with unexplained infertility, regardless whether it is the first or the last treatment.

Efficacy of Treatments for Unexplained Infertility

Treatment	*Approximate Cycle Fecundability*
No treatment	2–4%
IUI	2–4%
Clomiphene	2–4%
Gonadotropins	5–7%
Clomiphene/IUI	5–10%
Gonadotropins/IUI	7–10%
IVF	25–45%

SUMMARY

Overall, the treatment effects of treatments for unexplained infertility other than IVF are relatively small. In many cases, treatment may only hasten pregnancy for couples who would ultimately conceive on their own, given time. Careful counseling is essential and must take into account the couple's age, the duration of infertility, and the outcome of any previous pregnancies; before treatment is recommended, an ovarian reserve test also is prudent.[141] Couples who choose treatment should be informed thoroughly about the relative costs, risks, prognoses, and logistical challenges associated with different treatments so that they may select the one that best meets their needs and preferences. Partners can have differing levels of concern about their infertility and tolerance for risk and uncertainty.[751] Together, the medical evidence and shared decision-making determine the choice of management.[752]

Adoption

With proper evaluation and treatment, the majority of couples evaluated for infertility will achieve pregnancy. For those who fail simpler specific treatments, ART and adoption are both realistic options. Couples considering adoption have a wide range of choices including social agency adoptions, private adoptions, and international adoptions. In some states, private adoption is not legal, but where it is, private adoption can be an effective, more rapid alternative to adoption through a social agency. In most cases, the biologic mother has the opportunity to know the adopting parents and may reconsider her decision and reclaim her child for a time before the adoption is finalized. Those who prefer anonymity or who wish to avoid such potentially devastating disappointments likely will make a different choice. Couples interested in adoption should be referred to those knowledgeable about adoption laws in individual states and all of the available options.

All references are available online at: http://www.clinicalgynendoandinfertility.com

28

Recurrent Early Pregnancy Loss

Spontaneous abortion or miscarriage is defined as the involuntary termination of pregnancy before 20 weeks of gestation (dated from the last menstrual period) or below a fetal weight of 500 g. Losses after 20 weeks are considered stillbirths or premature births and generally have different causes than losses that occur earlier in gestation.

Historically, recurrent pregnancy loss or "habitual abortion" was defined as three or more consecutive spontaneous miscarriages. Popular theory during the 1930s and 1940s held that risk for spontaneous miscarriage increased progressively with each successive loss. Calculations based on that assumption by Malpas and later by Eastman suggested that three consecutive miscarriages demonstrated a predisposition to pregnancy loss that raised the risk for spontaneous miscarriage in the next pregnancy to as high as 73–84%.[1,2] In that era, the "control" for numerous studies evaluating the effectiveness of various treatments for recurrent pregnancy loss (hormones, vitamins, psychotherapy) was theoretical rather than real; the observed incidence of miscarriage in treated women was compared to the predicted or expected incidence rather than to the actual incidence observed in untreated or placebo-treated women. Unfortunately, one of the consequences of such flawed study design was the erroneous conclusion that treatments, including diethylstilbestrol (DES), were effective when in fact they were not. Years later, clinical studies based on empiric observations demonstrated that the risk of miscarriage after three previous losses is actually much lower than predicted (30–45%) and varies with the number of previous live births (none, 40–45%; one or more, about 30%).[3-6]

There is no specific number of miscarriages or firmly established criterion that justifies evaluation for recurrent pregnancy loss or defines the scope of investigation. Decisions must be individualized and consider the female partner's age, the timing and circumstances surrounding earlier pregnancy losses, elements of the personal and family medical history, and the couple's level of anxiety. Today, recurrent pregnancy loss is usually defined as three or more pregnancy losses (not necessarily consecutive).[7] Most also consider clinical investigation and treatment appropriate in couples with two consecutive spontaneous miscarriages, preferably documented by ultrasound or histopathological examination. Evaluation is especially indicated when any of the following are also present:

- Embryonic heart activity observed before any earlier pregnancy loss.
- Normal karyotype on products of conception from an earlier loss.
- Female partner age over 35 years.
- Infertility.

The Risk of Recurrent Early Pregnancy Loss in Young Women[4–6]		
	Number of Prior Miscarriages	*% Risk of Miscarriage in Next Pregnancy*
Women who have had at least one liveborn infant	0	12%
	1	24%
	2	26%
	3	32%
	4	26%
	6	53%
Women who have not had at least one liveborn infant	2 or more	40–45%

The vast majority of all early pregnancy losses result from chromosomal abnormalities arising in the egg, the sperm, or during early embryonic development and are random events. Even repeated miscarriages can occur by chance alone, but at least some affected couples have a predisposing factor. Among all the factors that have been implicated, the only undisputed causes of recurrent pregnancy loss are genetic (balanced chromosomal translocation in either partner, maternal age-related increase in prevalence of aneuploid oocytes), anatomic (congenital and acquired uterine abnormalities), or immunologic (the thrombotic complications of antiphospholipid syndrome). Alloimmunopathology, inherited thrombophilias (Factor V Leiden and others), endocrinopathies (thyroid disorders, diabetes, luteal phase deficiency), infections (genital mycoplasmas), and environmental exposures (smoking, heavy alcohol or caffeine consumption) have been implicated but are not established causes of recurrent pregnancy loss. Even after a comprehensive evaluation, recurrent pregnancy loss remains unexplained in well more than half of affected couples.

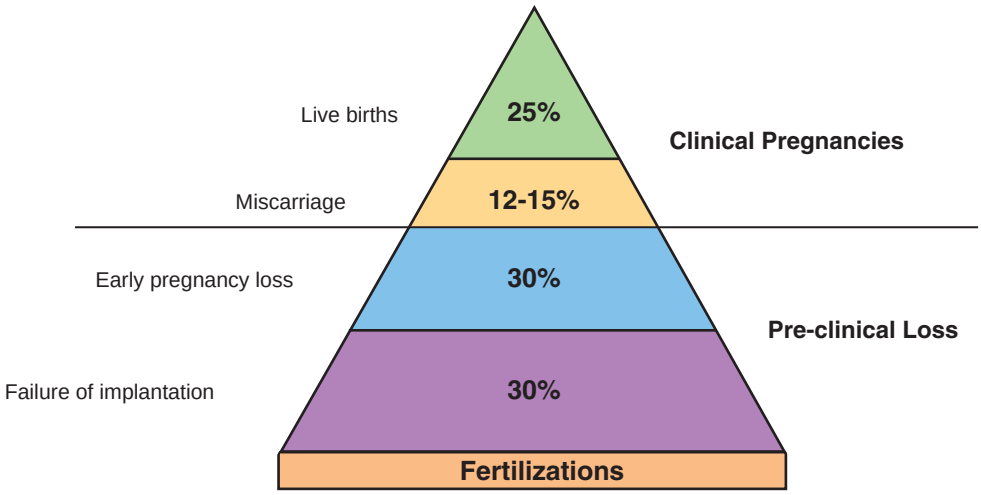
For all couples who have suffered recurrent pregnancy loss, education can provide important perspective; most couples welcome the offer of evaluation to identify any predisposing factor. When a likely cause can be defined, specific counseling and treatment can improve the prognosis for a successful pregnancy. When no specific cause can be found, reassurance and encouragement are no less valuable.

The Epidemiology of Pregnancy Loss

Early pregnancy loss is a very common event, even more so than most couples realize. Almost all chromosomally abnormal conceptions spontaneously abort, most before 10 weeks' gestation, and over 90% of conceptions having a normal karyotype continue.[8, 9] Miscarriage may thus be viewed as a natural selection process for quality control. Learning that miscarriage is common, normal, and inevitable in most cases does not heal the emotional wounds left by earlier losses or eliminate the anxiety that affected couples have when contemplating another attempt at pregnancy,[10, 11] but an accurate perspective is nonetheless important and often very helpful.

Overall, approximately 12–15% of clinically recognized pregnancies end in spontaneous miscarriage between 4 and 20 weeks of gestation. However, the true early pregnancy loss

rate, including both clinically recognized and unrecognized occult early miscarriages, is two to four times greater, depending on age. Careful studies in normally cycling healthy young women attempting pregnancy have shown that human chorionic gonadotropin (hCG) can often be detected transiently in the urine of women who are otherwise quite unaware that they had conceived and miscarried.[12-14] No less than 30% and as much as 60% of all conceptions abort within the first 12 weeks of gestation, and at least half of all losses go unnoticed. The reproductive loss that occurs even before a first missed menses is substantial.[15] Most recognized pregnancy losses occur before 8 weeks' gestation, and relatively few occur after 12 weeks.[16]



Numerous studies have documented that the risk of spontaneous miscarriage varies with past obstetrical history.[3, 16-18] In general, women in their first pregnancy, those whose only other pregnancy was electively terminated, and women whose only or last pregnancy was successful have a relatively low risk of spontaneous miscarriage (4–6%). Conversely, women whose only or last pregnancy ended in loss have a higher risk of miscarriage in their next pregnancy (19–24%).[16] Unless there has been a subsequent successful pregnancy, even a single loss increases risk for another spontaneous miscarriage in the next pregnancy. Taken together, available evidence also suggests that miscarriage risk increases with the number of pregnancy losses, but very gradually.[4-6] Overall, the risk is still less than 40% after four previous losses and no higher than about 50% even with six or more; risk may be modestly higher for women with recurrent pregnancy loss and no previous live-born children.

Independent of past obstetrical history, the risk of clinically recognized spontaneous miscarriage increases with age. Risk is relatively low before age 30 (7–15%) and only slightly higher for women aged 30–34 (8–21%), but then rises more sharply for ages 35–39 (17–28%) and women age 40 and older (34–52%).[19-23] Among women with a past history of pregnancy losses, advancing age adds to the risk related to previous losses; the miscarriage risk for women over age 40 (52%) is more than twice that for women under age 30 (25%).[6] If both recognized and occult pregnancy losses are considered, total pregnancy wastage in women over age 40 may reach or exceed 75%.[13, 23, 24]

In summary, approximately 12–15% of all clinically recognized pregnancies end in miscarriage, but the true incidence of miscarriage, including unrecognized early pregnancy losses, is two to four times higher (30–60%). Miscarriage risk increases with the number of previous pregnancy losses but rarely exceeds 40–50%. Risk for pregnancy loss also rises with increasing maternal age, moderately after age 35 and more rapidly after age 40.

Prognostic Value of Transvaginal Ultrasound Observations

Careful serial observations during early pregnancy have indicated that the risk for miscarriage decreases as the duration of pregnancy increases. The risk of pregnancy loss falls progressively after observation of a gestational sac (12%), a yolk sac (8%), and as embryonic crown-rump length increases (greater than 5 mm, 7%; 6–10 mm, 3%; greater than 10 mm, less than1%).[25] The observation of embryonic heart activity (by approximately 6 weeks' gestation) is another important developmental milestone and good prognostic indicator because most ill-fated pregnancies fail before then, but its predictive value varies with the past obstetrical history, clinical context, and age. In both normal and infertile asymptomatic young women, the timely appearance of embryonic heart activity decreases the risk of pregnancy loss from the global risk of 12–15% to between 3% and 5%.[26, 27] In women with past histories of recurrent pregnancy loss, the miscarriage rate after detection of embryonic heart activity is still three to five times higher (15–25%).[28, 29] In women with threatened abortions, demonstrable embryonic heart activity is again a good prognostic indicator overall (15% loss rate), but the incidence of subsequent loss is higher when there are other abnormal sonographic findings (slow or late appearing heart activity, size/date discrepancies, subchorionic hematoma).[30-33] Finally, the prognostic value of embryonic heart activity declines with increasing maternal age; whereas the risk for subsequent loss is low (less than 5%) in women age 35 and under, it is two to three times higher (approximately 10%) in women ages 36–39, and increased another 3-fold (29%) in women age 40 and over.[34]

Genetic Factors

Most spontaneous miscarriages result from chromosomal abnormalities in the embryo or fetus. Numerous studies in which large numbers of abortuses have been cultured and karyotyped have suggested that approximately 50% of all first trimester pregnancy losses, 30% of second trimester abortuses, and 3% of stillbirths are chromosomally abnormal.[21, 22, 35–37] However, these studies have very likely underestimated the prevalence of chromosomal abnormalities among abortuses because the data are biased by unrecognized maternal cell contamination and because normal euploid cells (from the mother or abortus) are less likely to fail culture than abnormal cell lines.[38-40] ***Analyses using newer techniques not dependent on cell culture (fluorescence in situ hybridization, FISH; comparative genomic hybridization, CGH), and more recent careful cytogenetic studies of early missed abortions suggest that the true incidence of chromosomal abnormalities in miscarried early pregnancies is closer to 75%.***[41, 42]

Over 90% of the chromosomal abnormalities observed among abortuses are numerical (aneuploidy, polyploidy); the remainder are split between structural abnormalities (translocations, inversions) and mosaicism.[42, 43] Overall, autosomal trisomies are the most common abnormality (usually involving chromosomes 13–16, 21, or 22), followed by monosomy X (45,X) and polyploidies.[21, 42, 44, 45] Among women with history of recurrent pregnancy loss, chromosomally normal (euploid) abortuses are more common, particularly in those age 35 and under.[45-48] The distribution of chromosomal abnormalities observed among the abortuses of women with recurrent pregnancy loss is otherwise no different from that seen in the general population when stratified by maternal or gestational age.[45] The high incidence of sporadic miscarriage and random chromosomal abnormalities means that some of the pregnancy losses in women with recurrent miscarriages result from chance. The likelihood of a euploid abortus increases with the number of previous miscarriages and after a previous abortion having a normal karyotype.[24, 47]

Parental Chromosomal Abnormalities

The overwhelming majority of chromosomally abnormal conceptions result from the chance union of one normal and one aneuploid gamete or from nondisjunction during early embryonic development. However, in 4–8% of couples with recurrent pregnancy loss, one or the other partner harbors a chromosomal abnormality that markedly increases the probability of a chromosomally unbalanced conceptus.[49-54] Balanced translocations (reciprocal, Robertsonian) are the most common abnormalities; sex chromosome mosaicism, chromosome inversions, and other structural abnormalities can also be observed.[55, 56]

In a balanced reciprocal translocation, pieces of two different autosomes (one from each of two different pairs) are translocated (exchanged). In a balanced Robertsonian translocation, the centromeres of two acrocentric chromosomes (numbers 13, 14, 15, 21, 22) fuse to form a single chromosome consisting of the long arms of the two affected chromosomes; the short arms (containing little or no essential genetic material) are lost. In both cases, the translocation carrier is genetically balanced and phenotypically normal. Unfortunately, when their oogonia or spermatogonia undergo meiosis to yield haploid oocytes or sperm, a large proportion of the gametes end up genetically unbalanced and abnormal, having either a deficiency or an excess of genetic material. Depending on how the chromosomes segregate during meiosis, the gametes may be chromosomally normal (containing only the normal copy of each of the two affected chromosome pairs), abnormal but balanced (containing the translocated member of each of the two affected chromosome pairs), or abnormal and unbalanced (containing two copies or no copies of an affected chromosome or chromosome segment). When such chromosomally unbalanced gametes combine with a normal gamete from an unaffected partner, the conceptus will have a trisomy and/or a monosomy and will almost always abort; an unbalanced conceptus may occasionally survive, but those that do are at high risk for malformations and mental retardation.[57]

In theory, one-fourth of the gametes produced by reciprocal translocation carriers should be normal, one-fourth should be abnormal but balanced, and one-half should be abnormal and unbalanced, yielding a 50% probability of a normal pregnancy (normal or balanced conceptus) and a 50% probability of an abnormal pregnancy (abortion or a viable but anomalous fetus), assuming union with a chromosomally normal gamete from the unaffected partner. Similarly, given the three different ways in which a Robertsonian translocation chromosome and the normal members of the affected chromosome pairs may align and segregate during meiosis, one-sixth of the gametes produced by carriers should be normal, one-sixth should be abnormal but balanced, and the remaining two-thirds should be abnormal and unbalanced, yielding a 33% probability of a normal pregnancy (normal or balanced conceptus) and a 67% probability of an abnormal pregnancy (abortion or a viable but anomalous fetus), again assuming a union with a chromosomally normal gamete from the unaffected partner. However, when a Robertsonian translocation involves both members of a single pair of chromosomes, the carrier will produce no normal gametes because all will have either two copies or no copy of the affected chromosome.

Some reciprocal translocations are predisposed to specific rather than random segregation patterns and may yield a skewed distribution of normal, balanced, and unbalanced gametes.[58-60] The probability of a successful pregnancy and the risk of a chromosomally abnormal but viable fetus vary with the specific chromosomes involved and the size and location of the translocated segments.[43, 57] Abnormalities of some chromosomes (chromosome 21) are better tolerated than others and risk of an unbalanced but viable conceptus is higher when the exchanged chromosomal segments are small. By their very nature, reciprocal translocations tend to be rather unique, so there is usually no easy way to accurately predict the probability of specific pregnancy outcomes for an individual affected couple. At best, the karyotype of the affected partner can allow one to predict the most likely segregation patterns for a specific translocation and to estimate the risk of unbalanced

offspring. When the male partner is the translocation carrier, the distribution of normal, balanced, and unbalanced sperm and the prognosis for successful conception can be more accurately defined.[60] When the female partner is the carrier or the distribution of gametes is otherwise unknown, the couple's own reproductive history (and that of any other similarly affected family member) is the best gauge. One exception is a specific recurring translocation involving chromosomes 11 and 22, t(11;22)(q23;q11), the most common reciprocal translocation in humans; over 100 unrelated affected families have been reported, and the reproductive performance of carriers has been well defined.[61-63]

Chromosomal inversions occur less frequently than translocations and may or may not have reproductive implications, depending on their size and location. Pericentric inversions (those that involve the centromere) often have no clinical consequences; a pericentric inversion of chromosome 9, inv(9)(p11q13), is so common (1–1.5% in the general population) that some consider it a normal variant with no importance.[64, 65] However, the crossovers and recombinations that can occur with paracentric inversions (those not located at the centromere) frequently result in an excess of genetic material resulting in abortion or an anomalous fetus.[43]

As might be anticipated, the most common reproductive history in translocation carrier couples includes both a normal child and early pregnancy losses (6–7%); other histories involving only spontaneous miscarriages or combinations of malformed children, stillbirths, and abortions are slightly less common (4–5%).[52] The probability of identifying a balanced chromosomal translocation in a couple with three or more previous pregnancy losses is not significantly higher than in those having had only two. In some couples, family history (recurrent pregnancy loss, stillbirths, or birth defects) suggests the possibility of an occult chromosomal abnormality after only one spontaneous miscarriage. Couples with miscarriages interspersed with normal pregnancies and outcomes should be evaluated in the same fashion as couples with consecutive miscarriages.[66]

Balanced chromosomal translocations can be found in either partner, and both must be karyotyped to exclude the possibility.[52] Any balanced translocation so identified may have arisen de novo or have been inherited from one of the carrier's own parents. If the translocation was inherited, any of the carrier's siblings and, in turn, their offspring might also be affected.[67] Any pregnancy in an affected couple becomes a candidate for prenatal diagnostic studies, regardless of the mother's age or previous reproductive history.[66] Consequently, counseling of translocation carrier couples with recurrent pregnancy loss should consider karyotyping the carrier's parents and, where appropriate, other potentially affected individuals in the kindred. For minor children who may be carriers, karyotyping is best postponed until they reach an age where they are able to grant informed consent.

It is entirely possible and even likely that some couples with recurrent pregnancy loss may harbor a genetic abnormality that predisposes to a higher risk of miscarriage but cannot be detected using standard cytogenetic techniques. Possibilities include isolated gonadal or germline mosaicism (including a trisomic cell line) and single gene defects.[43]

Aging and Gamete Aneuploidy

The mechanisms responsible for the age-related increase in the incidence of miscarriage and the use of ovarian reserve tests in the evaluation of reproductive age and prognosis are discussed in detail in Chapter 27. The genetic factors that contribute to the increase in pregnancy wastage associated with reproductive aging and the utility of ovarian reserve testing in women with recurrent pregnancy loss are only summarized briefly here.

Several lines of evidence suggest that age-related instability or degradation of the cellular mechanisms that govern meiotic spindle formation and function results in an increasing incidence of meiotic segregation errors and a rapid rise in the numbers of aneuploid oocytes during the later reproductive years.[68-74] The best available estimates indicate that the prevalence of aneuploid oocytes is relatively low before age 35 (less than 10%) but increases abruptly thereafter, reaching 30% by age 40, 50% by age 43, and nearly 100% after age 45.[68] These observations offer a logical explanation for the overall age-related increase in the incidence of miscarriage and the higher prevalence of aneuploidy among the abortuses of aging women.[19-22] Indeed, most trisomies observed among abortuses can be traced to maternal meiotic errors and oocyte aneuploidy.[75]

Some women with otherwise unexplained recurrent pregnancy loss have a diminished ovarian reserve that may help to explain their poor reproductive performance.[76, 77] The prevalence of abnormal ovarian reserve tests in women with unexplained recurrent pregnancy loss is higher than in women with other defined causes of recurrent pregnancy loss[76] and comparable to that observed in the general population of infertile women.[77] These observations suggest that women at advanced stages of ovarian follicular depletion are at higher risk for miscarriage, regardless of their age. For them, the curve that describes the age-related rise in the incidence of spontaneous miscarriage is shifted left, and the sharp rise in miscarriage risk that normally begins at about age 37 starts earlier.[19-22] Some women will suffer premature ovarian follicular depletion because they are born with a smaller than normal ovarian follicular pool and are genetically destined to be among the 10% of women who experience an early menopause.[78-82] Women who have had a trisomic abortus reach menopause at an earlier average age.[83] Other women may have their ovarian follicular pool depleted by disease that destroys ovarian tissue or requires its removal. Either way, the end result is the same—accelerated follicular depletion, declining fertility, and increasing risk for miscarriage begin at an earlier than normal age. Women with a demonstrated low ovarian reserve have an extremely high rate of pregnancy loss, regardless of age.[84]

Besides offering information that may help to explain recurrent pregnancy loss, ovarian reserve testing may identify young women at increased risk for fetal aneuploidy in subsequent pregnancies who would otherwise not be considered candidates for prenatal diagnostic studies.[85-89] The incidence of Down syndrome is increased in women with elevated serum levels of follicle-stimulating hormone (FSH), regardless of age and regardless of whether the low ovarian reserve the high FSH reveals came about naturally or resulted from ovarian surgery.[85, 87-89]

The prevalence of skewed X chromosome inactivation, defined as preferential inactivation (more than 90%) of one of the two X chromosomes in female cells, is increased in women with recurrent pregnancy loss,[90-94] although this finding was not confirmed in two studies.[95,96] This observation has prompted speculation that X-linked mutations lethal to the male cause skewed X chromosome inactivation in female carriers and predispose to abortion of male conceptuses and an increased prevalence of female live births.[92, 97] However, investigation of this issue has not confirmed the predicted excess of chromosomally normal male abortuses.[93] Observations of an increased prevalence of trisomic abortuses among women with recurrent pregnancy loss and skewed X chromosome inactivation have suggested the alternative hypothesis that X chromosome mutations or X-autosome translocations result in skewed X chromosome inactivation and a smaller than normal ovarian follicular pool or accelerated follicular depletion that predisposes to oocyte aneuploidy and recurrent pregnancy loss.[93, 98]

Ill-fated chromosomally abnormal conceptuses can also result from fertilization of a normal euploid oocyte by an aneuploid sperm. The sperm of men whose partners have a history of unexplained recurrent pregnancy loss exhibit a higher prevalence of abnormal morphology, chromosome aneuploidy, DNA fragmentation, and abnormal tests of sperm

function such as hypo-osmotic swelling.[99-106] The incidence of sperm aneuploidy rises with paternal age, if only slightly,[58, 107] and the incidence of miscarriage in young women with older male partners is higher than in those whose partners are young.[108] Taken together, these observations suggest that poor semen quality, like a low ovarian reserve in women, can predispose to both infertility and early pregnancy loss, two different points on a continuum of reproductive failure having some causes in common. However, sperm aneuploidy rarely rises above approximately 1–2%. Compared to the influence of oocyte aneuploidy on miscarriage risk, chromosomally abnormal sperm have relatively little importance as a predisposing factor in recurrent pregnancy loss.

Karyotyping the Abortus

Many view karyotyping the products of conception following miscarriage as an unnecessary and expensive luxury. Others consider it crucially important for differentiating couples who are candidates for thorough evaluation from those who are not. Without karyotyping, women who repeatedly miscarry generally are assumed to be losing normal pregnancies when, in fact, most are not. Some have even advocated karyotyping the first or second abortus, reasoning that women who abort chromosomally normal pregnancies should be screened for treatable causes of pregnancy loss sooner rather than later. Conversely, those who miscarry a chromosomally abnormal pregnancy might be spared unnecessary and costly evaluation and empiric treatments.[9]

Unfortunately, the karyotype of an abortus cannot provide information so definitive; a karyotype may be useful but has limitations and pitfalls that must be carefully considered. Most early failed pregnancies lose viability well before onset of clinical symptoms of miscarriage or other recognition of the inevitable loss; products of such conceptions may therefore fail to grow in culture. Tissue specimens passed spontaneously are more likely to fail culture than those obtained by curettage.[45]

A normal abortus karyotype might be interpreted as suggesting that genetic factors are not likely responsible, focusing attention on evaluation for other possible causes of recurrent pregnancy loss.[9] Unfortunately, a normal 46,XX karyotype can also represent maternal cell contamination of the tissue specimen and preferential growth of the normal maternal cell line in culture, particularly when no specific care has been taken to dissect, isolate, and submit only chorionic villi for cell culture.[38-40] Results from an embryoscopic and cytogenetic study of early missed abortions challenge directly the notion that a normal karyotype effectively excludes genetic causes for a failed pregnancy. Whereas 75% of the abortuses were chromosomally abnormal, fully two-thirds of the remaining 25% having a normal karyotype (17% of the total) exhibited gross developmental abnormalities as severe as those observed in aneuploid abortuses.[42] These observations strongly suggest that over 90% of all early missed abortions involving a recognizable embryo result from genetic errors and imply that a substantial proportion of failed early pregnancies result from gross genetic flaws in organizational and morphogenic processes not detectable with conventional cytogenetic techniques or even more modern methods (FISH, comparative genomic hybridization). Arguably, pregnancies that fail before any recognizable embryogenesis (empty sacs or “blighted ova”) are even more likely to result from chromosomal abnormalities, the inference being that well more than 90% of all early pregnancy losses can result from genetic causes.

An abnormal abortus karyotype revealing trisomy, monosomy, or polyploidy explains that specific pregnancy loss, suggests it likely resulted from chance alone, and generally has been considered as evidence that the risk for recurrence is not significantly increased.[24, 48] Whereas such findings may suggest that no formal evaluation is therefore needed, one must

assume that couples with other specific causes for recurrent pregnancy loss have at least the same random chance for conceiving an aneuploid pregnancy as anyone else; they might be overlooked if evaluation is offered only to those having a chromosomally normal abortus. Moreover, an aneuploid abortus might also reflect the influence of advanced maternal age or an otherwise unsuspected diminished ovarian reserve, in which case the risk for recurrence in a subsequent pregnancy clearly is increased.[84]

A karyotype of an abortus that demonstrates an unbalanced chromosomal translocation obviously suggests that a parent may be a balanced carrier of the same translocation, a suspicion easily confirmed by performing karyotypes on both partners in affected couples.

The more expensive methods for genetic screening to detect variants associated with miscarriages, FISH and comparative genomic hybridization discussed below, can be applied to couples having recurrent miscarriages. However, these are expensive tests, and although variants and polymorphisms can be detected, the chances of having a healthy child, despite a higher risk of miscarriage, are still high, perhaps as high as non-carrier couples.[109]

Preimplantation Genetic Diagnosis and Aneuploidy Screening

Preimplantation genetic diagnosis describes a number of techniques for preconceptional genetic evaluation of embryos resulting from in vitro fertilization (IVF). Preimplantation genetic diagnosis can be used to detect numerical (aneuploidy) and structural chromosomal abnormalities (translocation, inversions), to identify oocytes or embryos with inherited single gene disorders (cystic fibrosis, thalassemia, hemophilia, Duchenne muscular dystrophy, and numerous others),[110, 111] or to determine gender.[112, 113] The technique requires one or more cells that may be obtained at different stages of development. The chromosomal composition of the oocyte may be inferred from that of the extruded polar bodies.[110] One or two blastomeres may be removed from cleavage stage embryos. Biopsy of the trophoectoderm can also be performed at the blastocyst stage. In the most common scenario (cleavage stage embryo biopsy), a laser or a dilute solution of acid Tyrode's solution is used to create a small hole in the zona pellucida and one or two cells are aspirated, typically on the third day after oocyte retrieval and fertilization when embryos are at the 6–8 cell stage.[113]

FISH, fluorescence in situ hybridization, is a technique for detection of numerical chromosomal abnormalities using probes labeled with different colored fluorochromes that bind to specific gene sequences on specific chromosomes. In the context of recurrent pregnancy loss, it has been used to screen embryos resulting from IVF for the most common aneuploidies observed in abortuses (XY, 13, 14, 15, 16, 18, 21, 22) and also to distinguish chromosomally normal, balanced, and unbalanced embryos in couples who carry a balanced chromosomal translocation.[113-115] As a method for aneuploidy screening, FISH has both advantages and limitations. FISH is relatively easy to perform and yields results in time for transfer of genetically selected embryos 2 days after embryo biopsy (5 days after oocyte retrieval and fertilization), at the blastocyst stage. Although it allows evaluation of only a limited number of chromosomes (typically between 5 and 9), FISH can still detect over 80% of all chromosomal abnormalities because it typically includes all of the chromosomes involved in most aneuploidies.[116] Because probes hybridize to a specific locus or the centromere, FISH provides information only about the presence or absence of a very small segment of the chromosome; partial aneuploidies can go undetected.[117] Also, nuclear fragmentation in biopsied blastomeres is relatively common and can result in lost chromosomes, yielding erroneous diagnoses of aneuploidy.[118]

Comparative genomic hybridization is a related technique in which test DNA (extracted from a single blastomere) and normal male reference DNA (extracted from lymphocytes) are first amplified, then labeled with different colored fluorochromes (green/red), and simultaneously hybridized to template metaphase chromosomes from normal male lymphocytes; the green/red fluorescence ratio reflects the relative copy number for each chromosome in test DNA compared to the normal reference DNA.[117] Comparative genomic hybridization allows analysis of all 24 chromosomes (X, Y, 22 autosomes) and detection of abnormalities not recognized by more limited analysis with FISH.[118, 119] This technique has been applied to the genetic evaluation of fetal losses; the diagnostic yield is improved compared with conventional karyotyping, which is often handicapped by culture failures or maternal contamination.[120, 121]

Regardless whether FISH or comparative genomic hybridization is employed, preimplantation genetic diagnosis generally involves the analysis of only one or two blastomeres, assuming that those selected accurately represent the entire embryo. Unfortunately, a number of studies have now demonstrated that mosaicism is extremely common in early human embryos cultured in vitro. The prevalence of embryonic mosaicism increases with maternal age and with the stage of development; approximately half of all cleavage stage embryos and up to 90% of blastocysts exhibit some degree of chromosomal mosaicism.[122-126] Diagnostic errors, therefore, are inevitable and, to some extent, unavoidable, but can be minimized by analyzing two or even three blastomeres.[126, 127]

Overall, the results of preimplantation genetic diagnosis studies using both FISH and comparative genomic hybridization indicate that only approximately 35–45% of embryos are normal for all of the chromosomes examined.[128-131] Data derived from numerous studies reveal that older women and women with a history of recurrent pregnancy loss produce more aneuploid embryos than younger women and those with normal reproductive histories.[132-136] Transfer of preimplantation genetic diagnosis-selected embryos can improve implantation rates and decrease abortion rates in women at higher risk for pregnancy loss.[126-128, 137, 138] The ultimate impact of preimplantation genetic diagnosis for aneuploidy screening on the live birth rate in older women and in women with history of recurrent pregnancy loss is not yet clear, although one cost/benefit analysis concluded that IVF alone is the most cost-effective option under the age of 40, but over age 40, IVF alone and IVF with preimplantation genetic diagnosis are equal in cost.[139] Whereas it may be reasonable to consider preimplantation genetic diagnosis aneuploidy screening for indications of advanced maternal age or recurrent pregnancy loss in couples with other specific indications for IVF, the results achieved with preimplantation genetic diagnosis so far do not justify IVF with preimplantation genetic diagnosis for all couples with advanced maternal age or history of recurrent pregnancy loss. Up to now, the overwhelming majority of preimplantation genetic diagnosis has been performed in a very few centers worldwide, but further improvements in the technologies and wider application of the techniques are likely.

For couples with recurrent pregnancy loss in whom one partner carries a balanced chromosomal translocation, IVF with preimplantation genetic diagnosis and transfer of only normal and balanced embryos can achieve pregnancy rates comparable to those observed in unselected infertile couples with substantially decreased risk of spontaneous miscarriage, although pregnancy rates are inversely proportional to the proportion of abnormal gametes.[115, 140-144] When the male partner carries the balanced translocation, sperm FISH analysis can be used to determine the proportion of chromosomally unbalanced sperm and to predict the probability of conceiving a successful pregnancy.[60] Data suggest that when there are numerous good-quality embryos and less than approximately 65% unbalanced sperm, translocation carrier couples have a reasonable probability of success with IVF and preimplantation genetic diagnosis, but otherwise not.[60] Sperm FISH analysis could prove valuable to affected couples weighing the options of IVF with preimplantation genetic diagnosis and therapeutic insemination with donor sperm. Unfortunately, there is no way

to obtain similar information for female balanced translocation carriers; depending on the nature of the translocation and on the reproductive history, some women who carry a balanced translocation may prefer to apply their available resources to IVF with donor oocytes rather than attempt IVF with preimplantation genetic diagnosis.

Summary of Key Facts Relating to Genetic Factors

Overall, 50–75% of spontaneous miscarriages result from numerical chromosomal abnormalities in the embryo or fetus and occur by chance; trisomies are the most common. In approximately 5% of couples with recurrent pregnancy loss, karyotypes will reveal a balanced chromosomal translocation that markedly increases the risk of miscarriage due to the high prevalence of aneuploidy in the gametes of the affected parent. Reproductive aging in women is associated with an increasing risk of miscarriage, which reflects a rising prevalence of oocyte aneuploidy. Ovarian reserve testing in women with unexplained recurrent pregnancy loss can reveal evidence of premature reproductive aging. Karyotype of an abortus can explain the loss (aneuploidy), provide evidence for a chromosomal translocation in a parent (when an unbalanced translocation is observed), or suggest a non-genetic cause (when normal). However, a normal karyotype does not entirely exclude genetic causes for the miscarriage, and a normal female karyotype (46,XX) can result from maternal cell contamination of cultured tissue specimens. IVF with preimplantation genetic diagnosis and selected transfer of euploid embryos is an established treatment for couples with recurrent pregnancy loss when one partner carries a balanced chromosomal translocation. IVF with preimplantation genetic diagnosis (using FISH) for reasons of advanced maternal age or in couples with unexplained recurrent pregnancy loss can increase implantation rates and decrease miscarriage risk, but have not increased live birth rates. Consequently, the associated costs in couples without other specific indications for IVF cannot be justified.

Anatomic Factors

The anatomic uterine abnormalities that can predispose to a higher risk of pregnancy loss include congenital malformations, uterine leiomyomas, and intrauterine adhesions. Each has been considered in detail elsewhere as factors that may also adversely affect fertility (Chapters 4 and 27); discussion here is limited to their importance and management in women with a history of recurrent pregnancy loss.

The principal methods for evaluation of the uterus include traditional hysterosalpingography (HSG), transvaginal ultrasonography, and sonohysterography. Magnetic resonance imaging (MRI) and endoscopy (hysteroscopy and laparoscopy) are generally reserved when necessary for better defining the nature of anomalies identified or suggested by simpler methods. Each method and its limitations, pitfalls, and relative accuracy have been described at length in the context of the infertility evaluation (Chapter 27). However, their relative value for the evaluation of infertility and recurrent pregnancy loss are somewhat different. HSG has some advantages over ultrasonographic techniques for the evaluation of uterine factors in infertile women because it also provides information about tubal patency that transvaginal ultrasonography and sonohysterography cannot. However, for the evaluation of recurrent pregnancy loss, transvaginal ultrasonography and sonohysterography offer distinct advantages over HSG; both image the external uterine fundal contour

and, therefore, better distinguish septate and bicornuate uteri,[145-147] and both are generally easier to perform and better tolerated than HSG. Compared to HSG, sonohysterography has greater sensitivity and specificity for detection of intracavitary mass lesions (submucous myomas, endometrial polyps) and similar efficacy for diagnosis of intrauterine adhesions.[148-153] Three dimensional transvaginal ultrasonography, with and without saline contrast, can provide high-definition images comparable to those generated with MRI.[154, 155] As in women with infertility, endoscopic methods can usually be reserved for excision of cavitary mass lesions or intrauterine septa identified by simpler methods.

Congenital Uterine Malformations

Developmental uterine anomalies have long been associated with pregnancy loss and obstetric complications. The reported prevalence of uterine malformations in women with recurrent pregnancy loss has varied widely with differences in diagnostic methods and criteria.[156-158] The best available data suggest that the prevalence of major uterine anomalies (excluding arcuate uteri) in the general population is approximately 2% and about three times greater (6–7%) in women with history of recurrent pregnancy loss, supporting the notion that uterine malformations may indeed be the proximate cause for miscarriages in a small proportion of women with recurrent pregnancy loss.[156, 159-162] Pregnancy losses with uterine congenital abnormalities usually occur later in pregnancy in the second trimester; however, the presence of an uterine anomaly after repeated early losses deserves consideration of surgical repair.[163, 164] The pathogenesis of pregnancy wastage in women with congenital uterine malformations is uncertain but generally has been attributed to a reduced intrauterine volume or poor vascular supply.[165]

A unicornuate uterus results from failure of development of one müllerian duct. Pregnancy outcomes in women with unicornuate uteri are generally poor; approximately half of all recognized pregnancies fail.[166] Most unicornuate uteri are associated with a noncommunicating contralateral uterine horn, some of which have a functional cavity and should be removed to reduce risk for ectopic pregnancy, even if not otherwise necessary (pain, mass, endometriosis).[167] Since approximately 40% of unicornuate uteri are associated with an ipsilateral renal agenesis, further evaluation with an intravenous pyelogram or renal sonogram is also indicated.[167] No surgical procedure can enlarge the unicornuate uterus. Anecdotal reports of successful pregnancies after cervical cerclage are numerous, but the efficacy of cerclage in women with unicornuate uteri has not been carefully studied. ***Available evidence suggests that most pregnancies in women with unicornuate uteri are best managed expectantly with cervical cerclage reserved for those with previous second trimester pregnancy losses or evidence of progressive cervical shortening.***[166]

Uterine didelphys results from complete failure of fusion of the müllerian ducts and normal differentiation of each to form a cervix and hemiuterus. The reproductive outcomes of women with uterine didelphys are slightly better than those of women with unicornuate uteri, possibly because of improved collateral blood supply between the two fused horns. Nevertheless, approximately 40% of pregnancies in women with uterine didelphys end in spontaneous miscarriage.[166] In general, the only surgery indicated in women with uterine didelphys is the removal of an obstructing longitudinal vaginal septum (75% prevalence).[168] ***Unification procedures are usually unnecessary and meddlesome, but can benefit some women with numerous miscarriages or previable births. When surgery is performed, the recommended technique unifies the two fundi and leaves the two cervices intact.***[169]

A bicornuate uterus results from incomplete fusion of the müllerian ducts at the level of the fundus, creating two separate uterine cavities with a common lower segment and a single cervix; externally, the uterus has a midline cleft with a depth that varies with the severity of the fusion anomaly. Data from collected series of women with bicornuate uteri reveal miscarriage and overall fetal loss rates of approximately 30% and 40%, respectively.[166] Preterm delivery risks decrease as the size of the common lower uterine cavity increases.[170] ***Although the benefits of unification procedures have not been systematically evaluated, surgery generally is considered unnecessary and best reserved for those with a well established history of otherwise unexplained recurrent pregnancy loss or previable births.***[166] The Strassman abdominal metroplasty is the surgical procedure of choice and comparisons between pregnancy outcomes before and after unification suggest that surgery can benefit carefully selected women.[171, 172] The incidence of cervical incompetence associated with congenital uterine anomalies is reportedly highest for those having a bicornuate uterus, and there is evidence from case series that cervical cerclage can improve fetal survival rates.[173, 174]

A septate uterus results from incomplete resorption of the medial septum separating the two otherwise normally fused hemiuteri. Septum resorption normally occurs only after urologic development is completed; the prevalence of urinary tract anomalies, therefore, is not increased in women with septate uteri. The septate uterus is by far the most common uterine developmental anomaly, accounting for 80–90% of all major malformations in both women with recurrent pregnancy loss (3.5% prevalence) and in the general population.[145, 160, 162, 175] It is also the malformation most highly associated with poor pregnancy outcomes.[145, 176] Data from numerous case series indicate that the miscarriage rate associated with septate uteri is approximately 65%.[166] Uterine septa associated with recurrent pregnancy loss are not broader or longer than those observed in women with normal reproductive histories. However, the size of the unaffected cavity (bounded by the leading edge of the septum above and the internal cervical os below) is smaller in women with recurrent pregnancy loss,[156] an observation lending credence to the hypothesis that implantation on a poorly vascularized septum predisposes to pregnancy loss.[177-179] Although uterine septa are not always associated with a poor pregnancy outcome, their discovery in women with recurrent pregnancy loss provides an indication for surgical correction. ***Hysteroscopic septoplasty is a relatively brief and straightforward outpatient endoscopic procedure associated with low morbidity and dramatically improved postoperative pregnancy outcomes (80% term delivery, 5% preterm delivery, 15% pregnancy loss).***[145, 166] Hysteroscopic septoplasty can be performed using microscissors, any of a variety of electrosurgical instruments, or laser, and excellent results can be achieved with all methods.[145] With few exceptions, only incision rather than excision is required because the septum typically retracts, leaving little if any residual. The procedure usually is complete when both tubal ostia can be viewed at the same time. It is useful to remember that an arcuate uterus is generally regarded as a normal variant and that residual septa measuring less than 1 cm in size have no adverse effect on pregnancy outcome.[180]

Nearly 70% of women exposed to diethylstilbestrol (DES) in utero have a developmental uterine abnormality.[181] The T-shaped uterine cavity is the single most common malformation; others include a hypoplastic uterus, constriction rings, and irregular intrauterine filling defects. Although the use of DES in pregnancy was banned in 1971 because of an observed association with vaginal clear cell adenocarcinoma and most exposed women are now beyond their reproductive years,[182] affected women are still occasionally encountered. In utero DES-exposed women are at increased risk of adverse pregnancy outcomes, including a 2-fold higher risk of spontaneous miscarriage (approximately 24%) and a dramatic 9-fold higher risk of ectopic pregnancy.[183] Structural changes in cervical collagen content may predispose affected women to cervical incompetence, and data from nonrandomized clinical trials suggest cerclage warrants serious consideration in women with history of second trimester loss or preterm delivery.[184, 185]

Uterine Leiomyomas

There is no substantial evidence implicating uterine myomas as a cause of recurrent pregnancy loss. What evidence does exist derives from case series comparing reproductive outcomes before and after myomectomy.[186, 187] All of the mechanisms proposed to explain how myomas might predispose to recurrent pregnancy loss relate to the consequences of poor regional blood flow.[188] Numerous studies have examined the effect of uterine fibroids on fertility (Chapters 4 and 27), but none has specifically examined the effect of myomas on pregnancy outcome in *fertile* women. The best available data come from a series of studies designed to examine the effect of uterine myomas on outcomes achieved with in vitro fertilization (IVF) in infertile women. On balance, these data suggest that pregnancy outcomes, like pregnancy and implantation rates, are adversely affected by submucous myomas, but not by subserosal or intramural myomas under 5–7 cm in size.[189-192] Consequently, management recommendations for women with recurrent pregnancy loss and uterine myomas are comparable to those for infertile women with uterine myomas.

In general, when submucous myomas are single and small, the likely benefits of hysteroscopic myomectomy outweigh the few associated risks.[193-197] When submucous myomas are multiple or large, hysteroscopic myomectomy is more technically challenging and has greater risks, including sterility resulting from severe postoperative intrauterine adhesions. When submucous myomas extend deeply into the myometrium, treatment options include subtotal hysteroscopic myomectomy and abdominal myomectomy. When myomas have definite but limited impact on the uterine cavity, the decision to delay or to proceed with surgical treatment may vary, depending on age, reproductive history, size and location of myomas, and the complexity of any other treatments required. ***When myomas do not encroach on or distort the uterine cavity, surgery is not indicated in the absence of other specific symptoms attributable to myomas that demand treatment in and of themselves.***

Intrauterine Adhesions (Asherman's Syndrome)

Recurrent pregnancy loss is one possible result of intrauterine adhesions, but menstrual disorders (hypomenorrhea, amenorrhea, dysmenorrhea) and infertility are the more common clinical presentations.[198, 199] Any insult severe enough to remove or destroy endometrium can cause intrauterine adhesions, and the gravid uterus seems particularly susceptible to injury.[200, 201] Considering that pregnancy loss is among the most common indications for uterine curettage, intrauterine adhesions can first result from and then become a contributing cause of recurrent pregnancy loss. Mechanisms by which intrauterine adhesions can cause recurrent pregnancy loss include a decreased functional intrauterine volume and endometrial fibrosis and inflammation that can predispose to placental insufficiency.[172] The pregnancy outcomes of women with intrauterine adhesions are generally poor (40–80% ending in spontaneous miscarriage and approximately 25% in preterm delivery) and much improved after adhesiolysis (50–90% ending in term delivery, 7–23% ending in miscarriage); the prognosis generally correlates with the severity of disease.[199, 200, 202-205]

Hysteroscopy provides the means to confirm a diagnosis of intrauterine adhesions suggested by sonohysterography or HSG and is also the method of choice for treatment, being both safer and more effective than blind curettage. The hysteroscopic appearance of adhesions of varying severity, techniques for lysis, operative risks, and adjuvant therapies are described in detail in Chapter 27.

Summary of Key Facts Relating to Anatomic Factors

Congenital and acquired uterine abnormalities predispose to an increased risk of pregnancy loss and can be identified by sonohysterography or traditional HSG; magnetic resonance imaging may be required to accurately differentiate septate and bicornuate uteri. The septate uterus is the most common müllerian anomaly, the one most closely correlated with pregnancy loss, and the malformation most easily and successfully corrected; hysteroscopic septoplasty is indicated in women with recurrent pregnancy loss and having a septate uterus. Abdominal metroplasty procedures are rarely indicated for women with a uterus didelphys or a bicornuate uterus. Cervical cerclage may help to improve pregnancy outcomes in women with bicornuate uteri and in those with a unicornuate uterus or a uterus didelphys who have a history of previable delivery or exhibit progressive cervical shortening during early pregnancy. The prevalence of urinary tract abnormalities is increased in women having a unicornuate or bicornuate uterus or a uterus didelphys, but not in those having a septate uterus. Uterine leiomyomas are often identified in women with recurrent pregnancy loss, but only submucous myomas and larger intramural fibroids that clearly encroach upon or displace the uterine cavity are relevant. Intrauterine adhesions are an uncommon but established cause of recurrent miscarriage; pregnancy outcomes are much improved after hysteroscopic lysis.

Immunologic Factors

Both autoimmune and alloimmune mechanisms have been implicated as causes of recurrent pregnancy loss. Autoimmune disorders involve an immune response directed against a specific part of the host or self; those that have been linked to recurrent pregnancy loss include certain classic autoimmune diseases like systemic lupus erythematosus and the antiphospholipid syndrome. Alloimmune disorders involve an abnormal maternal immune response to fetal or placental antigens; possibilities include maternal cytotoxic antibodies, absent maternal blocking antibodies, and disturbances in natural killer cell function and distribution.

Autoimmune Disorders

Systemic lupus erythematosus has long been associated with pregnancy loss. Data from a number of case series suggest that the risk for loss is approximately 20%, all excess risk being attributable to losses occurring during the second and third trimester of pregnancy.[206] Early spontaneous miscarriages are no more common in women with systemic lupus than in the general population, but the incidence of later losses (6%) is two to four times higher.[207] Almost all fetal deaths that occur in women with systemic lupus are associated with antiphospholipid antibodies; they are the most sensitive indicator of fetal distress or death.[208-210] Active disease at conception, onset of systemic lupus erythematosus during pregnancy, and renal disease also increase the risk for pregnancy loss.[207] Careful monitoring and timely interventions can improve pregnancy outcomes.[211, 212] Treatment aimed at preventing pregnancy loss in women with systemic lupus and antiphospholipid antibodies is similar to that for women with antiphospholipid syndrome, discussed later. In general, women with active systemic lupus erythematosus should be advised to delay conception until remission can be established, those with moderate renal insufficiency must be advised

of the increased risk for pregnancy loss, and women with severe renal insufficiency should be encouraged to avoid pregnancy; even successful pregnancies are at increased risk for preeclampsia and preterm delivery.[207]

Antiphospholipid syndrome is an autoimmune disorder having specific clinical and laboratory features; diagnosis requires at least one of each.[213,214] The clinical diagnostic criteria include thromboembolic events (arterial, venous, small vessel) and pregnancy loss (three or more consecutive losses at less than 10 weeks' gestation, a fetal death after 10 weeks, premature birth at less than 34 weeks associated with severe preeclampsia or placental insufficiency). There are also three laboratory diagnostic criteria. One is the lupus anticoagulant, revealed by delayed clotting in phospholipid-dependent coagulation tests (activated partial thromboplastin time, kaolin clotting time, dilute Russell's viper venom time), corrected by addition of excess phospholipid but not by platelet-poor plasma.[215] The second is the demonstration of moderate to high levels of anticardiolipin antibodies (IgG or IgM); low levels may be observed in 3–5% of normal individuals and are of uncertain significance.[216] More recently, a high titer of antibodies to β2-glycoprotein 1 is also considered sufficient to establish the diagnosis.[217] Abnormal laboratory test results must be observed on at least two separate occasions at least 12 weeks apart.

International Consensus Definition for the Diagnosis of Antiphospholipid Syndrome[218]

DIAGNOSIS REQUIRES ONE OF THE CLINICAL CRITERIA AND ONE OF THE LABORATORY FINDINGS

Clinical Criteria:

1. Vascular Thrombosis
2. Pregnancy Morbidity
   - **A.** One or more losses after the 10th week of a morphologically normal fetus.
   - **B.** One or more premature births of a normal neonate before the 34th week because of preeclampsia or eclampsia or placental insufficiency.
   - **C.** Three or more unexplained consecutive early miscarriages.

Laboratory Tests:

1. Lupus anticoagulant present on two or more occasions at least 12 weeks apart.
2. Anticardiolipin antibody of IgG or IgM isotype in medium to high titer on two or more occasions at least 12 weeks apart.
3. Anti-β2-glycoprotein 1 antibody of IgG or IgM isotype in 99th percentile titer on two or more occasions at least 12 weeks apart.

Although the prevalence of antiphospholipid syndrome among all women with recurrent pregnancy loss is quite low (3–5%),[219] the disorder is nonetheless a potentially treatable cause of recurrent pregnancy loss. ***Tests for the detection of a lupus anticoagulant and antiphospholipid antibodies specified above are minimally invasive, relatively inexpensive, and, therefore, justified in the evaluation of most if not all women with recurrent pregnancy loss.***[207] Tests to detect other antibodies associated with this syndrome (anti-annexin, antiphosphatidylserine, antiphosphatidylethanolamine, antibodies to plasminogen and plasminogen activator, and others) have been advocated in the evaluation of women with recurrent pregnancy loss but provide no important additional information.[207] Assays for antiphospholipid antibodies other than lupus anticoagulant, anticardiolipin, and

anti-β2-glycoprotein 1 have not been standardized or extensively studied, their clinical relevance (independent of lupus anticoagulant, anticardiolipin, and anti-β2-glycoprotein 1) has not been established, and many apparently healthy women have low levels of circulating antiphospholipid antibodies.[207, 215, 216, 220]

The evidence implicating antiphospholipid antibodies as a predisposing factor in recurrent pregnancy loss is largely circumstantial but nonetheless fairly compelling. A small but still important number of women with recurrent pregnancy loss have circulating antiphospholipid antibodies.[207] Women with antiphospholipid antibodies identified by screening in early gestation have an increased rate of pregnancy loss.[221, 222] They also exhibit an unusually high rate of loss in subsequent pregnancies, even when treated.[223–226] Additional evidence comes from animal models; passive immunization with IgG antiphospholipid antibodies from women with antiphospholipid syndrome induces abortion in mice. Antiphospholipid syndrome can be found in association with systemic lupus erythematosus or in women with no other evidence of autoimmune disease.[227–229]

In contrast to observations in most series of women with recurrent pregnancy loss, between one-third and three-fourths of pregnancy losses related to antiphospholipid syndrome are fetal deaths (after 10 weeks' gestation).[223, 230, 231] Frequently, fetal death is preceded by observations of poor fetal growth, oligohydramnios, heart rate abnormalities and preeclampsia or eclampsia, all of which might reflect hypoxemia resulting from placental insufficiency. Evidence indicates that antiphospholipid antibodies are directed against platelets (promoting adhesion) and the vascular endothelium (where alterations in prostacyclin/thromboxane metabolism cause vasoconstriction),[232, 233] both mechanisms predisposing to thrombosis. Circulating antiphospholipid antibodies have also been linked to reduced levels of an antithrombotic phospholipid-binding protein on the surface of trophoblasts and endothelial cells (annexin V).[234, 235] Not surprisingly, spiral arteriolar vasculopathy can often be demonstrated or, in severe cases, placental infarctions.[236] The binding of antiphospholipid antibodies to trophoblast cells reduces proliferation and invasion, as well as the release of hCG. In decidual tissue, antiphospholipid antibodies reduce the production of vascular endothelial growth factor (VEGF) and matrix metalloproteinases, and the formation of new blood vessels.[237]

Thrombosis in the placental circulation is a plausible mechanism for late pregnancy loss in women with antiphospholipid syndrome, but does not explain losses that occur before 10 weeks' gestation when maternal arterial connections with the intervillous space become established. Recent evidence suggests that another mechanism relating to abnormalities of early trophoblast invasion may be involved, in both early and late pregnancy losses.[238] In normal early pregnancy, extravillous trophoblast invades the decidual vessels, first forming plugs that later dissociate as the trophoblast migrates along the maternal arterial circulation converting the uteroplacental vessels into a low-resistance circuit.[239] Defective trophoblast invasion of the uteroplacental arteries has been well described in association with later pregnancy complications, including preeclampsia and fetal growth restriction. Antiphospholipid antibodies on the surface of the trophoblast or maternal vessel walls may inhibit endovascular trophoblast invasion, preventing formation of plugs that normally serve to limit intervillous blood flow and prevent pressure-induced or oxidative damage to the trophoblast during early placental development. Alternatively, antiphospholipid antibodies may damage the trophoblast directly. Either way, abnormal endovascular trophoblast invasion could explain early miscarriages in women with antiphospholipid syndrome and, in less severe cases, the development of later pregnancy complications relating to uteroplacental vascular insufficiency.[238] The pathophysiology involves inflammation at the maternal-fetal interface, preventing normal trophoblast development and function.[240]

Treatments for antiphospholipid syndrome have included antiplatelet agents (aspirin), anticoagulants (heparin), and immunosuppressive therapies (prednisone, intravenous immunoglobulins). Whereas most studies have found heparin therapy more effective than aspirin and combined treatment with aspirin and heparin superior to treatment with

either alone,[222, 225, 241–243] two trials found that heparin did not further improve the outcomes achieved with aspirin.[244, 245] A typical combined treatment regimen includes aspirin (75–85 mg/day), beginning with attempts at conception, and unfractionated heparin (5,000–10,000 subcutaneous twice daily), beginning at first indication of pregnancy. Live birth rates for women with antiphospholipid syndrome who receive combined treatment with aspirin and unfractionated heparin during pregnancy (70–80%) are much improved over those observed in women who receive aspirin treatment or no treatment (20–40%).[241, 242, 246–248] However, treatment does not eliminate the high risk for pregnancy complications (preterm labor, premature rupture of membranes, intrauterine growth restriction and fetal demise, preeclampsia, and placental abruption) and poses additional risks for the mother (gastric bleeding, osteopenia).[242, 246] Randomized clinical trials have demonstrated that aspirin and heparin treatment of women with two or more consecutive pregnancy losses in the *absence* of evidence for the antiphospholipid syndrome (unexplained recurrent miscarriages) yields no benefit.[249, 250]

Low molecular weight heparin offers a number of advantages over unfractionated heparin. It has an increased antithrombotic ratio (preventing abnormal clotting with fewer bleeding side effects) and is associated with a lower incidence of both thrombocytopenia and osteopenia; the relatively longer half-life of low molecular weight heparin also permits less frequent dosing and requires less frequent monitoring, both improving compliance.[251] Although experience with low molecular weight heparin in pregnancy is relatively limited, interventional trials among women with recurrent pregnancy loss and antiphospholipid syndrome and other acquired thrombophilias suggest it is both safe and effective.[225, 252–254] Prednisone may have some efficacy in the treatment of women with recurrent pregnancy loss and antiphospholipid syndrome, but its risks (diabetes, hypertension, preterm delivery) outweigh its benefits.[255, 256] Combined treatment with aspirin and heparin is effective and also safer. Intravenous immunoglobulins have also been used to treat women with recurrent pregnancy loss and antiphospholipid syndrome; their efficacy has not been compared directly to heparin/aspirin or low molecular weight heparin/aspirin.[257, 258]

Although an increased prevalence of antithyroid antibodies and antinuclear antibodies has been observed in women with recurrent pregnancy loss, their relevance is uncertain because neither predicts subsequent pregnancy outcome, and there is no logical and proven effective treatment to offer.[207, 259, 260] Treatment with levothyroxine in thyroid antibody-positive women undergoing assisted reproduction technologies had no impact on outcomes.[261] Of course, if TSH levels are elevated, even in the range of subclinical hypothyroidism, as discussed later in this chapter, treatment is indicated. Tests to detect antinuclear and antithyroid antibodies have no clinical utility in euthyroid women with recurrent pregnancy loss. However, women with thyroid antibodies have a significant risk of becoming hypothyroid as pregnancy progresses and also an increased risk of postpartum thyroiditis.[262]

An uncommon autoimmune condition, celiac disease, may be linked with recurrent miscarriages and may require appropriate antibody testing in its subclinical form.[263] Another unusual cause of recurrent miscarriages is antibody formation related to a rare blood group, P type. Patients with this condition treated with plasmapheresis have achieved successful pregnancies.[264]

Summary of Key Facts Relating to Autoimmune Factors

Autoimmune diseases like systemic lupus erythematosus and the antiphospholipid syndrome are identifiable and treatable immunologic disorders associated with recurrent pregnancy loss. A variety of mechanisms can explain how antiphospholipid antibodies predispose to placental thrombosis or interfere with normal development of the uteroplacental circulation to cause both early and late pregnancy losses. At the present time, assays for the lupus anticoagulant, anticardiolipin antibodies, and anti-β2-glycoprotein 1 antibodies are the only validated immunologic

tests having clinical utility in the evaluation of women with recurrent pregnancy loss. As yet, assays for other antiphospholipid antibodies have no proven value. Combined aspirin and heparin therapy has proven effectiveness and is the preferred treatment for women with recurrent pregnancy loss associated with antiphospholipid syndrome.

Alloimmune Disorders

In theory, normal pregnancy requires maternal immunologic recognition and response to paternally derived antigens on embryonic tissues, and abnormalities in the maternal alloimmune response may predispose to or cause recurrent pregnancy loss. Suggested mechanisms have included maternal production of cytotoxic antibodies, maternal failure to produce blocking antibodies to prevent a maternal cell-mediated immune attack (possibly because mother and father are too antigenically similar), and cytokine dysregulation of immune mechanisms operating at the maternal-fetal interface.

Antipaternal lymphocytotoxic antibodies have been proposed as a cause of recurrent pregnancy loss.[265] However, they are also present in many women with normal pregnancies and are more prevalent in fertile couples than in those with recurrent pregnancy loss.[266, 267] These observations have led most to conclude that lymphocytotoxic antibodies reflect the number and duration of pregnancies and have no effect on subsequent pregnancy outcome.[267, 268]

In contrast to the cytotoxic antibody theory, the maternal blocking antibody theory of recurrent pregnancy loss holds that maternal failure to recognize and respond to fetal antigens by production of blocking factors (presumably antibodies) leaves the embryo exposed to a lethal cell-mediated immune rejection.[269] A poor maternal response to paternal cells in mixed lymphocyte culture was regarded as evidence for the putative immune deficiency in women with recurrent pregnancy loss.[270] However, blocking antibodies were not always detectable in women with normal pregnancies,[270] frequently were detected in women with recurrent pregnancy loss,[271] and their presence does not accurately predict subsequent pregnancy outcome.[268] Again, these effects are more likely to reflect the results of previous pregnancies than to explain their outcomes.[267, 272]

Increased sharing of major histocompatibility complex human leukocyte antigens (HLA) between mother and father was seen as one factor that might hamper maternal recognition of paternally derived fetal antigens and production of presumed essential blocking antibodies.[269] Numerous studies have examined HLA sharing in couples with recurrent pregnancy loss, but results have varied widely.[272] The best evidence that increased HLA sharing might predispose to recurrent pregnancy loss came from a study in the Hutterites, a highly inbred religious sect wherein HLA sharing was associated with increased risk for pregnancy loss, but in outbred populations at low risk for HLA homogeneity, HLA sharing does not predict pregnancy outcome and testing has no clinical utility.[273, 274] Indeed, in a subsequent study of the Hutterites, genotyping for HLA loci (the previous study was with serological measurements) could not detect any influence of HLA haplotypes on fetal loss.[275]

Human leukocyte antigen-G (HLA-G) blocks killer cell activity by binding to killer cell receptors. Sequencing of the HLA-G gene indicated a variety of differences in women with recurrent pregnancy losses, polymorphisms that could account for lowered levels of HLA-G and an increased risk of miscarriage.[276, 277] In time, this type of genetic profiling could delineate specific molecular mechanisms for fetal rejection.

The idea that successful pregnancy requires some type of maternal immune suppression remains attractive. The most recent concept to emerge has implicated dysregulation of

local immune functions at the maternal-fetal interface as a cause of recurrent pregnancy loss. Investigations have focused on a unique population of large granular lymphocytes related to natural killer cells that predominate in the decidua and on the patterns of cytokine secretion by maternal immune cells in response to trophoblast antigens.

The large granular lymphocytes present in the early pregnancy decidua appear to be regulated by both hormone changes and trophoblast invasion.[278, 279] Increased numbers of these decidual natural killer cells have been observed in mouse models of recurrent pregnancy loss, and the abortion rate in mice increases when natural killer cells are activated and decreases when they are depleted.[280, 281] In women with recurrent pregnancy loss, increased decidual natural killer cell activity has been observed,[282, 283] and increased numbers of peripheral natural killer cells have been associated with subsequent pregnancy failure.[284, 285] However, whether changes in peripheral natural killer cells mirror those occurring in the decidua and whether changes observed in the decidua are the result or the cause of pregnancy loss remain unclear.[286, 287]

Antigen-stimulated immune responses involving T-helper lymphocytes are of two basic varieties that reflect the environment in which the cells mature and differentiate; those exposed to interferon (INF)-γ become T-helper lymphocyte-1 cells, and those exposed to interleukin (IL)-4 become T-helper lymphocyte-2 cells.[288, 289] T-helper lymphocyte-1 cell responses are associated with inflammation and typically characterized by the release of INF-γ and IL-12, IL-2, and tumor necrosis factor (TNF)-α. T-helper lymphocyte-2 cell responses are associated with antibody production and the cytokines IL-10, IL-4, IL-5, and IL-6.[290, 291] TNF-α can be secreted by both T-helper lymphocyte-1 and T-helper lymphocyte-2 cells but is usually associated with a T-helper lymphocyte-1 response.[291] TNF-α variant polymorphisms have been identified in women with recurrent miscarriages.[292, 293] Whereas most women with normal pregnancies appear to have a predominant T-helper lymphocyte-2 immune response to undefined trophoblast antigens, some women with recurrent pregnancy loss exhibit a T-helper lymphocyte-1 inflammatory response that may be harmful to an implanting embryo.[294–298]

Approximately 15–20% of women with unexplained recurrent pregnancy loss exhibit an abnormal T-helper lymphocyte-1 cellular immune response to trophoblast antigens in vitro, compared to less than 3% of women with normal reproductive histories.[294–296, 299–301] CD4(+) CD25(+) T cells play a role in preventing fetal rejection; the proportion of these cells is lower in decidua and blood obtained from women with recurrent miscarriages.[302, 303] Once again, whether the cytokine secretory patterns of peripheral lymphocytes accurately reflect what is occurring at the maternal-fetal interface and whether T-helper lymphocyte-1 response patterns are the cause or the effect of pregnancy loss remain uncertain and controversial.

Studies have assessed cytokine expression in trophoblast tissue, assuming that altered expression would be a marker for fetal rejection. Indeed, interleukin expression is increased in placental tissue obtained from patients with recurrent miscarriages,[304] but again it is difficult to know whether this a primary or secondary event. Searches for cytokine polymorphisms in women with recurrent miscarriages have not been fruitful.[305–308]

The idea that abnormal maternal cellular or humoral immune responses to trophoblast antigens may be a cause of recurrent pregnancy loss has stimulated development of two distinctly different immunotherapies for women with unexplained and presumed alloimmune-mediated recurrent pregnancy loss, one immunostimulatory (paternal leukocyte immunization) and the other immunosuppressive (intravenous immunoglobulins).

Paternal leukocyte immunization is based on the beneficial effects that donor or third-party leukocytes have on allograft rejection in transplant patients and evidence that it may decrease the number of circulating natural killer cells in women with recurrent pregnancy loss.[309] Numerous small randomized and nonrandomized trials and meta-analyses yielded widely varying and conflicting results. Altogether, they suggested that, at best, a relatively

small treatment effect (8–10% improvement in live birth rate) had to be weighed against the costs and risks of treatment (injection site reactions, fever, myalgias, platelet and erythrocyte alloimmunization).[272] ***A large randomized controlled mulitcenter trial found no evidence that paternal leukocyte immunization was effective in the treatment of unexplained recurrent pregnancy loss, regardless of maternal age, number of previous miscarriages, and whether the couple had a previous live birth, and concluded that paternal leukocyte immunization should not be offered as treatment for recurrent pregnancy loss.***[310] ***A comprehensive meta-analysis of treatment trials drew the same conclusion.***[311]

Intravenous immunoglobulins are prepared by isolating and pooling immunoglobulin proteins from the serum of a large number of volunteer blood donors (up to 150 donors per vial) and generally may be regarded as immunosuppressive, acting through a number of different mechanisms.[312] Intravenous immunoglobulins treatment is costly, requires a number of intravenous infusions over the course of pregnancy, and carries risks including transmission of infections (viruses, prions) and anaphylaxis.[313] ***A substantial number of randomized trials of intravenous immunoglobulins therapy in women with unexplained recurrent pregnancy loss have now been conducted. Taken together, the results failed to demonstrate that intravenous immunoglobulins are effective in improving pregnancy outcomes for women with unexplained recurrent pregnancy loss.***[272, 311, 314]

A Cochrane review of randomized trials (12 trials, 641 women) concluded that immunotherapy (including paternal leukocyte immunization, intravenous immune globulin, third party donor leukocytes, and trophoblast membranes) did not improve the live birth rate compared with placebo treatment.[315]

New treatment options include TNF-inhibiting factor and granulocyte colony-stimulating factor (GCSF). The administration of the cytokine GCSF in a small clinical trial improved the live birth rate in women with recurrent miscarriages.[316]

Summary of Key Facts Relating to Alloimmune Factors

Maternal immune recognition and response doubtless play an important role in normal pregnancy and alloimmune disorders may be a cause of otherwise unexplained recurrent pregnancy loss. At present, cytokine dysregulation of immune mechanisms operating at the maternal-fetal interface is the most likely pathophysiologic mechanism involved. However, all current methods for the evaluation of suspected alloimmunopathology, including HLA testing, immune cell evaluation (mixed lymphocyte culture, natural killer cell assays) and cytokine testing (to differentiate those with T-helper lymphocyte-1 and T-helper lymphocyte-2 patterns of immune responses to trophoblast antigens in vitro) must be considered investigational. Furthermore, neither of the two principal immunotherapies advocated for the treatment of presumed alloimmune disorders in women with unexplained recurrent pregnancy loss (paternal leukocyte immunization, intravenous immunoglobulins) has proven effectiveness.

Inherited Thrombophilias

Interest in inherited thrombophilias as a potential cause of recurrent pregnancy loss emerged quite naturally after antiphospholipid syndrome (an acquired thrombophilia) was established as a significant and treatable cause of recurrent pregnancy loss. The pathophysiologic concept is the same: in some women with recurrent pregnancy loss, the thrombogenic changes of pregnancy exaggerate an inherent predisposition to thrombosis,

resulting in reduced uteroplacental blood flow, placental thrombosis, and pregnancy loss. On balance, the results of recent investigations support the hypothesis, but many questions still remain. ***Whereas there is little argument that pregnancy is a thrombogenic state, the importance of inherited thrombophilias as a cause of recurrent pregnancy loss, the indications for evaluation, and the benefits and risks of treatment are not yet established.***

Viewed simply, pathologic thrombosis results from an imbalance between coagulation and fibrinolysis, reflecting an imbalance between clotting factors, anticoagulant proteins (protein C, protein S, antithrombin III), and fibrinolytic mechanisms. Normal pregnancy is a hypercoagulable state characterized by increased levels of factors V, VII, VIII, X and fibrinogen, decreased levels of protein S, increased resistance to activated protein C, higher concentrations of plasminogen activator inhibitors (PAI), and an increased tendency to platelet aggregation, all promoting thrombosis.

Coagulation Factors:
- **Factors that favor clotting when increased**
  - **Fibrinogen**
  - **Factors VII, VIII, X**
- **Factors that favor clotting when decreased**
  - **Antithrombin III**
  - **Protein C**
  - **Protein S**

Fibrinolysis Factors:
- **Factors that favor clotting when increased**
  - **Plasminogen activator inhibitor-1 (PAI-1)**
- **Factors that favor clotting when decreased**
  - **Antiplasmin**

A variety of genetic mutations may result in an inherited predisposition to thrombosis. Among these, the Factor V Leiden mutation (G→A at nucleotide 1691)[317, 318] and another involving the prothrombin gene (G→A at nucleotide 20210)[319] are the most common. A third relatively common mutation involves the gene encoding the enzyme methylene tetrahydrofolate reductase (C→T at nucleotide 677); homozygotes are predisposed to hyperhomocystinemia, a known risk factor for thrombosis.[320] Other established inherited thrombophilias include deficiencies of antithrombin III, protein S, and protein C. A deficiency of factor XII (involved in both coagulation and fibrinolysis) is yet another abnormality recently implicated as predisposing to thrombosis and pregnancy loss.[321]

Protein C is a key component in the anticoagulant pathway and, when activated, inhibits the actions of clotting factors V and VIII; protein S serves as a cofactor to facilitate the actions of activated protein C. Resistance to the anticoagulant properties of activated protein C can be inherited or acquired and, in either case, results in increased thrombin generation and a hypercoagulable state. Approximately 95% of activated protein C resistance is related to the Factor V Leiden mutation, which yields an altered form of Factor V that is resistant to the action of activated protein C.[238, 322] The prevalence of the Factor V Leiden mutation varies between 2 and 10% in Western populations. The risk of systemic thrombosis is increased 5-fold in heterozygotes and is 80 times higher in those who are homozygous for the mutation.[238] In some individuals, resistance to activated protein C is acquired rather than inherited.

All risk factors for thrombosis increase risk for pregnancy complications related to thrombosis, including miscarriage, preeclampsia, placental abruption, intrauterine growth restriction, and stillbirth. The association between thrombophilias and pregnancy loss varies with the type of pregnancy loss (early, late) and the thrombophilia. A meta-analysis including data from 31 studies found that although thrombophilias are associated with both early and late pregnancy loss, the association is stronger for second trimester and later

losses than for early miscarriage.[323] The observation is not surprising when one considers that maternal intervillous blood flow does not develop to any appreciable degree before approximately 8 weeks' gestation (at the earliest); thrombosis related to a thrombophilia, therefore, is less likely to explain earlier pregnancy losses.[324] The association between thrombophilias and pregnancy loss is also stronger for Factor V Leiden, the prothrombin gene mutation, and protein S deficiency than for other thrombophilic defects.[323, 325]

Numerous studies have compared the prevalence of various thrombophilias in women with recurrent pregnancy loss to those in parous controls, with conflicting results. Some have observed an increased prevalence of thrombophilias among women with recurrent pregnancy loss and others have not.[326–332] In one study of 1,000 consecutive Caucasian women with recurrent pregnancy loss, the prevalence of the Factor V Leiden mutation in women with early (3.3%) and late pregnancy loss (3.9%) was similar to that observed in a parous control group (4.0%), but the prevalence of acquired activated protein C resistance was significantly greater among women with both early (8.8%) and late miscarriages (8.7%) than in normal parous controls (3.3%).[333] These observations suggest that acquired thrombophilias may be as or even more important than inherited defects as a cause of recurrent pregnancy loss.

Another group of studies has taken the opposite approach and examined the past reproductive performance of women known to carry thrombophilias and, in general, found a relatively strong relationship between maternal thrombophilic defects and adverse late pregnancy outcomes.[334, 335] In a prospective study, the live birth rate observed among untreated women with a history of recurrent pregnancy loss or late pregnancy loss and heterozygous for the Factor V Leiden allele (38%) was substantially lower than among women with similar reproductive histories having a normal Factor V genotype (69%).[324] Whereas pregnancy losses among women with recurrent pregnancy loss and Factor V Leiden occurred both early and late, all losses observed among women without the thrombophilia occurred before 12 weeks' gestation. ***The available evidence suggests that thrombophilias predispose to a higher risk of both early and late pregnancy loss.[336, 337] However, they also demonstrate that reproductive performance is entirely normal for many women who carry a thrombophilia. At the present time it cannot be established who among women with recurrent pregnancy loss that carry a thrombophilia is truly at increased risk for pregnancy loss.***

The risk of systemic thrombosis rises when more than one inherited thrombophilia is present, and evidence indicates that the same may be true for the risk of adverse pregnancy outcomes relating to maternal thrombophilias.[334, 335, 337] The genotype of the embryo or fetus may also have important influence on the risk of pregnancy loss in women with thrombophilias; placental infarction has been observed more often when the fetus also carries the Factor V Leiden allele than when it has a normal Factor V genotype.[338]

The indications for screening women with recurrent pregnancy loss for the growing number of recognized thrombophilias are not yet firmly established. ***Currently, screening seems reasonable for women with otherwise unexplained recurrent pregnancy loss with a suspicious loss (after 8 weeks' gestation or detection of embryonic heart activity) or history of other pregnancy complications that may have resulted from thrombosis or placental insufficiency (preeclampsia, intrauterine growth restriction, placental abruption).*** In addition to a lupus anticoagulant, anticardiolipin, and anti-β2-glycoprotein 1 for diagnosis of antiphospholipid syndrome (an acquired thrombophilia), screening includes tests for the Factor V Leiden and G20210A prothrombin gene mutations. They are the two most common inherited causes of venous thromboembolism and the thrombophilias most highly associated with adverse pregnancy outcomes, but race must also be considered. Their prevalence is relatively high among those of European descent, but very low in Asians, Africans, and Native Americans.[339] Measurement of activated protein C resistance is a more global test for detection of both inherited and acquired forms of activated protein C resistance.

Screening for the methylene tetrahydrofolate reductase mutation (serum homocysteine) and antithrombin III, protein S, and protein C deficiencies also warrants consideration, based on past and family medical history.

Indications for treatment in women with recurrent pregnancy loss associated with a thrombophilia are even less well established. In uncontrolled treatment trials, improved live birth rates have been observed in women with recurrent pregnancy loss and one or more thrombophilic defects after treatment with heparin, alone or combined with aspirin.[253, 254, 340, 341] Although it seems likely that not all such women truly require treatment, any criteria for more selective or targeted thromboprophylaxis must derive from randomized controlled trials not yet available. Routine empiric aspirin treatment in untested women with recurrent early pregnancy loss has no proven benefit.[327]

Summary of Key Facts Relating to Inherited Thrombophilias

Inherited thrombophilias resulting from genetic mutations in clotting factors have emerged as a potentially important cause of recurrent pregnancy loss, but a great many women with these mutations have completely normal reproductive performance. Why some with thrombophilias miscarry and others do not is unknown; women with more than one type of mutation or whose fetus inherits the mutation may be at greater risk. At present, which women with recurrent pregnancy loss should be screened for thrombophilias and how they should be evaluated remain unanswered questions. Selected screening for the most common abnormalities in women with otherwise unexplained recurrent pregnancy loss with a suspicious loss after 8 weeks' gestation or after detection of fetal heart activity is reasonable, but routine screening of all women with recurrent pregnancy loss cannot be justified. Whereas preliminary data suggest that treatment with heparin may improve pregnancy outcomes in women with recurrent pregnancy loss who carry a thrombophilia, empiric aspirin treatment in untested women has no proven benefit.

Endocrine Factors

Endocrine factors that may predispose to an increased risk of pregnancy loss include thyroid disease, diabetes mellitus, polycystic ovary syndrome (PCOS), and luteal phase deficiency.

Hypothyroidism

Risk of pregnancy loss may be increased for women with uncorrected overt or even subclinical hypothyroidism.[342] Mild or subclinical disease generally was not considered in the past as having important clinical consequences.[343, 344] However, the results of a study of pregnancy outcomes in women with hypothyroidism challenge that notion. The incidence of pregnancy loss was very low in treated hypothyroid women having normal thyroid indices, but markedly increased in women with elevated thyroid-stimulating hormone (TSH) levels, including both women with untreated subclinical disease and those with overt disease who received inadequate exogenous thyroid hormone replacement.[342] The evidence indicates that patients with hypothyroidism, even subclinical hypothyroidism, have an increased rate of spontaneous miscarriage.[342, 345–348] ***These observations suggest that subclinical hypothyroidism is not entirely benign and further justify earlier recommendations to***

include TSH screening in the evaluation of women with recurrent pregnancy loss. Diagnosis and treatment of even subtle thyroid disease can have important benefits for women with recurrent pregnancy loss.

Remember (see Chapter 20) that once pregnant, women being treated for hypothyroidism require an increase (20–50%) in thyroxine during pregnancy, beginning as early as the 5th week of pregnancy.[349–351] When previously diagnosed hypothyroid women become pregnant, it is best to empirically increase the levothyroxine dose by about 30% as soon as pregnancy is diagnosed, with further adjustments according to the TSH levels.[351, 352] ***TSH should be monitored monthly and again in the postpartum period, and dosage should be adjusted to keep the TSH level in the lower half of the normal range, <2.5 μU/mL in the first trimester and <3.0 μU/mL in the rest of pregnancy.*** Postpartum, the dose should be immediately decreased to the prepregnancy level. The need for proper monitoring and adequate treatment cannot be overemphasized. Women with thyroid autoimmunity (presence of thyroid antibodies) must be monitored with TSH levels for at least 6 months after delivery because of their increased risk for postpartum thyroiditis.

Diabetes Mellitus

Diabetic women with good metabolic control are no more likely than nondiabetic women to suffer pregnancy loss, but diabetic women with elevated blood glucose and glycosylated hemoglobin (A1c) levels in the first trimester are at significantly increased risk for spontaneous miscarriage. Among women with poor diabetic control, miscarriage risk rises with the level of hemoglobin A1c.[353-356] ***In women with recurrent pregnancy loss, evaluation with blood glucose and hemoglobin A1c levels is indicated for those with known or suspected diabetes but is otherwise unwarranted.*** Diabetic women with recurrent pregnancy loss and elevated hemoglobin A1c concentrations are best advised to postpone new attempts to conceive until levels return to the normal range.

Polycystic Ovary Syndrome

A number of studies have shown a correlation between elevated luteinizing hormone (LH) levels and recurrent pregnancy loss.[357, 358] In the past, the observation was attributed to the adverse effects of LH itself or the hyperandrogenism induced by LH hypersecretion in women with PCOS.[359, 360] However, suppression of LH secretion with a gonadotropin-releasing hormone (GnRH) agonist before ovulation induction with low-dose exogenous gonadotropins had no effect on pregnancy outcome in women with PCOS.[361] Hyperinsulinemia and high levels of PAI activity have been implicated as the proximate cause for the increased incidence of miscarriage (30–50%) observed among women with PCOS.[357, 360, 362] Metformin is an insulin-sensitizing drug with proven clinical utility for ovulation induction in anovulatory women with PCOS[363-366] and has also been demonstrated to decrease PAI activity.[367–370] Accordingly, metformin treatment before conception and throughout pregnancy has been evaluated as one means by which risk of pregnancy loss might be reduced in women with PCOS.

Results of retrospective studies suggested that metformin treatment can reduce or eliminate the increased risk of miscarriage in women with PCOS.[371–374] However, randomized trials comparing metformin and clomiphene, alone and in combination, have found that clomiphene is clearly superior to metformin and that combined treatment is no better than treatment with clomiphene alone.[375–377] In the largest single trial, the live birth rate achieved with clomiphene treatment was significantly higher than that of metformin (22.5% vs. 7.2%) and the results of combined treatment were not significantly different (26.8%).[376] Although some have advocated metformin treatment to reduce the increased risk for miscarriage in women with PCOS, which might relate to an underlying metabolic disorder,[378] ***no differences in the***

miscarriage rates of women who did or did not receive metformin treatment have been observed in the large randomized trials.[375-377] ***A meta-analysis of 17 randomized trials concluded that metformin treatment had no effect on the risk of miscarriage.***[379]

Metformin may be added to the treatment regimen when clomiphene alone fails to restore normal ovulatory function. Metformin treatment can be discontinued after conception or after the end of the first trimester, or continued throughout pregnancy in hopes of also reducing the risk for developing gestational diabetes.[380–382] Metformin is classified as a category B drug (no teratogenic effects demonstrated in animal studies) and no teratogenic or other serious adverse effects have been observed in limited studies in women thus far.[381, 383, 384] However, treatment with metformin during pregnancy has been associated with an increased prevalence of pre-eclampsia and increased perinatal mortality in some studies,[385] although not in others.[386] Currently, *routine* metformin treatment during pregnancy is not recommended for women with PCOS.[387]

Luteal Phase Deficiency

The many different causes of poor luteal function and methods for the clinical diagnosis of luteal phase deficiency are discussed at length in the context of the evaluation of female infertility in Chapter 27. Discussion here is limited to a summary of the pathophysiology, diagnosis, and treatment of luteal phase deficiency in women with recurrent pregnancy loss.

Several lines of evidence indicate that the success of early pregnancy is dependent on progestational support from the corpus luteum until approximately 7 weeks' gestation (menstrual dates). Serial measurements of 17-hydroxyprogesterone (produced by the corpus luteum but not by the trophoblast), progesterone, estradiol, and human chorionic gonadotropin (hCG) during early spontaneous pregnancies and pregnancies achieved via in vitro fertilization (IVF) using donor oocytes in women with ovarian failure indicate that the luteal-placental shift does not occur suddenly but takes place gradually over the interval spanning the fifth through the ninth weeks of pregnancy.[388, 389] Classic studies of pregnancy outcomes in women seeking pregnancy termination who underwent surgical luteectomy at various times during early gestation strongly suggest that pregnancy usually becomes independent of the corpus luteum at approximately 7 weeks' gestation; luteectomies performed earlier in pregnancy uniformly induced spontaneous miscarriage.[390, 391]

Progesterone concentrations in both normal and abnormal early pregnancies reflect the combined contributions of the corpus luteum and the developing trophoblast, range widely, and overlap to a large extent.[388, 392–395] ***Measurements of serum progesterone levels to determine the quality of luteal function in early pregnancy and to identify pregnancies at risk that might be salvaged by support with exogenous progesterone therapy are futile.*** A low progesterone concentration during early pregnancy can reflect a defective corpus luteum, an intrinsically abnormal conceptus, or both. The alternative approach, diagnosis in a nonconception cycle and treatment to correct luteal phase deficiency before the next conception, is possible but subject to all of the limitations and pitfalls of the various diagnostic methods. Serum progesterone concentrations fluctuate widely and cannot be confidently interpreted because corpus luteum progesterone secretion is pulsatile.[396] Endometrial biopsy is invasive, painful, and costly, even if traditional histologic dating were still considered a valid diagnostic tool.[397] Consequently, an abnormally short luteal phase duration (less than 13 days), best defined by the interval from detection of the midcycle LH surge to the onset of menses, is the most objective and reliable diagnostic criterion. When that criterion is met, a serum prolactin determination is indicated to exclude hyperprolactinemia and to define the best treatment option. Estimates of the prevalence of luteal phase deficiency among women with recurrent pregnancy loss have varied greatly with the method of diagnosis[49–51, 53, 54, 398]; its true prevalence and importance as a cause of recurrent pregnancy loss is unknown, but probably quite low (less than 10%).

We view luteal phase deficiency as a subtle form of ovulation dysfunction best corrected with the same treatments used for ovulation induction in anovulatory infertile women (Chapter 31). Considering that women with luteal phase deficiency already ovulate, albeit poorly, they generally do not require aggressive treatment. Prolactin and TSH determinations are indicated, and clomiphene citrate is a logical choice in euprolactinemic euthyroid women. There is a large volume of experimental and clinical evidence to indicate that follicular phase FSH levels are low in cycles with a short luteal phase[399–403] and that clomiphene can effectively correct the abnormality.[404, 405] Some prefer to treat luteal phase deficiency with exogenous progesterone supplementation beginning 2 to 3 days after ovulation,[406] but this approach often delays menses, creating false expectations of pregnancy, increasing stress, and inviting disappointment.

Summary of Key Facts Relating to Endocrine Factors

Endocrine factors are a relatively uncommon cause of recurrent pregnancy loss. Thyroid disorders are easy to identify and treat and should be excluded by measurement of TSH; even subtle abnormalities may adversely affect pregnancy outcome. Evaluation of blood glucose and hemoglobin A1c levels is indicated for women with known or suspected diabetes mellitus, but is otherwise not warranted. The risk of miscarriage is increased in women with polycystic ovary syndrome and may be substantially reduced by treatment with metformin; for women with polycystic ovary syndrome and hyperinsulinemia who require ovulation induction, metformin is the best initial treatment. Luteal phase deficiency cannot be diagnosed during pregnancy; a consistently short luteal phase duration is the most reliable diagnostic criterion. Clomiphene citrate is an effective treatment for luteal phase deficiency and avoids the confusion, anxiety, and disappointment associated with delayed menses often resulting from exogenous progesterone therapy.

Infectious Causes

Overall, data regarding the possibility that cervicovaginal infections might be a cause of early pregnancy loss are relatively scarce. Despite periodic reports that have implicated specific infectious agents as risk factors for miscarriages, there remains no compelling evidence that bacterial or viral infections are a cause of recurrent pregnancy loss. *Chlamydia trachomatis* infection was implicated in one study that found a high prevalence of anti-Chlamydia antibodies among women with recurrent pregnancy loss, possibly reflecting an excessive maternal immunologic reaction to the organism,[407] but a later prospective study found no association between anti-Chlamydia antibodies and spontaneous miscarriage.[408] Others have reported associations between spontaneous miscarriage and genital Ureaplasma (*U. urealyticum*) or Mycoplasma (*M. hominis*) infection.[409] *Toxoplasma gondii*, *Listeria monocytogenes*, *Campylobacter* species, herpes virus, and cytomegalovirus have also been implicated.

There is an association between miscarriage risk and bacterial vaginosis. In one large study, diagnosis of bacterial vaginosis at the first prenatal visit before 14 weeks' gestation was associated with a 5-fold increased risk of pregnancy loss before 20 weeks' gestation.[410] Another study involving infertile women attempting pregnancy via IVF observed no difference between conception rates in women with and without bacterial vaginosis, but those with bacterial vaginosis who conceived were twice as likely to miscarry as those without bacterial vaginosis who conceived.[411] A third large study found that bacterial vaginosis did not predict early miscarriage but was associated with a modest increase in risk of pregnancy loss after 13 weeks' gestation.[412] Other evidence that chronic subclinical endometritis is

relatively common in women with symptomatic lower genital tract infections, including cervicitis and bacterial vaginosis, suggests a mechanism that might explain the association between bacterial vaginosis and pregnancy loss.[413–415]

The available data cannot justify routine serologic testing for past Chlamydia exposure, cervical cultures, or endometrial biopsy in the evaluation of women with recurrent pregnancy loss. However, as in infertile women, further evaluation and treatment are appropriate and prudent in women with recurrent pregnancy loss having clinical cervicitis, chronic or recurrent bacterial vaginosis, or other symptoms that suggest pelvic infection. Uncontrolled studies have suggested that empiric antibiotic treatment may decrease the risk of miscarriage in women with genital mycoplasma infections[416] and in unselected women with recurrent pregnancy loss.[417, 418] ***Considering the relatively low cost and negligible risks involved, a 2-week course of empiric antibiotic treatment (azithromycin, erythromycin, or doxycycline) is more reasonable than numerous and repeated cultures.***

An analysis of the two large clinical trials that established the efficacy of the bivalent vaccine against human papillomavirus (HPV) types 16 and 18 could detect no increase in the miscarriage rate comparing the vaccinated women with the control group.[419]

Summary of Key Facts Relating to Infectious Causes

Routine serologic testing, cervical cultures, and endometrial biopsy to detect genital infections in women with recurrent pregnancy loss cannot be justified. Evaluation should be limited to women with clinical cervicitis, chronic or recurrent bacterial vaginosis, or other symptoms of pelvic infection. Empiric antibiotic treatment in women suspected of harboring a genital mycoplasma infection is less costly and less complicated than serial cultures.

Environmental Factors

Smoking, alcohol, and heavy coffee consumption have been implicated as environmental factors predisposing to pregnancy loss.

Numerous studies have examined the relationship between cigarette smoking and miscarriage risk.[420, 421] In sum, they support the conclusion that cigarette smoking increases the risk of spontaneous miscarriage in a dose-dependent manner;[422–428] the adverse effects of smoking become apparent in smokers who consume as few as 10 cigarettes per day.[421] The mechanisms responsible are uncertain but the vasoconstrictive and antimetabolic actions of some components of cigarette smoke, including nicotine, carbon dioxide, and cyanide, may predispose to placental insuffiency.

Alcohol is a known teratogen with dose-dependent adverse effects.[420, 428–430] Alcohol consumption exceeding two drinks per day has been estimated to double the risk of spontaneous miscarriage.[420] Although low-level consumption has no measurable impact on the risk of pregnancy loss,[431] there is no clearly established safe threshold. The adverse effects of alcohol may be additive with those of cigarette smoking.

Most, but not all, studies that have examined the relationship between maternal caffeine consumption and miscarriage risk have found that heavy caffeine consumption (greater than 300 mg/day, equivalent to about three cups of coffee) is associated with a modest increase (less than 2-fold) in risk for spontaneous miscarriage.[432–439]

Couples experiencing recurrent pregnancy loss are sometimes concerned that environmental toxins may have contributed to their reproductive difficulty. Their questions are difficult to answer because information regarding the effects of potential toxins on pregnancy is not readily available. Anesthetic gases, perchorethylene (a dry-cleaning solvent), other organic solvents,[440] and exposure to heavy metals (mercury, lead) have been implicated as causative agents of miscarriage.[428] Exposure to video terminals is not a factor.[441] Exercise programs do not increase risk, and bed rest will not decrease the risk of recurrent pregnancy loss. Isotretinoin (Accutane) is definitely associated with an increased incidence of spontaneous miscarriage.[442] An increased miscarriage risk has been observed among painters and factory workers, but not among dental assistants and laboratory or gardening workers.[443, 444] The use of electric blankets and heated water beds is also not associated with an increased risk of spontaneous miscarriage.[445]

It seems appropriate to list obesity under environmental factors because excess weight is largely a consequence of lifestyle. A body mass index (BMI) equal to or greater than 25 is associated with a greater risk of miscarriage, and Chinese and British studies have linked obesity with recurrent miscarriages.[446–448]

Summary of Key Facts Relating to Environmental Factors

Smoking increases the risk of miscarriage and should be discouraged. Alcohol consumption exceeding two drinks per day and caffeine consumption exceeding 300 mg per day may increase risk for pregnancy loss and are best avoided. Women who have experienced a miscarriage should be cautioned regarding known environmental toxins. An increased risk of miscarriage is one more reason to vigorously confront obesity (Chapter 19).

Unexplained Recurrent Pregnancy Loss

Even after a thorough and systematic evaluation, well more than half of all women with recurrent pregnancy loss have no identified predisposing factors that can explain their poor reproductive history, and the large majority does well in the next pregnancy. Those with a previous second-trimester loss have a poorer prognosis and are at increased risk for preterm delivery, stillbirth, and neonatal death.[51–54, 449, 450] Frequent communication, cautious optimism, and emotional support during the first trimester of the next pregnancy have their own distinct therapeutic value.[51] With determined efforts, 70–75% of women with unexplained recurrent pregnancy loss ultimately achieve a successful pregnancy.[6] Careful monitoring is warranted because women with recurrent pregnancy loss are also at increased risk for ectopic pregnancy.[451]

Many clinicians offer or recommend empiric exogenous progesterone supplementation during early pregnancy to women with unexplained recurrent pregnancy loss. Any concerned clinician wants to do all that can be done to improve the chances for a successful pregnancy. Considering that two-thirds or more of next pregnancies among women with unexplained recurrent pregnancy loss are likely to succeed, with or without treatment, it is also easy to understand why so many are also convinced that treatment has value when there is no compelling evidence for its effectiveness.[452–454] Low-dose aspirin treatment is another commonly recommended treatment for women with unexplained recurrent early pregnancy loss, even though randomized trials have demonstrated that it has no benefit.[243, 327] In the doses commonly administered, exogenous progesterone supplementation and

low-dose aspirin pose few risks and are difficult to condemn, but without clear evidence for their effectiveness, neither can be recommended.

Summary of Key Facts Relating to Unexplained Recurrent Pregnancy Loss

Even thorough evaluation will reveal no evidence for a predisposing factor in more than half of all women with recurrent pregnancy loss. Under such circumstances, the longer term prognosis for achieving a successful pregnancy is very good. Emotional support and careful monitoring during early pregnancy help to improve pregnancy outcomes. Empiric treatments with exogenous progesterone or aspirin in women with unexplained recurrent pregnancy loss have no proven value.

Summary of Evaluation and Treatment for Recurrent Pregnancy Loss

As a quick reference, the following table summarizes our recommended evaluation and treatments for factors known to predispose to recurrent pregnancy loss. ***Established tests and treatments are shown in bold type. Tests and treatments that must be applied selectively and those not yet firmly established are shown in standard type.***

Category	*Evaluation*	*Treatments*
Genetic	**Karyotype, both partners** Ovarian reserve test Comparative genomic hybridization	**Counseling** **Donor gametes where appropriate** Preimplantation genetic diagnosis
Anatomic	**Sonohysterography or HSG** Magnetic resonance imaging IVP or renal ultrasound	**Hysteroscopic septoplasty** **Hysteroscopic myomectomy** **Hysteroscopic adhesiolysis** Abdominal metroplasty Abdominal myomectomy Cervical cerclage
Immunologic	**Lupus Anticoagulant** **Anticardiolipin Antibody** **Anti-β2-glycoprotein 1 Antibody**	**Aspirin and Heparin**
Thrombophilias	Factor V Leiden Prothrombin Gene Mutation Activated Protein C Resistance Homocysteine Protein C Protein S Antithrombin III	Heparin
Endocrine	**TSH** **Luteal Phase Duration** Blood Glucose, Hgb A1c Prolactin	**Thyroxine** **Clomiphene citrate** **Metformin** **Dopamine agonists**
Infectious	**As indicated by symptoms**	Empiric antibiotics
Environmental	**History**	Behavior modification

All references are available online at: http://www.clinicalgynendoandinfertility.com

29 Endometriosis

Endometriosis is a benign disease defined by the presence of endometrial glands and stroma outside of the uterus and is associated with both pelvic pain and infertility. The ectopic endometrial tissue usually is located in the pelvis but can appear anywhere in the body. The disease exhibits a broad spectrum of clinical signs and symptoms, is prone to progression and recurrence, and often presents vexing clinical management problems for women and their physicians. The pathogenesis and natural history of endometriosis remain poorly understood, but investigations employing modern molecular methods are yielding new insights into the mechanisms of the disease and suggesting new approaches to its diagnosis and treatment.

The Epidemiology of Endometriosis

The true overall prevalence of endometriosis is unknown, primarily because surgery is the only reliable method for diagnosis and generally is not performed on women without symptoms or physical findings that strongly suggest the disease; accordingly, estimates vary with the indication for surgical treatment. The prevalence of asymptomatic endometriosis is 1–7% in women seeking elective sterilization, 12–32% among women of reproductive age with pelvic pain, 9–50% in infertile women, and approximately 50% among teens with chronic pelvic pain or dysmenorrhea.[1-6] The overall prevalence of endometriosis in reproductive aged women probably is between 3% and 10%.[7-10]

The mean age at time of diagnosis of endometriosis ranges between 25 and 35 years.[11, 12] Endometriosis is rare in premenarcheal girls but may be identified in half or more of adolescents and young women under age 20 with complaints of chronic pelvic pain or dyspareunia.[3, 4, 13] Most cases in young women under age 17 are associated with müllerian anomalies and cervical or vaginal obstruction.[14] Fewer than 5% of women who require surgery for endometriosis are postmenopausal and most of those have received estrogen

therapy.[15, 16] The prevalence of asymptomatic endometriosis may be somewhat lower in Blacks and higher in Asians than in White women.[6, 17]

Early menarche and short menstrual cycles have been associated with increased risk for endometriosis.[7, 17] The correlation between the risk of disease and the volume or duration of menses is less consistent.[7, 18, 19] Interestingly, the prevalence of endometriosis is inversely related to body mass index.[12, 17, 20] Pregnancy has a protective effect that decreases with time; whereas risk decreases with parity and prolonged periods of lactation, risk increases with the number of years since last childbirth.[21, 22] Assorted epidemiologic studies have suggested that heavy consumption of alcohol and caffeine also may increase risk and that regular exercise and smoking may decrease risk for endometriosis.[7, 17] Primate data have suggested that exposure to polychlorinated biphenyl (PCB) or dioxin may be linked with endometriosis, but studies in women have yielded inconsistent results.[23] Other data suggest that exposures *in utero* can play a role in development of the disease; the incidence of endometriosis is increased in women having prenatal exposure to diethylstilbestrol.[24]

Pathogenesis of Endometriosis

Although the classic peritoneal lesions of endometriosis were first described in the 1800s, the disease forever will be linked to John Sampson who, in 1921, described a series of perforating hemorrhagic ovarian cysts he called "chocolate cysts," coining the term "endometriosis" to describe the peritoneal implants he first envisioned as seedlings derived from disease in the ovary.[25] There is no generally accepted thesis regarding the origin of endometriosis. Several pathogenic mechanisms have been proposed, including retrograde menstruation and implantation, coelomic metaplasia, direct transplantation, and vascular dissemination. No one mechanism explains all cases of endometriosis and each probably contributes, at least to some extent.

The retrograde menstruation and implantation theory holds that endometrial tissue shed during menstruation is transported via the fallopian tubes into the peritoneal cavity where it implants on the surfaces of pelvic organs. Sampson's classic paper proposing that endometriosis is 'due to menstrual dissemination of endometrial tissue into the peritoneal cavity' was published in 1927.[26] Several lines of evidence support the implantation theory as the primary mechanism involved in the pathogenesis of endometriosis.

- When laparoscopy is performed during menses, blood in the peritoneal fluid can be observed in 75-90% of women with patent fallopian tubes.[27-29]
- Viable endometrial cells recovered from the peritoneal fluid during menses can be grown in cell culture[28, 30] and can attach to and penetrate the mesothelial surface of the peritoneum.[31, 32]
- Endometriosis is more prevalent in women with obstructing müllerian anomalies than in women with malformations that do not obstruct menstrual outflow.[33]
- The incidence of endometriosis is increased in women with an early menarche, short menstrual cycles, or menorrhagia.[7, 17, 18, 34]
- Endometriosis is observed most commonly in the dependent portions of the pelvis, on the ovaries, in the anterior and posterior cul-de-sacs, and on the uterosacral ligaments, the posterior uterus, and the posterior surface of the broad ligaments.[35-37]
- Experimental endometriosis can be induced in nonhuman primates after surgically-induced peritoneal menstruation[38, 39] or retroperitoneal injection of menstrual endometrium,[40] and in women who receive peritoneal injections of their menstrual tissue.[41]

Although the evidence for the implantation theory may seem convincing, the coelomic metaplasia theory offers an alternative explanation for the same observations. ***The coelomic metaplasia theory holds that endometriosis results from spontaneous metaplastic change in mesothelial cells derived from the coelomic epithelium (located in the peritoneum and the pleura). The induction theory is a variation on the same theme and envisions that coelomic metaplasia is induced by exposure to menstrual effluent or other stimuli.*** In his original paper, Sampson himself allowed that foci of peritoneal endometriosis also might be 'due to some specific irritant present in the cyst contents which stimulates the peritoneal endothelium to a metaplasia with the development of endometrial tissue typical both in structure and function.'[25] A number of observations suggest that endometriosis results from spontaneous or induced coelomic metaplasia, at least in some cases.[42]

- Endometriosis has been described in a premenarcheal girl,[43] in women who never have menstruated,[44, 45] and also occurs in adolescent girls having had relatively few menstrual cycles.[46]
- Because intact endometrial cells have no access to the thorax in the absence of an anatomical defect, the implantation theory cannot explain cases of pleural and pulmonary endometriosis (almost all of which occur on the right side).[47] Metaplasia in the pleura (derived from the coelomic epithelium, like the peritoneum and the müllerian ducts), induced by steroid hormones or chemical stimuli released by degenerating endometrial cells into the peritoneal fluid (which communicates with the thoracic cavity via the right hemi-diaphragm), is the more plausible explanation.
- Metaplasia in misintegrated coelomic epithelium (adjacent to the mesenchymal limb buds during early embryogenesis)[48, 49] can explain endometriosis in unusual peripheral sites like the extremities (thumb, thigh, knee).[50-52]
- Rare cases of endometriosis have been observed in men treated with high doses of estrogen (urinary bladder, abdominal wall).[53-55]
- Ovarian surface epithelium and stromal cells, co-cultured with estradiol in a three-dimensional collagen gel lattice, form endometrial glands and stroma.[56]
- Eutopic (inside the uterus) and ectopic (outside the uterus) endometrium are distinctly different, both morphologically and functionally, which is difficult to reconcile with the notion that endometriotic implants represent autotransplants of normal endometrial tissue.[57]

Other mechanisms are invoked to explain cases of extrapelvic endometriosis.[58, 59] Although coelomic metaplasia might explain endometriosis in the pelvis, the thoracic cavity,[47] the urinary and digestive tracts,[60, 61] the inguinal canal,[62] and the umbilicus,[63] evidence indicates that ***vascular or lymphatic dissemination of endometrial cells*** also may be involved.[64-67] Alternatively, circulating stem cells originating from the bone marrow might differentiate into endometriotic tissue in various locations.[68] Inadvertent ***direct transplantation of endometrial tissue*** at the time of cesarean section, other pelvic surgery, or episiotomy repair seems the most plausible explanation for endometriosis found in abdominal scars[69, 70] and in the perineum.[71]

Regardless whether pelvic endometriosis results from implantation of viable endometrial tissue regurgitated into the peritoneal cavity at time of menses or from coelomic metaplasia induced by hormones or other chemical stimuli derived from degenerated endometrial cells, several key questions remain. Why does endometriosis develop only in some women when retrograde menstruation occurs in most women? What explains the widely varying presentations of the disease? Why is there such a poor correlation between the extent of disease and the severity of associated symptoms? Studies of the immune function and genetics of women with endometriosis are suggesting answers.

The Immunobiology of Endometriosis

Endometriosis has been associated with headaches, arthalgias and myalgias, allergies, eczema, hypothyroidism, fibromyalgia, chronic fatigue syndrome, and susceptibility to vaginal candidiasis,[72] suggesting a possible link between endometriosis and autoimmune disease. A cross-sectional survey of members of the Endometriosis Association found that members with endometriosis had a higher prevalence of hypothyroidism, chronic fatigue syndrome, rheumatoid arthritis, systemic lupus erythematosis, Sjogren syndrome, and multiple sclerosis, compared with the published rates in the general U.S. female population; allergies and asthma also were more common.[73] Although provocative, the results of the study were highly susceptible to recall and ascertainment bias. Others have found no association between endometriosis and autoimmune disease.[74]

A higher prevalence of antinuclear antibodies has been reported for women with endometriosis.[75, 76] The most common autoantibodies identified have been directed against endometrial antigens,[77-82] including transferrin and laminin-1, which also is found in embryonic tissues.[77, 83] Such immunologic autoreactivity probably results from inflammation and develops as a consequence of chronic local tissue destruction.[72]

Endometriosis is associated with changes in both cellular and humoral immunity, suggesting that impaired immune function may contribute to the development of the disease. Altered immune function may predispose some women to develop endometriosis, or influence the severity of disease in affected women. Numerous immune-mediated mechanisms have been implicated. Although the peritoneal fluid of women with endometriosis contains increased numbers of immune cells, evidence suggests that their actions do more to promote the disease than to prevent it.

Macrophages are a key element of the innate immune response, the part of the immune system that is not antigen-specific and does not involve immunologic memory. Macrophages defend the host by recognition, phagocytosis, and destruction of offending microorganisms and also serve as scavengers, helping to clear apoptotic cells and cellular debris. Macrophages secrete a variety of cytokines, growth factors, enzymes, and prostaglandins that help to mediate their own functions while stimulating the growth and proliferation of other cell types. Macrophages are a normal inhabitant of the peritoneal fluid and their numbers and activity are much increased in women with endometriosis.[84-87] ***Instead of acting as scavengers to eliminate ectopic endometrial cells, activated peritoneal macrophages and circulating monocytes in women with endometriosis appear to promote the disease by secreting growth factors and cytokines that stimulate proliferation of ectopic endometrium and inhibit their scavenger functions.***[88, 89]

Natural killer (NK) cells are another important component of the innate immune system and function in two ways. NK cells have receptors for immunoglobulin G (IgG) and kill IgG-bound cells in a process known as antibody-dependent cellular cytotoxicity. NK cells also have killer-activating and killer-inhibiting receptors that, when occupied, direct or inhibit cytotoxic activity. Whereas studies of the numbers of peritoneal NK cells in women with endometriosis have yielded conflicting results,[87, 90, 91] those investigating NK cell function have consistently observed decreased cytotoxic activity, which is most pronounced in women with advanced stages of the disease.[91-93] One of the mechanisms responsible appears to involve over-expression of killer-inhibiting receptors in both peripheral and peritoneal cells in women with endometriosis.[94, 95]

Lymphocytes mediate the acquired immune response. B lymphocytes mature in the bone marrow and secrete immunoglobulins, which are antigen-specific antibodies directed against extracellular microorganisms. T lymphocytes help B cells to make antibodies and also eliminate intracellular pathogens by activating macrophages and by killing virus-infected

or malignant cells. T cells are of two types, cytotoxic/suppressor T cells (involved in the cellular immune response) and helper T cells (involved in the humoral immune response). The numbers of both T cell types are increased in the peritoneal fluid of women with endometriosis and in the stroma of ectopic endometrium.[87, 96, 97]

Cytokines and growth factors are a large family of soluble proteins and glycoproteins secreted by leukocytes and other cells into the extracellular environment where they act on the same (autocrine) or nearby cells (paracrine), serving as messengers both within and outside the immune system in regulation of chemotaxis, mitosis, angiogenesis, and differentiation. Whereas an impaired cellular immune response may result in ineffective clearance of refluxed endometrial cells from the peritoneal cavity, cytokines and growth factors appear to promote implantation and growth of ectopic endometrium by facilitating its attachment to the peritoneal surface and by stimulating proliferation and angiogenesis.

Interleukin-1 (IL-1) is a cytokine involved in inflammatory and immune responses and is secreted by activated monocytes and macrophages, T and B lymphocytes, and NK cells. IL-1 has been identified in the peritoneal fluid of women with endometriosis and IL-1 receptor expression is increased in endometriosis-derived stromal cells.[91, 92, 98] IL-1 may promote the development of endometriosis by stimulating the release of angiogenic factors (vascular endothelial growth factor, interleukin-6, interleukin-8)[99, 100] and by helping endometrial cells that enter the peritoneal cavity to escape immunosurveillance by inducing the release of a soluble form of intercellular adhesion molecule-1 (ICAM-1) from endometriotic cells that competes for immune recognition sites on NK and other immune cells.[101, 102]

Interleukin-8 (IL-8) is a potent angiogenic cytokine produced by mesothelial cells, macrophages, endometrial and other cells. Peritoneal fluid levels of IL-8 are elevated in women with endometriosis and correlate with the severity of disease.[103] IL-8 is expressed in endometriotic implants and is up-regulated by IL-1.[100, 104] IL-8 stimulates adhesion of endometrial stromal cells to extracellular matrix proteins, matrix metalloproteinase activity, and endometrial stromal cell proliferation in a dose-dependent manner, all of which may help to promote the implantation and growth of ectopic endometrium.[105-107]

Moncyte chemotactic protein-1 and RANTES (regulated on activation, normal T-cell expressed and secreted) are two chemo-attractant cytokines that recruit macrophages into the peritoneal cavity. Both are secreted by a variety of leukocytes and by mesothelial and endometrial cells, and production of both is increased in ectopic endometrium.[108-110] In women with endometriosis, peritoneal fluid concentrations are increased and correlate with the severity of disease.[111, 112] IL-1 up-regulates monocyte chemotactic protein-1 expression in eutopic endometrial epithelial cells in women with endometriosis and in cultured ectopic endometrial cells,[113, 114] an action further stimulated by estrogen.[115] RANTES production by endometriotic implants is stimulated by other peritoneal fluid cytokines.[98]

Tumor necrosis factor-α (TNF-α) is an inflammatory cytokine produced by activated lymphocytes, macrophages, and NK cells, among others. TNF-α is expressed in eutopic endometrial epithelial cells and is up-regulated by IL-1. Peritoneal fluid concentrations are increased in women with endometriosis and correlate with the stage of the disease. The observations that TNF-α increases adherence of cultured stromal cells to mesothelial cells suggests it may facilitate attachment of ectopic endometrium to the peritoneum in women with endometriosis.

To implant and grow, ectopic endometrium must establish a blood supply. Vascular endothelial growth factor (VEGF) is an important mediator of local angiogenesis produced by monocytes and macrophages. The growth factor stimulates proliferation of vascular endothelial cells and also acts as a chemo-attractant for monocytes. VEGF is produced primarily in endometrial glands and is up-regulated by a variety of factors including estrogen and IL-1. Peritoneal fluid VEGF concentrations are increased in women with endometriosis

and are highest in advanced stages of the disease. VEGF also is expressed in endometriotic lesions, more so in active red lesions than in inactive "powder-burn" implants.

Summary

A wide variety of immunologic abnormalities has been described in women with endometriosis. The peritoneal fluid of affected women contains increased numbers of immune cells, but instead of acting to efficiently remove refluxed endometrial debris from the peritoneal cavity, they appear to promote the disease via two basic mechanisms.

Defects in cellular immune mechanisms (mediated by macrophages and NK cells) impair normal recognition and clearance of refluxed endometrial debris via the immune/inflammatory response, thereby affording endometrial cells and tissue fragments the opportunity to attach to peritoneal mesothelial cells and to invade into the extracellular matrix where they can persist and proliferate. Immune cells in the peritoneal fluid of women with endometriosis also secrete a wide variety of humoral factors (cytokines and growth factors) that stimulate attachment and proliferation of ectopic endometrium and local angiogenesis.

Altered immune function may thus predispose to the development of endometriosis or to more severe disease. Although it is not yet clear whether the immunologic abnormalities observed in women with endometriosis are the cause or the consequence of the disease, they almost certainly play an important role in its pathogenesis.

The Genetics of Endometriosis

In both humans[116-118] and nonhuman primates,[119] endometriosis tends to cluster within families, suggesting that genetic factors probably influence susceptibility to developing endometriosis. The disease frequently is observed in monozygotic and dizygotic twin pairs[120-122] and exhibits a similar age of onset in affected non-twin sisters.[123] ***Endometriosis is six to seven times more prevalent among the first-degree relatives of affected women than in the general population.***[124-127] All of these observations suggest that endometriosis has a genetic foundation and that a predisposition to the disease is inherited as a complex genetic trait for which the phenotype reflects interactions between allelic variants of susceptibility genes and environmental factors.[128, 129] Gene-expression profiling has identified candidate susceptibility genes relating to implantation failure, infertility, and progesterone resistance.[130, 131]

Molecular Mechanisms

Genes predisposing to the development of endometriosis might include any that direct the molecular processes controlling the survival of detached endometrial cells, their attachment to and invasion of peritoneal surfaces, proliferation, neovascularization, or the inflammatory response. The ectopic endometrium of women with endometriosis exhibits abnormal expression of numerous gene products that are relevant to the pathogenesis of the disease. Compared to normal endometrium, ectopic endometrial implants produce excessive amounts of estrogen, prostaglandins, and cytokines.[132-134] Similar, but more subtle, abnormalities are observed in the eutopic endometrium of women with endometriosis. ***These observations suggest that abnormalities intrinsic to the endometrium of women who develop endometriosis predispose to cell survival, ectopic implantation, proliferation, and chronic inflammation.***

Retrograde menstruation occurs in most women, but endometriosis develops in only a few. The survival of refluxed endometrial debris might result from immune dysfunction, as discussed above, or may reflect a molecular abnormality in the eutopic endometrium of women with endometriosis,[135] as discussed below. In any case, eutopic endometrial cells from women with endometriosis are resistant to apoptosis, the normal but complex gene-regulated physiologic process of programmed cell death that contributes to endometrial breakdown and turnover during the late secretory and menstrual phases of the cycle.[136, 137] Ectopic endometrium appears even more resistant to apoptosis.[138] ***Resistance to apoptosis may improve the survival of endometrial cells entering the peritoneal cavity and also help to explain why ectopic endometrium is resistant to macrophage-mediated immune surveillance and clearance***.

In normal women, levels of endometrial estrogen and prostaglandin E_2 (PGE_2) production are low because activity of the enzymes aromatase (mediating estrogen synthesis) and cyclooxygenase-2 (COX-2, mediating prostaglandin synthesis) are low. Moreover, during the secretory phase of the cycle, progesterone stimulates 17β-hydroxysteroid dehydrogenase (17βHSD) activity, which converts estradiol to the less potent estrogen, estrone.[10] In women with endometriosis, aromatase and COX-2 activity are increased in eutopic endometrium, and greatly elevated in ectopic endometrial tissue implants. Tissue levels of estradiol are high, due to increased aromatase and decreased 17βHSD activity, and tissue levels of PGE_2 are increased because COX-2 activity is elevated.[10] ***The eutopic and ectopic endometrium of women with endometriosis thus differs from normal endometrium in at least three distinct and important ways, exhibiting 1) high local estrogen production, 2) high local prostaglandin production, and 3) resistance to the actions of progesterone.***

Estrogen Production

There is little question that estrogen plays an important role in the pathogenesis of endometriosis. With rare exception, the disease arises only after menarche and regresses after menopause.[139] Estrogen stimulates endometriosis, whereas aromatase inhibitors cause its regression.[140] ***Substantial evidence indicates that both estrogen production and metabolism are altered in endometriosis in ways that promote the disease.***[10, 141, 142]

Estrogen in women with endometriosis derives from three major sources. As in normal women, estrogen is secreted by the ovary into the circulation and released directly into the peritoneal cavity at ovulation, and is produced in adipose and skin via conversion of circulating androgens. However, in women with endometriosis, substantial amounts of estrogen also are synthesized locally, in endometriotic tissue, which expresses a complete set of steroidogenic enzymes, including aromatase.[143]

The overproduction of estrogen in endometriotic stromal cells is linked with another of the functional abnormalities observed in the tissue, high local production of prostaglandins. PGE_2 stimulates the expression of all of the genes encoding the steroidogenic enzymes required for synthesis of estradiol from cholesterol in endometriotic stromal cells, including, in particular, *STAR* (encoding the steroidogenic acute regulatory protein, STAR) and *CYP19A1* (encoding aromatase).[143] Stromal cells in both normal and ectopic endometrium express all of the PGE_2 receptor subtypes (EP_1, EP_2, EP_3, and EP_4).[141] In endometriotic stromal cells, activation of the EP_2 receptor (by PGE_2) increases intracellular levels of cyclic adenosine monophosphate (cAMP), which induces *STAR* and *CYP19A1* expression.[141, 143, 144] PGE_2 or a cAMP analog increases STAR and aromatase levels and activity in endometriotic stromal cells, but *not* in normal endometrial stromal cells.[134, 143, 144] These observations indicate that PGE_2- and cAMP-dependent steroidogenesis in endometriotic stromal cells requires downstream effectors that are absent or opposed by other inhibitory mechanisms in normal endometrial stromal cells, which exhibit no steroidogenic activity.

The key downstream effector molecule in endometriotic tissue is steroidogenic factor-1 (SF-1), which is absent in normal endometrium. In endometriotic stromal cells exposed to PGE_2, SF-1 binds to and assembles an enhancer transcriptional complex, which interacts with the promoters of *STAR* and *CYP19A1* and induces their expression.[143] In normal endometrial stromal cells, the absence of any steroidogenic response to PGE_2 can be attributed to the absence of SF-1 and to the presence of transcriptional inhibitors of the *STAR* and *CYP19A1* gene promotors. These repressors include the Wilms' tumor 1 transcription factor (WT1),[145] and CCAAT/enhancer binding protein β (C/EBPβ), levels of which are much higher in normal endometrium than in endometriotic tissue. In the absence of SF-1, a transcriptional complex of inhibitors binds to and suppresses steroidogenic promoters in endometrial cells.[143]

Prostaglandin Production

Prostaglandins are locally produced hormones involved in inflammation and pain. Both PGE_2 and prostaglandin $F_{2\alpha}$ ($PGF_{2\alpha}$) are overproduced in the uterine and endometriotic tissues of women with endometriosis.[146] $PGF_{2\alpha}$ stimulates both vasoconstriction and myometrial contractions, resulting in pain and dysmenorrhea. As in most cells, production of prostaglandin H_2 (PGH_2) in the myometrium, endometrium, and endometriotic tissue is catalyzed by cyclooxygenase (COX), which has two isoforms.[147] COX-1 primarily drives basal prostaglandin synthesis and COX-2 is involved in inflammation. PGH_2 is metabolized to other prostaglandins by cell-specific enzymes; in the uterus, PGH_2 is converted to $PGF_{2\alpha}$ (by prostaglandin F synthase) and PGE_2 (by prostaglandin E synthase).[146]

In both ectopic and eutopic endometrium in women with endometriosis, COX-2 is upregulated to a greater extent than in endometrial stromal cells from disease-free women,[148, 149] resulting in increased PGE_2 production that induces local estrogen synthesis (as discussed above) and causes inflammation, resulting in pain[134, 144, 150, 151]; prostaglandin E synthase activity also may be increased.[10] COX-2 expression and PGE_2 production in uterine and endometriotic tissues are stimulated by IL-1β, PGE_2 (an autocrine action), VEGF, and estradiol (via estrogen receptor β).[10, 152-154] Altogether, these complementary and redundant mechanisms maintain high levels of PGE_2 production in endometriotic tissue.

The high levels of PGE_2 in endometriotic tissue induce a classical inflammatory response, characterized by increased production of cytokines, metalloproteinases, and chemokines. Increased levels of inflammatory cytokines (IL-1β, IL-6 and TNF-α) promote adhesion of shed endometrial tissue to peritoneal surfaces, and proteolytic membrane metalloproteinases promote their implantation.[110, 130, 132, 155-158] Increased levels of chemokines (monocyte chemoattractant protein 1, IL-8, and RANTES) attract increased numbers of granulocytes, NK cells, and macrophages,[109, 110, 130, 132, 155, 156, 159] and auto-regulatory positive feedback loops perpetuate the process.

Progesterone Resistance

Whereas estrogen clearly aggravates endometriosis, the effects of progesterone are less clear. Progesterone stimulates proliferation of normal endometrial stromal cells during the secretory phase, at least transiently.[10] Although progestins can be effective for relieving pain in women with endometriosis,[160, 161] a variety of selective progesterone receptor (PR) modulators having mixed agonistic and antagonistic actions also are.[162, 163] Moreover, endometriotic tissue produces substantial amounts of progesterone and contains much lower levels of PR than normal endometrium.[144, 164]

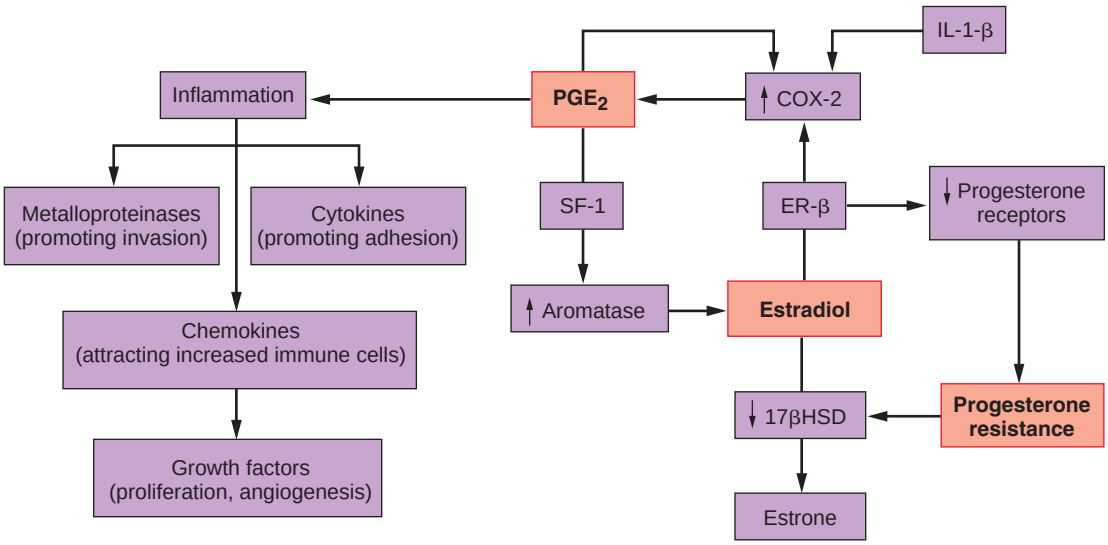
In the normal menstrual cycle, progesterone induces differentiation of both endometrial epithelial and stromal cells.[165-167] In the epithelium, progesterone stimulates glycodelin production, and in the stroma, progesterone induces decidualization and stimulates prolactin production. Progesterone also stimulates prolactin production in endometriotic stromal cells, but to a significantly lesser extent, suggesting some degree of progesterone resistance.[168] In the normal endometrium, progesterone acts as an anti-estrogen in a paracrine fashion, by stimulating retinoic acid production in the stroma,[169-171] which then induces 17βHSD activity in the epithelium,[172-175] resulting in the conversion of estradiol to the less potent estrogen, estrone. However, in endometriotic stromal cells, progesterone does not induce retinoic acid production,[176, 177] and epithelial 17βHSD activity remains low; tissue estradiol levels are elevated, because high local aromatase activity drives production and low 17βHSD activity impairs its metabolism.[168]

Gene-expression profiling of the endometrium of women with and without endometriosis has identified a number of genes linked to the actions of progesterone that are downregulated during implantation when progesterone levels normally peak, such as glycodelin, suggesting that the eutopic endometrium of women with endometriosis also is progesterone-resistant.[130, 131] Abnormalities in PR regulation might explain the phenomenon.[164] During the normal menstrual cycle, levels of both PR-A and PR-B increase progressively during the proliferative phase and decrease after ovulation, indicating that estradiol stimulates PR production, but in endometriotic tissues, PR-A levels are much reduced and PR-B is undetectable.[10] Decreased production of a binding protein required for PR function also may contribute to progesterone resistance.[178]

Epigenetic Changes

The very high levels of SF-1 observed in endometriotic tissue (>12,000 times greater than in normal endometrium) appear related to a cyosine–phosphate–guanine (CpG) island at its promoter, which is heavily methylated in normal endometrial stromal cells and unmethylated in endometriotic stromal cells.[179] Whereas an inhibiting transcription factor binds to the methylated SF-1 promoter, preventing its interaction with transcriptional activators, a stimulating transcription factor binds to the unmethylated SF-1 promoter in endometriotic cells, and activates it.[180]

The more than 100-fold higher levels of ER-β in endometriotic tissue, compared to normal endometrium, also are associated with hypomethylation of a CpG island, at the promoter



region of the ER-β gene, causing high levels of expression. In endometriotic stromal cells, ER-β occupies the ER-α promoter and suppresses its activity, yielding high levels of ER-β to bind to the PR promoter, which down-regulates PR expression.[181]

Summary

The working molecular model of the pathogenesis of endometriosis centers on increased survival of refluxed endometrial cells, (resistance to apoptosis) and functional abnormalities in eutopic and ectopic endometrial cells (high local production of estrogen and prostaglandins, progesterone resistance) originating from epigenetic changes (hypomethylation of the promoters for SF-1 and ER-β), all combining to induce a chronic inflammatory response in a feed-forward, self-perpetuating cycle.

High local production of PGE_2 stimulates aromatase expression (via SF-1), resulting in increased local production of estradiol, which stimulates COX-2 activity (via ER-β), thereby maintaining the stimulus for increased PGE_2 production. PGE_2 also induces an inflammatory reaction, with increased local production of cytokines (promoting adhesion), metalloproteinases (promoting invasion), and chemokines (attracting increased numbers of immune cells, which secrete growth factors that stimulate proliferation and angiogenesis). Increased expression of ER-β suppresses PR expression, causing progesterone resistance, manifested as decreased 17βHSD activity, which decreases metabolism of high local estradiol levels generated via increased aromatase activity.

Ultimately, inflammatory and immune responses, angiogenesis, and apoptosis are altered in ways that promote the survival, attachment, and proliferation of ectopic endometrial tissue.

Mechanisms of Pain

Pain is the most common symptom associated with endometriosis. The mechanisms involved are difficult to determine, for a number of reasons. Pain itself is hard to measure, especially when it is chronic. The hormonal environment influences the perception of pain. The placebo effect on pain is substantial and varies among studies. Chronic pelvic pain has a tendency to involve surrounding organ systems over time. The perception and tolerance of pain also vary widely among women.

The pain associated with endometriosis has been attributed to three primary mechanisms.

- The direct and indirect effects of focal bleeding from endometriotic implants.
- The actions of inflammatory cytokines in the peritoneal cavity.
- Irritation or direct infiltration of nerves in the pelvic floor.

Among these, neural irritation or invasion has received the majority of recent attention. Tender nodularity in the cul-de-sac and along the uterosacral ligaments has approximately 85% sensitivity and 50% specificity as a clinical criterion for the diagnosis of deeply infiltrating endometriosis.[182] Severe dysmenorrhea and deep dyspareunia are commonly associated symptoms; those having disease adjacent to or within the rectal wall also may have dyschesia.[183] ***The intensity of pain associated with deeply infiltrating endometriosis relates to the depth of penetration and to the proximity or direct invasion of nerves.***[182,184-186]

Whereas neural inflammation or invasion might explain the pain of women with deeply infiltrating endometriosis, it cannot be the mechanism that produces pain in women who have only superficial disease. ***The pain associated with mild disease more likely relates to inflammation resulting from cyclic focal bleeding in and around peritoneal implants, or from the actions of inflammatory cytokines released by the larger numbers of macrophages and other immune cells in the peritoneal fluid of women with endometriosis.*** However, there is no relationship between stage, site, or the morphologic characteristics of pelvic endometriosis and pain.[187] The explanation for why many women with advanced endometriosis have little or no pain and those with mild disease may have incapacitating pain remains unclear. The cause may relate to the fact that severe disease is generally chronic and may be less metabolically active. There also is evidence to suggest that persistent neural input from endometriotic tissues may cause central sensitization of the nociceptive system (neurons that receive painful stimuli), manifested by somatic hyperalgesia (increased sensitivity to pain) and areas of referred pain in some women with endometriosis.[188]

One additional mechanism that may be involved in the pain associated with endometriosis relates to evidence that the hormonal milieu influences pain perception. Numerous studies have examined measures of pain perception across the menstrual cycle. A meta-analysis including 16 such studies concluded that somatic sensory pain thresholds and tolerance are near their lowest levels just prior to and during menses.[189]

Mechanisms of Infertility

Endometriosis is strongly associated with infertility; between 20% and 40% of infertile women have the disease. Numerous observations support a causal relationship between endometriosis and infertility:

- The overall prevalence of endometriosis is greater in infertile than in fertile women.[7, 8]
- Infertile women are more likely than fertile women to have moderate to severe disease[8]; the prevalence of minimal endometriosis in infertile women with normal and azoospermic male partners is comparable.[190]
- Although reduced to a similar extent in untreated women with minimal and mild endometriosis and women with unexplained infertility, monthly fecundity decreases further with increasing severity of disease.[191-195]
- Monthly fecundity of women with minimal and mild endometriosis receiving treatment with exogenous gonadotropin stimulation and intrauterine insemination (partner sperm), is less than half that observed in women without the disease.[196]
- Monthly fecundity achieved with donor sperm insemination in women with minimal and mild endometriosis is significantly lower than in women with a normal pelvis.[197-199]
- Overall, the success rates achieved with in vitro fertilization (IVF) in women with endometriosis (all stages) are approximately half those observed in women with tubal disease.[200]
- Women with minimal and mild endometriosis are significantly more likely to conceive after surgical treatment than untreated women.[201, 202]

Taken together, these observations support the conclusion that endometriosis decreases fertility to an extent that correlates roughly with the severity of disease.

The subfertility associated with endometriosis has been attributed to two primary mechanisms: 1) distorted adnexal anatomy that inhibits or prevents ovum capture after ovulation,

and 2) excess production of prostaglandins, metalloproteinases, cytokines, and chemokines, resulting in chronic inflammation that impairs ovarian, tubal, or endometrial function, leading to disorders of folliculogenesis, fertilization, or implantation. The first mechanism offers a logical explanation for infertility in women with advanced stages of endometriosis. The second mechanism may operate in women with milder disease, but whether minimal and mild endometriosis even should be regarded as a cause of infertility remains controversial.

There is reasonably good experimental evidence that endometriosis decreases fertility when it results in grossly distorted pelvic anatomy. In monkeys with peritoneal autografts of adipose tissue or endometrium, cumulative pregnancy rates were significantly lower in animals that developed moderate or severe endometriosis (12%) than in others with minimal or mild disease or in controls (40%), and none of the animals with ovarian adhesions conceived.[203] Decreased fertility in women with advanced endometriosis also could result from premature depletion of the ovarian follicular pool (due to ovarian surgery or destruction).[204-206]

Evidence that endometriosis causes abnormalities of follicular development, tubal transport, or endometrial function is relatively weak, deriving from observations of increased apoptosis in the granulosa cells of women with endometriosis,[206] adverse effects of peritoneal fluid from women with endometriosis on sperm motility[207] and tubal ciliary function *in vitro*,[208] and abnormalities in the expression of markers of endometrial receptivity[209] that could result from an intrinsic progesterone resistance.[10, 210, 211]

Results of the many observational studies of IVF outcomes in women with endometriosis have varied, but offer some useful insights. A 2002 meta-analysis of observational studies concluded that infertile women with endometriosis were less likely to achieve success than women with tubal factor infertility (OR = 0.56, CI = 0.44–0.70); outcomes were worse in severe than in mild disease.[200] The ovarian response to gonadotropin stimulation in women with endometriosis was less robust than in women with tubal disease; both the peak estradiol concentration and number of oocytes retrieved were lower. Fertilization and implantation rates also were decreased, compared to those in women with all indications for IVF or to women with isolated tubal factor infertility.[200] The results further suggested that the adverse effects of endometriosis on fertility were not related solely to anatomical factors. Compared directly, women with severe endometriosis had a lower peak estradiol level, oocyte yield, pregnancy rate, and implantation rate than women with mild disease. In contrast, fertilization rates in women with advanced endometriosis were higher than in those with mild disease, possibly because women with severe endometriosis have a more chronic disease that is less metabolically active.[200] Results of the analysis remain controversial, because subsequent studies have yielded conflicting data, observing outcomes not significantly different from those in women with unexplained infertility.[212, 213] Data from the 2007 U.S. National Report on Assisted Reproductive Technology (ART) indicate that live birth rates for patients with endometriosis (34.3%) were comparable to those for women with diagnoses of tubal factor (30.7%), male factor (35.8%), and unexplained infertility (31.8%).[214]

The overall lower implantation and pregnancy rates observed in women with endometriosis after IVF could reflect poor oocyte quality and subsequent embryogenesis or decreased endometrial receptivity. Studies of IVF outcomes in donor oocyte recipients offer a means to distinguish the two possibilities. Donor oocytes from healthy women yield similar results in recipients with and without endometriosis, but oocytes from women with endometriosis yield poorer results than those from healthy donors in disease-free recipients.[215-219] Embryos derived from oocytes retrieved from women with endometriosis also have fewer blastomeres and exhibit a higher incidence of arrested and abnormal development than those derived from women without the disease.[218, 220] Taken together, these observations suggest that the lower implantation and pregnancy rates observed in women with endometriosis more likely result from abnormalities of oocyte quality and subsequent embryogenesis than from decreased endometrial receptivity.

A wide assortment of studies has been aimed at identifying differences in the follicular environment in women with and without endometriosis that might explain the presumed poor oocyte quality in women with the disease. Studies have compared the numbers and types of resident immune cells, the production of various hormones, growth factors, and cytokines, and the expression of numerous genes in follicular fluid and cultured granulosa cells obtained from women with and without endomtriosis, but no consistent differences have been observed.[221]

Although the outcomes of donor oocyte IVF cycles in women with and without endometriosis do not suggest that the disease has important adverse effects on endometrial receptivity, gene-expression profiling has identified a number of gene products that may be abnormally up- or down-regulated in the endometrium of women with endometriosis during the putative implantation window, including various cell adhesion molecules, matrix metalloproteinases, transcription factors, growth factors, enzymes, and steroid hormone receptors.[130, 131, 222] If endometriosis does have adverse effects on endometrial receptivity, evidence suggests they can be overcome by IVF treatment regimens in most women.

Retrospective studies have observed an increased risk for early pregnancy loss in women with endometriosis.[223-225] However, in appropriately controlled studies, miscarriage rates in untreated women with endometriosis have been in the ranges normally expected and no higher than in treated women.[226-228]

Diagnosis of Endometriosis

Classically, the diagnosis of endometriosis requires histologic evidence of ectopic endometrial glands and stroma, but a tissue diagnosis generally is unnecessary because the physical characteristics of the disease are well-described and easily recognized. Unfortunately, despite substantial advances in our understanding of the pathogenesis of endometriosis, there is not yet any reliable noninvasive alternative to laparoscopy for diagnosis of the disease.

Clinical Diagnosis

The clinical symptoms of endometriosis include dysmenorrhea, pain, dyspareunia, cyclic bowel or bladder symptoms, subfertility, abnormal bleeding, and chronic fatigue. A 2008 cross-sectional survey of 1,000 women with endometriosis found that dysmenorrhea (79%) and pain (69%) were the most common symptoms leading to diagnosis.[229] Comparing women with minimal and mild endometriosis to those with more advanced stages of disease, dyspareunia was significantly more common in women with limited disease (51% vs. 39%), whereas subfertility (22% vs. 30%) and an ovarian mass (7% vs. 29%) led to a diagnosis more often in those with advanced endometriosis.[229] Interestingly, the time to diagnosis was similar among all women. A large case-control study conducted in the U.K. comparing the prevalence of specific symptoms in women with and without endometriosis observed that a greater proportion of women with endometriosis had abdominal/pelvic pain, dysmenorrhea, or menorrhagia (73% vs. 20%).[230] Compared to controls, women with endometriosis had increased risks of abdominal/pelvic pain (OR = 5.2, CI = 4.7–-5.7), dysmenorrhea (OR = 8.1, CI = 7.2–9.3), menorrhagia (OR = 4.0, CI = 3.5–4.5), subfertility (OR = 8.2, CI = 6.9–9.9), dyspareunia and/or postcoital bleeding (OR = 6.8, CI = 5.7–8.2) ovarian cysts (OR = 7.3, CI = 5.7–9.4), and for diagnosis of irritable bowel syndrome (OR = 1.6, CI = 1.3–1.8) and pelvic inflammatory disease (OR = 3.0, CI = 2.5–3.6).[230] These data demonstrate that whereas specific symptoms are associated with endometriosis, the same symptoms are not

uncommon in women without the disease. They also reveal that endometriosis can coexist with or be misdiagnosed as irritable bowel syndrome or pelvic inflammatory disease. It is not surprising that diagnosis can be delayed, often for a period of years.[231]

Dysmenorrhea and pain that are new in onset, progressive, or severe strongly suggest, but do not reliably predict endometriosis.[232] The dysmenorrhea associated with endometriosis often begins before onset of menstrual flow and usually persists throughout menses, sometimes even beyond. The pain usually is diffuse, located deep in the pelvis, dull and aching, and may radiate to the back and thighs or be associated with rectal pressure, nausea, and episodic diarrhea.[233] Pain may be more common, severe, and associated with dyspareunia and painful defecation in women with deeply infiltrating disease involving the cul-de-sac and recto-vaginal septum.[183, 234-236] Dyspareunia associated with endometriosis usually is new in onset and most intense with deep penetration immediately prior to menstruation.[237, 238] One-half to two-thirds of women with endometriosis and pain have inter-menstrual pain.[187]

The severity of endometriosis does not correlate with the number and severity of symptoms; women with advanced disease may have few or no symptoms and those with minimal or mild disease may have incapacitating pain.[187, 237, 238] However, in women with deeply infiltrating endometriosis, the severity of pain generally correlates with the depth and volume of disease.[183, 234, 237] Extra-pelvic endometriosis may be associated with a wide assortment of cyclic symptoms that reflects the organs involved (abdominal scars,[239, 240] the gastrointestinal and urinary tracts,[61, 241-243] the diaphragm,[244] the pleura,[245] and peripheral nerves[246]).

Physical findings in women with endometriosis vary widely and, when present, relate to the location and extent of disease.[237] The external genitalia are typically normal. Occasionally, speculum examination may reveal characteristic blue-colored implants or red proliferative lesions that bleed on contact, both usually in the posterior fornix. Whereas deeply infiltrating endometriosis involving the recto-vaginal septum frequently is palpable, it is not often visible and may have no obvious signs.[235] The uterus often is retroverted and can exhibit decreased mobility or fixation. Women with ovarian endometriomas can have a tender, fixed, adnexal mass. Focal tenderness, thickening, induration, and nodularity of the uterosacral ligaments are the most common, and frequently the only, physical finding.[247, 248] Physical examination has its greatest diagnostic sensitivity when performed during menstruation, but even then a normal examination does not exclude the diagnosis.[249] **Overall, *compared to the gold standard surgical diagnosis of endometriosis, physical examination has relatively poor sensitivity, specificity, and predictive value.***[250]

CA-125

CA-125 is a cell surface antigen expressed by derivatives of the coelomic epithelium (including the endometrium) and is well-established as a useful marker for the monitoring of women with epithelial ovarian cancer. Levels of CA-125 often are elevated in women with advanced endometriosis,[251-253] but also during early pregnancy and normal menstruation, and in women with acute pelvic inflammatory disease or leiomyomata. Serum CA-125 concentrations vary somewhat across the menstrual cycle; in general, levels are highest during the menstrual phase and lowest during the midfollicular and periovulatory phases of the cycle.[249, 254, 255] However, studies of cycle-dependent assay sensitivity and reproducibility have yielded conflicting results, so there is no one best time to perform the test.[250]

Serum CA-125 has been advocated as a screening test for diagnosis of endometriosis, but a meta-analysis including 23 studies using surgically diagnosed disease as the gold standard concluded that the marker performs rather poorly.[256] Cutoff values that provide 90% overall specificity have less than 30% sensitivity, and if adjusted to achieve even 50% sensitivity,

specificity falls to 70%. As a screening test for advanced stages of endometriosis, values associated with 90% specificity have less than 50% sensitivity.[256] ***Overall, the serum CA-125 concentration does not have the necessary sensitivity to be an effective screening test for the diagnosis of endometriosis.***

Serum CA-125 levels may have some value in the preoperative evaluation of women known or suspected to have advanced disease. In one study involving 685 women having surgery for endometriosis, the mean serum CA 125 concentrations were 19, 40, 77, and 182 IU/mL in women with minimal, mild, moderate, and severe disease, respectively; preoperative bowel preparation was suggested for women with levels over 65 IU/mL (upper limit normal 35 IU/mL), as they were more likely to have dense omental adhesions, ruptured endometriomas, or cul-de-sac obliteration.[257] Serum CA-125 levels also may be helpful for differentiating ovarian endometriomas from other benign cysts, especially when combined with transvaginal ultrasonography.[258, 259] Whereas the serum CA-125 generally is not a reliable predictor of the effectiveness of medical therapy,[260, 261] a sustained elevation of serum CA-125 after surgical treatment predicts a relatively poor prognosis.[262, 263]

Imaging

Transvaginal ultrasonography can be helpful for identifying women with advanced endometriosis. ***Transvaginal ultrasonography can detect ovarian endometriomas, but cannot image pelvic adhesions or superficial peritoneal foci of disease.*** Endometriomas can have varying ultrasonographic features but appear typically as cystic structures with diffuse low-level internal echoes surrounded by a crisp echogenic capsule. Some have internal septations or thickened nodular walls.[259, 264] When the characterisitic features are present, transvaginal ultrasonography has 90% or higher sensitivity and almost 100% specificity for detection of endometriomas.[265, 266] Transvaginal or transrectal ultrasonography can be especially helpful when deeply infiltrating disease involving the bladder, the uterosacral ligaments, or the recto-vaginal septum is suspected.[267-274]

Like transvaginal ultrasonography, magnetic resonance imaging (MRI) can be helpful for detection and differentiation of ovarian endometriomas from other cystic ovarian masses, but cannot reliably image small peritoneal lesions.[269, 275, 276] For detection of peritoneal implants, MRI is superior to transvaginal ultrasonography but still identifies only 30-40% of the lesions observed at surgery. For detection of disease documented by histopathology, MRI is approximately 70% sensitive and 75% specific.[277] The principal advantage MRI has over ultrasonography is its ability to distinguish more reliably between acute hemorrhage and degenerated blood products. Whereas endometriomas usually exhibit a relatively homogeneous high signal intensity on T1-weighted images and a hypointense signal on T-2 weighted images (“shading”), acute hemorrhage generally has low signal intensity on both T-1 and T-2 weighted images.[250] However, a short interval of observation, during which hemorrhagic cysts will typically regress, accomplishes the same goal. Gadolinium contrast offers no additional diagnostic value.[278] MRI also can be used to aid in the diagnosis of recto-vaginal disease.[279]

Diagnosis by Therapeutic Trial

Medical treatment of dysmenorrhea certainly is appropriate before considering surgical evaluation and treatment for suspected endometriosis, particularly in adolescents.[280] A trial of treatment with a nonsteroidal anti-inflammatory drug (NSAID), ideally combined with an estrogen/progestin or progestin only contraceptive, is reasonable when the symptoms

do not suggest an acute process. For adult women with suspected endometriosis, some have advocated a trial of medical therapy with a gonadotropin-releasing hormone (GnRH) agonist when there is no other indication for surgical treatment (e.g., suspicious adnexal mass),[281] based on the premise that empiric medical treatment in patients with chronic pelvic pain and a high probability of endometriosis often can avoid a diagnostic surgical procedure.

Evidence in support of empiric medical therapy with a GnRH agonist derives primarily from one clinical trial, in which women with moderate or severe chronic pelvic pain unrelated to menstruation and unrelieved by treatment with NSAIDs and antibiotics were randomized to receive depot leuprolide acetate (3.75 mg i.m. monthly for 3 months) or placebo before diagnostic laparoscopy. Those treated with leuprolide were amenorrheic and had greater relief from symptoms before surgery, which revealed endometriosis in 78/95 subjects (82%).[282]

Although the rigorous clinical criteria employed proved fairly specific (82%) for the diagnosis of endometriosis and although treatment was more effective than placebo, the response to leuprolide treatment did *not* improve diagnostic accuracy; women without surgically proven endometriosis were as likely to receive symptomatic relief from treatment as those with documented disease. It is possible that treatment eliminated or obscured disease in women without documented endometriosis or that some who experienced relief from symptoms had isolated deeply penetrating disease that escaped detection.[282] However, it is at least equally likely that treatment suppressed symptoms related to another cause,[283] that the amenorrhea and symptoms of estrogen deficiency in treated women led them to accurately suspect they were receiving active drug, influencing their reported response, or that leuprolide-induced hypoestrogenism raised pain thresholds.[284] ***The results of the study demonstrate the diagnostic accuracy of rigorous clinical criteria and the efficacy of empiric leuprolide treatment in women with chronic pelvic pain, but do not support the conclusion that the clinical response to treatment has diagnostic value.***[281, 282]

Surgical Diagnosis

Laparoscopy with histologic examination of excised lesions is the gold standard for the diagnosis of endometriosis. The optimal time during the menstrual cycle to perform laparoscopy is not clear, but to avoid under-diagnosis, surgery generally should not be performed during or within 3 months after hormonal medical treatment.[267] Greater awareness of the varied appearance of endometriotic lesions has doubled the frequency with which endometriosis is diagnosed with laparoscopy when a careful and systematic examination is performed.[285, 286]

The classic peritoneal implant is a blue-black "powder burn" lesion (containing hemosiderin deposits from entrapped blood) with varying amounts of surrounding fibrosis, typically observed on the ovaries and on peritoneal surfaces in the cul-de-sac, uterosacral ligaments, and ovarian fossa.[287] ***However, the majority of implants are "atypical," appearing white and opaque, red and flame-like, or vesicular.*** Less commonly, disease may be found in ovarian adhesions, yellow-brown patches, in peritoneal defects, or involving the appendix.[285-289] Red lesions are highly vascular, proliferative, and represent an early stage of disease.[290] Pigmented lesions represent more established or advanced disease. Both are metabolically active and more commonly associated with symptoms. White lesions are less vascular and active, and less often symptomatic.[290, 291] Studies involving serial laparoscopy have suggested a natural progression in the appearance of endometriotic lesions over time, and revealed that a variety of lesions may be observed at any one time in an individual.[292, 293]

Strict histologic criteria will confirm the surgical diagnosis of endometriosis in approximately 50–65% of excised lesions.[287, 294] When diagnosis is in doubt, biopsy of suspicious areas should be performed to prevent misdiagnosis and avoid inappropriate or unnecessary treatment[295]; lesions that can be confused with endometriosis include endosalpingiosis, mesothelial hyperplasia, hemosiderin deposition, hemangiomas, adrenal rests, inflammatory changes, and splenosis.[296] Conversely, a negative laparoscopy is highly reliable for excluding endometriosis.[295] Microscopic evidence of endometriosis in normal-appearing peritoneum is common in asymptomatic infertile women with and without other apparent disease (6–13%),[290, 297] but of uncertain clinical significance because it may exist in most women but progress only in some.[298]

Endometriomas usually appear as smooth, dark cysts, typically associated with adhesions and containing a dense brown chocolate-like fluid.[234, 299] Larger endometriomas frequently are multilocular. Careful visual inspection of the ovaries generally is highly reliable for detection of endometriomas,[300] but when disease is suspected strongly and not readily apparent, careful exploration by ovarian puncture and aspiration can be helpful.[301] Whereas ovarian endometriomas usually are accompanied by numerous other visible peritoneal lesions,[302] deeply infiltrating endometriosis is largely retroperitoneal, often not readily apparent, and frequently isolated; it may even represent a distinct entity, arising from müllerian rests within the recto-vaginal septum.[303, 304]

Classification and Staging Systems

A uniform classification system for endometriosis that considers both the distribution and severity of disease is useful, because both treatment and prognosis in women with endometriosis are determined, to some extent, by the extent of disease. A valid uniform classification also is crucial for comparing the results of treatment trials performed in different centers.

In 1979, the American Fertility Society (now the American Society for Reproductive Medicine; ASRM) introduced a classification system based on surgical findings at laparoscopy or laparotomy that was modeled after those used for grading malignant disease.[305] The system assigned a point score, based on the size, depth, and location of lesions and associated adhesions. The classification system was revised in 1985,[306] and again in 1996, to acknowledge the varying morphology of endometriosis and to improve consistency in scoring and the prognostic value for women with pain or infertility.[307] The current version of the revised classification system is the most widely accepted classification tool, but still has serious limitations.[308] Chief among them is the relatively poor correlation with pregnancy rates.[309, 310] Further revisions of the classification scheme are anticipated as our understanding of the pathogenesis of infertility advances, and likely will include empirically derived weights and thresholds to define stages of disease; other factors may be incorporated if they prove to have prognostic value.[311]

The classification of endometriosis used most commonly in clinical practice is descriptive and relatively simple:

- Minimal endometriosis — isolated superficial disease on the peritoneal surface with no significant associated adhesions.
- Mild endometriosis — scattered superficial disease on the peritoneal surface and ovaries, totaling less than 5 cm in aggregate, with no significant associated adhesions.
- Moderate endometriosis — multifocal disease, both superficial and invasive, that may be associated with adhesions involving the fallopian tubes and/or the ovaries.
- Severe endometriosis — multifocal disease, both superficial and invasive, including large ovarian endometriomas, usually associated with adhesions, both filmy and dense, involving the fallopian tubes, ovaries, and cul-de-sac.

In 2009, a new staging system was proposed, called the Endometriosis Fertility Index (EFI), by combining the factors that best predicted pregnancy (without IVF) after analysis of clinical and surgical data (275 variables) collected from 579 infertile patients with endometriosis.[310] The EFI score (0–10, with 0 representing the poorest and 10 the best prognosis) was validated by collecting comparable data from an additional 222 patients, calculating the score for each, and observing good correlation between actual and predicted (estimated from life table analysis) pregnancy rates. The key element of the new staging system, absent from its forerunners, is a numerical measure of functional anatomy, based on careful assessment of the tubes (extent of serosal injury, mobility, and patency), fimbriae (extent of injury, architecture) and ovaries (size, extent of surface injury). The EFI score predicts cumulative pregnancy rates over 3 years after surgery, which range from a low of 10% (EFI 0–3) to a high of 75% (EFI 9–10).[310] Based on the original report, the method has promise as a clinical tool for developing treatment plans in infertile patients with a surgical diagnosis of endometriosis. Ultimately, the clinical utility of the EFI will hinge on additional independent validation and its general acceptance by clinicians.

Summary

Careful clinical evaluation can identify women likely to have endometriosis but cannot establish the diagnosis. Although a serum CA-125 concentration may provide corroborative evidence of disease, the sensitivity of the test is too low to make it an effective screening tool. Transvaginal ultrasonography and MRI are both highly sensitive and specific for detection of ovarian endometriomas but cannot reliably image peritoneal implants of disease. Although empiric medical treatment can help women with suspected endometriosis to avoid diagnostic surgery, a clinical response to treatment does not establish the diagnosis of endometriosis. In the majority of women, the diagnosis of endometriosis requires a careful and systematic laparoscopic examination. Histologic examination of excised lesions can confirm surgical impressions and is preferred, but not required, to establish the diagnosis with reasonable certainly.

Treatment of Endometriosis

The treatment of endometriosis depends on its clinical manifestations, which fall into two basic categories: pelvic pain and infertility. Because both can be very difficult to assess objectively, the results of treatment trials must be interpreted carefully.

To be truly informative, studies of the effects of treatments on pain should use a validated objective measure of pain, follow subjects over a prolonged interval because the recurrence of pain is time-dependent, and employ a placebo control because the placebo effect in studies of pain is often quite large (30–50%). Studies of the effects of treatments on the volume of endometriosis are equally difficult to conduct and interpret because disease may regress spontaneously and also may recur or advance after discontinuation of therapy. Studies of the effects of treatments on fertility must consider that most women with endometriosis and infertility are only subfertile, not sterile; they can conceive, but do so less efficiently. Ideally, cycle fecundity over a defined interval of time should be compared to that in a group of similarly affected but untreated women.

Treatment for endometriosis can be expectant or limited to the use of analgesics, or can involve one or a combination of medical treatments, conservative or definitive surgery, or

a combination of medical and surgical treatment. Expectant management generally is reserved for patients without significant symptoms and for those approaching menopause. However, even those with few symptoms may benefit from treatment aimed at preventing progression of the disease.[312] Because endometriosis normally regresses after menopause, due to the marked decrease in ovarian estrogen production, perimenopausal women with mild symptoms may choose expectant management, or treatment limited to non-narcotic analgesics, for the short-term. Young women with significant symptoms generally will require more aggressive medical or surgical treatment.

Medical Treatment

Women with pelvic pain, suspected endometriosis, and no other indication for surgical treatment can be managed effectively with empiric medical treatment without establishing a surgical diagnosis,[281] keeping in mind that a response to treatment does not establish the diagnosis of endometriosis, as emphasized in an earlier section of this chapter. Empiric medical therapy may involve treatment with nonsteroidal anti-inflammatory drugs (NSAIDs) or estrogen-progestin contraceptives for women with mild symptoms, or with a gonadotropin-releasing hormone (GnRH) agonist for those with moderate or severe pelvic pain. ***Although symptoms will improve in most, it is important to emphasize that medical therapies have no measurable effect on fertility and are not an effective treatment for patients with endometriomas or pelvic adhesions.***[267, 313, 314] Consequently, infertile women and those with suspected endometriomas or more advanced disease are better treated surgically.

Traditional medical therapies for endometriosis have been based on Sampson's theory of retrograde menstruation and implantation, and on the simple premise that ectopic endometrium may be expected to respond to treatment in much the same way as normal eutopic endometrium. Consequently, the objectives of treatment have been to reduce or eliminate cyclic menstruation, thereby decreasing peritoneal seeding and the likelihood that new implants will develop, and to suppress the growth and activity of the endometrium, anticipating that the same would occur in the endometriotic tissue derived from it. These simple operational concepts have shaped medical treatments for endometriosis for decades, but our growing understanding of the pathogenesis of endometriosis at the molecular level is now beginning to suggest new treatment strategies aimed at the mechanisms of disease.

Estrogen-Progestin Contraceptives

Estrogen-progestin contraceptives, taken in a cyclic or continuous fashion, have been a mainstay of the medical treatment of symptomatic endometriosis almost since their introduction.[315] Even today, they are the most commonly prescribed treatment for the disease. Continuous treatment has been dubbed "pseudopregnancy" because combined estrogen-progestin therapy induces amenorrhea and endometrial decidualization and resembles the high-estrogen, high-progesterone environment of pregnancy widely believed to improve or suppress endometriosis. Limited evidence indicates that estrogen-progestin contraceptives also may enhance apoptosis of eutopic endometrial tissue in women with endometriosis.[316]

Estrogen-progestin contraceptives are a good initial choice for women with mild symptoms who also need contraception. They can be expected to provide effective relief from pain associated with endometriosis in 75–90% of affected women, particularly when taken

continuously.[317-320, 321, 322] They also might help to prevent progression of endometriosis. There is no evidence that any formulation is superior. One advantage estrogen-progestin contraceptives have over other medical therapies is they can be taken indefinitely. Supplemental estrogen (conjugated estrogens 1.25 mg or micronized estradiol 2.0 mg daily, for 7–10 days) can be used to control episodic breakthrough bleeding, which is more common with continuous than with cyclic therapy.

Progestins

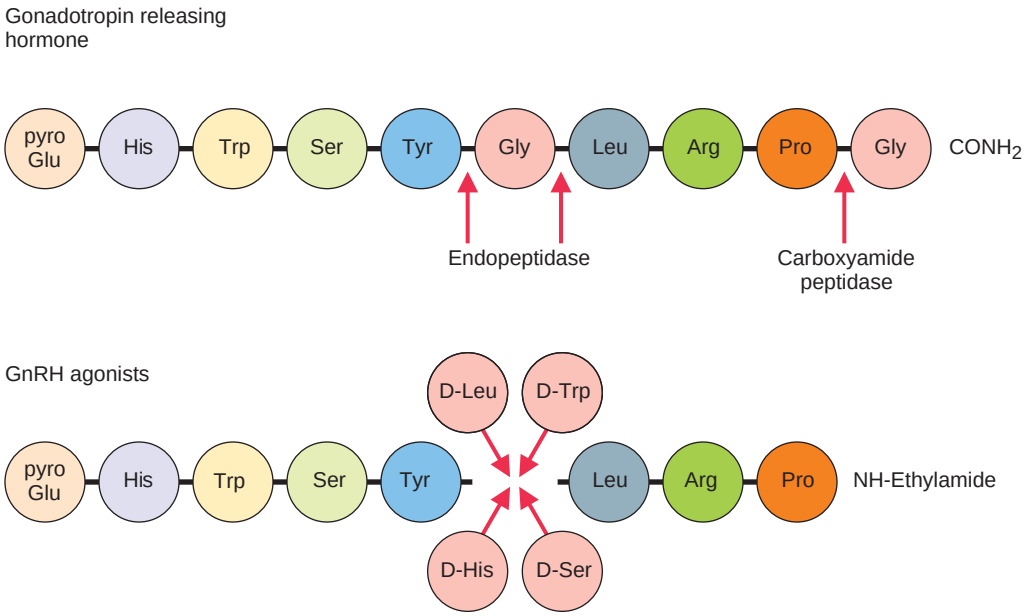
Progestins have long been used to treat symptomatic endometriosis because they inhibit endometrial growth (and presumably, the growth of endometriotic tissue), first inducing decidualization, then atrophy.[323, 324] In high doses, they also can inhibit pituitary gonadotropin secretion and ovulation, inducing amenorrhea. A large variety of different progestins is available, including those derived from progesterone like medroxyprogesterone acetate and others derived from 19-nortestosterone, with norethindrone the prototype. Suppression of endometrial matrix metalloproteinases (now recognized as contributing to the pathogenesis of endometriosis) could be another useful action.[325] Although matrix metalloproteinase activity in eutopic endometrium of women with endometriosis is unusually resistant to progesterone suppression,[326] the higher doses used in the treatment of endometriosis may be sufficient to overcome the effect.

Several different progestins have been used effectively for the treatment of pain (dysmenorrhea, dyspareunia, intermenstrual pain) associated with endometriosis.[160,327-329,330] Medroxyprogesterone acetate can be administered orally (20–100 mg daily) or by injection (150 mg i.m. every 3 months). Side effects include nausea, weight gain, fluid retention, breast tenderness, irregular bleeding, and depression. Breakthrough bleeding is common (35–50%) but generally well tolerated and usually can be eliminated by treatment with short courses of supplemental estrogen (conjugated estrogens 1.25 mg or estradiol 2.0 mg daily for 7–10 days). Depression is not uncommon (approximately 5%) and can be severe enough to require discontinuation of treatment. A single randomized controlled trial demonstrated that high doses of medroxyprogesterone acetate (100 mg daily for 6 months) induced complete remission of all visible endometriosis in 50% of women, compared to 12% of those who received placebo, and incomplete remission in 13%, compared to 6% of placebo-treated controls.[328] Norethindrone acetate (5–15 mg daily) and megestrol acetate (40 mg daily) also have been used in the treatment of endometriosis and have side effects similar to those of medroxyprogesterone acetate.[331, 332] The levonorgestrel-releasing intrauterine device is another option that may have particular value for women with deeply infiltrating recto-vaginal endometriosis,[333-337] and limited evidence indicates that the etonogestrel subdermal implant (Implanon) also can be effective for decreasing pain associated with endometriosis.[338, 339]

Progestins can have adverse effects on serum lipoprotein levels. The 19-nortestosterone-derived progestins decrease HDL significantly; the effects of medroxyprogesterone acetate are less severe.[340, 341] However, these effects are unlikely to have any clinical importance over a relatively short interval of months. At higher doses, the suppressive effects of progestins on the hypothalamic-pituitary-ovarian axis may be sufficient to induce a hypogonadal state, resulting in spinal bone mineral depletion amounting to 2–4% over an interval of 6–12 months; longer term treatment may result in even greater loss, but recovery usually is rapid after treatment is discontinued and any impact on fracture risk is unlikely.[342] Although progestins are effective treatment for pain associated with endometriosis, their adverse effects on fertility limit their utility in infertile women seeking pregnancy.

Gonadotropin-Releasing Hormone Agonists

Gonadotropin-releasing hormone (GnRH) agonists are derived from native GnRH by substituting a D-amino acid for the native L-amino acid at position 6 in the decapeptide. The substitution yields an agonist resistant to degradation, increasing its half-life and time of receptor occupancy. Pituitary follicle stimulating hormone (FSH) and luteinizing hormone (LH) secretion requires a pulsatile hypothalamic GnRH stimulus, which allows receptor concentrations to be replenished between pulses; a constant intravenous infusion of GnRH generates an initial response ("flare"), followed by down-regulation of receptor concentrations, which desensitizes the pituitary to continued stimulation. Long-acting GnRH agonists (leuprolide, nafarelin, goserelin, buserelin, triptorelin) have the same effect, inducing a hypogonadotropic hypogonadal state that has been dubbed "pseudomenopause" or "medical oophorectomy," although both terms are misnomers.[343, 344] In menopause, the ovaries produce no estrogen because they are depleted of follicles and cannot, and in castrates, the ovaries are altogether absent; in both cases, serum gonadotropin levels are markedly elevated. In contrast, women under treatment with GnRH agonists do not produce estrogen because their ovaries receive no effective gonadotropin stimulation; levels of both FSH and LH are very low. GnRH agonists are effective for the treatment of endometriosis because they induce a hypogonadal state, which deprives existing disease of estrogen support, and amenorrhea, which prevents new peritoneal seeding.



GnRH agonists can be administered intramuscularly, subcutaneously, or intranasally, the route varying with the specific drug. Approximately 75% of women are rendered hypogonadal within 4 weeks of treatment and almost all are by 8 weeks.[345] The side effects of GnRH agonsits are those of hypogonadism and include hot flashes, progressive vaginal dryness, decreased libido (both estrogen and androgen production is suppressed), depression, irritability, fatigue, headache, changes in skin texture, and bone mineral depletion. Over 80% report vasomotor symptoms and 30% report vaginal symptoms and headache.[346] GnRH agonist therapy does not have adverse effects on serum lipids and lipoprotein concentrations like those associated with danazol or high dose progestin treatment.[347, 348] ***The decrease in bone mineral associated with standard GnRH agonist treatment regimens (6 months) is significant; bone loss occurs in both the lumbar spine (trabecular bone) and femoral neck (cortical bone) and can approach or even exceed 1% per month.***[349-351]

After discontinuation of treatment, bone loss is slowly recovered,[351-353] but not completely in all women.[354-356] The loss of trabecular bone may cause damage to bone structure that cannot be effectively reversed.[357]

In efforts to prevent the bone mineral depletion that accompanies GnRH agonist therapy, a number of different "add-back" treatment strategies have been developed. Low-dose combined estrogen-progestin add-back regimens (conjugated estrogens 0.625 mg and medroxyprogesterone acetate 2.5 mg daily or norethindrone 5.0 mg daily)[358, 359] are based on the notion that the level of estrogen required to support endometriosis is greater than is required to prevent vasomotor symptoms or bone mineral depletion. Results achieved with low-dose combined estrogen-progestin add-back regimens support this "estrogen threshold hypothesis."[360] However, estrogen-only add-back is inadvisable; a clinical trial (oral estradiol, 1 mg daily) was terminated early due to observations of recurrent pain in treated subjects.[361] Numerous other add-back regimens have been described, including progestins alone (norethindrone 2.5–5 mg daily)[359] tibolone (2.5 mg daily),[350] bisphosphonates (cyclic etidronate 400 mg daily for 2 weeks every 2 months;[362] alendronate 10 mg daily[363]), and the selective estrogen receptor modulator, raloxifene (60 mg daily).[364] ***Combined estrogen-progestin add-back treatment regimens protect bone and have the added advantage of preventing hot flushes and the development of genitourinary atrophy.*** Progestin-only add-back regimens have been less consistently effective.[365, 366] Endometriosis itself is not associated with bone loss.[367] Evidence from randomized trials indicates that hormone add-back regimens protect bone and reduce symptoms of estrogen deficiency without sacrificing control of pain associated with endometriosis.[345, 366]

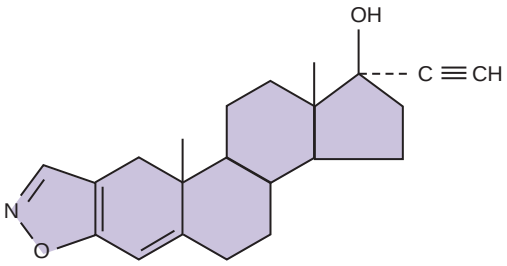
GnRH agonists are proven effective in relieving pain in women with endometriosis; all agonists appear equally efficacious.[346, 368, 369] A number of clinical trials comparing GnRH agonist treatment with and without steroid hormone add-back therapy have concluded that combination therapy is as effective as treatment with agonist alone but results in fewer side effects relating to estrogen deficiency.[370-374] Whereas many choose not to begin add-back therapy until pain is under control, evidence suggests that delaying add-back therapy is unnecessary in most women.[374] Pain of lesser or equal intensity may recur soon after cessation of treatment; the recurrence rate is at least 10–20% per year.[160, 375-377] The overall cumulative recurrence rate 5 years after treatment with a GnRH agonist is approximately 55%, lower for women with minimal and mild endometriosis (37%) than for those with advanced disease (74%).[378]

In a large uncontrolled trial, leuprolide treatment decreased the volume of endometriosis in almost 90% of women.[379] Numerous other studies comparing a GnRH agonist with danazol have concluded they have equal efficacy[369, 375, 380-383] Another comparing a GnRH agonist (leuprolide) to a progestin (lynestrenol) observed a greater decrease in disease in women treated with the agonist.[384] Treatment with a GnRH agonist can sometimes decrease the size of endometriomas, but does not eliminate them.[385]

Danazol

Danazol is the first drug ever approved for the treatment of endometriosis in the U.S. It is an orally administered isoxazol derivative of 17α-ethinyltestosterone that acts primarily by inhibiting the midcycle urinary LH surge and inducing a chronic anovulatory state, but also inhibits a number of steroidogenic enzymes and increases free testosterone levels.[386-389] The many different effects of danazol combine to yield a high androgen, low estrogen environment that inhibits growth of endometriosis. The amenorrhea that commonly results from danazol treatment also decreases new peritoneal seeding.[390]

Danazol has proven effective for reducing endometriosis-related pain (dysmenorrhea, deep dyspareunia, intermenstrual pain) in up to 90% of treated women;[328, 391] the median time to recurrence of pain after discontinuation of treatment is approximately 6 months.[392] Studies examining the effect of danazol on endometriotic implants (typically assessed after 6 months of treatment) have consistently observed a decrease in the volume of disease ranging between 40% and 90%.[369, 375, 393-396] One randomized controlled trial observed regression of disease in 60% of danazol treated women and in 18% of those who received placebo.[328]



Danazol

Although danazol is effective for the treatment of pain associated with endometriosis, the recommended doses (400–800 mg daily) have substantial androgenic and hypoestrogenic side effects that limit its clinical utility. The most common are weight gain, fluid retention, fatigue, decreased breast size, acne, oily skin, hirsutism, atrophic vaginitis, hot flushes, muscle cramps, and emotional lability. Some can be expected, occurring in up to 80% of women taking danazol, but less than 10% have side effects sufficient to warrant discontinuation of treatment.[397] Danazol has been associated with virilization of a female fetus *in utero* and should not be given when there is any possibility of pregnancy.[398] The androgenic actions of danazol also can deepen the voice irreversibly.[399, 400] Adverse changes in the lipid profile reflect the drug's androgenic effects; total cholesterol and low density lipoprotein (LDL) levels are increased and the high density lipoprotein (HDL) concentration is lowered, but these effects pose no significant risk over a relatively short term of treatment. Rarely, danazol treatment may result in liver damage or arterial thrombosis.[397,401] Lower doses of danazol are better tolerated but also possibly less effective.[402, 403] Danazol also has been administered vaginally, but experience is limited.[404]

Aromatase Inhibitors

Although not approved for the treatment of endometriosis, aromatase inhibitors offer a new and promising approach to management of the disease.[10, 140, 405] Aromatase inhibitors effectively suppress estrogen production in the periphery (e.g., brain, adipose) and in endometriotic tissues, as well as in the ovary.[140]

In numerous case reports and small series, aromatase inhibitors (anastrozole 1 mg daily, letrozole 2.5 mg daily) have been found effective for the management of pain associated with endometriosis.[405-411] A randomized trial comparing treatment with a GnRH agonist alone to combined treatment with an agonist and anastrozole in the postoperative management of women receiving surgical treatment for endometriosis found that addition of an aromatase inhibitor prolonged the time to symptom recurrence significantly; 14/40 women (35%) treated with a GnRH analog alone and 3/40 (7.5%) also treated with anatrozole had recurrent symptoms during 24 months of follow-up.[412] A 2008 systematic review including eight studies involving treatment with aromatase inhibitors for pain associated with endometriosis concluded that limited evidence supports their effectiveness.[413]

Aromatase inhibitors can be expected to cause significant bone loss with prolonged use and cannot be used alone in premenopausal women because they stimulate FSH release, causing development of multiple ovarian cysts. To avoid the complication, they must be used in combination with a GnRH agonist or norethindrone acetate (5 mg daily) in premenopausal women. In one study, combined treatment with letrozole and norethindrone was found more effective than norethindrone alone for reducing pain and deep dyspareunia in women with deeply infiltrating recto-vaginal endometriosis, but also was associated with more adverse effects and did not improve patient satisfaction or influence recurrence of pain.[414]

Summary

Established medical therapies for the treatment of pain associated with endometriosis include estrogen-progestin contraceptives, progestins, GnRH agonists, and danazol. Evidence indicates that pain relief and recurrence rates are similar for all and that no one medical treatment is best. Consequently, treatment decisions must be individualized, after carefully considering the severity of symptoms, the extent of disease, the desire for future pregnancy, age, side effects, and costs. Aromatase inhibitors are another promising new therapeutic option aimed at one of the key pathogenic mechanisms of endometriosis. Medical therapy generally is not effective for the management of endometriomas larger than 1 cm.

There is no substantial evidence that medical treatment of endometriosis improves fertility. Moreover, because all medical treatments for endometriosis inhibit ovulation, fertility is all but eliminated during treatment. Medical therapy may even adversely affect fertility, due to the time forfeited during treatment, particularly in older women who have a narrowing window of opportunity to achieve their reproductive goals.

Surgical Treatment

The objectives of surgical treatment for endometriosis are to restore normal anatomical relationships, to excise or destroy all visible disease to the extent possible, and to prevent or delay recurrence. For women having moderate or severe endometriosis that distorts the reproductive anatomy and hoping to restore or preserve fertility, surgery is the treatment of choice because medical treatment cannot achieve the goal. When disease is less severe, medical treatment can effectively control pain in the large majority of women but has no effect on fertility; surgery is at least as effective as medical treatment for relieving pain and also may improve fertility.

Although surgery for the treatment of endometriosis can be performed via laparotomy or laparoscopy, technical advances in instrumentation and technique generally allow the endoscopic approach in all but those who require extensive enterolysis or bowel resection; highly skilled surgeons can accomplish even these objectives via laparoscopy.[415, 416] Laparoscopy offers the advantages of better visualization, less tissue trauma and desiccation, smaller incisions, and a speedier postoperative recovery.[417] Postoperative adhesions and complications also may be less than after laparotomy.[418] Most importantly, the results achieved with laparoscopy are equivalent to or better than with laparotomy.[419]

Minimal and Mild Disease

Peritoneal implants of endometriosis may be ablated with unipolar or bipolar electrosurgical instruments or lasers, or excised by sharp dissection.[420] Opinions regarding the superiority of one method over another are strongly held but not substantiated by the results of any direct comparisons. Those favoring excision over ablation emphasize that because the depths of disease and ablation cannot be determined, the risk that treatment may be inadequate is greater when disease is ablated rather than excised. Adhesions associated with endometriosis that distort the reproductive anatomy should be excised, even though adhesion reformation occurs in the large majority of cases.[421] ***Excision is preferable to simple lysis because adhesions frequently contain disease.*** Strict adherence to the same microsurgical principles that governed reconstructive pelvic surgery before the advent of modern laparoscopic techniques improves operative outcomes—use of magnification, minimum tissue trauma and exposed suture, and meticulous hemostasis.

Only a few trials have compared the results of laparoscopic surgery with no treatment, other treatments, or placebo in the management of pelvic pain associated with endometriosis.[422-424] In one informative trial, outcomes after laparoscopic laser ablation of endometriosis and uterosacral nerve ablation were compared to those after diagnostic surgery and expectant management. Six months after surgery, pain was eliminated or improved in over 60% of women who received surgical treatment of their disease (minimal, mild, or moderate) and in less than 25% of women whose disease was not ablated. After an average of 6 years of follow-up, two-thirds of the women in the original study could be contacted to assess the long-term results of their earlier operation. Pain had recurred in almost 75% of subjects, with a median interval to recurrence approaching 20 months (range 5–60 months), but over 50% of women reported satisfactory pain relief; half of those not receiving satisfactory pain relief had undergone definitive surgical treatment (hysterectomy with or without bilateral oophorectomy).[425] Other trials have observed success in achieving pain relief in 70–100% of women with endometriosis.[426, 427]

Unfortunately, as with medical treatment, recurrent disease and pain after local excision or ablation of endometriosis are common; symptoms recur in at least 10–20% of treated women per year.[428, 429] A study involving second look laparoscopy after surgical treatment of minimal or mild endometriosis found that disease most often recurs in the same or adjacent areas of the pelvis, suggesting either incomplete excision during the initial surgery or favored implantation in certain locations.[430] A case series of 120 patients who had local excision of endometriosis found that 20% required further surgery by 2 years, and 45–55% by 5–7 years.[431] The incidence of recurrent disease may be higher after surgical treatment in the luteal phase than in the follicular phase because refluxed endometrial cells may be more likely to implant at sites of unhealed peritoneal trauma when the interval from surgery to next menstruation is short.[432]

The effects of surgery on fertility in women with minimal and mild endometriosis have been examined in two randomized controlled trials. In a multi-center Canadian trial, women with unexplained infertility had diagnostic laparoscopy and those with minimal or mild endometriosis were randomized to treatment (excision or ablation of disease) or expectant management and followed for 36 weeks or until 20 weeks of gestation if pregnancy occurred during follow-up.[228] The chance of pregnancy in treated women was twice that in untreated women (OR = 2.03, CI = 1.28–3.24). Overall, 50/172 (0.29) women randomized to treatment achieved an ongoing pregnancy, compared to 29/169 (0.17) of those managed expectantly, yielding a treatment effect of 0.12 and a number needed to treat (the inverse of the treatment effect) of 8.3, rounded upward to 9. ***The results suggest that approximately nine infertile women with minimal or mild endometriosis must undergo surgical treatment to achieve one additional pregnancy, a small but potentially important effect.*** However,

in a second smaller Italian trial of similar design (n = 96), no difference between treatment and no treatment was observed.[433] A meta-analysis combining the data from both studies concluded that surgical treatment of minimal and mild endometriosis may improve fertility (OR = 1.64, CI = 1.05–2.57); the number needed to treat was 12.[202]

Moderate and Severe Disease

The optimum surgical management of ovarian endometriomas is somewhat controversial. Endometriomas have been treated by wedge resection, enucleation (stripping), and by drainage with and without ablation of the internal cyst wall.[434] Careful histologic studies have demonstrated that endometriosis always can be found in the cyst wall, involving as little as 10% to more than 90% of the surface (median 60%), but does not penetrate more deeply than approximately 1.5–2.0 mm.[435] A 2003 meta-analysis of data derived from four comparative trials observed that endometriomas recurred in 39/212 women (18%) treated by coagulation or laser vaporization, compared with 19/295 women (6%) who had cyst enucleation.[436] A 2005 systematic review concluded that laparoscopic excision of the cyst wall was associated with a lower rate of recurrence of the endometrioma (OR = 0.41, CI = 0.18–0.93), a decreased requirement for further surgery (OR = 0.21, CI = 0.05–0.79), reduced recurrence of dysmenorrhea (OR = 0.15, CI = 0.06–0.38), dyspareunia OR = 0.08, CI = 0.01–0.51) and non-menstrual pelvic pain (OR = 0.10, CI = 0.02–0.56), and with an increased pregnancy rate in women who previously were infertile (OR = 5.21, CI = 2.04–13.29).[437] Careful surgical technique is important because ovarian function can be compromised by excision of excessive tissue or damage to hilar vessels[438]; the risk of ovarian failure after excision of bilateral ovarian endometriomas is approximately 2.5%[439]. Studies involving second-look surgery also suggest that postoperative adnexal adhesions are more likely after wedge resection than with other surgical treatments and in women whose original surgery also involved adhesiolysis.[434, 440]

Deeply infiltrating endometriosis involving the recto-vaginal septum requires extensive surgery. Because disease in this location typically includes smooth muscle as well as endometrial glands and stroma, some view it as a distinctly different entity, as a nodule of adenomyosis arising by metaplasia in müllerian rests rather than as endometriosis extending downward from the peritoneal surface.[441] Surgery involves thorough dissection and exposure of the anterior rectum, the posterior vagina, and nodular disease. Often, a portion of the posterior vagina must be excised, and sometimes, a short segment of rectum must be resected, followed by anatamosis.[415, 416, 441] In experienced hands, surgical excision of the disease generally achieves excellent results. Over 3 years, postoperative recurrence rates for dysmenorrhea, deep dyspareunia, and pelvic pain range from 15–30% and are lowest when any involved vagina or rectum also is removed.[441]

Although no studies have compared the effectiveness of surgical treatment with no treatment or medical treatment in infertile women with moderate to severe endometriosis, the cumulative pregnancy rate 1–3 years after surgical treatment is approximately 50% for women with endometriomas[419, 442-444] and about 30% for women with complete cul-de-sac obliteration in various case series.[419, 445] These cumulative success rates are lower, but not dramatically lower, than those observed in surgically treated infertile women with minimal and mild disease (44–62%) and, intuitively, significantly higher than might reasonably have been anticipated without treatment considering that most women with moderate to severe endometriosis have grossly distorted reproductive anatomy.[419] The use of adhesion barriers reduces adhesion formation after surgical treatment in infertile women,[446] but there is no convincing evidence that adhesion barriers or other adhesion prevention strategies improve pregnancy rates after surgical treatment.[446, 447] In a prospective cohort study involving 169 infertile women under age 38 with symptomatic deeply infiltrating endometriosis,

the pregnancy rate achieved with IVF was significantly higher in women who chose to have preliminary surgical treatment.[448]

In women with advanced symptomatic endometriosis who have completed childbearing and those in whom medical and conservative surgical treatment fails, definitive surgical treatment deserves serious consideration. In highly selected women having no significant ovarian disease, hysterectomy alone can be considered, although the risk of recurrent disease requiring additional treatment is approximately 6-fold higher when oophorectomy is not performed.[449] Other risk factors for persistent or recurrent endometriosis and pain include incomplete excision of disease and postoperative estrogen therapy in women with extensive or residual disease.[449, 450] However, when all visible endometriosis is removed, the risk for recurrent pain in women who receive immediate or delayed hormone treatment is similar.[451] ***Ovarian remnant syndrome*** involves persistent or recurrent disease and pain associated with residual functional ovarian tissue. The syndrome is not altogether rare and occurs most frequently when the ovaries are enlarged or densely adherent to the pelvic sidewalls and dissection is technically difficult.[452-454]

Adjuvant Procedures

Adjuvant presacral neurectomy and laparoscopic uterosacral nerve ablation (LUNA) have been advocated for the management of dysmenorrhea and severe central pelvic pain unresponsive to medical or previous surgical treatment for endometriosis. Presacral neurectomy involves interrupting the sympathetic innervation of the uterus at the level of the superior hypogastric plexus and LUNA involves the destruction of the midportion of the uterosacral ligaments. The results of a number of randomized controlled trials examining outcomes after the two procedures suggest they can be effective in some, but not all, women, and generally do not relieve symptoms of dyspareunia and intermenstrual pain.[455-459] ***In sum, results of controlled trials have not provided compelling evidence that these procedures add value to conservative surgery.***[460] Operative complications and postoperative bowel or bladder dysfunction, although uncommon, do occur and can be debilitating.[456, 457] Considering their uncertain benefits and potential risks, routine presacral neurectomy or LUNA at time of conservative surgery for endometriosis cannot be recommended[460]; both procedures should be reserved for highly selected individuals, who must be counseled very carefully.

Pre- and Post-operative Medical Treatments

The value of pre- and postoperative medical treatment in the management of moderate and severe endometriosis has been controversial. Some advocate preoperative medical treatment with a GnRH agonist, believing it may offer certain advantages, including a decrease in the volume of disease requiring surgical treatment, elimination of functional ovarian cysts that may pose technical problems, greater convenience of surgical scheduling, and a better overall outcome.[461, 462] However, with the possible exception of deep recto-vaginal endometriosis wherein preoperative medical treatment may decrease the likelihood of recurrent disease and symptoms,[462] there is no convincing evidence that medical treatment before surgery improves either pain control or fertility compared to surgical treatment alone.[420, 463]

Postoperative medical suppressive therapy also has been controversial. Whereas some studies have observed a longer pain-free interval or higher pregnancy rates when surgical treatment is followed by an interval of medical treatment with a GnRH agonist,[464] danazol,[465] progestins,[466, 467] or estrogen-progestin contraceptives,[468, 469] others have found

no differences between the prevalence of recurrent pain or pregnancy rates 1–3 years after surgical treatment in women who did and did not receive postoperative medical treatment.[470-473] Considering that the highest pregnancy rates after conservative surgery in infertile women generally are observed in the first year after operation, most clinicians have been reluctant to use medical treatment that will prevent pregnancy after surgical treatment. ***When the primary objective of surgical treatment for endometriosis is relief from pain and pregnancy is not an immediate goal, postoperative medical treatment may have value, particularly in women with extensive disease and those with residual disease than could not be completely excised.***[463, 464]

After conservative surgical treatment of endometriosis in infertile women, a choice between expectant management and active treatment must consider age, the surgical results, and the influence and severity of any other infertility factors. Considering the modestly increased fecundability of women with minimal and mild endometriosis after surgical treatment,[202] young women with limited disease and otherwise unexplained infertility of relatively short duration might be treated expectantly, but not for longer than 6–9 months. A more aggressive approach involving immediate further empiric treatment with a combination of clomiphene citrate or exogenous gonadotropins and intrauterine insemination or even IVF is justified in those with longer durations of infertility or more advanced endometriosis, and in older women.[216, 474-477]

After radical surgery (hysterectomy and bilateral salpingo-oophorectomy) for persistent or recurrent endometriosis, hormone therapy can begin immediately in most women with negligible risk of inducing growth of residual disease and recurrent symptoms.[451] However, in those with extensive disease, an interval without hormone treatment or progestin-only treatment may be prudent. Progestins may have value both for their direct suppressive effects on any residual foci of endometriosis and for alleviating the otherwise inevitable vasomotor symptoms that accompany removal of the ovaries. ***Low-dose combined estrogen-progestin treatment is strongly recommended over treatment with estrogen alone, even though the uterus is absent, because the numerous reports of adenocarcinoma arising from endometriosis in women treated with unopposed estrogen cannot be ignored.***[449, 478-482]

A Patient Support Organization

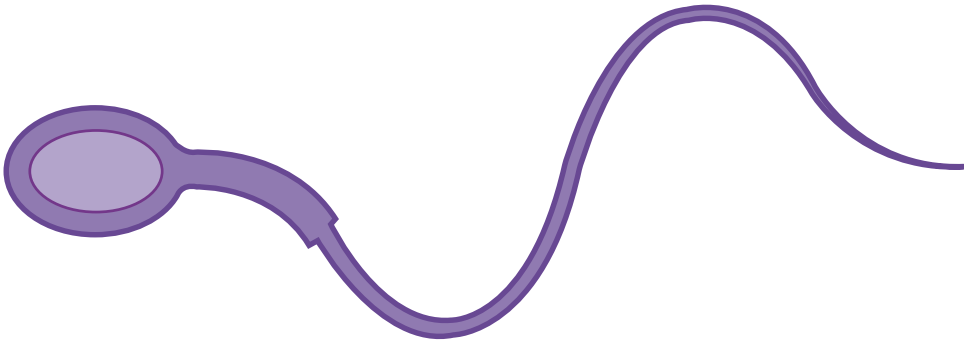
The Endometriosis Association is an international organization that provides education and support for women with endometriosis.

Endometriosis Association

http://www.endometriosisassn.org

All references are available online at: http://www.clinicalgynendoandinfertility.com

30 Male Infertility



Our understanding of male reproductive function and the importance of male factors in infertility has advanced significantly over the last two decades. In the past, the female partner was the primary focus of attention and male factors were regarded as a relatively uncommon cause of infertility. We now recognize that abnormalities in the male are the sole cause of infertility in approximately 20% of infertile couples and are an important contributing factor in another 20–40% of couples with reproductive failure.[1]

Correct diagnosis and specific treatment can help many infertile men to achieve a natural conception with their partners. In others, mild but important semen abnormalities can be overcome by treatments such as intrauterine insemination (IUI). When all else is futile or fails, modern assisted reproductive technologies (ART) still may provide the means to achieve success. In vitro fertilization (IVF) by intracytoplasmic sperm injection (ICSI), involving the injection of a single sperm directly into a mature oocyte, offers men previously considered hopelessly infertile a realistic chance to father children. Artificial insemination using donor sperm, once the only option available for many couples with male factor infertility, remains an important and highly effective treatment strategy, but now may be regarded as the treatment of last resort.

Physicians who care for infertile couples must know how to conduct a basic evaluation of male reproductive function and how to recognize men who require more extensive or sophisticated evaluation and treatment beyond the scope of their own expertise. This chapter will consider the regulation of testicular function, describe the causes of male infertility, discuss the analysis of semen and other methods for evaluation of infertile men, and review current concepts regarding the treatment of male factor infertility.

Regulation of Testicular Function

The testes have two distinct components, the seminiferous tubules (the site of spermatogenesis) and the Leydig cells (the source of testosterone). The seminiferous tubules are composed of germ cells, called spermatogonia, and Sertoli cells, which produce inhibin. Tight junctions between the Sertoli cells form a diffusion barrier known as the blood-testis barrier (similar to the blood-brain barrier), which protects the germ cells from antigens, antibodies, and environmental toxins.[2] The seminiferous tubules are therefore essentially avascular, so regulatory molecules must enter by diffusion. The Leydig cells are located in the connective tissue between the seminiferous tubules.

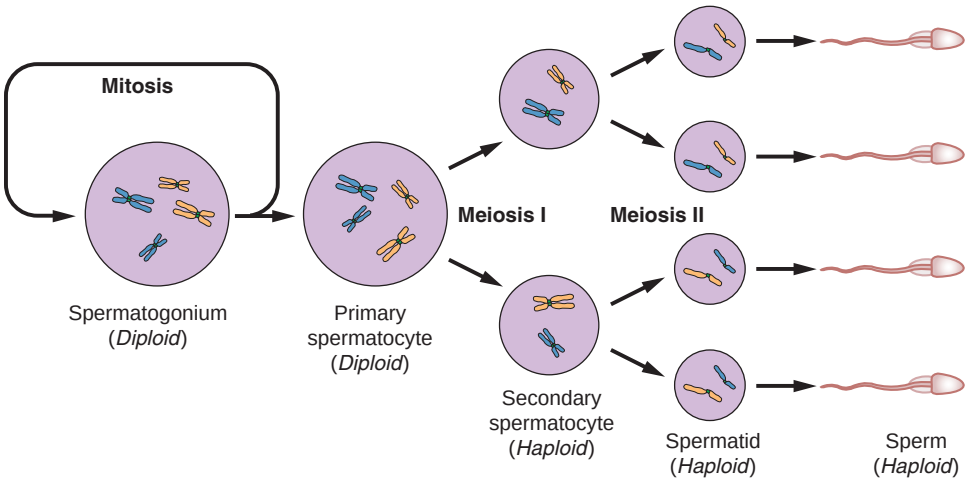
Spermatogenesis

After migration of the germ cells to the genital ridge during embryogenesis, there are approximately 300 thousand spermatogonia in each gonad. Each undergoes a series of mitotic divisions and, by puberty, there are about 600 million in each testis. Continued proliferation during adult life supports the production of approximately 100–200 million sperm each day and more than 1 trillion during a normal reproductive life span.[3] A spermatogonia-specific transcription factor identified in mice, Plzf, is required for maintenance of the spermatogonial stem cell pool.[4, 5]

As spermatogenesis begins, the diploid (46 chromosomes) spermatogonia grow to become primary spermatocytes before entering meiosis. The first meiotic division yields two haploid (23 chromosomes) secondary spermatocytes, each of which gives rise to two spermatids during the second meiotic division. Thereafter, each spermatid gradually matures to become a mature spermatozoan. Approximately 3 million spermatogonia begin development each day, but about half of all potential sperm production is lost during meiosis.[6]

As spermatids develop into mature sperm, the nucleus moves to an eccentric position at the head of the spermatid and becomes covered by an acrosomal cap.[7] The core of the sperm tail consists of nine outer fibers around two inner fibers, surrounded in the middle section by mitochondria. The tail fibers are attached to each other by arms containing the protein dynein, which is an ATPase. Hydrolysis of ATP (adenosine triphosphate) in the adjacent mitochondria provides the energy for sperm motility, which is produced by a sliding action between the fibers in the sperm tail.

The spermatogenic process is directed by genes located on the Y chromosome[8] and takes approximately 70 days to complete from the spermatocyte stage.[9] Another 12–21 days are required for the transport of sperm from the testis through the epididymis to the ejaculatory duct.[10] During passage through the epididymis, sperm mature further to develop the capacity for sustained motility.[11] ***The long time required for sperm development and transit implies that the results of a semen analysis reflect conditions existing many weeks earlier***. Final maturation, or capacitation, of sperm may occur after ejaculation into the female genital tract. Normal spermatogenesis requires the lower temperature of the scrotum, but slight increases in scrotal temperature, such as are associated with the wearing of athletic supporters, do not appear to have any measureable adverse effect.[12] Semen includes secretions contributed by the prostate, the seminal vesicles, and the distal vasa deferentia.



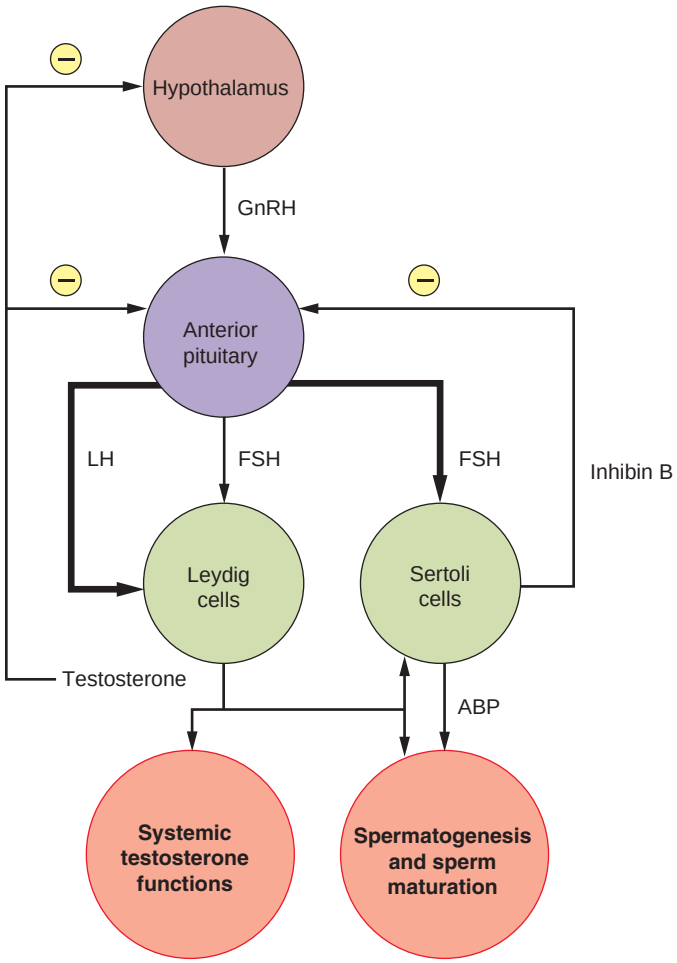
Hormone Regulation

Normal testicular function requires the actions of both pituitary gonadotropins, follicle-stimulating hormone (FSH) and luteinizing hormone (LH). LH stimulates the Leydig cells in the testicular interstitium to synthesize and secrete testosterone (approximately 5–10 mg per day). The actions of LH are supported indirectly by FSH, which induces the appearance of LH receptors on testicular Leydig cells[13] and stimulates synthesis of androgen binding protein (ABP) in Sertoli cells.[14] Testosterone is secreted both into the circulation and into the lumen of the seminiferous tubules where it is highly concentrated to the levels needed to support spermatogenesis in the germinal epithelium and sperm maturation in the epididymis; concentrations within the testes are 50–100 times higher than in blood.[15, 16] The actions of testosterone in support of spermatogenesis are mediated by the Sertoli cells, which line the seminiferous tubules and contain androgen receptors.[16]

Rising serum testosterone levels exert feedback inhibition on LH secretion, acting both at the hypothalamic level to slow the pulsatile release of hypothalamic gonadotropin-releasing hormone (GnRH),[17, 18] probably via a mechanism involving endogenous opiates,[19] and at the pituitary level to decrease pituitary gonadotrope sensitivity to GnRH stimulation.[20] Numerous studies involving infusions of testosterone, estradiol, or dihydrotestosterone (which cannot be converted to estrogen) or the administration of estrogen antagonists in normal subjects,[21,22] in individuals with androgen insensitivity,[23] and in men with idiopathic hypogonadotropic hypogonadism[24] have established that testosterone exerts its negative feedback effects on LH secretion both directly, and indirectly via conversion to estradiol in the brain. Evidence that estradiol is involved in LH feedback control derives from the observation that LH levels are elevated in men with aromatase deficiency[25] and after treatment with aromatase inhibitors.[26]

In contrast to its effects on LH secretion, physiologic levels of testosterone do not suppress FSH secretion. Rather, the regulation of pituitary FSH secretion is controlled by inhibin. FSH levels rise progressively after orchiectomy, the observation that led ultimately to the discovery of inhibin. Inhibin B is synthesized and secreted by Sertoli cells in response to FSH stimulation and specifically inhibits GnRH-stimulated pituitary FSH secretion.[27–29] In the castrate male monkey, treatment with recombinant human inhibin can restore normal FSH levels in the absence of testosterone.[30] Sertoli cell inhibin B secretion is modulated indirectly by LH via testosterone, which inhibits Sertoli cell inhibin B gene expression.[31] Inhibin A is not produced in any significant amount in men. Evidence from studies *in vitro* suggests that other autocrine/paracrine regulatory mechanisms involving locally produced growth factors, neuropeptides, vasoactive peptides, and immune-derived cytokines also are involved, much like the complex interactions that operate in the ovarian follicle.[32–36] The Sertoli cells of the testis are analogous to the granulosa cells of the ovary, and the Leydig cells are comparable to the theca cells.

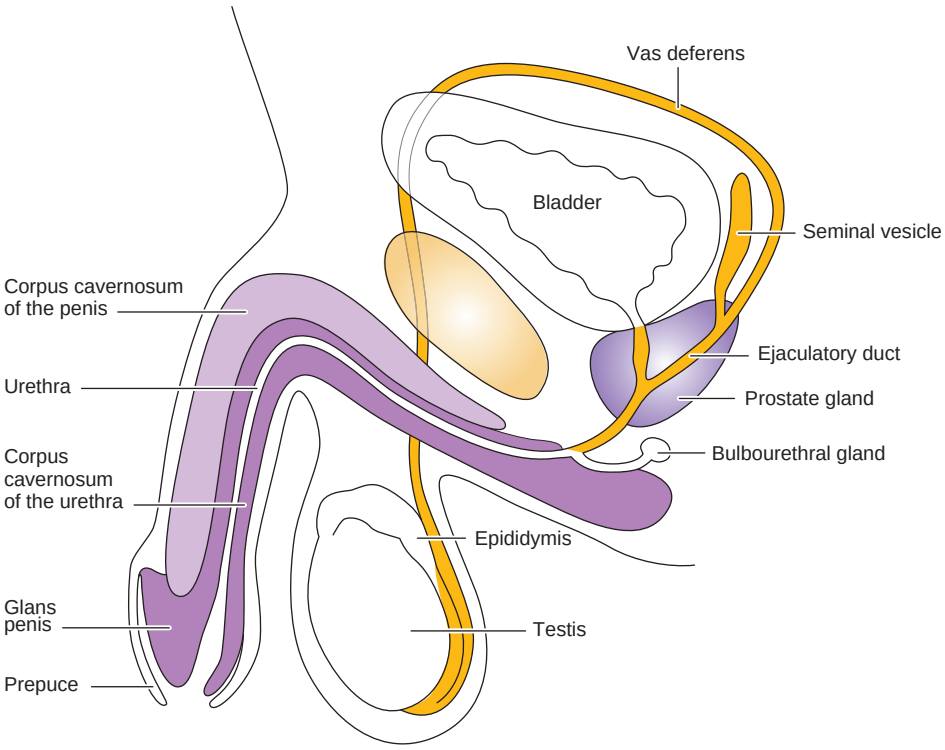
The extent to which FSH and LH are needed to initiate and maintain spermatogenesis has been difficult to define because observations in various natural and experimentally-induced conditions have yielded conflicting evidence. The presence of sperm in the ejaculate of a man with an inactivating mutation in the LH β-subunit gene and in other men with isolated LH deficiency suggests that FSH alone can initiate spermatogenesis,[37] although the possibility of some residual LH activity or FSH-stimulated Leydig cell testosterone production via Sertoli cell factors cannot be excluded.[38] Conversely, low level sperm production in men with inactivating mutations of the FSH receptor[39] and other forms of isolated FSH deficiency [40, 41] suggest that LH-driven testosterone production alone can initiate spermatogenesis, although the possibility of residual FSH activity in the presence of high circulating FSH concentrations must be acknowledged. Evidence that high doses of exogenous testosterone can stimulate complete spermatogenesis in immature monkeys, albeit at low levels, further suggests that FSH is not an absolute requirement,[42] but descriptions of azoospermic men with mutations in the FSH β-subunit gene suggest the opposite.[43, 44] In men with hypogonadotropic hypogonadism of prepubertal onset, normal spermatogenesis can be stimulated by combined treatment with human chorionic gonadotopins (hCG, having potent LH-like actions) and human menopausal gonadotropin (containing FSH), but not by treatment with hCG alone.[45]

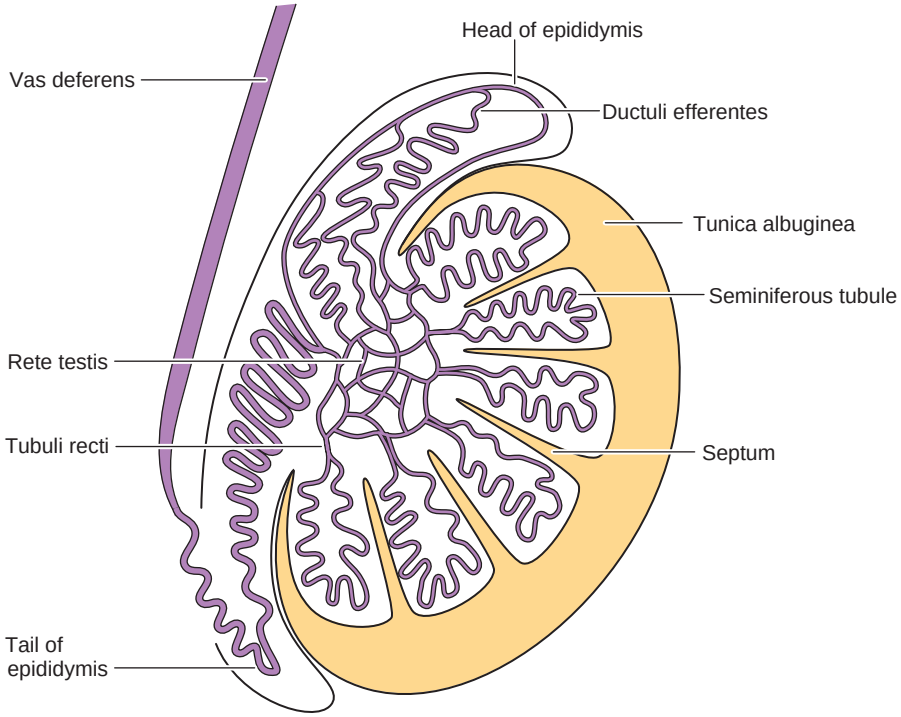


The requirements for the maintenance of spermatogenesis are similarly controversial. The observation in monkeys that exogenous FSH can maintain testicular volume and the numbers of spermatogonia after complete suppression of gonadotropin secretion by treatment with a GnRH antagonist suggests that FSH alone can maintain spermatogenesis in primates, at least to some degree.[46, 47] The description of a unique individual with an activating mutation of the FSH receptor (function in the absence of FSH stimulation) and normal inhibin B

levels (a marker for FSH-stimulated Sertoli cell function)[48] who had undergone hypophysectomy for removal of a benign pituitary tumor (eliminating all endogenous gonadotropin secretion) and remained fertile while receiving only physiologic exogenous testosterone replacement therapy (normally inadequate to support spermatogenesis in hypophysectomized men) serves to further illustrate the importance of FSH in maintaining spermatogenesis.[49] In contrast, the restoration of fertility after treatment with only exogenous hCG in azoospermic men with isolated gonadotropin deficiency (low levels of both FSH and LH) suggests that although LH-stimulated testosterone production may be insufficient to *initiate* spermatogensis, it is sufficient to *maintain* spermatogenesis.[50] In men who develop hypogonadotropic hypogonadism after puberty, during adulthood (e.g., due to a pituitary tumor), spermatogenesis stops but usually can be restored by treatment with hCG alone.[45]

Regardless whether FSH or LH-stimulated testosterone alone is sufficient to initiate or to maintain spermatogenesis, both clearly are required for qualitatively and quantitatively normal sperm production. The importance of FSH has been demonstrated in a variety of experiments in nonhuman primates and men involving the selective suppression of FSH by immunization against FSH or by high dose chronic exogenous hCG treatment. FSH suppression induces both qualitative and quantitative abnormalities of semen quality that can be reversed by simultaneous treatment with exogenous FSH but not with testosterone.[51–55] Moreover, in male contraceptive trials involving treatment with high doses of testosterone, alone or in combination with levonorgestrel to suppress spermatogenesis, azoospermia developed only in men whose serum FSH concentration was suppressed to undetectable levels.[56, 57] The importance of testosterone in spermatogenesis is evident from observations that FSH alone can induce proliferation of the seminiferous epithelium in prepubertal monkeys, but only treatment with both FSH and hCG increases testicular volume and the numbers of Sertoli cells and spermatogonia.[58, 59] Also, in men with idiopathic hypogonadotropic hypogonadism (due to absent GnRH stimulation), exogenous pulsatile GnRH stimulation or a combination of exogenous FSH and LH or hCG can induce spermatogenesis and achieve fertility,[60–62] but treatment with FSH, alone or in combination with low doses of testosterone (insufficient to achieve the high local concentrations of testosterone required to support spermatogenesis), cannot.[63]





Aging and Male Reproductive Function

Although aging has adverse effects on male reproductive function, the impact of age is less obvious than it is in women. Semen quality and male fertility, as well as androgen production and serum testosterone levels, decrease very gradually as age increases.

Aging and Male Fertility

The relationship between age and fertility in men is more difficult to define than in women, largely due to the fundamental difference in gametogenesis between the two sexes. In women, the number of oocytes at birth inexorably declines as age advances until it is functionally exhausted at menopause, and fertility declines with the number of oocytes remaining (Chapter 27). In men, mitotic divisions in the spermatogonia throughout life replenish the supply of germ cells and spermatogenesis continues well into advanced ages, allowing men to reproduce even during senescence. Although fertility in men does appear to decline as age increases, the effects of age are much less distinct. The issue may be growing in importance because an increasing number of men are choosing to father children at older ages. In the U.S., birth rates for men between the ages of 35 and 54 increased by nearly 30% between 1980 (68.2 per 1000 men) and 2000 (88.3 per 1000 men).[64, 65]

Semen volume, sperm motility, and the proportion of morphologically normal sperm, but not sperm concentration, appear to decrease gradually as age increases.[66, 67] However, semen characteristics generally do not accurately predict fertilizing capacity;[68–71] neither do endocrine parmaters.[72, 73] A study in a convenience cohort of nearly 100 men ages

22–80 with no known fertility factors observed decreases in semen volume (–0.03 mL per year), total motility (–0.7% per year), progressive motility (–3.1% per year), and total (progressively) motile sperm count (–4.7% per year).[67] Another study that examined the relationship between age and semen quality among over 400 male partners of women pursuing pregnancy via IVF using donor oocytes found that total motile sperm count decreased by approximately 2.5 million sperm per year.[74]

On balance, the available evidence indicates that pregnancy rates decrease and time to conception increases as male age increases.[66, 75] In studies of the effect of male partner age on pregnancy rates, female partner age and declining coital frequency with increasing age are obvious and important confounding factors.[76] A study examining the effect of paternal age on pregnancy and live birth rates in couples undergoing assisted reproductive technologies found that pregnancy rates declined with age of the male partner and that each additional year of paternal age was associated with 11% increased odds of not achieving a pregnancy and 12% increased odds of not having a live birth; in first treatment cycles, each additional year of paternal age was associated with 5% increased odds of not achieving a pregnancy.[77] Another study of the risk of infertility associated with paternal age, involving over 6,000 randomly selected European women ages 25 to 44, observed that the risk of infertility was increased 2- to 3-fold among women ages 35 to 39 when the male partner was 40 years or older.[78] A study of IVF outcomes involving almost 2,000 women with tubal factor infertility (absent or obstructed fallopian tubes) determined that advanced paternal age (40 years and greater) increased the risk for treatment failure approximately 2-fold for women ages 35–40 years and more than 5-fold for women age 41 and older.[79] Others have observed that pregnancy rates for men over 50 are 23–38% lower than for men under age 30,[66] and that the probability of achieving pregnancy within a year is approximately 50% lower for men over age 35 than for those under age 25.[80] Results of a British study (adjusted for partner age and coital frequency) indicate that time to conception is 5-fold longer for men over age 45 than for men under age 25, even when analysis is restricted to men with young partners.[75] Two other studies have suggested that male fertility may start to decline before age 40.[81, 82] The effect of paternal age is perhaps best assessed in couples using oocyte donation, which makes male age the dependent variable (because almost all oocyte donors are ages 18–35). Unfortunately, data from such studies are conflicting, with some indicating that male age has limited or no impact on pregnancy, implantation, and live birth rates,[74, 83, 84] and others finding that paternal age is inversely related to reproductive outcomes,[85, 86] including a decrease in live birth rates and an increase in miscarriage rates.

There are several possible biological mechanisms that might contribute to an age-related decline in male fertility. One involves cellular or physiologic changes in the male reproductive tract. The testes and prostate exhibit morphological changes with aging that might adversely affect both sperm production and the biochemical properties of semen.[87] Autopsy studies of men who died from accidental causes have observed narrowing and sclerosis of the seminiferous tubules, decreased spermatogenic activity, and reduced numbers of germs cells and Leydig cells as age increases.[88, 89] Another possible mechanism is age-related changes in the hypothalamic-pituitary-testicular axis. Average FSH levels in men increase after age 30,[90] suggesting that the endocrine environment may begin to change during midlife.[91] Decreased semen volume may relate to reduced androgen-stimulated fluid production in the prostate and seminal vesicles because testosterone levels decrease with advancing age.[92] Whatever the mechanism(s), decreasing fertility with increasing male age in healthy couples suggests that normal sperm overproduction may not fully buffer the effects of increasing age. ***However, because there is little or no overall measurable decline in male fertility before age 45–50, the available data suggest that male factors likely contribute relatively little to the overall age-related decline in fertility in women.***

Paternal Age and Pregnancy Outcomes

Because male germ cells pass through more mitotic replications than those of females, there is greater opportunity for error. Older men also are more likely than younger men to have smoked (and for longer periods of time) and to have been exposed to gonadotoxins that may cause DNA damage.[66, 67] ***Increased paternal age has been associated with an increase in numerical and structural chromosomal abnormalities,[93–97] with increased DNA fragmentation,[98] and with a higher frequency of point mutations.[99]*** There also is evidence to suggest that increasing male age may raise the risk of spontaneous abortion in young women.[85, 86, 100]

A number of studies have observed that advanced paternal age is associated with an increase in the prevalence of ***birth defects*** (e.g., neural tube defects, cardiac defects, and limb defects) and ***congenital diseases*** (e.g., Wilms tumor).[101–105] In a large population-based retrospective cohort study that included over 5 million births, the observed overall prevalence of birth defects was 1.5%; compared with infants born to fathers ages 25–29 years, the adjusted odds ratios for birth defects were 1.04 for infants with fathers aged 30–35 years, 1.08 for those aged 40–45 years and 45–50 years, and 1.15 for infants who fathers were over 50 years old.[106] These data suggest that the risk for birth defects increases only slightly, if at all, with increasing paternal age.

Advanced paternal age has been associated with an increase in new ***autosomal dominant mutations*** (e.g., achondroplasia and Alpert, Waardenburg, Crouzon, Pfeiffer, and Marfan syndromes).[107] At least in theory, the observation might reflect a decrease in the activity of antioxidant enzymes in the semen and sperm of older men, increasing their susceptibility to mutation.[108] DNA repair mechanisms also may be impaired in older men. Although the relative risk of autosomal dominant disorders is increased markedly, the absolute risk is still very small (<1%) because autosomal dominant diseases are rare.[109]

Evidence indicates that advanced paternal age is associated with an increased risk for ***schizophrenia*** in offspring;[110–112] overall, the incidence is increased 2- to 3-fold for children whose fathers are over age 45 years, possibly as a consequence of mutations emerging during spermatogenesis.[113] Similarly, increasing paternal age has been associated with an increased risk for ***autism*** in children, which may reflect *de novo* mutations or errors in genetic imprinting.[114–118]

In older fathers, mutations resulting in ***X-linked disease*** also may be more common; examples include hemophilia A and Duchenne muscular dystrophy.[119] The "grandfather effect" describes their transmission from carrier daughters to affected grandsons. Overall, advanced paternal age does not appear to be associated with any significant increase in the risk of fetal autosomal or ***sex chromosome aneuploidy***.[119–124] However, available data are limited and confounded by female partner age. Moreover, results from one study examining the chromosomal complement of paternal gametes suggest that the incidence of sex chromosome aneuploidy may increase with age.[124] A population-based study involving more than 4 million children observed that paternal age was associated with a small but significant increase in risk of leukemia and central nervous system cancers.[125]

It still is not clear whether the risk for miscarriage increases with paternal age, because results of studies conducted thus far are conflicting, with some finding evidence for an association with both early and late fetal loss,[100, 126, 127] and others not.[128, 129] Whereas one retrospective study of 558 pregnancies conceived using donor oocytes observed no association between paternal age and live birth rate,[74] another found that risk for miscarriage increased with paternal age.[83] ***On balance, the weight of available evidence suggests that advanced paternal age may be associated with a small increase in the risk of spontaneous abortion.*** Limited data suggest that advanced paternal age does not significantly increase the risk for fetal growth restriction,[130, 131] or stillbirth.[132]

Androgen Deficiency in the Aging Male

Serum total and free testosterone levels decrease in men as age increases.[133] However, unlike the profound estrogen deficiency and associated symptoms that occur after menopause in women, the age-related decline in androgen levels in men is more gradual and smaller,[134] and the clinical consequences of decreasing androgen levels are not yet clear.

Serum testosterone concentrations exhibit a distinct diurnal variation in young men (with highest levels in the morning), but vary relatively little in elderly men.[135] In the cross-sectional European Male Aging Study, involving 3,220 men ages 40 to 79 years, serum total testosterone concentrations fell by an average of 0.4% per year and free testosterone levels by 1.3% per year.[136] Longitudinal studies have observed a somewhat greater age-related decline in testosterone concentrations and found that levels decrease at a fairly constant rate.[133, 137, 138] Because sex hormone binding globulin (SHBG) concentrations increase gradually with age, free testosterone levels decrease more than total testosterone concentrations.[134] In the Massachusetts Male Aging Study, free testosterone levels fell by an average of almost 3% per year.[139] SHBG levels also may rise in association with increased abdominal obesity, further contributing to the decrease in free testosterone.[140]

As testosterone levels fall steadily, an increasing percentage of aging men become hypogonadal, as defined by testosterone concentrations (total testosterone <300–325 ng/dL; free testosterone <5ng/dL) and/or by signs and symptoms of hypogonadism. In one population-based observational survey, the prevalence of hypogonadism ranged from 3% to 7% among men ages 30 to 69 years, and was 18% in men over age 70.[141] In a longitudinal study, serum total testosterone levels in the hypogonadal range were observed in 20% of men in their 60s, in 30% of those in their 70s, and in 50% of men in their 80s.[133]

In some men over age 50, decreasing serum androgen concentrations may be associated with clinical symptoms and signs of androgen deficiency suggesting “andropause.” Symptoms of androgen deficiency may include decreased libido,[142, 143] with or without erectile dysfunction,[144] reduced strength, energy, or stamina,[145] irritability and perceptions of a lower quality of life,[146] sleep disturbance, depressed mood, and lethargy,[141] and changes in cognitive function.[147, 148] Symptoms may be accompanied by physical changes, including osteopenia or osteoporosis,[149] decreased muscle mass,[150] increased visceral adipose,[151] testicular atrophy and gynecomastia. Epidemiologic studies have observed that low serum testosterone concentrations are associated with development of central obesity, increased insulin levels, the metabolic syndrome, diabetes, and increased mortality.[152–155] Validated questionanaires now are available for use in evaluating older men.[156, 157] However, scores do not predict or correlate well with measured free and total testosterone levels,[158] and therefore lack specificity for the diagnosis of androgen deficiency in the aging male (ADAM).[159–161]

Men with symptoms or signs of androgen deficiency merit evaluation by measuring the serum total testosterone level, ideally during the morning hours to minimize the influence of pulsatile and circadian rhythms in testosterone secretion. Frankly low concentrations (<200 ng/dL) should be confirmed by repeated measurements.[162] Serum total testosterone includes not only free testosterone, but also testosterone bound to albumin and SHBG. “Bioavailable testosterone” is the sum of free and albumin-bound testosterone and the measurement that has correlated best with bone mineral density[145] and sexual[163] and cognitive function[148] in epidemiologic studies. Because the serum total testosterone level occasionally may be misleading, some prefer to measure free or bioavailable testosterone, but the accuracy of free testosterone assays has been challenged[164] and neither assay is widely available. A free testosterone index (FTI) calculated from measurements of total testosterone and SHBG (total testosterone/SHBG) provides an indirect measure of the amount of bioavailable testosterone. ***It is important to emphasize that in men with documented***

androgen deficiency, a normal or low serum LH suggests a secondary hypogonadism that merits additional evaluation by measurement of serum prolactin and magnetic resonance imaging (MRI) to detect any hypothalamic or pituitary mass lesion.

Treatment

A consensus of expert opinion published in 2002 suggested that a total testosterone level under 200 ng/dL (6.9 nmol/L) is evidence of hypogonadism that warrants treatment in symptomatic men, that those with concentrations between 200 ng/dL and 400 ng/dL (6.9–13.9 nmol/L) may benefit from treatment, and that higher levels all but exclude androgen deficiency.[165] A bioavailable testosterone level below the normal range for normal young adult men or an FTI less than 0.153 (nmol/nmol) also is consistent with the diagnosis of androgen deficiency.[133] ***Evidence-based guidelines issued by the Endocrine Society in 2006 recommended that, in the absence of pituitary or testicular disease, testosterone therapy be reserved for men with clearly and consistently low serum total testosterone concentrations (<200 ng/dL) and clinically important symptoms of androgen deficiency.***[162]

The potential risks of testosterone treatment include fluid retention, gynecomastia, increased red blood cell mass, worsening of sleep apnea, promotion of benign or subclinical malignant prostate disease, and possible added risk for cardiovascular disease.[166] ***Accordingly, the Endocrine Society guidelines recommend against testosterone treatment in men with prostate or breast cancer, a palpable prostate nodule or induration, prostate-specific antigen (PSA) greater than 3 ng/mL without further urologic evaluation, eryrthrocytosis (hematocrit >50%), untreated obstructive sleep apnea, severe lower urinary tract symptoms (International Prostate Symptom Score > 19), or class III or IV heart failure.***[162]

Any of the commercial formulations of testosterone may be used for treatment. Androgen therapy may involve parenteral testosterone esters (75 mg per week or 150 mg every 2 weeks), implanted pellets (225 mg every 4–6 months), scrotal (40 cm^2, one patch daily) or peripheral skin patches (5 mg, one patch daily) or testosterone gel (5 g per day); treatment should be individualized. At present, there are no data to indicate that any one formulation is clearly superior. ***The therapeutic goal is to raise serum testosterone concentrations over pretreatment values without exceeding the normal range for young adult men. The target serum testosterone concentration should be lower than that for younger men (e.g., 300–400 ng/dL) to decrease the potential risk of testosterone-dependent disease.***[162] Dehydroepiandrosterone may be converted to testosterone and is commercially available as an oral dietary health supplement; standard doses (50–100 mg daily) generally do not raise serum testosterone concentrations, although higher doses may.[167]

In randomized, placebo-controlled studies, the effects of testosterone therapy on bone density have been inconsistent. In one, no overall increase in hip or spine bone density was observed, but treatment had greatest effect in men with the lowest pre-treatment testosterone levels.[168] In another, testosterone treatment did not increase bone density, but prevented the decrease observed in men receiving placebo.[169] In a third, testosterone treatment (with and without finasteride, which blocks conversion of testosterone to dihydrotestosterone) increased spine bone density by 9–10% and hip bone density by 2–3%.[170] All three studies[168, 169, 171] and a subsequent systematic review[162] found that testosterone treatment increased fat free mass and decreased fat mass. However, the increase in lean mass did not result in any consistent improvement in muscle strength or physical performance.[168, 169, 171] Testosterone treatment also was not accompanied by any demonstrable improvement in quality of life measures or sexual function, as judged by questionnaires.[168, 169]

Androgen therapy must be monitored because the long-term health risks and benefits of treatment have not been established. A baseline physical examination (breasts, heart, lungs, prostate), serum prostate specific antigen (PSA), and complete blood count should be obtained; prostate biopsy is recommended when the digital rectal examination or serum PSA is abnormal. Within 3 months after therapy begins, men receiving androgen therapy should be evaluated for weight gain and any signs of emerging peripheral edema, gynecomastia or breast tenderness, sleep disorders, or prostate enlargement. Recommended monitoring also includes hemoglobin or hematocrit and a serum PSA. A rapid rise in PSA (>1 ng/mL) soon after treatment begins suggests the possibility of an undetected prostate cancer and is reason to discontinue treatment pending a thorough prostate evaluation.[172] Serum testosterone also should be measured to ensure that treatment is achieving the target concentration, but the subjective clinical response is the most important gauge of the effectiveness of androgen therapy. Men with a good clinical response, no apparent adverse effects, and normal testosterone levels may continue treatment, but should return for similar monitoring after another 6 months, and at least annually thereafter. If osteoporosis was one of the indications for treatment, bone mineral density also should be re-evaluated approximately 1–2 years after treatment starts.

In clinical trials of testosterone treatment in elderly men, only a few cases of prostate cancer have been observed, but statistical power was insufficient to support a conclusion that testosterone treatment does not increase risk for prostate cancer. A meta-analysis including 19 trials found that testosterone treatment was associated with a higher prevalence of elevated PSA values and prostate cancer, although biopsy was more commonly performed in men receiving treatment.[173] There is little evidence that short-term treatment has adverse effects on the prostate,[174] but the effects of long-term treatment remain uncertain. Similarly, it is not clear whether physiologic testosterone therapy increases the risk of sleep apnea, because data are conflicting.[173, 175–177] However, testosterone treatment in elderly men clearly can cause erythrocytosis. In individual studies, up to one-third of treated men have developed an abnormally elevated hematocrit,[170, 178] and a meta-analysis concluded that testosterone treatment is associated with more than a 4-fold increased risk for erythrocytosis.[173] Taken together, evidence indicates that testosterone treatment in hypogonadal men has little effect on serum concentrations of total and low-density lipoprotein cholesterol.[179]

Causes of Male Infertility

Male infertility may result from a variety of causes. Some, like ductal obstruction and hypogonadotropic hypogonadism, can be defined accurately and treated effectively. Others, like primary testicular failure, can be defined but are not amenable to treatment, and still others, like seminiferous tubule dysfunction cannot be corrected but can be overcome by intrauterine insemination (IUI) or ART. Although rare, male infertility also may be the first indication of a serious underlying medical condition. ***Unfortunately, much of male infertility is idiopathic, reflecting our still very poor understanding of the mechanisms that govern testicular function.***

The list of known causes of male infertility is long and varied, but can be divided into 4 major categories: 1) hypothalamic-pituitary disorders (1–2%), which may be congenital, acquired, or result from systemic illness; 2) primary gonadal disorders (30–40%), both congenital and acquired; 3) disorders of sperm transport (10–20%); and 4) idiopathic (40–50%).

Causes of Male Infertility

Hypothalamic-Pituitary Disorders
Idiopathic isolated gonadotropin deficiency
Kallmann syndrome
Single gene mutations (e.g., involving the GnRH receptor, FSHβ, LHβ, or transcription factors involved in pituitary development)
Hypothalamic and pituitary tumors (e.g., craniopharyngioma, macroadenoma)
Infiltrative diseases (sarcoidosis, histiocytosis, transfusion siderosis, hemochromotosis)
Hyperprolactinemia
Drugs (GnRH analogs, androgens, estrogens, glucocorticoids, opiates)
Critical illness or injury
Chronic systemic illness or malnutrition
Infections (e.g., meningitis)
Obesity

Primary Gonadal Disorders
Klinefelter syndrome
Y chromosome deletions
Single gene mutations and polymorphisms (e.g., involving the androgen, estrogen, or FSH receptor)
Cryptorchidism
Varicoceles
Infections (e.g., viral orchitis, leprosy, tuberculosis)
Drugs (e.g., alkylating agents, alcohol, antiandrogens, cimetidine)
Radiation
Environmental gonadotoxins (e.g., heat, smoking, metals, organic solvents, pesticides)
Chronic illness (renal insufficiency, cirrhosis, cancer, sickle cell disease, amyloidosis, vasculitis, celiac disease)

Disorders of Sperm Transport
Epididymal obstruction or dysfunction
Congenital bilateral absence of the vas deferens (relating to *CFTR* mutations)
Infections causing obstruction of the vas deferens (e.g., gonorrhea, chlamydia, tuberculosis)
Vasectomy
Kartagener syndrome (primary ciliary dyskinesia)
Young syndrome
Ejaculatory dysfunction (e.g., spinal cord disease, autonomic dysfunction)

Hypothalamic-Pituitary Disorders

Any hypothalamic or pituitary disease or disorder causing a deficiency of gonadotropin-releasing hormone (GnRH) or gonadotropins can cause male infertility. The most common congenital cause is idiopathic isolated gonadotropin deficiency due to absent or defective GnRH secretion (resulting in sexual infantilism).[180] When accompanied by one or more extragonadal abnormalities, such as anosmia, red-green color blindness, midline facial defects (e.g., cleft palate), neurosensory hearing loss, synkinesis (mirror movements), or renal anomalies, the disorder is known as ***Kallmann syndrome.***[181–183] A variety of mutations have been identified in affected men, involving genes encoding cell surface adhesion molecules or receptors, which are required for normal migration of GnRH neurons from the olfactory placode to the hypothalamus; examples include *KAL1*,[184, 185] fibroblast growth

factor 1 (also known as *KAL2*),[186] and prokineticin-2 (*PROK2*) and its receptor (*PROKR-2*).[187] Other genetic causes of hypogonadotropic hypogonadism include rare mutations affecting the GnRH receptor[188] the β-subunit of FSH[43] or LH,[37, 189, 190] or transcription factors involved in pituitary development during embryogenesis, such as *LHX3*,[191] *LHX4*, *HESX1*,[192] and *PROP-1*.[193]

Hypogonadotropic hypogonadism also can result from hypothalamic disease or treatments that inhibit GnRH secretion, abnormalities of the pituitary stalk that interfere with GnRH delivery, and pituitary disease that prevents normal gonadotropin secretion. ***Hypothalamic or pituitary tumors*** can distort the pituitary stalk or compress and suppress pituitary gonadotropes.

Infiltrative diseases of the hyopothalamus or pituitary (sarcoidosis, histiocytosis, transfusion siderosis, hemochromotosis) can inhibit GnRH or pituitary gonadotropin secretion.[194, 195] ***Hyperprolactinemia*** from any cause,[196] and ***treatment with GnRH analogs*** (e.g., for prostate cancer) ***androgens*** (e.g., anabolic steroids),[197] ***glucocorticoids,***[198] or ***opiates***[199–201] can suppress gonadotropin secretion. ***Critical illness***[202] or ***injury*** (e.g., head trauma)[203] and ***chronic systemic illness*** (e.g., diabetes mellitus) or ***malnutrition*** also have been associated with hypogonadotropic hypogonadism. ***Infections*** (e.g., meningitis) are another rare but recognized cause of hypopituitarism.[204]

Obesity in men is associated with hypogonadotropic hypogonadism, involving several mechanisms.[205] Serum free testosterone concentrations are inversely related to body weight and body mass index, independent of changes in SHBG levels,[206–208] and estrogen concentrations are elevated due to increased aromatase activity in adipose.[209] ***Obstructive sleep apnea*** is a separate but related additional factor, resulting in hypoxia.

Primary Gonadal Disorders

Primary gonadal failure (hypergonadotropic hypogonadism) is a major cause of azoospermia and oligospermia and can result from a variety of congenital or acquired disorders, including Klinefelter syndrome, Y chromosome deletions, single gene mutations, cryptorchidism, varicoceles, and other less common causes.

Klinefelter Syndrome

Klinefelter syndrome is one of the most common causes of primary testicular failure, affecting approximately 1 in 1,000 males,[210, 211] and is characterized by sex chromosome aneuploidy. Although an extra X chromosome (47,XXY) is the most common form, some men with Klinefelter syndrome have a greater or lesser number of X chromosomes (e.g., 48,XXXY, 46,XY/47,XXY);[212] 46,XX males, resulting from translocation of the testis-determining gene (*SRY*) to an X chromosome, also have Klinefelter syndrome. The phenotype varies with the number of extra X chromosomes, and possibly also with the number of trinuceotide CAG repeats on the androgen receptor gene (a polymorphism); as the length of the repeat sequence increases, receptor activity decreases. A longer CAG repeat sequence has been associated with taller stature, lower bone mineral density, gynecomastia, and decreased penile length.[213, 214]

Men with Klinefelter syndrome generally have small, firm testes, resulting from damage to both seminiferous tubules and Leydig cells. Serum concentrations of FSH and LH are elevated and testosterone levels are decreased to varying extent. Affected men

have severely reduced sperm counts and are under-virilized.[212, 215] Cryptorchidism is more common in men with Klinefelter syndrome and causes more severe testicular damage.[216]

The length of the arms and legs is increased in men with Klinefelter syndrome, due both to testosterone deficiency and to an independent abnormality of the long bones. Men with Klinefelter syndrome also exhibit a number of psychosocial abnormalities,[217] which have been described as a marked lack of insight, poor judgment, and an impaired ability to learn from adverse experience.[218] They also may have difficulty with complex speech and a decreased attention span.[219] Later in life, they have an increased risk for developing pulmonary diseases, breast cancer,[220] mediastinal germ cell tumors,[221] varicose veins and leg ulcers,[222] systemic lupus erythematosus,[223] and diabetes mellitus.[224]

Other chromosomal abnormalities associated with primary gonadal failure include the 46,XY/45,X karyotype, causing a syndrome characterized by short stature and other features of Turner syndrome.[225] Because the testes may be streaks, dysgenetic, or normal, the phenotype varies from female to normal male. In those with a streak and a dysgenetic testis (mixed gonadal dysgenesis), the risk of gonadoblastoma is increased (approximately 20%), and gonadectomy is therefore indicated.

Y Chromosome Deletions

Microdeletions of the long arm of the Y chromosome are now recognized as a relatively common cause of severe oligospermia and azoospermia, affecting up to 20% of men with infertility.[226] Most map to the Yq11 region (named azoospermia factor, or *AZF*), which contains three regions, AZFa, AZFb, and AZFc. Deletions of the AZFa or AZFb regions typically result in azoospermia. Mutations in the AZFc region cause infertility of varying severity, ranging from oligospermia to azoospermia and are the largest known recurrent deletions in humans.[227, 228] The *DDx3Y* and *USP9Y* genes, both located in the AZFa region, have been implicated as having an important role in spermatogenesis; azoospermia is consistently observed when both are deleted.[229, 230] Y chromosome deletions also have been identified in men with cryptorchidism, varicocele, and obstructions of the vas deferens.[231, 232]

Because all Y chromosome abnormalities will be transmitted to sons of affected men conceived via intracytoplasmic sperm injection (ICSI), genetic testing and counseling should be offered to affected men before their sperm are used for that purpose. Given the importance and potential consequences of Y chromosome deletions, there is a need to standardize the tests for their detection.[233]

Single Gene Mutations and Polymorphisms

Normal male sexual differentiation and spermatogenesis require both normal androgen production and normal androgen receptors (Chapter 9). The androgen receptor plays an important role in the differentiation of spermatids and their release from the seminiferous epithelium. Consequently, it is not surprising that defects in androgen synthesis or androgen sensitivity are associated with infertility.[234, 235]

As discussed above, the number of trinucleotide CAG repeats in exon 1 of the androgen receptor gene is inversely correlated with its transcriptional activity.[213, 214] In one study in normal fertile men, those with short repeat sequences had the highest sperm concentrations.[236] However, studies in men with idiopathic infertility have yielded inconsistent results, with some finding an association between longer CAG repeat segments and male

infertility,[237–239] and others not.[240] ***A meta-analysis including 33 studies concluded that men with idiopathic infertility had significantly longer CAG repeat lengths than fertile men, suggesting that even subtle abnormalities in androgen action may adversely affect male fertility***.[241]

Evidence suggests that ***disorders of estrogen synthesis or action*** also may be associated with infertility in men. Impaired spermatogenesis has been observed in mice and in men lacking a functional estrogen receptor (alpha),[242, 243] and in mice with an inactivating mutation in the aromatase enzyme.[244] Polymorphisms involving variations in the number of TA tandem repeats in the promoter region of the estrogen receptor gene also have been related to sperm production, with higher numbers of TA repeats being associated with lower sperm counts.[245] Inactivating mutations in the FSH receptor gene are a rare cause of male infertility.[39, 246]

Several other autosomal and X-linked genes have been identified as important regulators of spermatogenesis. Men with myotonic dystrophy (an autosomal disorder associated with impaired motor function, cataracts, premature balding, mild mental retardation, and hypogonadism) also can exhibit abnormal spermatogenesis.[247] Mutations in the *SYCP3* gene (involved in regulation of the synapse between homologous chromosomes during meiosis) have been implicated as a potential cause of male infertility.[245] Others include polymorphisms of *DAZL* (an autosomal homolog of the *DAZ*, deleted in azoospermia, gene),[248–252] *PRM1* and *PRM2* (protamines involved in chromatin compaction), *TNP1* and *TMP2* (transition nuclear proteins), and *USP26* (de-ubiquitinating enzyme family).[245]

Cryptorchidism

Cryptorchidism results from a failure of testicular descent during fetal development, which is an androgen-dependent process. Consequently, it is common in men with abnormalities of testosterone production, such as Kallmann syndrome, androgen resistance, and defects in testosterone synthesis. Cryptorchidism can be unilateral or bilateral and, in either case, is associated with impaired spermatogenesis and an increased risk for developing testicular tumors. Even in the absence of cryptorchidism, the incidence of testicular cancer is increased in infertile men.[253, 254]

In men with cryptorchidism, serum FSH levels often are elevated, but LH concentrations generally are normal, reflecting normal Leydig cell function. ***The severity of the semen abnormality relates to the duration of time the testes have been outside of the scrotum.*** Because the testes are more easily retractable early in life, very young boys may appear transiently to have cryptorchidism but, in most, the testes descend and remain in the scrotum by age 1.[255] Men having low serum inhibin B levels, increased FSH concentrations, and decreased sperm density after repair of cryptorchidism are at increased risk for infertility.[256]

Varicoceles

Varicoceles result from dilation of the panpiniform plexux of the spermatic veins in the scrotum. They are more prevalent in infertile men (up to 30%) than in fertile men (10–15%) and are 10 times more commonly found on the left than on the right, probably because the left spermatic vein is longer and joins the left renal vein at a right angle.[257] ***Although increased testicular temperature, delayed removal of local toxins, hypoxia, and stasis are viewed as the mechanisms likely responsible for the association between varicoceles and infertility, no causal relationship has been established.***[258–260]

Other Causes of Primary Gonadal Failure

Certain ***infections*** are associated with male infertility. Mumps orchitis is widely recognized as a cause of male infertility. Although rare in prepubertal males, it occurs in up to 25% of adult men with mumps, some of whom become infertile. The mechanism may involve damage to the germinal epithelium, ischemia, or immune dysfunction.[261, 262] Gonorrhea and chlamydial infections also can cause orchitis. Other infections associated with male infertility include tuberculosis, which may cause epididymal obstruction, leprosy,[263] and human immunodeficiency virus (HIV).[264, 265]

Drugs that can adversely affect spermatogenesis or Leydig cell function include alkylating agents (e.g., cyclophosphamide, chlorambucil), anti-androgens (e.g., flutamide, cyproterone, spironolactone), ketoconazole, cimetidine, and anabolic steroids.[266] Doses of ***radiation*** as low as 0.015 Gy (15 rads) can suppress spermatogenesis and doses above 6 Gy generally cause permanent azoospermia and infertility.[267]

Environmental exposures that may act as ***gonadotoxins*** include heat, smoking, metals, organic solvents, and pesticides. A modest increase in scotal temperature can adversely affect spermatogenesis and a febrile illness can result in dramatic, if also transient, decreases in sperm density and motility. Hyperthermia also may explain the infertility associated with spinal cord injuries, and chronic sauna or spa exposure.[268] In theory, environmental sources of heat, including tight-fitting underclothing, hot baths and spas, and occupations that require long hours of sitting (long-distance driving) might decrease fertility, but none has ever been substantiated in clinical studies.[12] Smoking or heavy use of marijuana, alcohol, or cocaine can decrease semen quality and testosterone levels.[269–271]

Chronic illness,[272] such as chronic renal insufficiency[273] cirrhosis, or malnutrition,[274] can result in primary gonadal failure. Infertility also is common in men with sickle cell disease, probably due to testicular ischemia.

Disorders of Sperm Transport

Even when sperm production is normal, epididymal obstruction or dysfunction can result in infertility. The cause of infertility is clear in men with obstruction, but relatively little is known about epididymal function. Isolated asthenospermia is presumed to result from epididymal dysfunction, and intrauterine exposure to diethylstilbestrol may be one cause.[275]

Congenital or acquired abnormalities of the vas deferens can cause obstruction and infertility. Approximately 1–2% of infertile men have ***congenital bilateral absence of the vas deferens*** (CBAVD), almost always related to mutations in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene.[276] Most affected men do not exhibit any respiratory and pancreatic disease. ***Infections*** (gonorrhea, Chlamydia, tuberculosis) and ***vasectomy*** are other causes of vasal obstruction. ***Primary ciliary dyskinesia (Kartagener syndrome)*** is a genetic disease that adversely affects cilia structure and function and generally presents as recurrent sinus and pulmonary infections, bronchiectasis, situs inversus, and male infertility due to oligo-asthenospermia.[277, 278] ***Young syndrome*** is another genetic disease, in which inspissated secretions in the vas and epididymis result in obstructive azoospermia.[279, 280]

Ejaculatory dysfunction resulting from spinal cord disease or injury, sympathetomy, or autonomic disease is another cause of infertility relating to disorders of sperm transport.

The Male Infertility Evaluation

The evaluation of the infertile male should be directed towards achieving all of the following goals[281]:

- To identify and to correct specific causes of infertility, when possible.
- To identify individuals whose infertility cannot be corrected but may be overcome by IUI or use of various forms of ART.
- To identify individuals having a genetic abnormality that may affect the health of any offspring that may be conceived through the use of ART.
- To identify individuals whose infertility can neither be corrected nor overcome with ART, for whom adoption or the use of donor sperm are options worthy of consideration.
- To identify any important underlying medical condition that may require specific medical attention.

Evaluation of the male partner should begin at the same time as in the female partner, generally when pregnancy fails to occur after 1 year of reasonably regular unprotected intercourse. Earlier evaluation is indicated for men with any obvious infertility risk factor, those whose partner is age 35 or older, (where it is important to identify all potential infertility factors as quickly and efficiently as possible), and men who have reason to question their fertility.

In the male partner, the most relevant parts of the medical history and physical examination include the following:[281]

History

- Duration of infertility and previous fertility.
- Coital frequency and any sexual dysfunction.
- Results of any previous evaluation or treatment for infertility.
- Childhood illnesses and developmental history.
- Previous surgery, its indications and outcome, and systemic medical illnesses.
- Past episodes of or exposures to sexually-transmitted infections.
- Exposures to environmental toxins, including heat.
- Current medications and allergies.
- Occupations and use of tobacco, alcohol, and other drugs.

Physical Examination

- Examination of the penis, to include the location of the urethral meatus.
- Palpation of the testes and measurement of their size.
- The presence and consistency of both the vasa and epididymides.
- Presence of any varicocele
- Secondary sexual sex characteristics, including body habitus, hair distribution, and breast development.
- Digital rectal examination.

A history of cryptorchidism or mumps orchitis suggests the possibility of testicular atrophy.[261, 282] The timing and extent of secondary sexual development may alert one to the possibility of an endocrinopathy. Ductal obstruction can result from sexually-transmitted infections. Diabetes mellitus (bladder neck dysfunction resulting in retrograde ejaculation) and cystic fibrosis (highly associated with congenital absence of the vas deferens) are medical illnesses that may hinder fertility in men. Inguinal hernia repair, renal transplant, and scrotal surgery are associated with risks for unrecognized injury to the vas deferens.[283] Retroperitoneal surgery may disrupt neural pathways and cause ejaculatory dysfunction;

treatment with alpha-blockers, phentolamine, methyldopa, guanethidine, or reserpine may have similar effects.

When the infertility evaluation is directed by the gynecologist or primary clinician, physical examination of the male may be deferred pending the results of the first semen analysis when there is no history of any male genital abnormality, trauma, surgery, or sexual dysfunction. However, an abnormal reproductive history or semen analysis is indication for additional formal evaluation that may be conducted by the gynecologist having the necessary training and experience, but is most often performed by the urologist or other specialist in male reproduction.

Semen Analysis

If a male infertility factor exists, it almost always will be revealed by an abnormal semen analysis, although other male factors (sexual dysfunction) may be involved even when semen quality is normal. ***The initial evaluation for male factor infertility should include at least one properly performed semen analysis. If abnormal, another semen analysis should be obtained after at least 4 weeks.***[281] Semen parameters can vary widely over time, even among fertile men,[284–287] and also exhibit seasonal variations.[288–290] Considering that the overall objective is to gain a sense of the usual semen quality, over time, more than one analysis is helpful, because a single semen sample yields only a point estimate that may or may not be representative. However, with relatively few exceptions, a normal initial semen analysis generally excludes an important male factor when there is no complaint or suspicion of sexual dysfunction. Conversely, abnormal semen parameters suggest the need for additional endocrine, urologic, or genetic evaluation.

Standard but detailed instructions for semen collection should be provided, to include a defined abstinence period of 2–3 days. Shorter intervals of abstinence decrease the semen volume and sperm density but generally have little or no impact on sperm motility or morphology.[291] ***Longer abstinence intervals increase semen volume and sperm density, but also increase the proportion of dead, immotile or morphologically abnormal sperm.***[292] Ideally, the semen specimen should be collected by masturbation directly into a clean container. If necessary, semen may be collected via intercourse using a specially manufactured silastic condom that does not contain spermicidal agents like those in condoms intended for contraceptive purposes. Collection after withdrawal during intercourse risks loss of the initial portion of the specimen, which generally contains the highest concentration of sperm. If possible, the semen specimen should be collected in a private room within or near the laboratory. When necessary, the specimen can be collected at home but should be kept at room or body temperature during transport. Regardless of the method of collection, the semen sample should be examined within an hour after collection.

Normal Reference Values

The normal reference values in wide use are based on comparisons of the values observed in the male partners of fertile and infertile couples without specific exclusion of female infertility factors,[293–295] and therefore do not necessarily represent the average ranges observed in fertile men. Unfortunately, there is considerable overlap between the semen parameters observed in fertile and infertile men.[296] The normal reference ranges certainly do not represent the absolute minimum values needed for conception; many men with values outside the normal ranges are fertile and many with normal values are nonetheless infertile.[296–299] ***Values outside of normal ranges suggest a male infertility factor that may***

require additional clinical or laboratory evaluation, but each parameter must be considered in the context of the whole. A mildly low sperm density may have little significance when semen volume, sperm motility, and the proportion of abnormal sperm are normal. Conversely, a normal sperm density offers little reassurance when semen volume is frankly low or the proportion of motile or normal sperm is grossly abnormal. ***Overall, the odds of male infertility increase with the number of major semen parameters (concentration, motility, morphology) in the subfertile range; the probability is two to three times higher when one is abnormal, five to seven times higher when two are abnormal, and 16 times greater when all three are abnormal.***[296]

Although detailed procedures for semen analysis have been established by the World Health Organization (WHO), the methods and accuracy of semen analyses as they are performed in physician offices, hospitals, and specialty andrology laboratories may vary. Ideally, to ensure accurate and reliable results, semen analyses should be performed in a laboratory having an established quality control program that conforms to the standards outlined in the Clinical Laboratory Improvement Amendments (CLIA; www.hcfa.gov/medicaid/clia/cliahome.htm).[300, 301] The traditional WHO normal reference values are as follows[302–304]:

Semen Analysis: Normal Reference Values

Volume	**1.5–5.0 mL**
pH	**>7.2**
Viscosity	**<3 (scale 0–4)**
Sperm concentration	**>20 million/mL**
Total sperm number	**>40 million/ejaculate**
Percent motility	**>50%**
Forward progression	**>2 (scale 0–4)**
Normal morphology	**>50% normal**[302]
	>30% normal[303]
	>14% normal[304]
Round cells	**<5 million/mL**
Sperm agglutination	**<2 (Scale 0–3)**

Over time, the methods and normal reference values for determining sperm concentration and motility have changed little, but those for sperm morphology have changed rather substantially. Using the most recent and rigorous standard, even fertile men have relatively few normal sperm. The rationale for the change in the morphology standard and its clinical relevance are discussed below (see Sperm Morphology).

In 2010, the WHO published revised ***lower reference limits*** for semen analyses, which represent the fifth centile in a population of over 1,900 men from eight countries on three continents whose partners conceived within 12 months:[305]

Semen Analysis: Lower Reference Limits (95% CI) in Fertile Men

Volume	**1.5 (1.4–1.7) mL**
Sperm concentration	**15 (12–16) million/mL**
Total sperm number	**39 (33–46) million/ejaculate**
Total motility	**40 (38–42) %**
Progressive motility	**32 (31–34) %**
Normal morphology	**4 (3–4) %**
Vitality	**58 (55–63) %**

These data provide reliable, clinically relevant reference values for use in the evaluation of infertile men and in assessing their prognosis for achieving pregnancies with their partners.

Ejaculate Volume and pH

A low or absent ejaculate volume suggests the possibility of failed emission, incomplete collection, a short abstinence interval, congenital bilateral absence of the vas deferens (CBAVD), ejaculatory duct obstruction, hypogonadism, or retrograde ejaculation. Other semen parameters can help to differentiate the cause.

The majority of semen volume comes from the seminal vesicles which share a common embryology with the vasa deferentia. Seminal vesicle secretions are alkaline and contain fructose. Because the seminal vesicles are hypoplastic or absent in most men with CBAVD, they generally produce a low-volume acidic (pH less than 7.2) ejaculate that contains little or no fructose and reflects the greater contribution of acidic prostatic secretions.[306–308] Men with ejaculatory duct obstruction produce an ejaculate having similar characteristics because the ejaculatory ducts are formed by the union of the vasa with the ducts exiting the seminal vesicles, proximal to the prostate; semen fructose concentrations decrease with increasing severity of ejaculatory duct obstruction.[309–311] ***When both ejaculatory ducts are completely obstructed, the semen is acidic (containing only prostatic secretions) and contains neither fructose nor sperm***. Hypogonadal men with either primary or secondary testicular failure also may exhibit low ejaculate volumes because the secretions of the seminal vesicles and prostate are stimulated by androgens; volume is therefore decreased when androgen levels are low.

A post-ejaculatory urinalysis can detect retrograde ejaculation and should be considered whenever the ejaculate volume is less than 1 mL, except when hypogonadism, CBAVD, collection problems, or a short abstinence interval offers an obvious explanation. When indicated, the post-ejaculatory urinalysis involves centrifugation for 10 minutes at no less than 300 *g*, followed by microscopic examination of the pellet (400X). ***In men with no or low semen volume and azoospermia (no sperm in the ejaculate), the observation of any sperm on post-ejaculatory urinalysis suggests retrograde ejaculation.*** More substantial numbers of sperm must be observed in men with low volume oligospermia before making the diagnosis of retrograde ejaculation because sperm found in the urine may simply have been washed from the urethra during urination.[306]

Sperm Concentration and Total Sperm Count

Azoospermia describes the absence of sperm on standard microscopic examination. The prevalence of azoospermia is approximately 1% in all men[312] but up to 10–15% in infertile men.[313] ***To establish the diagnosis, the semen specimen should be centrifuged at high speed (3,000 g for 15 minutes) and the pellet examined at high magnification (400X)[304]; the absence of sperm should be documented on at least two separate occasions***. Azoospermia is generally classified as obstructive (normal sperm production) or non-obstructive (decreased or absent spermatogenesis).

Obstructive azoospermia may result from a blockage anywhere in the ductal system, from the efferent ductules to the ejaculatory ducts, as the consequence of severe infection, iatrogenic injury during scrotal or inguinal surgery, or congenital anomalies (CBAVD); approximately 40% of azoospermic men have an obstruction.[306] Non-obstructive azoospermia is caused by intrinsic testicular disease (primary testicular failure) or endocrinopathies and other conditions that suppress spermatogenesis (secondary testicular failure). Men with non-obstructive azoospermia may have low level sperm production that is insufficient to drive epididymal transport and to permit sperm to enter the ejaculate.[314] Careful examination of a centrifuged semen sample will identify sperm in the ejaculates of up to one-third

of men with a preliminary diagnosis of non-obstructive azoospermia.[315] The observation has practical significance because men in whom even a modest number of sperm can be recovered from the ejaculate may not require surgical sperm retrieval for IVF (testicular sperm extraction; TESE).

Oligospermia is defined traditionally by a sperm density less than 20 million/mL and is considered severe when the sperm concentration is below 5 million/mL. ***The probability of conception increases with increasing sperm concentrations up to approximately 40–50 million/ mL, but does not rise further with higher sperm densities.***[296, 297] The results of a large U.S. study comparing semen parameters in fertile and infertile men with normal partners indicate that the likelihood of male infertility is increased approximately 5-fold (5.3, 95% confidence interval 3.3–8.3) when sperm density is less than 13.5 million/mL.[296] In an earlier European study of similar design, the density representing the tenth percentile for fertile men was 14 million/mL.[316] These values are consistent with the lower reference limit for fertile men recommended recently by the WHO (15 million/mL).[305] Oligospermia may be associated with a varicocele, hypogonadism, or specific microdeletions in the Y chromosome. ***Endocrine and genetic evaluation is indicated for men with severe oligospermia*** (discussed below).

Total sperm count is simply the product of multiplying the semen volume and sperm concentration. The total sperm count may be normal in oligospermic men when volume is high, and also normal when volume is low but density is high. The two parameters fluctuate and must be considered together in making judgments regarding semen quality. Numerous studies have suggested that the average sperm count in men has been decreasing steadily over the past few decades,[317, 318] raising concerns that environmental toxins and chemicals having estrogen-like activity (xenoestrogens) might be responsible. However, numerous others have observed no evidence of any significant change.[319–324] Most importantly, the prevalence of infertility has not increased significantly over the same intervals, indicating that any decrease in semen quality that may have occurred has had no global clinical impact.

Sperm Motility, Forward Progression, Total Motile Count, and Vitality

Sperm motility is estimated as a percentage of the total sperm population exhibiting any motion. Forward progression generally is graded on an arbitrary scale (grade 0–4) and most often reported as the percentages exhibiting rapid (grade 3–4), slow (grade 2), and non-progressive motility (grade 0–1). Total progressive motility generally represents an estimate of the percentage of sperm exhibiting purposeful forward motion (grades 2–4). ***The probability of conception rises with increasing motility up to approximately 60%.***[296] According to one large U.S. study, the likelihood of male infertility is increased approximately 5-fold (OR 5.6, 95% CI 3.5–8.3) when progressive motility is less than 32%.[296] In another, the threshold separating fertile and infertile men was 45% and the tenth percentile motility for fertile men was 28%.[316] Again, these values compare well with the lower reference value for progressive motility now recommended by the WHO (32%).[305]

The total motile sperm count is calculated from the total sperm count and the percentage of progressively motile sperm and represents an estimate of the total number of active sperm in the ejaculate. Allowing for the inevitable procedural losses associated with processing a semen sample for IUI (up to approximately 50%), the total motile sperm count can be used to estimate the likely processed total motile sperm count, which correlates with the probability of pregnancy achieved with IUI in the treatment of male factor infertility (see Treatment, below).[325–329]

In general, poor sperm motility (asthenospermia) suggests testicular or epididymal dysfunction. Asthenospermia has been associated with sperm autoantibodies (predisposing to aggregation), genital tract infections (leukocytes in the semen), partial obstruction of the ejaculatory ducts or at the site of a vasectomy reversal (reanastomosis), varicoceles, and prolonged abstinence intervals.

Large numbers of viable nonmotile sperm suggest the rare possibility of primary ciliary dyskinesia (Kartagener syndrome), in which sperm tails have a structural abnormality and cannot flagellate. The cilia of the respiratory tract usually also are involved; affected individuals are infertile and predisposed to chronic respiratory tract infections. Diagnosis is made by examination of sperm using electron microscopy.

When no motile sperm are observed, a sperm vitality test can differentiate viable nonmotile sperm from dead sperm. One method involves mixing fresh semen with a supravital dye (eosin Y or trypan blue); sperm with intact membrane function do not take up the stain. Another method, the hypo-osmotic sperm swelling test, involves incubation of sperm in a hypo-osmotic solution; the tails of sperm with normal membrane function swell and coil as fluid is transported across the membrane. In men with few or no motile sperm, the hypo-osmotic swelling test can be used to identify living nonmotile sperm for ICSI.[330]

Sperm Morphology

Sperm morphology reflects the quality of spermatogenesis. Morphological abnormalities (teratospermia) are categorized by location, involving the head, neck (midpiece), or tail. Cytoplasmic droplets in the midpiece that occupy more than approximately one half of the area of a normal sperm head represent another specific defect. Sperm classified as normal must be normal in all respects. Teratospermia has been associated with varicocele and with both primary and secondary testicular failure. It may be observed in association with abnormalities in sperm concentration and motility or occur as an isolated abnormality.

The most recent WHO reference values (since 1999) for the evaluation of sperm morphology are very similar to those known as the Kruger (Tygerberg) or "strict" criteria,[331, 332] which arose from efforts to identify predictors of fertilization in IVF cycles. When sperm morphology was judged according to a strict normal standard, fertilization efficiency *in vitro* correlated with the percentage of morphologically normal sperm.[331, 333, 334] Conventional fertilization rates were highest when the percentage of normal sperm was 14% or higher, very poor (7–8%) when less than 4% of sperm had normal morphology, and intermediate when values fell between the two threshold values.[331] After several studies confirmed the predictive value of strict sperm morphology in IVF,[335–342] severe teratospermia (0–4% normal sperm by strict criteria) became widely accepted as an indication for ICSI in IVF cycles. However, others have observed no differences in the fertilization, pregnancy, and live birth rates achieved with ICSI and conventional fertilization and argue that isolated teratospermia is not a valid indication for performing ICSI.[343–346] ***Controversy continues, but strict sperm morphology remains the best available predictor of sperm function (the capacity to fertilize a mature oocyte).***

It was logical to anticipate that if strict sperm morphology could predict fertilization efficiency under optimized conditions *in vitro*, it also might have value for predicting the likelihood of successful fertilization *in vivo* and help to discriminate fertile and infertile men. A number of studies have examined semen parameters in couples with no known infertility factors who were attempting pregnancy,[297, 347] or compared the semen parameters of fertile and infertile men;[296, 316, 340, 348] two have included only men whose partners had no apparent infertility factors.[296, 316] ***Whereas sperm concentration and progressive motility had value***

for distinguishing fertile from infertile men, strict sperm morphology (as determined by an individual having extensive training and experience) was the one most discriminating value.[296, 316] In the larger of the two studies, the likelihood of male infertility was increased approximately 4-fold (OR = 3.8, 95% CI = 3.0–5.0) when strict sperm morphology was less than 9% normal.[296] The 9% threshold value had a sensitivity of 43% and a specificity of 81% for identifying infertile men; lowering the threshold value to 5% normal forms decreased sensitivity to only 19%, but increased specificity to 94%.[296] In a smaller study of similar design, the threshold value that identified infertile men was 10% and the value corresponding to the tenth centile among fertile men was 5% normal forms.[316]

Strict sperm morphology is perhaps most relevant for couples with mild oligospermia or asthenospermia or with unexplained infertility (normal ovulatory function, female reproductive anatomy, and semen parameters). In such couples, IUI (with or without ovarian stimulation) and IVF are the treatment options that offer the greatest likelihood for success (Chapter 27). Most,[349–352] but not all,[353, 354] studies have observed that cycle fecundability in IUI cycles correlates with the proportion of morphologically normal sperm and is generally poor when strict morphology is less than 5% normal. ***Although no threshold value excludes the possibility of pregnancy with expectant management or IUI, the relationship between strict sperm morphology and cycle fecundability with IUI merits careful consideration and discussion when planning treatment for couples with male factor and unexplained infertility.*** Other important considerations include age of the female partner, the duration of infertility, and the comparative costs, logistics, risks, and prognosis associated with alternative treatment strategies, including IVF with and without ICSI.

It is important to emphasize that strict sperm morphology values, like other semen parameters, vary among specimens within individuals, among technologists within laboratories, and among laboratories.[332, 355] A rigorous quality control program helps to ensure accuracy and consistency.[356, 357] Unfortunately, relatively few of the laboratories that perform routine semen analyses have sufficient test volume and the highly trained and experienced personnel required to provide a valid assessment of strict sperm morphology. ***Consequently, earlier WHO standards for sperm morphology (1987, 1992) that classify more sperm as normal are still widely used in most hospital laboratories.***[302, 303] Although morphology assessments using the earlier standards have little value, results for semen volume, sperm concentration, and motility are still informative and can reveal an obvious male factor. However, a more sophisticated semen analysis, including strict sperm morphology, merits serious consideration before implementing treatment for couples with male factor or unexplained infertility.

Round Cells and Leukocytospermia

Epithelial cells, prostate cells, immature sperm (round spermatids, spermatocytes), and leukocytes all appear as “round cells” and cannot be differentiated in a routine semen analysis. ***When the round cell count exceeds 5 million/mL, additional studies should be performed to differentiate leukocytes from immature sperm and to identify those men having true leukocytospermia (greater than 1 million leukocytes/mL) who may require additional evaluation for genital tract infection or inflammation.*** Any of a variety of special stains, biochemical tests, and immunohistochemical techniques can be used to identify the proportion of round cells that is leukocytes.[358, 359] Although leukocytospermia has been implicated as a cause of poor sperm motility and function,[360] more recent studies have failed to demonstrate any association between the leukocytospermia in men with chronic prostatitis and abnormal semen parameters.[361] Nevertheless, documented leukocytospermia generally is regarded as an indication for semen culture (Mycoplasma hominis, ureaplasma

urealyticum, Chlamydia). When cultures are performed, the penis should be washed carefully with betadine before sample collection to reduce the likelihood of contamination from skin flora. For reasons unknown, leukocytospermia unrelated to infection or inflammation also may be observed in the semen of men with spinal cord injuries.[362]

Semen Viscosity

The viscosity of semen is evaluated routinely and graded on an arbitrary scale (grade 0–4). Seminal hyperviscosity is relatively uncommon and its causes have not been clearly defined. Not surprisingly, hyperviscosity has been associated with asthenospermia.[363, 364] Although genital tract infections and sperm autoantibodies have been implicated as causes of seminal hyperviscosity, evidence for the association is lacking.[365] Like abnormalities of pH and fructose levels, increased semen viscosity suggests the possibility of dysfunction in the accessory glands (prostate, seminal vesicles),[366] but in practice, the parameter has relatively little importance.

Specialized Tests

Although all of the major semen parameters (concentration, motility, morphology) have impact on fertility when clearly abnormal, they do not measure or answer what is arguably the most important question: can the sperm effectively attach to, penetrate, and fertilize the partner's ova? Strict sperm morphology is a useful indirect measure of sperm function by virtue of its correlation with fertilization rates *in vitro*, but the parameter leaves much to be desired and generally is available only in specialty andrology laboratories associated with IVF centers.

Unfortunately, although a wide assortment of specialized tests and procedures has been developed to evaluate sperm attachment to the zona pellucida, penetration of the oocyte membrane, or the release of acrosomal enzymes, we still have no reliable validated test of sperm function. Because we do not yet understand, cannot measure, and have no way to correct a suspected sperm function abnormality, attention has focused on ICSI as a way to negate or circumvent sperm function abnormalities. However, the need for a reliable sperm function test persists, because IVF and ICSI are not practical options for a great many infertile couples, and all would like to use their available time and resources in the most efficient and effective way possible[367]; the time and expense associated with treatments involving IUI might be avoided if there was good evidence to indicate that only ICSI offered a realistic likelihood for success.

Sperm Autoantibodies

The blood-testis barrier normally isolates sperm from immune recognition (sperm develop after immunocompetence is established) but if it is disrupted and sperm are exposed to blood, an antigenic response may result. Risk factors for antisperm antibodies include ductal obstruction, previous genital infection, testicular torsion or trauma, and sterilization reversal (vavovasostomy or vasoepididymostomy).[368] Sperm autoantibodies can be found in the serum, but evidence indicates they have no clinical significance.[369, 370] In contrast, antibodies bound to sperm may be clinically relevant because they may interfere with sperm motility or prevent fertilization.[371, 372]

Marked sperm clumping or agglutination, like isolated asthenospermia, may signal the presence of sperm autoantibodies, but neither is observed commonly. Some also regard unexplained infertility as an indication for antisperm antibody testing. The two most widely used tests for detection of sperm autoantibodies involve the use of beads or latex particles with attached antibodies (raised against human immunoglogulins) that bind to antibodies on the surface of sperm.[304] The threshold for a positive test is not well-established, but antibodies generally are considered clinically important when more than 50% of sperm are coated. However, the levels of antibody can fluctuate, even without treatment.[373] New research on sperm proteomics may help to link specific sperm proteins with their functions and to identify relevant sperm autoantibodies.[374]

Pregnancy rates are reportedly lower for men with demonstrable antisperm antibodies than for those without antibodies and, among those with antisperm antibodies, pregnancy rates are lower when more than 50% of sperm are antibody-bound.[375] Antisperm antibodies have been associated with poor postcoital test results, but routine postcoital testing is no longer performed because results have no proven value (Chapter 27). Because IUI was among the most popular and effective treatments for antisperm antibodies[376] (as it was for presumed cervical factor infertility) and IUI has become a core element of most treatments for unexplained infertility other than IVF (Chapter 27), the results of antisperm antibody testing, like those from postcoital testing, rarely offer any information that affects treatment decisions or outcomes. ***Sperm autoantibody testing is seldom any longer performed because when IUI fails or IVF is otherwise indicated, ICSI can effectively circumvent any adverse effects of antisperm antibodies.***[377]

Sperm Penetration Assay

The zona pellucida surrounding the oocyte blocks entry of more than one sperm and fertilization by sperm of a different species, but if removed by gentle enzymatic digestion, sperm of another species can penetrate the egg. In the sperm penetration assay, zona-free eggs collected from superovulated golden hamsters are incubated with washed human sperm and the proportion of eggs penetrated or the number of sperm penetrations per egg by the sperm of the test subject is compared to that observed in a parallel incubation using sperm from a known fertile individual.[304, 378, 379] In theory, the test evaluates four specific sperm functions: capacitation, the acrosome reaction, fusion with the oolemma, and decondensation within the egg cytoplasm.

Unfortunately, the results of the sperm penetration assay are quite sensitive to varying culture conditions and the test procedure has been difficult to standardize. The test relies on spontaneous *in vitro* or chemically-induced acrosome reactions.[380] Test results also vary over time and even proven sperm donors may fail the sperm penetration assay on a given occasion.[381] The predictive value of the sperm penetration assay for IVF or natural conception among infertile couples has varied widely among studies and depends, in large part, on the experience of the individual laboratory.[382–385] Interestingly, test results also have not consistently correlated with strict sperm morphology, the most commonly accepted predictor of fertilization.[386–388] Perhaps most importantly, the test is cumbersome, costly, time-consuming, and not widely available.

Human Zona Binding Assay

Whereas sperm penetration of zona-free eggs may test the ability of sperm to penetrate the oocyte, it does not, by definition, test the ability of sperm to bind to and penetrate the

zona pellucida. The hemizona assay uses bisected zonae derived from human oocytes not previously exposed to sperm and compares the binding of test subject and fertile control sperm.[389, 390] Results have been used to predict fertilization *in vitro*,[391, 392] but the limited availability of human zonae and the technical aspects of the test effectively preclude application beyond use as an investigative tool.

Computer-Assisted Sperm Analysis

Computer-assisted sperm analysis (CASA) was developed in efforts to establish a precise, automated, and objective evaluation of sperm concentration and motion characteristics (velocity and head movement). The technology employs sophisticated instruments to generate digitized video images for analysis, but its accuracy is highly dependent on the methods of sample preparation, frame rate, and sperm concentration.[393, 394] Although some have found that sperm motility characteristics have predictive value for fertilization *in vivo* and *in vitro*, others have not.[395–397]

Acrosome Reaction

The acrosome is a membrane-bound structure located at the tip of the sperm head containing proteolytic enzymes necessary for penetration of the zona pellucida (Chapter 7), and acrosin is one of those enzymes.[398] The acrosome reaction involves the fusion of the acrosome and the plasma membrane, followed by release of the acrosomal enzymes and exposure of the sperm head, which must occur after sperm binding to the zona pellucida. The sperm of infertile men exhibit an increased prevalence of spontaneous acrosome loss and decreased acrosome reactivity in response to treatment with a calcium ionophore.[399] However, the clinical relevance of acrosin measurements and abnormal acrosome reactivity *in vitro* remains to be established.

Biochemical Tests

Biochemical tests of sperm function include measurements of sperm creatine phosphokinase and reactive oxygen species. Creatine phosphokinase is an important enzyme involved in the generation, transport, and use of energy within the sperm. Studies of the levels or forms of the enzyme in the sperm of fertile and infertile men have yielded conflicting results.[400, 401]

Normal oxygen metabolism generates reactive oxygen species, which may be toxic in excess. In both fertile and infertile men, leukocytes are the principal source of reactive oxygen species but sperm themselves also produce them. Increased levels have been observed in the semen of infertile men and implicated as a cause of otherwise unexplained male infertility.[402–404] Peroxidation of sperm lipid membranes and generation of toxic fatty acid peroxides may interfere with sperm functions.[405] Reactive oxygen species can be detected with chemiluminescent probes, but such tests remain investigational.

Sperm Chromatin Structure and DNA

A significant proportion of infertile men have increased levels of DNA damage that may adversely affect fertility[406–411, 407, 408, 412] even when all standard semen parameters are normal.[413]

Men with abnormal semen parameters often exhibit high levels of DNA fragmentation, but the same can be observed in men with normal semen parameters.[406, 410, 414–416] Recently developed tests of sperm chromatin structure and DNA fragmentation provide a measure of sperm chromatin and nuclear integrity,[417] but their clinical utility has not been established. A meta-analysis including 13 relevant studies involving over 2,000 treatment cycles concluded that the small association between sperm DNA integrity test results and pregnancy in IVF and ICSI cycles is not sufficient to warrant their routine use in the evaluation of infertile men.[418]

Endocrine Evaluation

Endocrine disorders involving the hypothalamic-pituitary-testicular axis are well recognized but uncommon causes of male infertility and are extremely uncommon in men having normal semen parameters. ***Indications for endocrine evaluation in infertile men include an abnormal semen analysis (particularly a sperm concentration less than 10 million/mL), sexual dysfunction (decreased libido, impotence), and other clinical symptoms or findings that suggest a specific endocrinopathy.[306] A basic endocrine evaluation of the infertile male involves measurements of serum FSH and total testosterone and will detect the vast majority of clinically significant endocrinopathies.[419]***

When the total testosterone level is low (<300 ng/dL), the assay should be repeated to confirm the finding, and a serum free testosterone, LH, and prolactin should be obtained.[306,420] Together, the levels of FSH, LH, and testosterone help to differentiate the clinical condition. In men with hypogonadotropic hypogonadism, generally all three hormone levels are distinctly low. In men with abnormal spermatogenesis, the FSH level may be normal or high and LH and testosterone levels are normal. Those with testicular failure exhibit high levels of FSH and LH and a low or normal testosterone concentration. Men with a prolactin-secreting pituitary tumor generally have normal or low gonadotropin concentrations, a low serum testosterone, and an elevated prolactin level. ***In those with hypogonadotropic hypogonadism, with or without hyperprolactinemia, magnetic resonance imaging of the hypothalamic-pituitary region is indicated to exclude a mass lesion.***

In infertile men with severe oligospermia (<5 million/mL), low testosterone levels (<300 ng/ dL), and normal gonadotropin concentrations, evaluation might be expanded to include a serum estradiol and calculation of the testosterone (ng/dL):estradiol (pg/mL) ratio, because those with low values (<10) may benefit from treatment with an aromatase inhibitor.[421, 422]

Urologic Evaluation

If not performed earlier, grossly abnormal semen parameters are indication for a thorough physical examination by a urologist or other specialist in male reproduction; some men also may require further urologic evaluation.

In normal men, the testes are firm and measure 15–25 mL in volume.[423] Small soft testes suggest testicular failure. Epididymal fullness suggests obstruction in men with azoospermia.[424] The diagnosis of CBAVD is made by physical examination alone and does not require scrotal sonography or exploration.[307, 308] Palpation of the spermatic cord (erect and supine, with and without valsalva) may reveal a varicocele,[425] which can be graded (grade 1–3) according to severity.[426] Digital rectal examination defines the size and symmetry of

the prostate and may reveal the presence of midline cysts or dilated seminal vesicles suggesting ejaculatory duct obstruction.

Transrectal ultrasonography is indicated for the diagnosis of ejaculatory duct obstruction in men with severe oligospermia or azoospermia, palpable vasa, low-volume ejaculates, and normal testis volume, particularly when the semen is acidic and contains little or no fructose.[306, 427] Vasography offers an alternative method for diagnosis of ejaculatory duct obstruction, but transrectal ultrasonography is less invasive and avoids the risk of vasal injury.[428] Observations of midline cysts, dilated seminal vesicles or ejaculatory ducts suggest, but do not establish, the diagnosis of ejaculatory duct obstruction. Conversely, the absence of any such findings does not exclude the possibility. Seminal vesicle aspiration and vesiculography under transrectal ultrasound guidance provides the means to make a definitive diagnosis; any sperm retrieved can be cryopreserved for use in IVF with ICSI.[429] Definitive treatment requires transurethral resection of the ejaculatory ducts.[430]

Transcrotal ultrasonography can help to clarify uncertain physical findings or to confirm the presence of a scrotal mass. It also may reveal non-palpable varicoceles, but there is no evidence to indicate they have clinical importance.[281]

Renal ultrasonography is indicated for men with unilateral or bilateral vasal agenesis. Approximately 25% of men with unilateral vasal agensis and 10% of men with CBAVD have unilateral renal agenesis.[431]

Testis biopsy may be performed for diagnostic purposes in azoospermic men. Those with elevated serum FSH levels do not require a diagnostic biopsy because a high FSH concentration is diagnostic for abnormal spermatogenesis. Although biopsy may be performed to determine the likelihood that sperm can be retrieved for IVF with ICSI, results may not be all that helpful because sperm production can be limited to specific foci within the testes. In contrast, diagnostic biopsy is indicated for azoospermic men with normal testicular size, at least one palpable vas deferens and a normal serum FSH level, because the normal FSH does not guarantee that spermatogenesis is normal. When biopsy is performed, a portion of testicular tissue can be cryopreserved for use in a future IVF/ICSI treatment cycle to avoid the need for a second procedure. A biopsy that reveals normal spermatogenesis implies obstruction at some level, which then must be defined by surgical exploration with or without vasography (see Surgical Treatment, below).[306]

Genetic Evaluation

Genetic abnormalities may cause infertility by interfering with sperm production or transport. Currently, those most relevant to male infertility and its treatment include 1) mutations within the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene which are highly associated with CBAVD, 2) chromosomal anomalies resulting in testicular dysfunction (Klinefelter syndrome; 47,XXY), and 3) Y chromosome microdeletions associated with abnormalities of spermatogenesis. These conditions have implications that extend beyond their association with azoospermia and severe oligospermia because they can have consequences for the offspring of affected couples. Ideally, genetic counseling should be offered both before and after genetic testing.

Mutations of the CFTR gene are highly associated with CBAVD; almost all men with cystic fibrosis have CBAVD and at least two-thirds of men with CBAVD have a demonstrable CFTR mutation.[432, 433] The gene encodes a protein involved in the formation of the seminal vesicles and the reproductive ductal system in men. Although approximately 4%

of Caucasian men carry a known *CFTR* gene mutation, clinical CBAVD is much less common because penetrance is low in heterozygous individuals.[434] Common speculation, and the prudent clinical assumption, is that virtually all men with CBAVD may have such mutations, some of which have a low carrier frequency and have simply not yet been defined. It is important to note that the spectrum of vasal aplasia includes not only CBAVD, but also unilateral absence of the vas deferens, bilateral partial absence of the vas or epididymides, and epididymal obstruction. ***Men with CBAVD or less severe forms of vasal aplasia, and their female partners, should be screened for CFTR mutations before any attempts at pregnancy via ART to determine the risk for transmitting cystic fibrosis or CBAVD to offspring.***

The overall prevalence of ***chromosomal anomalies*** in infertile men is approximately 7% and inversely related to sperm concentration; the prevalence is highest in azoospermic men (10–15%), lower in oligospermic men (approximately 5%) and very low in men with normal semen quality (less than 1%).[435, 436] By far, the most common chromosomal anomaly in infertile men is Klinefelter syndrome (47,XXY, 46,XY/47,XXY), which accounts for about two-thirds of chromosomal abnormalities in infertile men.[437] Structural chromosomal abnormalities (translocations, inversions) make up the majority of the remainder.[438] The partners of affected men are at increased risk for miscarriage and birth of children with aneuploidy and congenital anomalies. ***Karyotyping should be offered to men with non-obstructive azoospermia or severe oligospermia (less than 5 million/mL) before their sperm are used for IVF with ICSI.***[306] Embryo biopsy and preimplantation genetic diagnosis using fluorescence *in situ* hybridization or other techniques to evaluate chromosomal composition can be used to identify normal embryos suitable for transfer.[439]

Approximately 7% of azoospermic and severely oligospermic infertile men harbor a ***Y chromosome microdeletion*** that cannot be detected with a standard karyotype but can be identified using more sophisticated genetic techniques.[440] Most such microdeletions occur in regions of the long arm of the Y chromosome (Yq11), designated as AZF (azoospermic factor) a (proximal), b (central), and c (distal), which appear to include genes necessary for normal spermatogenesis.[441] Many men with microdeletions in the AZFc region are only severely oligospermic,[442, 443] and those who are azoospermic generally produce sufficient sperm to allow their recovery by testis biopsy. In contrast, the prognosis for sperm recovery in men with microdeletions in the AZFa or AZFb region is very poor.[444–446] Microdeletions in the AZFd region are associated with normal spermatogenesis and their clinical significance is unknown.[447]

Sons of men with Y chromosome microdeletions can be expected to inherit the defect and its clinical consequences.[448–450] Until recently, infertility was the only known clinical consequence of Y microdeletions, but a 1.6-Mb deletion that removes part of the AZFc region (known as the gr/gr deletion) now has been associated with an increased risk for developing testicular germ cell tumors.[451] ***Screening for Y chromosome microdeletions should be offered to all men with non-obstructive azoospermia or severe oligospermia (less than 5 million/mL) who are candidates for IVF with ICSI.***[306]

Medical Treatment for Male Infertility

With a few specific and important exceptions, male infertility generally is not amenable to medical treatment. Careful evaluation can identify those men with treatable conditions who may benefit from medical therapy.

Hypogonadotropic Hypogonadism

Men with hypogonadotropic hypogonadism represent one group in which medical treatment can be successful, after its cause has been defined. Most affected men have a congenital isolated gonadotropin deficiency associated with abnormal puberty, due to absent or abnormal pulsatile hypothalamic GnRH secretion. The endocrinopathy may be genetic in origin, resulting from failure of GnRH neuronal migration during embryogenesis (Kallman syndrome),[452] or idiopathic. When the disorder has onset after puberty, affected men are virilized but hypogonadal, impotent, and azoospermic.

Postpubertal hypogonadotropic hypogonadism is uncommon, but may arise as the consequence of a hypothalamic or pituitary tumor or an inflammatory process (sarcoidosis, hemochromatosis, autoimmune hypophysitis).[453] Pituitary tumors, specifically prolactinomas, are the most common cause. Men with prolactinomas often present with impotence and androgen deficiency. In contrast to the microadenomas commonly identified in women, tumors in men are more often large (macroadenomas). The hypogonadism associated with hypothalamic or pituitary mass lesions may result from hyperprolactinemia and disruption of pulsatile GnRH secretion or from direct effects on the pituitary via compression of pituitary gonadotropes.

Hypogonadotropic hypogonadism due to hyperprolactinemia is generally uncommon in infertile men (approximately 1%)[454] but is amenable to treatment with dopamine agonists when it is identified. Treatment with bromocriptine or cabergoline can effectively restore normal prolactin and testosterone levels and, subsequently, improve libido, potency, semen quality, and fertility in hyperprolactinemic hypogonadal men.[455–458] Increased testosterone levels and potency are observed within approximately 3–6 months after normal prolactin levels are achieved; changes in semen quality generally take longer.[457, 458] Improved semen quality may be expected, but not all men achieve normal semen parameters. In eugonadal infertile men with mild hyperprolactinemia, dopamine agonist treatment promptly restores normal prolactin levels but has little effect on semen quality.[454, 459] Although prolactin levels are modestly higher in infertile than in fertile men, there is no evidence to indicate that dopamine agonist treatment in euprolactinemic men with idiopathic oligospermia or asthenospermia can improve semen quality or fertility.[460]

Because concerns regarding delayed puberty generally arise well before any active interest in fertility, most men with congenital hypogonadotropic hypogonadism are treated first with hCG (to stimulate Leydig cell testosterone production) or exogenous testosterone; either can induce secondary sexual development, but neither can initiate and support normal spermatogenesis.

In contrast, hCG alone (2,000–5,000 IU three times weekly) often can restore meaningful spermatogenesis in men with adult onset of hypogonadotropic hypogonadism.[461, 462] Treatment always should begin with hCG alone (as a substitute for LH), without FSH, for several reasons: 1) hCG stimulates Leydig cells to produce testosterone, resulting in the high intratesticular testosterone concentrations required to stimulate and support spermatogenesis; 2) whereas hCG alone may be sufficient to stimulate spermatogenesis, FSH alone is not[63]; and 3) the annual costs of hCG treatment are substantially lower than the costs of treatment with human menopausal gonadotropins (hMG containing both FSH and LH) or recombinant FSH. During treatment, the serum testosterone concentration should be measured every 1–2 months for the first 3–4 months, aiming for a level between 400 and 900 ng/dL. If not achieved within that interval, dosage should be adjusted accordingly. Some require doses as low as 500 IU and others as much as 10,000 IU. The sperm count also should be monitored at regular intervals. In most, a normal sperm count can be restored within 6 months, but some require treatment for up to 24 months.[463] However, even low sperm concentrations do not preclude fertility.[464]

In men with congenital hypogonadotropic hypogonadism and those with postpubertal onset who do not respond to treatment with hCG alone, normal spermatogenesis can be induced by combined treatment with hCG and hMG or pure FSH (75–150 IU three times weekly).[63,465] Treatment with exogenous testosterone and FSH is ineffective because it cannot generate the highly concentrated intratesticular testosterone concentrations required for normal spermatogenesis.[63] The sperm count should be monitored at least monthly to detect a meaningful trend, because concentrations can fluctuate. As with hCG only treatment, maximum sperm concentrations generally are achieved within 6–24 months. Once spermatogenesis is established by combined treatment with hCG and FSH/hMG, it can be maintained by hCG alone for extended intervals, although semen quality gradually declines again without further FSH treatment.[466]

Men with hypogonadotropic hypogonadism unrelated to hyperprolactinemia or a hypothalamic or pituitary mass lesion also can be treated with exogenous pulsatile GnRH therapy,[467, 468] ***generally administered subcutaneously via a portable programmable pulsatile infusion pump in doses individually titrated to maintain normal adult male serum LH levels.***[469, 470] Treatment with pulsatile exogenous GnRH can successfully restore normal levels of gonadotropin secretion and thereby induce testosterone production and spermatogenesis. Although it is a very specific treatment for men with deficient endogenous GnRH secretion, exogenous pulsatile GnRH therapy is costly, cumbersome, and may require extended periods of time to achieve the desired result. Moreover, GnRH currently is not available in the United States. Normal serum gonadotropin levels can be achieved within little more than a week of treatment and normal serum testosterone concentrations within just a few weeks, but stimulation of spermatogenesis takes considerably longer.[471, 472] Evidence of spermatogenesis may be observed within a year after treatment with pulsatile exogenous GnRH begins, but up to 2 years of therapy may be required to achieve maximum testicular growth, spermatogenesis, and fertility. The best predictors of response are a postpubertal onset of hypogonadotropic hypogonadism, absence of cryptorchidism, and a serum inhibin B concentration greater than 60 pg/mL.[468, 473] Pulsatile GnRH and gonadotropin treatment have comparable efficacy for stimulating spermatogenesis.[474, 475]

Eugonadotropic Hypogonadism

Men with severe oligospermia (<5 million sperm/mL), low serum testosterone levels (<300 ng/dL) and an abnormally low serum testosterone (ng/dL)/estradiol (pg/mL) ratio (<10) may benefit from medical treatment with an aromatase inhibitor. In such men, treatment (testolactone 50–100 mg twice daily, anastrozole 1 mg daily) can normalize ratios and improve semen quality.[421, 422]

Hypergonadotropic Hypogonadism

There is no evidence that any form of medical treatment can improve semen quality and fertility in infertile men with hypergonadotropic hypogonadism. For men with complete spermatogenic failure, the only treatment options are insemination with donor sperm or adoption. For those with severe oligospermia, IVF with ICSI still may offer a realistic chance for success, but preliminary genetic evaluation is strongly recommended, as described earlier.

Retrograde Ejaculation

Men with documented retrograde ejaculation may benefit from medical treatment with sympathomimetics (imipramine 25 mg twice daily or 50 mg at bedtime, pseudoephedrine 60 mg or ephedrine 25–50 mg four times daily, phenylpropanolamine 50–75 mg twice daily), directed at control of the internal sphincter. Alternatively, sperm can be recovered directly from the bladder after masturbation; for best results, urine pH and osmolality (300–380 mOsm/L) must be carefully controlled by alkalinizing the urine (sodium bicarbonate 650 mg four times daily, beginning 1–2 days before collection) and managing fluid intake.[476, 477] When such efforts prove cumbersome or ineffective, the bladder can be drained and filled with buffered medium (approximately 100 mL) immediately before ejaculation. In men with ejaculatory failure, electroejaculation may be required.[478] If sufficient numbers of motile sperm can be recovered, IUI may be performed, and if not, IVF and ICSI may be necessary.

Leukocytospermia

Leukocytospermia has been associated with other abnormal semen parameters[479] and antibiotic treatment (doxycycline, erythromycin, trimethoprim-sulfamethoxazole, or a quinolone) clearly is indicated for men with symptomatic genital tract infections. ***However, antibiotic treatment does not improve semen parameters in men with asymptomatic leukocytospermia[480] and often fails even to decrease the numbers of leukocytes to normal levels (less than 1 million/mL).***[481, 482] Leukocytospermia often is episodic and does not accurately predict genital tract infection.[483] Moreover, there is little evidence it has any adverse effects on fertility.[484] Treatment therefore probably is best limited to men with documented genital tract infections.

Idiopathic Male Infertility

Most infertile men are eugogonadotropic, normally virilized, and otherwise healthy, but have low sperm density or other semen abnormalities for which the cause is unknown. Idiopathic male subfertility is common and a wide assortment of empiric medical treatments has been described; androgens, gonadotropins, and antiestrogens have received the most attention. ***Unfortunately, no medical treatment has proven reliably effective for improving semen parameters or fertility in men with idiopathic subfertility.***

Androgen therapy has been advocated as a means to stimulate spermatogenesis, directly by increasing intratesticular androgen concentrations, and indirectly via a "rebound" increase in pituitary gonadotropin secretion after an interval of androgen-induced suppression. However, the results of a meta-analysis of 11 randomized clinical trials involving almost 1,000 men indicate that neither treatment strategy reliably improves semen parameters or fertility.[485] There is no substantial evidence that androgen therapy is effective treatment for idiopathic male infertility.[486]

Results of studies involving the use of ***exogenous FSH*** to stimulate spermatogenesis directly have been conflicting. Whereas two randomized trials in subfertile men found no evidence that such treatment improves semen quality or fertility,[487, 488] others suggest that exogenous FSH may improve semen quality in a subset of men with idiopathic oligospermia in whom testicular biopsy reveals maturation arrest and serum FSH and inhibin B levels are normal.[489–491]

Empiric treatment (3–6 months) with either ***clomiphene citrate*** (25 mg daily) or ***tamoxifen*** (20 mg daily) commonly is offered to stimulate increased pituitary gonadotropin secretion and spermatogenesis in men with idiopathic subfertility. The results of numerous studies are inconsistent. Whereas treatment appears to benefit some men, there is no reliable method for identifying those who might respond. ***Overall, antiestrogen treatment is not effective***. A randomized clinical trial conducted by the World Health Organization involving nearly 200 men and over 1,300 couple-months of observation found no differences among men treated with clomiphene or placebo.[492] Moreover, a meta-analysis including 10 randomized trials involving over 700 men concluded that evidence is insufficient to indicate that antiestrogen treatment improves semen quality or male fertility.[493]

Intrauterine Insemination

Artificial insemination has been used to treat infertile couples for almost 200 years and is an accepted form of treatment for men with severe hypospadias, retrograde ejaculation, neurologic impotence, and sexual dysfunction. Artificial insemination also has been used as a means to overcome oligospermia, asthenospermia, low ejaculate volumes, sperm autoantibodies, and cervical factors. Therapeutic insemination using donor sperm is an established and highly effective treatment for severe and uncorrectable male factor infertility, inherited genetic disorders in the male partner, and single or lesbian women who desire pregnancy. Before the advent of IVF and ICSI, therapeutic donor insemination was the only viable treatment option for couples with severe male factor infertility, and it remains highly effective when ART is rejected or fails.[494]

Artificial insemination may be performed by depositing sperm into the cervical os or directly into the uterus, but IUI is now almost universally performed, for several reasons. First, when trying to overcome the limitations of decreased sperm density or motility in the treatment of male factor infertility, cervical insemination offers no significant advantage over what can be achieved by intercourse. Second, whereas the potential for reactions to the proteins, prostaglandins, and bacteria in semen severely limits the volume of untreated semen (and thus the numbers of sperm) that can be delivered to the upper female genital tract, IUI with a "washed" sperm concentrate (devoid of seminal plasma) delivers most of the sperm in an ejaculate. ***Most importantly, IUI yields substantially better overall results than cervical insemination***. In one meta-analysis including 12 separate studies involving nearly 700 women and over 2,000 insemination cycles, the overall pregnancy rate per cycle was 18% for women receiving IUI, compared to 5% for women who received cervical insemination; considering only the 10 studies in which frozen donor sperm were used, the pregnancy rate per cycle with IUI was more than twice that of cervical insemination (OR = 2.63, 95% CI = 1.85–3.73).[495] An earlier analysis including seven studies yielded similar results (OR = 2.4, 95% CI = 1.5–3.8).[496]

It is difficult to gauge the effectiveness of IUI using the sperm of infertile men because almost all of the many published series examining IUI cycle outcomes have included couples with a variety of infertility factors and have employed combined treatment with IUI and empiric ovarian stimulation. There are ample data from retrospective studies of outcomes in therapeutic donor insemination cycles (IUI with or without ovarian stimulation), but the results achieved using infertile partner sperm cannot be expected to equal those using normal donor sperm. ***Considering all of the relevant variables, the available data suggest that cycle fecundity ranges between 3% and 10% when IUI is performed using infertile partner sperm,[325, 497–500] and is approximately three times higher (9–30%) when donor sperm are used.[500–506]***

Regardless whether infertile partner sperm or frozen donor sperm are used, the methods for sperm preparation, the timing and technique of IUI, and the influence of other infertility factors on prognosis are largely the same. The number, motility, and morphology of frozen donor sperm generally are not limiting because sperm donors are highly selected for their semen quality, but semen parameters definitely do affect the prognosis for success with IUI using infertile partner sperm.

Semen Parameters and Prognosis

Not surprisingly, the likelihood for success with IUI using infertile partner sperm depends, to some extent, on the severity of the semen abnormality. Sperm density, motility, and morphology all have influence on success rates.

The probability of successful IUI increases with the number of total motile sperm inseminated. ***Best results are achieved when the number of total motile sperm exceeds a threshold of approximately 10 million.***[325, 327, 497] Higher counts do not further increase the likelihood for success[497, 507] and IUI is very seldom successful when fewer than 1 million total motile sperm are inseminated.[498, 508] Combining the yield from two ejaculates obtained approximately 4 hours apart may increase the numbers of sperm available from oligospermic men.[509]

As might be expected, considering that strict sperm morphology has some predictive value for successful conventional fertilization *in vitro*,[331, 332] the percentage of morphologically normal sperm (strict criteria)[304] appears to have similar predictive value for IUI. Numerous studies have examined the correlation between strict sperm morphology and IUI cycle outcomes. Most,[351, 510, 349, 350, 352] but not all,[353, 354] have found a strong relationship between the two. Like the results observed in IVF cycles, the probability for success with IUI rises with the percentage of morphologically normal sperm. ***Success rates with IUI are highest when 14% or more of the sperm have normal morphology, intermediate with values between 4% and 14%, and generally quite poor when fewer than 4% of sperm are normal.***[349] In general, therefore, couples with male factor infertility involving severe teratospermia (<4% normal sperm) may be best advised to apply their available resources to IVF and ICSI when that is possible.

Other Prognostic Factors

Needless to say, the prognosis for success with IUI in the treatment of male factor infertility is best when there are no other coexistent infertility factors. Most specifically, the prognosis is greatly influenced by the age of the female partner, the consistency and quality of her ovulatory function, and the condition of her reproductive anatomy. The extent to which these additional factors should be evaluated before treatment with IUI begins must be individualized.

Maternal Age

Maternal age is a key variable in all infertile couples. Even when donor sperm are used, the proabability for success declines progressively with increasing maternal age.[502, 505, 511–513] Cycle fecundability and cumulative pregnancy rates in women under age 35 inseminated with donor sperm (0.20 per cycle, 88% after up to seven cycles) equal those observed in normal fertile couples but are lower for women between ages 35 and 40 (0.12 per cycle,

65%) and those over age 40 (0.06 per month, 42%).[514] Ovarian reserve testing (Chapter 27) merits consideration when the female partner is over age 35, has a family history of early menopause, previous ovarian surgery, chemotherapy, or radiation, and when she is a smoker or previously has responded poorly to exogenous gonadotropin stimulation. Women with a poor ovarian reserve have a significantly reduced probability for success with IVF and, by inference, likely have a relatively poor chance for success with IUI.

Ovulatory Function

At a minimum, it is certainly prudent to assess ovulatory function by some objective means when the treatment plan does not include empiric ovarian stimulation (discussed below). Ovulatory disorders are common, even in women seeking therapeutic donor insemination,[512, 514–516] Ovulation induction increases success rates with therapeutic donor insemination for women with ovulatory dysfunction,[515] although cycle fecundability remains lower than in women with spontaneous ovulatory cycles.[511, 517]

Uterine and Tubal Factors

Hysterosalpingography (HSG) is recommended for women over age 35 and when the medical history or physical examination raises suspicion of endometriosis or uterine or tubal infertility factors because IUI is less likely to succeed in couples with combined male factor and tubal factor infertility; IVF, with or without ICSI, usually a is better treatment option. In the absence of such suspicions, the likelihood of abnormal HSG results is quite low.[518] If not performed before treatment begins, HSG is recommended for women who fail to conceive after 4–6 therapeutic donor insemination cycles. Laparoscopy and hysteroscopy are unnecessary for most women, but appropriate for those with an abnormal HSG or signs or symptoms of advanced pelvic disease.

Empiric Ovarian Stimulation

Empiric ovarian stimulation with clomiphene citrate or exogenous gonadotropins is commonly combined with IUI in the treatment of couples with male factor infertilty, based on observations that cycle fecundability (probability of pregnancy per cycle) is higher after combined treatment than after IUI or ovarian stimulation alone in couples with unexplained infertility.[499, 519, 520] Although the value added by ovarian stimulation when IUI is performed using infertile partner sperm is unproven, data derived from large case series of therapeutic donor insemination cycles provide some useful insight.

The cycle fecundability observed in spontaneous and clomiphene-stimulated therapeutic donor insemination cycles is similar (6–13%),[502, 521, 522] suggesting that clomiphene stimulation has little or no added value. It is possible, but unproven, that clomiphene treatment might have benefits limited to cycles in which it achieves multifollicular development and ovulation. In contrast, exogneous gonadotropin stimulation appears to increase cycle fecundability in therapeutic donor insemination cycles approximately two-fold (14–24%).[502, 521–523] However, the risks (multiple ovulation, ovarian hyperstimulation), costs, and logistical demands associated with gonadtropin treatment are also substantially higher.

The outcomes observed in large case series strongly suggest that exogenous gonadotropin stimulation increases cycle fecundability in therapeutic donor insemination cycles,

particularly because combined treatment generally has been added only after inseminations in spontaneous cycles proved unsuccessful. Equivalent or better outcomes have been observed after gonadotropin stimulation even though comparisons were biased against combined treatment because the selected population already had demonstrated intrinsically lower fertility.[502, 521, 522] In one randomized trial, cycle fecundability in gonadotropin-stimulated cycles (14%) was more than twice that observed in clomiphene-stimulated cycles (6%).[523] By inference, gonadotropin stimulation also might be expected to improve cycle fecundability when IUI is performed using infertile partner sperm. However, because the poorer quality of infertile partner sperm may be the limiting factor, gonadotropin stimulation may have less value than in therapeutic donor insemination cycles.

When ovulatory function is normal, treatment with IUI alone is reasonable and appropriate. When IUI in spontaneous cycles or indicated clomiphene-induced cycles fails (approximately 3–4 cycles) or when the female partner is over age 35, exogenous gonadotropin stimulation may be expected to improve the likelihood for success.

Sperm Preparation

There are a variety of methods for extracting sperm from the seminal plasma for IUI. The most common methods include conventional washing, the "swim-up" procedure, and density gradient centrifugation. The best choice among them may vary with the quality of the semen sample.[524, 525] The results of a randomized study comparing the pregnancy rates achieved with IUI after a variety of sperm preparation methods suggest that swim-up and density gradient centrifugation may offer a greater chance for success than conventional sperm washing.[524] Another study found that density gradient centrifugation yielded better results than conventional washing when the insemination specimen contains less than approximately 20 million sperm.[525] However, a recent meta-analysis including five trials involving over 250 couples and comparing three techniques concluded that evidence is insufficient to recommend any specific preparation technique.[526] Results achieved with IUI using cryopreserved donor sperm are comparable, regardless whether the sperm are prepared before freezing or after thawing.[503]

Both the conventional washing and swim-up methods allow sperm to remain in contact with dead or defective sperm and leukocytes, which produce high levels of reactive oxygen species that may cause oxidative damage to sperm membranes and motility.[527, 528] ***Whereas methods more sophisticated than conventional washing or swim-up may be used to prepare sperm (density gradient centrifugation, glass wool filtration, others), and often are when preparing sperm for IVF,[529] they generally are not required for IUI.***

Washing

The simplest method of washing sperm involves diluting the liquefied semen sample in buffered medium (available from a number of commercial suppliers) in a sterile tube (1:1–1:3, depending on volume), followed by low speed centrifugation (200–300 *g* for approximately 10 minutes) and removal of the supernatant. After two or more cycles, the final pellet is resuspended in a small volume (approximately 0.5 mL) of medium for insemination. Sperm washing yields the greatest numbers of sperm, but the final specimen also contains dead and abnormal sperm and other cellular debris. When sperm viability or motility is abnormally low or the round cell concentration in the semen is abnormally high, methods to exclude them from the insemination specimen deserve consideration.

Swim-Up

The swim-up method for preparing sperm adds another step to the washing process. The final pellet is gently overlaid with 0.5–1.0 mL of fresh medium and incubated at 37° C for 30–60 minutes, allowing the most motile sperm to swim up into the supernatant.[530] The method generates a cleaner specimen, devoid of dead sperm and other cellular debris, but also yields significantly lower numbers of sperm (albeit with high motility) and therefore may be inadvisable when the sperm concentration is already very low.

Density Gradient Centrifugation

The typical methodology for density gradient centrifugation involves overlaying the liquefied ejaculate on a column of higher density media that are layered to create a gradient of increasing density from the top to the bottom of the column, followed by low speed centrifugation for 15–30 minutes.[531] The most highly motile sperm traverse the gradient more rapidly and can be recovered from the soft pellet at the bottom. The method also appears to select a population of sperm with normal morphology.[532, 533] As with the swim-up procedure, the sperm yield is substantially lower than with conventional washing.

Timing and Technique

For obvious reasons and for best results, IUI should be timed to coincide with the time of spontaneous or induced ovulation. Normal sperm can survive in the female reproductive tract and retain the ability to fertilize an egg for at least 3 days, but an oocyte can be successfully fertilized for only approximately 12–24 hours after it is released.[534] In normal fertile couples, the probability of conception rises progressively over an interval of 5–6 days and peaks when intercourse occurs on the day before or day of ovulation.[82, 535, 536] The longevity of normal sperm in the female genital tract relates, in part, to their retention within the cervical mucus which, of course, is bypassed by IUI. Although unproven, there is reason to believe that sperm may have a significantly shorter functional lifespan after IUI. Logically, the lower numbers and motility of infertile partner sperm may be even more limiting. Cryopreservation damages sperm[537] and even frozen-thawed donor sperm lose viability and motility more rapidly than fresh normal sperm. ***The timing of IUI in the treatment of male factor infertility is therefore far more critical for success than the timing of natural intercourse in infertile couples, regardless whether infertile partner sperm or frozen donor sperm are used.***

The various methods that may be used to detect ovulation and to ensure that IUI is optimally timed are described at length elsewhere in this text (Chapter 27); only the most commonly used methods and their relationship to the time of ovulation are again briefly summarized here. Ovulation generally may be expected to occur on the day before the midcycle rise in basal body temperature (BBT)[82, 536] or 14–26 hours after the urinary LH surge is first detected.[538, 539] ***In natural and clomiphene-stimulated cycles, the most practical and reliable method for timing IUI involves urinary LH monitoring beginning approximately 3 days before expected ovulation and insemination on the day following detection of the LH surge. When ovulation is triggered by injection of exogenous hCG in natural or stimulated cycles, IUI generally is best performed approximately 34–40 hours later.***

Immediately before performing IUI, removal of any excess mucus that might clog the catheter tip is recommended. The tip of the insemination catheter is then simply inserted into the cervical os and advanced slowly into the uterine cavity. A large variety of specialized catheters having varying rigidity is readily available from commercial sources and any may be used. Designs involving a stiffer moldable outer sheath over a more atraumatic and flexible inner catheter are the most versatile. The insemination specimen (approximately 0.5 mL) should be introduced slowly over 10–30 seconds. Although there are no data to indicate that it matters, it is customary to have the patient remain supine for approximately 10–15 minutes after insemination.

Although some have suggested that two inseminations (12 and 34 hours after hCG-induced ovulation) yield a higher cycle fecundability than a single IUI,[540] other similarly designed studies have found no such advantage.[541–543] A meta-analysis including three randomized controlled parallel trials involving nearly 400 couples concluded that available data do not allow a confident conclusion.[544] Two studies of cycle fecundability after therapeutic donor inseminations have observed that two inseminations are no more effective than one.[545, 546]

Most women who pursue therapeutic donor insemination are otherwise fertile and conceive within 4–6 insemination cycles; cycle fecundability declines by half to two-thirds thereafter.[514, 547–549] Cumulative conception rates after up to 12 insemination cycles reach 75–80%,[502, 504, 514] but are approximately 50% lower for those having other infertility factors.[512] Combined with gonadotropin stimulation, donor insemination succeeds in more than half of treated couples achieving superovulation after up to three cycles.[549] As expected, success rates for insemination with infertile partner sperm are significantly lower, but may still approach 30% after up to six treatment cycles.[498] The number of treatment cycles offered must consider the influence of female partner age, coexisting infertility factors, the duration of infertility, the quality of the insemination specimen, and the number of mature preovulatory follicles when ovarian stimulation also is used.

Donor Sperm

In general, commercial and university-based sperm banks recruit healthy young donors having desirable general physical characteristics and consistently outstanding semen quality. However, it is important to understand that whereas sperm banks generally adhere to the guidelines established by the American Society for Reproductive Medicine,[550] they remain, at present, self-regulated. The choice of a sperm bank should therefore consider whether it has formally adopted the established guidelines.

Current guidelines require extensive screening of prospective sperm donors before acceptance. Semen quality, to include an evaluation of sperm viability and motility after a trial freeze and thaw,[537, 551] excludes approximately 75% of all candidates. Personal health history and physical examination, family medical history, genetic screening for cystic fibrosis and other carrier states (depending on ethnicity), and screening for sexually-transmitted infections (syphilis, gonorrhea, Chlamydia, cytomegalovirus, hepatitis B and C, human immunodeficiency virus [HIV] types I and II, and human T-lymphocytic virus [HTLV] types I and II) exclude another 5–10% of candidates. Sperm donors must be screened repeatedly for sexually-transmitted infections at intervals, generally every 6 months. Sperm banking practices changed forever in 1985 after documentation of HIV seroconversion in four of eight women inseminated with cryopreserved sperm from an asymptomatic HIV (then called HTLV-III) carrier.[552] ***Now, sperm specimens must be quarantined and cannot be released for use unless they have remained sequestered for at least the***

180 days preceding the most recent negative test for HIV. Even with rigorous adherence to current guidelines, human semen can never be regarded as completely safe. Although perhaps remote, the possibility remains that frozen donor sperm specimens may contain other microorganisms not generally considered as sexually-transmitted infections or as yet unknown viruses.

Ideally, donor-specific cycle fecundability rates, which may vary significantly, also would be defined after a reasonable number of inseminations.[551, 553] However, such information generally is not available, primarily because outcomes of donor inseminations are difficult to track accurately and vary substantially among recipients, depending on age and the presence or absence of other infertility factors. In the absence of such information, it is reasonable to select an alternative donor after four to six unsuccessful insemination cycles when there are no other coexisting infertility factors.

Although the quality of frozen donor sperm specimens generally is reliable, it should not be assumed. Frozen donor sperm specimens typically do not include all of the sperm in an ejaculate; a number of aliquots are prepared from each sample, depending on its quality. One study found a 10-fold variation in the number of motile sperm in random specimens obtained from seven different commercial sperm banks (4.3–39 million).[554] As when using infertile partner sperm, the likelihood of success with therapeutic donor insemination increases with the number of motile sperm in the specimen and is greatest when the count exceeds 20 million.[514] Most sperm banks guarantee a minimum number of motile sperm in each specimen, but that guarantee is not always met and only implies a refund of the purchase price when it is not. Consequently, it is prudent to determine the motile sperm count after thaw of donor sperm samples and to seek an alternative source when quality consistently falls below a reasonable minimum standard (10 million total motile sperm).

Surgical Treatment for Male Infertility

Although IVF with ICSI now provides the means to treat even the most severe forms of male factor infertility, including irreparable reproductive tract obstruction and non-obstructive azoospermia, the associated costs and risks are substantial. For men with obstructive azoospermia or a varicocele, specific surgical treatment offers a viable alternative, but proper patient selection is key.[555–558]

Vasovasostomy and Vasoepididymostomy

About one-half million American men undergo vasectomy every year and approximately 2–6% of vasectomized men later seek reversal of their sterilization procedure. Obstructive azoospermia also may result from iatrogenic injuries to the vas deferens, usually during hernia repair.[283]

In most vasectomized men, microsurgical vasovasostomy or vasoepididymostomy can restore patency of the ductal system and return sperm to the ejaculate. When microscopic examination of the fluid in the testicular end of the vas reveals no sperm even after lavage, vasoepididymostomy can be performed.[559] When sperm are found in the vasal fluid on at least one side, microsurgical vasovasostomy returns sperm to the ejaculate in

nearly all vasectomized men;[560] late obstruction after initial patency may be observed in up to 12% of men.[561] ***Over 2 years or more after vasovasostomy, pregnancy rates in the range of 50–60% may be expected, depending on whether other infertility factors also must be overcome.[562] The likelihood of pregnancy decreases modestly with time since vasectomy, but not dramatically; for most, surgery offers comparable or better results than can be achieved with IVF and ICSI.[563, 564]*** The results of re-operation may approach the same outcomes when the first attempt is technically unsuccessful.[565] Compared to vasovasostomy, vasoepididymostomy is less often successful, with patency rates between 50% and 85%, pregnancy rates between 40% and 50%, and a greater likelihood of re-occlusion, depending on the site of anastamosis.[555, 566] Newly described tubal intussusception techniques have simplified the procedure and yielded excellent patency rates.[567, 568] Cryopreservation of sperm collected at the time of vasectomy reversal offers those with failed procedures the opportunity to pursue pregnancy by IVF and ICSI without further intervention.[566, 569]

Transurethral Resection of the Ejaculatory Ducts

Ejaculatory duct obstruction is a cause of infertility in 1–5% of infertile men,[570] and should be suspected in men with normal, palpable vasa deferentia and semen analyses revealing low ejaculate volumes combined with low or normal sperm concentration and low or absent motility.[571] The condition also may present as hemospermia and painful ejaculation. Ejaculatory duct obstruction usually is congenital but also can result from chronic prostatitis or compression by prostate or seminal vesicle duct cysts, from duct calcification or blockage due to post-infectious or postoperative scar, and may be amenable to correction by transurethral resection,[572, 573]

Methods for evaluating suspected ejaculatory duct obstruction include trans-scrotal vasography, (antegrade or retrograde injection of contrast medium into the vas deferens or seminal vesicle), transrectal ultrasonography (to detect an enlarged seminal vesicle), and seminal vesicle sperm aspiration and ejaculatory duct chromotubation. A study comparing the accuracy of the methods found that diagnosis by trans-rectal ultrasonography was confirmed at surgery in fewer than half of cases and concluded that dynamic tests (vasography, chromotubation) can decrease unnecessary duct resection procedures and improve the outcomes of those that are indicated.[574]

Transrectal ultrasound-guided aspiration of cystic or dilated ejaculatory ducts or seminal vesicles and microscopic examination of the aspirate may yield sperm than can be cryopreserved.[429] Subsequent introduction of indigo carmine dye diluted in radiographic contrast and X-ray provides the means to define the lesion and to confirm successful surgical resection. When no sperm are found, vasography may be performed, and if ejaculatory duct obstruction is confirmed, a coincident epididymal obstruction is likely. Under such circumstances, microsurgical epididymal sperm aspiration or testis biopsy and cryopreservation of sperm for IVF and ICSI generally is a better option than simultaneous epididymovasostomy and transurethreal resection of the ejaculatory ducts.

Transurethral resection of an ejaculatory duct obstruction results in increased semen volume in approximately two-thirds of affected men and returns sperm to the ejaculate in about half of azoospermic men. Results are better in men with midline cysts and men with partial obstruction than in those with complete obstructions. Whereas IVF and ICSI is an obvious alternative to transurethral resection, successful surgery can allow many men to conceive naturally or by IUI without need for ART.[575]

Varicocele Repair

The prevalence of varicoceles is approximately 10–15% in the normal male population and about 25–40% in infertile men.[306, 426] The weight of available evidence indicates that varicoceles have an adverse effect on spermatogenesis. The pathophysiology involved is unclear but widely believed to involve venous reflux and increased testicular temperatures because spermatogenesis is exquisitely temperature sensitive. ***Since only palpable varicoceles have any documented association with infertility, other means of diagnosis (scrotal ultrasonography, thermography, Doppler ultrasonography, radionuclide scanning, and spermatic venography) generally are not indicated for infertile men with no palpable varicocele.*** Scrotal ultrasonography may be useful when physical examination is inconclusive and spermatic venography can help to better define the location of refluxing spermatic veins that recur or persist after repair.[306]

Varicocele repair is indicated primarily for men with palpable varicoceles and abnormal semen parameters having either a partner with normal fertility or treatable infertility or an interest in future fertility. Adolescent males with unilateral or bilateral varicoceles associated with decreased testicular size also may be candidates for varicocelectomy; those with normal testicular size should be followed carefully to detect any decrease in testicular size or semen quality.[576, 577] Similarly, young adult men with palpable varicoceles and normal semen may be at risk for progressive testicular dysfunction and should be monitored to detect any evidence of decreasing semen quality.[425, 578, 579]

Treatment options for men with abnormal semen quality associated with a palpable varicocele include surgical repair, IUI, and IVF with or without ICSI. The best choice among these options depends on the age of the female partner and presence of other infertility factors. Varicocelectomy offers the potential advantages of a permanent cure and natural conception.[557] Even when there are other clear female partner indications for IVF, varicocele repair may deserve consideration because surgery can restore sperm to the ejaculate in some men with non-obstructive azoospermia.[580, 581]

Varicoceles can be repaired with a variety of surgical approaches (retroperitoneal, inguinal, subinguinal, laparoscopic) or by percutaneous embolization. No one method has proven clearly superior. Most male reproductive specialists prefer microsurgical inguinal or subinguinal repair.[582] Percutaneous embolization of varicoceles requires expertise in interventional radiologic techniques and is not universally applicable. Surgical treatment corrects over 90% of varicoceles; results achieved with embolization are more variable.

Semen quality often improves after varicocele repair[583] and men with large varicoceles generally realize the greatest improvement.[584] ***However, the results achieved with varicocele repair have varied widely and convincing evidence for improved fertility is still lacking.***[557, 585–587] The results of two randomized trials in men with palpable varicoceles, abnormal semen parameters, and normal female partners are perhaps the most informative. In one, 60% of men who underwent surgical repair achieved pregnancy with their partner during the first postoperative year, compared to only 10% of untreated controls; after surgical repair of varicoceles in the still infertile men not initially treated, more than 40% achieved pregnancy during the following year.[588] In the second trial, men who underwent varicocele repair had improved semen parameters compared to untreated controls, but were no more likely to achieve pregnancy.[589] ***In general, the best candidates for varicocele repair are young men with large varicoceles and infertility of relatively short duration. Atrophic testes, elevated FSH levels, and severe oligospermia or azoospermia indicate severe epithelial damage and are associated with a poor prognosis after varicocele repair.***

One other potential benefit of varicocelectomy deserves mention. Even when varicocele repair is not followed by natural conception, improvement in semen parameters may be sufficient to allow IUI when IVF would otherwise be necessary, or IVF with conventional fertilization rather than with ICSI.[590]

Orchiopexy

Cryptotorchidism is associated with a high incidence of infertility even when it is unilateral; when both testes are undescended, azoospermia is all but certain. Occasionally, an undescended testis will escape detection until adulthood; if the contralateral testis is normal, fertility may be preserved. Even in adult men with bilateral cryptorchidism, orchiopexy can result in spermatogenesis and fertility; at the least, it preserves testicular hormone production.[591]

Vibratory Stimulation and Electroejaculation

Men with neurological conditions affecting the sympathetic system frequently have dysfunctional or absent emission. Examples include men with spinal cord injuries, demyelinating neuropathies, diabetes, and those who have had retroperitoneal lymph node dissections. In most, ejaculation can be achieved with vibratory stimulation, and in those who don't respond, electroejaculation can be used to obtain motile sperm for IUI or IVF and ICSI.[478, 592] Because ejaculation may be retrograde, additional techniques for recovery of sperm from the bladder may be required (described above).

Assisted Reproductive Technologies

IVF and ICSI have revolutionized the treatment of male infertility. As it was originally performed, IVF involved insemination of each oocyte with 2–6 million sperm; consequently, the method had only limited application when the male was severely oligospermic. With refinements in technique over time, the number of motile sperm used for insemination decreased to 50–100 thousand per oocyte, opening the door to wider application of ART in couples with male factor infertility. The advent of ICSI further expanded capabilities to overcome even the most severe forms of male infertility.[593, 594] Now, a male factor is the single one most common diagnosis among couples who undergo IVF. In the U.S. national summary of ART success rates for the year 2006, 18% of all cycles were performed for male factor indications and a male factor was one of multiple infertility factors in another 18% of cycles.[595] In 62% of all cycles involving fresh non-donor oocytes, ICSI was performed. ***Overall, the results achieved with IVF in couples with male factor infertility, with and without ICSI, are comparable to those observed in couples with other indications for IVF.***[595]

Sperm Retrieval

Although ICSI is now applied rather liberally in IVF cycles, even in couples without male factor infertility, it is most specifically indicated in couples with severe male factor

infertility where poor or failed fertilization is more likely. When few or no viable sperm can be recovered from the ejaculate, a variety of sperm retrieval techniques may be used to obtain sperm for IVF and ICSI. Even when substantial numbers of sperm are retrieved, ICSI is prudent because sperm obtained from chronically obstructed reproductive systems usually exhibit poor motility and decreased fertilizing capacity.

Epididymal Sperm Aspiration

Sperm may be obtained by microsurgical epididymal sperm aspiration at the time of vasoepididymostomy or as an isolated procedure in men with CBAVD or uncorrectable obstructions. The technique involves incision of an isolated dilated tubule, gradually moving more proximally, if necessary, until sperm are obtained.[596, 597] Sperm are collected into a micropipette by capillary action with gentle compression of the testis and epididymis and flushed into a container with a small volume of IVF culture medium. Recovered sperm are cryopreserved in multiple aliquots for use in IVF cycles, if required.[598]

Percutaneous epididymal sperm aspiration using a fine needle has also been used successfully to obtain sperm and achieve pregnancy,[599, 600] but the technique is less reliable, the small quantities of sperm obtained are sometimes inadequate to allow cryopreservation, and pregnancy rates achieved have generally been lower than with the open technique.

Testicular Sperm Extraction and Aspiration

In men with non-obstructive azoospermia and those in whom epididymal sperm aspiration techniques fail or are inapplicable, sperm may be retrieved directly from the testis.[601, 602] Open microsurgical testicular sperm extraction yields the greatest number of sperm with potential for cryopreservation. Percutaneous core biopsy or aspiration of the testis have also been described and are most applicable in men with normal spermatogenesis and obstructive azoospermia.[603]

Using the preferred open microsurgical technique, sperm can be retrieved from the majority of men. Magnification minimizes the risk of injury to the testicular blood supply, increases the probability of retrieving a blood-free biopsy specimen, and allows identification of larger caliber tubules that are more likely to yield sperm.[604, 605] Normal pregnancies have been achieved even in those with congenital or acquired testicular failure,[606–608] post-chemotherapy azoospermia,[609–611] and Klinefelter syndrome.[439, 612]

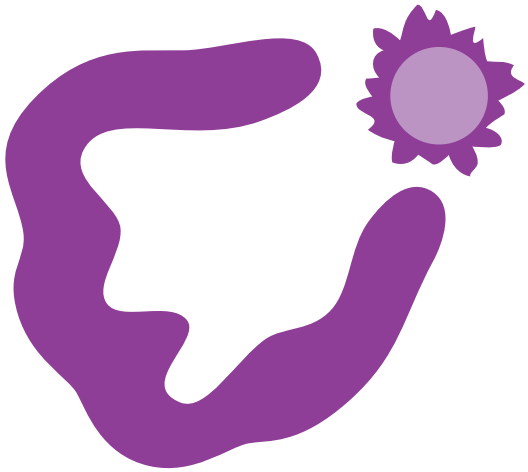
Genetic Risks Associated with ICSI

Because ICSI may override natural safeguards that serve to prevent fertilization by sperm with abnormal or damaged DNA, there is reason for concern that children born after ICSI might be at increased risk for chromosomal and other major congenital anomalies,[94, 613, 614] cancers,[615–617] or infertility.[448–450, 615] Most studies,[618–621] but not all,[613] have failed to identify any increased incidence of major congenital malformation among children born after ICSI (beyond that associated with conventional IVF),[621, 622] perhaps, at least in part, because

embryos derived from damaged paternal DNA are less likely to implant and naturally select out.[409, 417] One specific malformation that may be more prevalent in children born after ICSI is hypospadias, possibly due to paternal infertility.[623] ***Regardless, karyotyping and Y chromosome deletion analysis should be offered to all men with severe male factor infertility who are candidates for IVF with ICSI and additional studies are clearly needed to determine what if any increased risks might be imposed on children born after ICSI.***[624]

All references are available online at: http://www.clinicalgynendoandinfertility.com

31 Induction of Ovulation



Although it seems commonplace today, indeed even routine, the ability to induce ovulation and attain pregnancy in anovulatory infertile women remains one of the greatest achievements of reproductive endocrinology. Once limited to clomiphene citrate, the therapeutic armamentarium for ovulation induction now includes a wide variety of agents.

Ovulatory disorders can be identified in 18–25% of infertile women.[1] When anovulation is the only infertility factor, the prognosis for pregnancy generally is quite good because modern ovulation induction strategies are highly effective. When a specific cause for anovulation can be identified, treatment often restores normal cycle fecundity. Even when no specific cause can be found, as in most anovulatory women, empiric treatments with low costs and risks usually succeed. When those fail, other more complex forms of treatment are effective. One way or another, almost all anovulatory infertile women can be induced to ovulate. Unfortunately, many still do not conceive, often because there are other coexisting infertility factors.

Clinicians caring for infertile couples must have a thorough understanding of the methods for treatment of anovulatory infertility. This chapter reviews the principles that guide the choice of treatment, the results achieved with different therapies, and their associated risks.

Diagnosis of Anovulation

The diagnosis of anovulation generally is not difficult to establish. ***Women with irregular, unpredictable, or infrequent menses do not require specific diagnostic tests to prove what is already obvious.*** When anovulation is suspected but uncertain, a variety of methods can be used to evaluate ovulatory function, as discussed in Chapter 27 and summarized briefly here.

Ovulatory cycles typically are associated with a classic "biphasic" basal body temperature (BBT) pattern that is not difficult to recognize, when present.[2] ***BBT recordings having no sustained interval of temperature elevation preceding the onset of menses strongly suggest anovulation.*** Biphasic recordings exhibiting a short luteal phase (onset of menses less than 12 days after the midcycle rise in BBT) suggest a subtle, but still important, form of ovulatory dysfunction. Although uncommon, BBT recordings are not clearly biphasic in some ovulatory women.

A serum progesterone measurement is the simplest, most common, objective and reliable test of ovulatory function, as long as it is appropriately timed. A progesterone concentration less than 3 ng/mL implies anovulation, except when drawn immediately after ovulation or just before the onset of menses, when lower levels naturally might be expected.[3, 4] A normal ovulatory cycle is 25–35 days in duration and exhibits a luteal phase lasting approximately 14 days. Ideally, the serum progesterone level should be drawn approximately one week before the expected onset of menses, when the concentration is at or near its peak. ***Contrary to popular belief and practice, cycle day 21 is <u>not</u> always the best time to measure the serum progesterone concentration and the threshold level indicating ovulation is <u>not</u> 10 ng/mL.*** Cycle day 21 is a good choice for women with cycles lasting approximately 28 days, but a poor choice for women with 35 day cycles. A serum progesterone concentration greater than 10 ng/mL suggests normal luteal function, but not when the luteal phase is grossly short, and a level less than 10 ng/mL can be quite normal, because progesterone is secreted by the corpus luteum in distinct pulses, temporally linked to pulsatile luteinizing hormone (LH) secretion;[5] random sampling can coincide with a transient nadir in serum levels.

Other more complicated or sophisticated tests of ovulation, such as monitoring urinary LH excretion and serial transvaginal ultrasonography, can be useful once ovulation has been achieved, but are unnecessary for the diagnosis of anovulation.

Classification of Ovulatory Disorders

After evaluation for the causes of anovulation is completed, virtually all women can be classified according to the criteria adopted by the World Health Organization (WHO).[6] Hyperprolactinemic anovulation is considered as a fourth and specific category.

WHO Group I: Hypogonadotropic Hypogonadal Anovulation. The group accounts for approximately 5–10% of anovulatory women and includes those with low or low-normal serum follicle-stimulating hormone (FSH) concentrations, and low serum estradiol levels, due to absent or abnormal hypothalamic gonadotropin-releasing hormone (GnRH) secretion or pituitary insensitivity to GnRH. Examples include women with hypothalamic amenorrhea relating to physical, nutritional, or emotional stress, weight loss, excessive exercise, anorexia nervosa and its variants, Kallmann syndrome, and isolated gonadotropin deficiency. Women in the group may require hypothalamic-pituitary imaging to exclude a mass lesion.

WHO Group II: Eugonadotropic Euestrogenic Anovulation. This group is the largest, including 75–85% of anovulatory women, and is characterized by normal serum FSH and estradiol levels and normal or elevated LH concentrations.[7] The most common examples are women with polycystic ovary syndrome (PCOS), some of whom ovulate at least occasionally. Women with PCOS should be screened for type 2 diabetes mellitus before treatment, due to the fetal risks associated with untreated diabetes.[8] Weight loss generally is the best initial treatment for those who are obese because it can, by itself, restore ovulatory function.[9–11]

WHO Group III: Hypergonadotropic Anovulation. The group accounts for approximately 10–20% of anovulatory women and includes those with elevated serum FSH concentrations; most, but not all, have amenorrhea. The classic example is premature ovarian failure, due to follicular depletion, and few respond to treatment aimed at ovulation induction.

Hyperprolactinemic Anovulation. Approximately 5–10% of anovulatory women have hyperprolactinemia, which inhibits gonadotropin secretion. Consequently, serum FSH concentrations generally are low or low-normal and serum estradiol levels also tend to be relatively low. Most hyperprolactinemic women have oligomenorrhea or amenorrhea. When hyperprolactinemia cannot be attributed confidently to coexisting hypothyroidism or to medications, hypothalamic-pituitary imaging is indicated to exclude a mass lesion.

Pretreatment Evaluation and Treatment

The causes of anovulation are many and varied. Thyroid disease, hyperprolactinemia, adrenal disease, pituitary or ovarian tumors, eating disorders, extremes of weight loss or exercise, polycystic ovary syndrome (PCOS), and obesity all are commonly associated with ovulatory dysfunction. Treatment should be directed at the underlying cause, when that can be determined, because specific treatment is more likely to succeed and some conditions can have longer-term health consequences if not recognized and treated.

All anovulatory women deserve at least some preliminary evaluation, both to exclude important pathology that may require medical attention before ovulation induction begins and to identify the most likely successful form of treatment. Chapter 11 considers the causes and management of amenorrhea and galactorrhea. Chapters 12 and 13 discuss the pathophysiology and treatment of PCOS and hirsutism. Chapter 15 describes the evaluation of dysfunctional uterine bleeding. At a minimum, anovulatory women should be screened for thyroid disorders (serum TSH) and hyperprolactinemia (serum prolactin) because both require further evaluation and specific treatment.[12–14] Depending on the menstrual history, endometrial sampling also merits consideration, because chronic anovulation is associated with increased risk for endometrial hyperplasia.

Screening for impaired glucose tolerance and diabetes is recommended for all obese anovulatory women with PCOS; up to 35% exhibit impaired glucose tolerance and 7–10% meet criteria for type 2 diabetes mellitus.[15, 16] Screening is best accomplished by measuring the glucose level 2 hours after a 75 gm oral glucose load; concentrations between 140 and 199 mg/dL indicate impaired glucose tolerance and levels of 200 mg/dL or greater indicate noninsulin-dependent diabetes.

Anovulation offers an obvious potential explanation for infertility, but often is not the only infertility factor. ***Before ovulation induction begins, a screening semen analysis is prudent because male factors are an important contributing cause in 20–40% of infertile couples.***[17] Early recognition of a significant co-existing male factor helps to avoid wasted time, effort, expense, and associated frustrations.

Additional preliminary evaluation with hysterosalpingography (HSG) or transvaginal ultrasonography merits serious consideration, particularly in women with a history of previous pelvic infection or surgery, ectopic pregnancy, inflammatory bowel disease, pelvic pain or other symptoms of endometriosis, or an abnormal physical examination. In the absence of such risk factors, the likelihood of tubal disease is low and HSG can be deferred safely in young women and those who do not require complicated and costly forms of ovulation

induction. In older women with a narrowing window of opportunity, it is generally wise to evaluate objectively all relevant infertility factors before reatment begins to ensure that time is used to best possible advantage. In women who require ovulation induction with exogenous gonadotropins, the associated costs, logistics, and risks also justify a thorough preliminary evaluation. ***Preliminary HSG and transvaginal ultrasonography are recommended when the medical history or physical examination raises suspicion for co-existing uterine or tubal infertility factors, for women over age 35, and when ovulation induction requires treatment with exogenous gonadotropins.*** Laparoscopy and hysteroscopy are unnecessary for most women, but certainly appropriate for those with an abnormal HSG or signs or symptoms of pelvic disease.

The best initial treatment for obese anovulatory women is weight loss, when it can be achieved. Even modest weight loss (5–10% of body weight) often restores ovulatory cycles in obese anovulatory women with PCOS.[18–25] At a minimum, weight loss can increase sensitivity to ovulation-inducing drugs and decrease the complexity of treatment required. In one study, 60 of 67 obese anovulatory women (90%) who lost an average of 10 kg/m^2 in a diet and exercise program resumed spontaneous ovulation and 52 (78%) ultimately achieved pregnancy, 18 (27%) without other interventions.[26] A body mass index (BMI) less than 27 is a reasonable if also modest goal.

Clomiphene Citrate

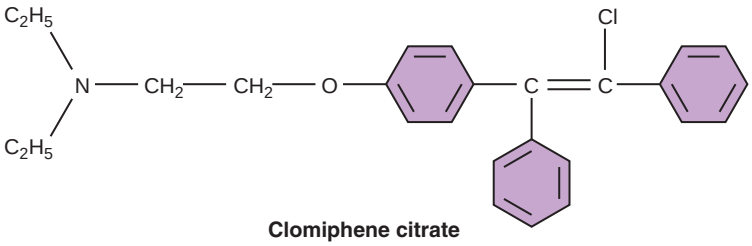
Clomiphene citrate was first synthesized in 1956, introduced for clinical trials in 1960, and approved for clinical use in the United States in 1967.[27, 28] In early clinical trials, 80% of anovulatory women treated with clomiphene achieved ovulation and half of those who ovulated also conceived.[27, 28] The collected clinical experience gained in the years since remains consistent with those early observations.

Pharmacology and Mechanism of Action

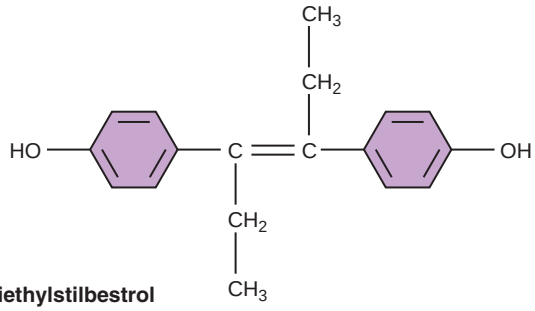
Clomiphene is a non-steroidal triphenylethylene derivative that acts as a selective estrogen receptor modulator (SERM), having both estrogen agonist and antagonist properties.[29] However, in almost all circumstances, clomiphene acts purely as an antagonist or antiestrogen; its weak estrogenic actions are clinically apparent only when endogenous estrogen levels are very low. Clomiphene is cleared through the liver and excreted in the stool; approximately 85% is eliminated within a week, but traces can remain in the circulation for longer.[30] Clomiphene is a racemic mixture of two different stereoisomers, enclomiphene (62%; originally known as *cis*-clomiphene) and zuclomiphene (38%; originally known as *trans*-clomiphene).[29, 31] Enclomiphene is the more potent isomer and the one responsible for its ovulation-inducing actions.[29, 32] The half-life of enclomiphene is relatively short, so serum concentrations rise and fall quickly during and after treatment.[30, 33] Zuclomiphene is cleared much more slowly; serum levels remain detectable for weeks after a single dose[30] and may even accumulate gradually over a series of cycles, but there is no evidence that residual zuclomiphene has any important clinical effects or consequences.[33]

Structural similarity to estrogen allows clomiphene to compete with endogenous estrogen for nuclear estrogen receptors at sites throughout the reproductive system. However, unlike estrogen, clomiphene binds to nuclear estrogen receptors for an extended interval of time and thereby depletes receptor concentrations by interfering with receptor recycling.[29] At the hypothalamic level, estrogen receptor depletion prevents accurate interpretation of

circulating estrogen levels; circulating estrogen levels are perceived as lower than they truly are. ***Reduced negative estrogen feedback triggers normal compensatory mechanisms that alter the pattern of gonadotropin-releasing hormone (GnRH) secretion and stimulate increased pituitary gonadotropin release which, in turn, drives ovarian follicular development.*** At the pituitary level, clomiphene also might increase the sensitivity of gonadotrophs to GnRH stimulation.[34]



Clomiphene citrate



Diethylstilbestrol

When administred to already ovulatory women, clomiphene increases GnRH pulse frequency.[35] In anovulatory women with polycystic ovary syndrome (PCOS) who already exhibit an increased GnRH pulse frequency, clomiphene increases only pulse amplitude.[36] Serum levels of both FSH and LH rise during clomiphene treatment and fall again soon after the typical 5-day course of therapy is completed.[37] In successful treatment cycles, one or more follicles emerge and grow to maturity. In parallel, serum estrogen levels rise progressively, ultimately triggering an LH surge and ovulation. In sum, clomiphene works primarily by stimulating the normal endocrine mechanisms that define the hypothalamic-pituitary-ovarian feedback axis. The importance of other effects it may have on insulin-like growth factors (decrease in IGF-I concentrations) and sex hormone-binding globulin (increase in serum levels) is uncertain.[38, 39]

Peripheral Actions

In addition to its desirable central actions, clomiphene can exert less desirable anti-estrogenic effects at peripheral sites in the reproductive system, which some have suggested might explain the difference between ovulation and pregnancy rates achieved with clomiphene treatment. ***Adverse effects of clomiphene on the endocervix, the endometrium, the ovary, the ovum, and the embryo have been described, but there is no compelling evidence to indicate that such effects have important clinical consequences in most women.***

On balance, the weight of evidence from controlled trials suggests that the quality and quantity of ***cervical mucus production*** can be decreased in clomiphene treatment cycles. Whereas some have observed no significant changes in mucus characteristics during treatment,[40] others have found clomiphene has dose-dependent adverse effects.[41, 42] The conflicting results have several possible explanations. The effect may be more apparent when the interval between the end of treatment and ovulation is short.[43] The effect might

often be negated by higher serum estradiol levels resulting from clomiphene-induced multifollicular development.[44] It also is possible that some individuals may be more sensitive to the effect.[45] Regardless, any adverse effect that clomiphene may have on cervical mucus now is largely moot (Chapter 27). In recent years, even the evaluation of cervical mucus has all but disappeared from clinical practice because controlled trials have demonstrated that postcoital testing (the traditional test of cervical factors) has little or no predictive value[46, 47] and because modern treatment regimens for persistent infertility now routinely incorporate intrauterine insemination (IUI), which bypasses the cervix altogether.[48, 49]

Impaired ***endometrial growth*** also has been reported in clomiphene-treated women. However, preovulatory endometrial thickness in clomiphene-induced cycles remains well within the range normally observed in spontaneous ovulatory cycles in the large majority of women.[50–54] Other subtle differences in endometrial morphology have been attributed to the effects of clomiphene, but their clinical relevance, if any, is uncertain.[55, 56] It is likely that clomiphene inhibits endometrial growth, at least in some women, for the same reasons that it may inhibit cervical mucus production, but the same caveats apply; the effect is inconsistent, may be offset by the higher estrogen levels in clomiphene-induced cycles, and probably has little clinical importance, except perhaps in those individuals exhibiting grossly poor endometrial growth (peak preovulatory thickness less than 5–6 mm).

Clomiphene does not appear to have any clinically relevant direct effects on the ***ovary*** or ***embryo.*** Although clomiphene can inhibit steroid hormone production by cultured avian,[57] ovine,[58] and human granulosa/luteal cells *in vitro*,[59] serum estrogen and progesterone concentrations in clomiphene-induced cycles are typically higher, not lower, than in spontaneous ovulatory cycles. Adverse effects on mouse ovum fertilization and embryo development have been observed *in vitro*,[60] but studies in women indicate that serum concentrations of enclomiphene and zuclomiphene never approach the levels required to induce such effects, even after several consecutive treatment cycles.[33]

Clinical Indications

Clomiphene citrate is the traditional drug of choice for ovulation induction in anovulatory infertile women with normal thyroid function, normal serum prolactin levels, and normal endogenous estrogen production, as determined by clinical observations (oligomenorrhea, estrogenic cervical mucus), a serum estradiol determination (greater than approximately 40 pg/mL), or a normal menstrual response to a progestin challenge (WHO Group II).[61] Although the drug also frequently is used empirically to stimulate multi-follicular development in ovulatory women with unexplained infertility (usually in combination with IUI),[48, 62–64] the focus here is on ovulation induction in anovulatory women; empiric clomiphene and other treatments for unexplained infertility are discussed in detail in Chapter 27.

Given its mechanism of action, it is not surprising that clomiphene typically is ineffective in women with hypogonadotropic hypogonadism (WHO Group I). Together, low or low-normal FSH levels and low serum estrogen concentrations indicate that the hypothalamic-pituitary-ovarian axis is not functioning normally in women with hypothalamic amenorrhea; if it were, FSH levels would be high because estrogen concentrations are low. If low endogenous estrogen concentrations cannot stimulate increased FSH secretion, there is little reason to think that a clomiphene-induced decrease in the level of negative estrogen feedback will succeed, and it rarely does. Alternative treatments that directly stimulate the pituitary (pulsatile exogenous GnRH) or the ovary (exogenous gonadotropins) usually are required.

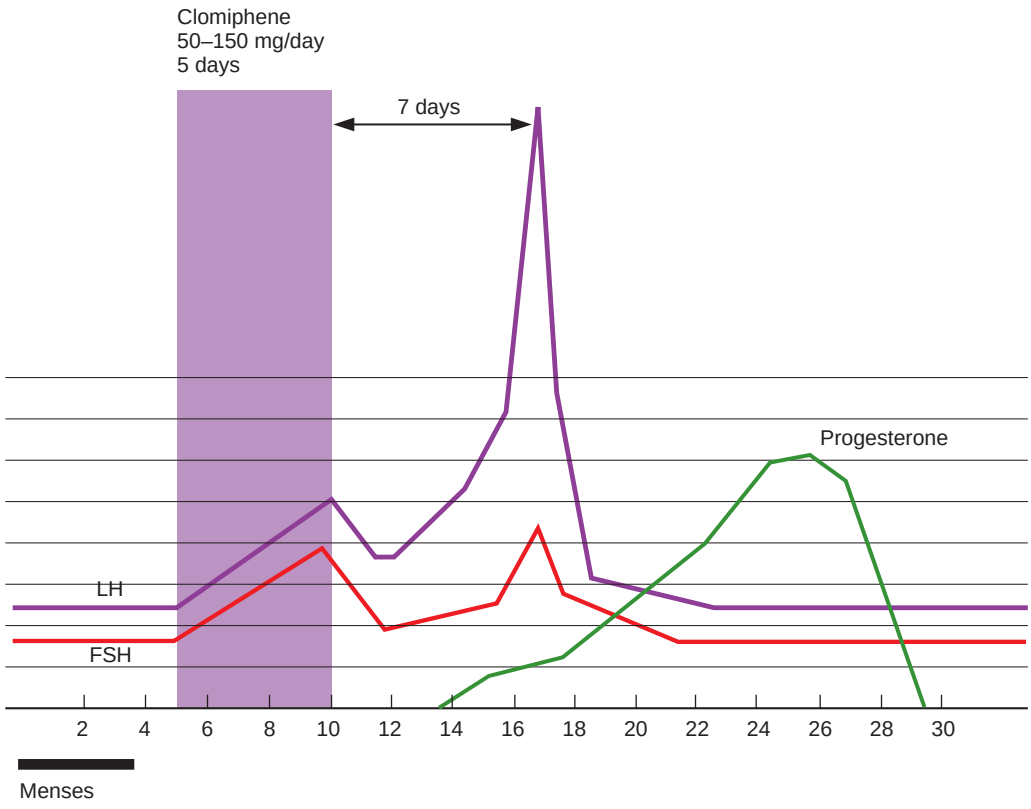
Because the corpus luteum derives from the ovulatory follicle, its functional capacity depends, in part, on the quality of preovulatory follicular development. Logically, inadequate follicular development can be expected to cause or predispose to poor luteal function,

if ovulation still occurs. Indeed, the most obvious example of poor luteal function, a short luteal phase, is associated with abnormally low follicular phase FSH levels.[65, 66] Consequently, clomiphene is both a logical and effective choice for treatment.[67–70] Progesterone levels typically are higher in clomiphene-induced ovulatory cycles than in normal spontaneous cycles, likely because preovulatory follicular development is optimized and because treatment often results in more than one corpus luteum.[71, 72]

Clomiphene citrate treatment is generally limited to women with demonstrated ovulatory dysfunction but may also be justified in normally ovulating women whose infertility remains unexplained, particularly when they are young and infertility is of short duration, and those unwilling or unable to pursue more aggressive treatments. The efficacy of clomiphene treatment in women with unexplained infertility can be attributed to optimizing follicular development or to the "superovulation" of more than a single ovum.[62, 63] ***Empiric clomiphene treatment is most effective when combined with intrauterine insemination (IUI), in an effort to increase the numbers of both ova and sperm (Chapter 27).***[48, 73]

Clomiphene Treatment Regimens

Clomiphene is administered orally, typically beginning on the third to fifth day after the onset of a spontaneous or progestin-induced menses. Ovulation and conception rates and pregnancy outcomes are similar when treatment starts anywhere between cycle days 2 and 5.[74] In women with amenorrhea, treatment can begin immediately if pregnancy has been excluded. The dose of clomiphene required to induce ovulation correlates with body weight but cannot be predicted confidently for an individual woman.[75, 76] Although obese women often require higher doses of clomiphene treatment, the results achieved ultimately are similar to those observed in lean women.[77, 78] No clinical or laboratory parameter has proven utility for predicting the dose of clomiphene needed to induce ovulation.[79]



Treatment usually starts with a single 50 mg tablet daily for a 5-day interval and, if necessary, increases by 50 mg increments in subsequent cycles until ovulation is achieved. ***Most women who respond to clomiphene will respond to either 50 mg (52%) or 100 mg (22%).***[80, 81] Lower doses (12.5–25 mg daily) deserve consideration for women who prove highly sensitive to the drug or develop large ovarian cysts that prevent continued treatment.[82] Although not approved by the U.S. Food and Drug Administration, higher doses of clomiphene (150–250 mg daily) sometimes can succeed when lower doses fail (150 mg, 12%; 200 mg 7%; 250 mg 5%).[80, 81] We believe that treatment with doses up to 150 mg is reasonable before considering alternatives.[83] Longer durations of clomiphene treatment (7–10 days) can succeed in some women when standard treatment does not and occasionally may be useful when there are no practical alternatives.[84–86]

The same methods used for diagnosis of anovulation can be used to evaluate the response to treatment. BBT recordings are simple and inexpensive, but can become tedious over time. ***A serum progesterone level greater than 3 ng/mL provides reliable evidence that ovulation has occurred,***[3, 4] ***but must be timed appropriately for confident and correct interpretation.*** Measuring the serum progesterone concentration between cycle days 22 and 25 will minimize the risk of sampling immediately after ovulation (occurring as late as cycle day 19–20 in cycles lasting up to 35 days) or before menses, when levels less than 3 ng/mL might be observed and misinterpreted. More sophisticated tests of ovulation involving greater costs and logistical demands are unnecessary to determine only if ovulation occurred, but may be justified once successful ovulation induction has been achieved.

Commercial test kits can detect the midcycle urinary LH surge and help to determine not only whether ovulation occurred, but when, and to accurately define the length of the luteal phase.[87] ***In clomiphene-induced ovulatory cycles in anovulatory women, the LH surge typically occurs 5–12 days after treatment ends, most often on cycle day 16 or 17 when clomiphene is administered on days 5–9.***[88] Ovulation generally occurs 14–26 hours after surge detection and almost always within 48 hours.[87] However, in clinical practice, both false negative and false positive results are relatively common.[89, 90] An endometrial biopsy yielding secretory endometrium also implies recent ovulation,[91] but the associated costs and discomfort cannot be justified for that purpose alone. Serial transvaginal ultrasonography can demonstrate the size and number of developing follicles, track endometrial growth, and provide presumptive evidence of ovulation,[92, 93] but is difficult to justify when less complicated and costly methods can provide the necessary information. Some advocate monitoring at least the first cycle of treatment to identify those who respond excessively.[94] A study comparing fecundity in clomiphene-induced cycles monitored with BBT, urinary LH excretion, or serial transvaginal ultrasonography found no clear advantage for any one of the three methods.[95]

Traditionally, when the serum progesterone concentration reveals persistent anovulation after clomiphene treatment, a progestin is prescribed (e.g., medroxyprogesterone acetate, 5–10 mg daily for 5–7 days) to induce menses before the next cycle begins at a higher dosage. Although effective, the sequence takes time and several months may pass before a patient is proven unresponsive to clomiphene. A "stair-step" treatment protocol is an alternative that can shorten the time required to achieve ovulation and to identify those who require different treatment. The regimen involves treatment with clomiphene (50 mg) on cycle days 5–9 after a spontaneous or induced menses, ultrasonography on day 11–14, immediate treatment at the next higher dose level (100 mg) if no dominant follicle (≥15 mm) has emerged, repeated ultrasonography 1 week later, and if still no dominant follicle is observed, immediate treatment at the highest dose level (150 mg) and ultrasonography again 1 week later.[96] In a case series involving 31 anovulatory infertile women with PCOS treated with a stair-step regimen, the time to ovulation was 23–35 days, cycle fecundability was 13%, and estimated costs were increased only modestly.[96]

In the past, monthly pelvic examinations to exclude residual ovarian enlargement were recommended before each new clomiphene treatment cycle began. It is still prudent to

postpone further treatment when symptoms lead to discovery of a large cyst or grossly enlarged ovaries, but clinical studies and experience indicate that routine baseline physical examination or ultrasonography is unnecessary.[88] Nevertheless, regular contact should be maintained to review the progress of treatment and to ensure that any additional evaluation needed is accomplished efficiently.

Results of Clomiphene Treatment

Clomiphene will induce ovulation successfully in 70–80% of properly selected women.[97, 98] The likelihood of response decreases with increasing age and body mass index and with the extent of any associated hyperandrogenemia in anovulatory women. Interestingly, women with amenorrhea are more likely to conceive than those with oligomenorrhea,[97] possibly because infertile women who menstruate also likely ovulate, albeit infrequently, and are more likely to have other co-existing infertility factors.

Among anovulatory infertile women who respond to clomiphene treatment, the overall cycle fecundability is approximately 15%. In women with no other infertility factors, cycle fecundability may reach as high as 22%, comparable to that observed in normal fertile couples after discontinuation of barrier contraception and those with male factor infertility receiving therapeutic donor inseminations.[97] ***Cumulative pregnancy rates of 70–75% can be expected over six to nine cycles of treatment.***[99, 100] Thereafter, cycle fecundability falls substantially. ***When pregnancy is not achieved within 3–6 clomiphene-induced ovulatory cycles, the infertility investigation should be expanded to exclude other infertility factors not yet evaluated, or to change the overall treatment strategy if evaluation is already complete. Prolonged treatment with clomiphene is inappropriate, particularly for women over age 35.***

Side Effects

Clomiphene treatment generally is very well tolerated. Minor side effects are relatively common, but rarely are persistent or severe enough to require that treatment be discontinued.

Transient hot flashes, usually limited to the short interval of treatment, occur in 10–20% of women.[83] Considering that clomiphene causes a central misperception that endogenous estrogen levels are low, vasomotor symptoms are not difficult to understand. Mood swings also are relatively common. Other mild and less common side effects include headache, breast tenderness, pelvic pressure or pain, and nausea. Visual disturbances (blurred or double vision, scotomata, light sensitivity) are uncommon (1–2%) and reversible, but reports of persistent “afterimages” (palinopsia) and light sensitivity (photopobia) make them nonetheless unnerving;[101] when such symptoms appear, prudence dictates that treatment be abandoned in favor of alternative methods for ovulation induction.

Risks

Clomiphene treatment has risks, but serious complications are rare. Inevitably, questions arise concerning the risks for multiple pregnancy, congenital anomalies, and other potential adverse outcomes associated with clomiphene treatment.

The principal risk associated with clomiphene treatment is an increased risk for conceiving a conceiving a ***multiple pregnancy***. The clomiphene-induced increase in FSH secretion is only transient, so normal selection mechanisms still operate to yield only a single mature follicle in most treatment cycles in anovulatory women. ***Nevertheless, multi-follicular development is relatively common and the overall risk of multiple pregnancy is increased to approximately 7–10%.***[102–104] The large majority of multiple pregnancies conceived in clomiphene-induced cycles are twins; the risk for triplets is 0.3–0.5%, for quadruplets 0.3%, and for quintuplets 0.1%.[28] ***The higher risk of multi-fetal gestation is another reason to treat with the lowest effective dose of clomiphene; higher doses do not improve results and only increase the risk of superovulation and multiple pregnancy, with all of the attendant antenatal and neonatal complications.***

There is no evidence that clomiphene treatment increases the overall risk of birth defects or of any one anomaly in particular.[105] Several large series have examined the question and have drawn the same conclusion.[102–105] In a series of 1,034 pregnancies resulting from clomiphene treatment, 14.2% ended in miscarriage, 0.5% in ectopic pregnancy, 0.1% in molar pregnancy, and 1.6% in stillbirth, and among 935 live born infants, malformations were detected in 21 (2.3%).[105] Earlier suggestions that the risk of neural tube defects might be higher in pregnancies conceived after clomiphene treatment were not confirmed in later investigations.[106, 107] A small study of pregnancy outcomes in women exposed inadvertently to clomiphene during the first trimester of pregnancy found no evidence of teratogenicity.[108] There also is no evidence that clomiphene treatment increases the risk of ***developmental delay or learning disability*** in children conceived during clomiphene treatment.[109]

Early studies suggested that the incidence of spontaneous ***miscarriage*** in pregnancies resulting from clomiphene treatment might be increased, but a number of others have observed miscarriage rates no different from those in pregnancies conceived without treatment.[80, 110]

The incidence of ***ovarian hyperstimulation syndrome*** (OHSS) in clomiphene-induced cycles is difficult to determine confidently because definitions of the syndrome vary widely among studies. In general, mild symptoms of ovarian hyperstimulation (transient abdominal discomfort, mild nausea, vomiting, diarrhea, and abdominal distention) are not altogether uncommon but require only expectant management. When induction of ovulation proceeds in the recommended incremental fashion to establish the minimum effective dose, the risk of clinically significant OHSS (massive ovarian enlargement, progressive weight gain, severe abdominal pain, intractable nausea and vomiting, gross ascites, oliguria) is remote.

The incidence of ***ovarian cancer*** is decreased among parous women and those using hormonal contraception for prolonged intervals, suggesting that "incessant ovulation" (repeated epithelial disruption and repair) predisposes to development of ovarian cancer and that treatment with ovulation-inducing drugs might increase the risk.[111] The results of case-control studies conducted in the 1990s lent credence to the notion and raised considerable concern,[112, 113] although their conclusions were challenged because of important methodologic flaws. One study compared infertile treated women to fertile women rather than to infertile untreated women, even though infertility and nulliparity were known risk factors for ovarian cancer.[112] Another included cancers of all types and tumors of low malignant potential despite their differing pathophysiology.[113] Since then, numerous studies have confirmed that the incidence of ovarian cancer is increased in infertile women, but have failed to find any substantive evidence that ovulation-inducing drugs increase the risk.[114–121] The results of studies of the risk for ***breast cancer*** have been conflicting, with some showing no association with ovulation-inducing drugs and others suggesting possible increases in risk in certain subgroups.[120] ***No causal relationship between ovulation-inducing drugs and ovarian or breast cancer has been established, but prolonged treatment with clomiphene nonetheless should be avoided, primarily because it has little hope of success.***

Adjuvant and Combination Treatments

Clomiphene failure describes women who do not ovulate in response to clomiphene treatment, not those who fail to conceive despite successful clomiphene-induced ovulation. In the latter group, additional evaluation is indicated to identify other potential co-existing infertility factors not already excluded. If or when that has been accomplished, persistent infertility is best regarded and treated as unexplained infertility (Chapter 27).

Although most properly selected women will ovulate in response to clomiphene treatment, many do not. Direct ovarian stimulation with exogenous gonadotropins, discussed later in detail, is an obvious alternative, but it is by no means the only option that merits consideration. Many clomiphene-resistant anovulatory infertile women will respond to supplemental or combination treatment regimens. Options include adjuvant treatment with glucocorticoids, exogenous human chorionic gonadotropin (hCG), or metformin, and preliminary suppressive therapy (hormonal contraceptives). ***It is helpful to be familiar with these less common ovulation induction strategies because many couples are understandably reluctant or unable to pursue the obvious alternative of gonadotropin treatment once fully advised of the associated costs, logistical demands, and risks.***

Failure to respond to one or more of these less commonly employed treatment regimens is not a prerequisite for exogenous gonadotropin therapy. They are simply useful alternatives for those unwilling or unable to pursue gonadotropin therapy and, for some, may be the only options when clomiphene treatment has failed. A choice among them is not entirely arbitrary, but should consider specific elements of the patient's history, the results of laboratory evaluation, and observations in previous unsuccessful clomiphene treatment cycles.

Clomiphene and Glucocorticoids

Numerous studies have examined the efficacy of adjuvant treatment with glucocorticoids in clomiphene-resistant anovulatory women and all have found that combined treatment with clomiphene and a glucocorticoid can successfully induce ovulation in many who fail to respond to clomiphene alone.[86, 122–126] Both continuous and more limited follicular phase treatment regimens (cycle days 5–14) have been described, using either prednisone (5 mg daily) or dexamethasone (0.5–2.0 mg daily). Whereas some studies have suggested that combined treatment with clomiphene and glucocorticoids is most effective in women having elevated serum dehydroepiandrosterone sulfate (DHEA-S) concentrations,[122, 123] others have found that treatment also can be effective in those with normal DHEA-S levels.[124, 126] and in unselected populations of clomiphene-resistant women.[86, 125]

In the largest randomized trial involving more than 200 clomiphene-resistant anovulatory infertile women, over 80% of those receiving combined treatment with clomiphene (200 mg daily cycle days 5–9) and dexamethasone (2 mg daily, cycle days 5–14) ovulated, compared to 20% of controls treated with clomiphene and placebo; the cumulative pregnancy rate in women receiving dexamethasone (40%) was 10-fold higher than in those who received placebo (4%).[126] The mechanism of glucocorticoid action remains unclear, but appears to involve more than simple androgen suppression. Other possibilities include direct effects on the developing oocyte and indirect effects on intrafollicular growth factors and cytokines, which may act synergistically with FSH.[127] Regardless, adjuvant treatment with glucocorticoids may be justified for three to six cycles when it is successful, but should be promptly discontinued when it is not. There is no evidence that glucocorticoid treatment has any important side effects or risks when used in the doses and durations described.

Clomiphene and HCG

Although there are few if any data to demonstrate its value, exogenous hCG has been used commonly as a surrogate LH surge to trigger ovulation in clomiphene-induced cycles, particularly when IUI is performed, as in couples with unexplained infertility and those with a co-existing male factor. Adjuvant hCG treatment can be useful, but has limited indications, distinct disadvantages, and potential consequences.

In anovulatory women who fail to ovulate in response to clomiphene alone, adjuvant hCG treatment is based on the premise that clomiphene may be successful in stimulating the emergence of a preovulatory follicle but ultimately fail to trigger an endogenous LH surge and to induce ovulation. Physiologically, and in practice, the scenario is most unlikely. Moreover, serial transvaginal ultrasonography is required to demonstrate the phenomenon and to ensure that the ovulatory stimulus is delivered at the appropriate time. If administered blindly and prematurely, before the dominant follicle is mature enough to respond, hCG is more likely to induce atresia than ovulation. The question of when to administer hCG presents a dilemma. Although hCG commonly is administered when the lead follicle reaches 18–20 mm,[128] clinical studies indicate that the peak preovulatory follicular diameter in successful clomiphene-induced ovulatory cycles ranges between 18 and 30 mm (mean 25 mm).[88, 129] Considering that the preovulatory follicle grows approximately 2 mm per day as it approaches maturity,[92, 93] the corresponding interval may thus span up to 6 days. Normally, the preovulatory follicle triggers its own ovulatory stimulus at the peak of maturity by generating and maintaining the estrogen levels that are required to induce the LH surge. The timing of the spontaneous LH surge is therefore always optimal, but that of hCG treatment can never be more than an educated guess.

When combined treatment with clomiphene and IUI is required, insemination usually is best performed on the day after detection of the spontaneous LH surge, using one of the now widely available commercial kits designed for the purpose, because ovulation generally occurs 14–26 hours after urinary LH surge detection.[87, 130] However, the lower limit of LH detection usually is between 20 and 40 IU/L and many ovulatory women exhibit peak LH levels below 40 IU/L or surges of brief duration that may escape detection;[89] false negative results are therefore not uncommon, and frustrating. ***Exogenous hCG can be useful for the few women who require IUI but repeatedly fail to detect the LH surge despite other objective evidence of successful ovulation induction.*** In such circumstances, we believe that hCG is best postponed until the preovulatory follicle reaches or exceeds 20 mm in mean diameter. Ovulation occurs 34–46 hours after hCG injection,[131] so IUI usually is performed approximately 36 hours later.

When the LH surge can be detected, adjuvant hCG treatment has no value and only adds unnecessary expense and inconvenience. Numerous studies have compared outcomes in clomiphene-induced cycles when IUI was performed after an endogenous LH surge or exogenous hCG injection; results are no better when hCG is administered and, in some cases, worse.[64, 89, 130, 132–134] The idea that hCG may still serve to ensure or improve the quality of luteal function even if is not required to trigger ovulation also is not supported by existing data. In spontaneous ovulatory cycles, hCG treatment superimposed on the endogenous LH surge has no effect on luteal phase duration or serum estrogen or progesterone concentrations;[135] the same is true in clomiphene-induced ovulatory cycles.[133] In sum, adjuvant hCG treatment is best limited to those few women who require IUI and ovulate but cannot reliably detect a midcycle LH surge.

Clomiphene and Metformin

Insulin resistance and hyperinsulinemia are common features of PCOS and an important contributing cause of the hyperandrogenism and chronic anovulation that characterize the disorder. Anovulatory infertile women with PCOS and hyperinsulinemia also are typically more resistant to clomiphene treatment.

Recognition of the pathophysiologic importance of insulin resistance in PCOS stimulated intense interest in the use of insulin-sensitizing agents for the treatment of the disorder. Metformin is a biguanide oral insulin-sensitizing agent that acts primarily by reducing hepatic gluconeogenesis, but also decreases intestinal absorption of glucose and increases peripheral glucose uptake and utilization. The effects of metformin on insulin levels and sensitivity, androgen concentrations and other metabolic and clinical measures are considered at length elsewhere in this text (Chapter 12); its adjunctive use an ovulation-inducing agent is the focus here.

Ttreatment with metformin alone or with other insulin-sensitizing drugs (thizolidinediones, D-chiro-inositol) can increase ovulation rates in some women with PCOS,[136–138] but there is no practical method for predicting those who will respond. Fasting insulin concentrations and glucose-insuin ratios do not predict response to metformin,[139] and overall, metformin appears most effective in patients who also respond to clomiphene.[138, 140] A meta-analysis of studies involving the use of metformin as an ovulation-inducing drug in women with PCOS concluded that its efficacy compared favorably with that of clomiphene,[139] but subsequent randomized trials comparing the two drugs, alone and in combination, have found that clomiphene is superior to metformin and that combined treatment is no better than treatment with clomiphene alone.[141–143] In the largest single trial, the live birth rate achieved with clomiphene treatment was significantly higher than that of metformin (22.5% vs. 7.2%) and the results of combined treatment were not significantly different (26.8%)[142] Metformin treatment also did not decrease the dose of clomiphene required to induce ovulation.[144]

In a few small studies involving clomiphene-resistant anovulatory women with PCOS, combined treatment has increased ovulation and pregnancy rates over those achieved with clomiphene alone.[145–148] A 2008 meta-analysis including 17 randomized trials concluded that combined treatment with metformin and clomiphene achieves higher ovulation and pregnancy rates than treatment with clomiphene alone.[138] Although there is no convincing evidence that combined treatment with metformin and clomiphene can increase live birth rates over those achieved with clomiphene alone,[149] the attempt seems justified for women having few alternatives besides ovarian drilling or treatment with exogenous gonadotropins. Limited evidence indicates that combined treatment with metformin and roziglitazone,[150] or with clomiphene and rosiglitazone,[151] is no more effective than metformin alone. Combined with the safety alert issued by the U.S. Food and Drug Administration concerning a possible increased risk of ischemic cardiovascular events in patients receiving treatment with thiazolidinediones,[152] these data argue against their adjuvant use for ovulation induction. ***In summary, combined treatment with metformin and clomiphene deserves consideration in women who prove clomiphene resistant before proceeding to ovarian drilling or treatment with gonadotropins.***

Metformin treatment is commonly associated with gastrointestinal side effects including nausea, vomiting, abdominal cramps, and diarrhea that can be severe enough to limit the dose administered or require discontinuation of treatment.[139, 153–156] Because side effects

tend to be dose-dependent and diminish with time, it is usually best to begin with a low daily dose (500 mg), increasing gradually at weekly intervals to a daily dose of 1500–2000 mg, as tolerance allows. Lactic acidosis can be a rare complication of metformin treatment, although recent systematic reviews have questioned whether there is a true causal relationship.[157, 158] Women at highest risk are those with chronic hypoxemic conditions related to cardiovascular, renal, hepatic and pulmonary disease and advanced age.

Although there is no evidence that metformin treatment during pregnancy is associated with any increased risk for major fetal malformations,[159] its safety during pregnancy is not yet established. Although some have advocated metformin treatment to reduce the increased risk for miscarriage in women with PCOS, which might relate to an underlying metabolic disorder,[160–162] no differences in the miscarriage rates of women who did or did not receive metformin treatment have been observed in large randomized trials.[141–143] Metformin treatment during pregnancy also has been advocated to reduce the risk for developing gestational diabetes and other pregnancy complications in women with PCOS.[163] In diabetic women, treatment with metformin during pregnancy has been associated with an increased prevalence of pre-eclampsia and increased perinatal mortality in some studies,[164] but not in others.[165] Currently, routine metformin treatment during pregnancy is not recommended for women with PCOS.[140]

Preliminary Suppressive Therapy

Considering that anovulation reflects a dysfunctional hypothalamic-pituitary-ovarian axis, it is reasonable to think that an interval of preliminary suppressive therapy might help to restore harmony and ovulatory function, at least temporarily. The idea is consistent with clinical observations of a few normal menstrual cycles immediately following discontinuation of estrogen-progestin contraception in some women who previously exhibited abnormal menstrual patterns. Limited data suggest that a 2-month interval of continuous estrogen-progestin contraception can effectively suppress serum LH and androgen levels and that ovulation rates up to 70% and cumulative pregnancy rates over 50% can be achieved with clomiphene treatment immediately thereafter in women who were previously clomiphene-resistant.[166, 167]

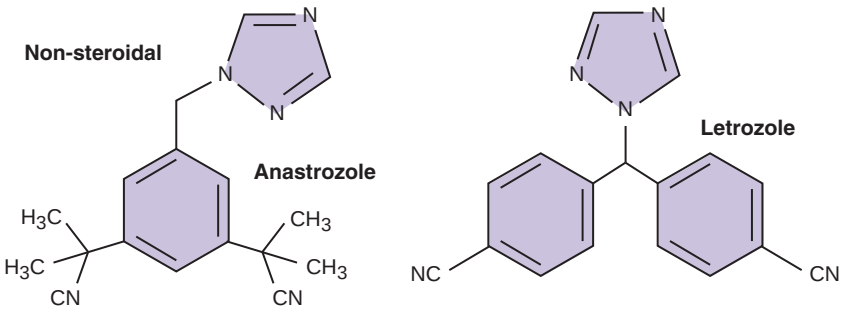
A long-acting gonadotropin-releasing hormone (GnRH) agonist, alone or in combination with an estrogen-progestin contraceptive, can be used for the same purpose. Combined suppressive therapy with a GnRH agonist and an estrogen-progestin contraceptive (3–6 months) achieves a greater and longer-lasting reduction in serum LH and androgen concentrations than treatment with estrogen-progestin contraception alone and also prevents the otherwise inevitable estrogen deficiency symptoms associated with use of a GnRH agonist. Spontaneous resumption of ovulatory cycles can follow,[168–170] potentially eliminating even the need for clomiphene treatment.

Aromatase Inhibitors

Aromatase inhibitors are used primarily in the treatment of postmenopausal breast cancer but are emerging rapidly as a new class of ovulation-inducing agents. Their use for ovulation induction has been controversial, primarily because they are not approved for that purpose and because early preliminary data suggested they might have significant fetal toxicity, which prompted the manufacturer to issue a caution against the use of letrozole in premenopausal women. Although two subsequent studies have found no evidence to

indicate that birth defects are more common in children conceived after treatment with aromatase inhibitors than in those conceived naturally or after treatment with clomiphene citrate,[171, 172] concerns persist due to the risk of inadvertent exposure in early pregnancy and evidence of teratogenic potential from animal studies.[173]

Anastrozole and letrozole are triazole (antifungal) derivatives that act as potent, competitive, nonsteroidal inhibitors of aromatase,[174, 175] the enzyme that catalyzes the rate-limiting step in estrogen production. They block estrogen production both in the periphery and in the brain, resulting in a compensatory increase in pituitary gonadotropin secretion that stimulates ovarian follicular development.[176–178] In this regard, their mechanism of action is similar to, but also distinct from, that of clomiphene. Although both stimulate increased gonadotropin secretion by decreasing the negative feedback effects of estrogen during treatment, clomiphene does so via depletion of central estrogen receptors whereas aromatase inhibitors decrease estrogen production directly.



At least in theory, the different actions of aromatase inhibitors and clomiphene may have functional, and clinical, importance. After treatment with an aromatase inhibitor ends, estrogen production in growing follicles increases promptly and rising serum concentrations exert negative feedback on gonadotropin secretion, thereby restoring the mechanism that normally serves to select and promote development of a single dominant follicle. After treatment with clomiphene ends, rising estrogen levels cannot immediately exert negative feedback due to the depletion of central estrogen receptors, resulting in a more sustained increase in gonadotropin levels that is more likely to support multifollicular development. The transient accumulation of androgen substrate during treatment with aromatase inhibitors also may increase FSH receptor expression[179, 180] and production of insulin-like growth factor-1 (IGF-1),[181–183] amplifying the actions of FSH. Moreover, because aromatase inhibitors do not interfere with the actions of estrogen in the periphery, they may be less likely than clomiphene to inhibit estrogen-stimulated cervical mucus production and endometrial proliferation.

Aromatase Inhibitor Treatment Regimens

In almost all studies conducted to date, letrozole (2.5–7.5 mg daily) and anastrazole (1 mg daily) have been administered for a 5-day interval in a manner very similar to that typical for clomiphene treatment (e.g., cycle day 3–7). One trial comparing outcomes achieved with 5 or 10 days of letrozole treatment (2.5 mg daily, beginning on cycle day 1) in clomiphene-resistant women (100 mg daily) observed significantly more preovulatory follicles (3.0 vs. 1.8) and higher pregnancy rates (17.4% vs. 12.4%) in women receiving the longer course of treatment.[184] A single-dose treatment regimen of letrozole (20 mg on cycle day 3) also has been described, with preliminary data suggesting it can achieve results similar to those observed with a lower multi-dose treatment protocol.[185]

The optimal dosage of letrozole and anastrazole has not been established firmly. In most trials involving anovulatory women, 2.5 mg of letrozole or 1 mg of anastrozole has been administered. In a trial comparing outcomes achieved with a 2.5 mg or 5 mg dosage of letrozole in ovulatory women with unexplained infertility, the 5 mg dose yielded significantly greater numbers of follicles and a higher pregnancy rate (26.3% vs. 5.9%).[186] In another trial comparing outcomes in women with unexplained infertility randomly assigned to receive treatment with letrozole (7.5 mg daily) or clomiphene (100 mg daily), followed by intrauterine insemination, the two treatments yielded similar numbers of preovulatory follicles (2.1 vs. 1.7) and pregnancy rates (11.5% vs. 8.9%); results also suggested that such higher doses of letrozole may result in greater and longer suppression of estrogen production that could limit endometrial proliferation.[187]

Altogether, the available data suggest that the optimal dose of letrozole probably ranges between 2.5 mg and 5 mg daily; outcomes achieved with doses of anastrazole greater than 1 mg daily remain to be evaluated.

Results of Treatment with Aromatase Inhibitors

Evidence for the efficacy of aromatase inhibitors for ovulation induction is accumulating rapidly. Early studies exploring the use of aromatase inhibitors focused on anovulatory women considered resistant to clomiphene, because they failed to ovulate or exhibited poor endometrial proliferation during treatment. More recent studies have sought to compare the effectiveness of aromatase inhibitors to that of clomiphene in unselected populations of anovulatory infertile women.

In an early proof of concept trial involving anovulatory clomiphene-resistant women, 9/12 patients (75%) ovulated after treatment with letrozole (2.5 mg daily) and hCG (lead follicle follicle ≥ 20 mm), three conceived (resulting in two singleton births), and normal endometrial proliferation was observed in all.[188] In a subsequent trial involving clomiphene-resistant (150 mg daily) anovulatory women with polycystic ovary syndrome (PCOS), 22/44 patients (50%) ovulated after treatment with letrozole (2.5 mg daily) and hCG (lead follicle > 18 mm) and six conceived; response did not relate to age, BMI, or menstrual pattern (amenorrhea vs. oligomenorrhea) and mean endometrial thickness was 10.2 mm.[189] In a third involving 64 women with clomiphene-resistant anovulation (100 mg daily), patients were randomized to receive treatment with letrozole (7.5 mg daily) or clomiphene (150 mg daily), followed by hCG (lead follicle ≥ 18 mm); women receiving letrozole had higher ovulation (62.5% vs. 37.5%) and pregnancy rates (40.1% vs. 18.8%) but the differences were not significant.[190]

Two randomized trials have compared the effectiveness of letrozole (2.5 mg daily) and anastrazole (1 mg daily) for ovulation induction in clomiphene-resistant anovulatory women with PCOS also receiving hCG (lead follicle follicle ≥ 18 mm). One involved 40 clomiphene-resistant patients (200 mg daily, or endometrial thickness ≤ 5 mm) and observed significantly greater endometrial thickness (8.2 vs. 6.5 mm), ovulation rates (84% vs. 60%), and pregnancy rates (19% vs. 10%) in women receiving letrozole.[191] In the larger trial involving 220 clomiphene-resistant patients (100 mg daily, or endometrial thickness < 5mm), ovulation rates (62% vs. 63%), pregnancy rates (12 % vs. 15%), and endometrial thickness (9.1 vs. 10.2 mm) in the two groups were similar.[192] ***Taken together, these observations suggest strongly that aromatase inhibitors can be effective in anovulatory women who fail to ovulate in response to clomiphene treatment.*** Aromatase inhibitors also might be considered for women who respond to clomiphene but exhibit grossly poor endometrial proliferation. ***Virtually all studies have included monitoring with ultrasonography and adjuvant treatment with hCG, so it remains unclear whether these additions***

to the treatment regimen are necessary to achieve success. If not, the substantially lower complexity, risks, and costs of treatment, compared to the alternative of gonadotropin therapy, make it easy to justify a trial of treatment with an aromatase inhibitor for clomiphene-resistant anovulatory women.

The results achieved with aromatase inhibitors in clomiphene-resistant anovulatory women suggested that aromatase inhibitors might be considered a first-line treatment for ovulation induction. Three trials have compared letrozole and clomiphene in treatment-naïve anovulatory women. The first involved 106 patients who were randomized to receive letrozole (2.5 mg daily) or clomiphene (100 mg daily), followed by hCG (lead follicle ≥ 18 mm); a significantly greater endometrial thickness (8.4 vs. 5.2 mm), ovulation rate (82% vs. 64%), and pregnancy rate (22% vs. 9%) was observed in women receiving letrozole.[193] In a second trial with the same design, ovulation occurred in 65/99 letrozole-induced cycles (66%) and in 71/95 (75%) clomiphene-induced cycles, but endometrial thickness (8 mm vs. 8 mm) and pregnancy rates (9% vs. 7%) were not different.[194] The largest of the three randomized trials involved a total of 438 women who received treatment with letrozole (5 mg daily) or clomiphene (100 mg daily) and hCG (lead follicle follicle ≥ 18 mm); ovulation rates (365/540, 68% vs. 371/523, 71%) and pregnancy rates (15% vs. 18%) were not different and endometrial thickness was significantly greater in women receiving clomiphene (9.2 vs. 8.1 mm).[195] In the only trial involving anastrazole, 115 patients received treatment with anastrazole (1 mg daily, cycle days 3–7, 243 cycles) and hCG (lead follicle follicle ≥ 18 mm) and outcomes were compared to those in matched historical controls treated with clomiphene (100 mg daily, 226 cycles); endometrial thickness was significantly greater in anastrazole cycles (10.1 vs. 8.2 mm), but ovulation rates (68% vs. 69%) and pregnancy rates (10.2% vs. 7.9%) were similar in the two groups.[196]

A few studies have examined the outcomes of pregnancies conceived after treatment with aromatase inhibitors. A cohort study found that pregnancies conceived in letrozole-induced cycles are significantly more likely to be singletons than those conceived in cycles involving treatment with clomiphene or gonadotropins.[197] One case series comparing the incidence of congenital malformations in 911 newborns of women who conceived after treatment with letrozole (14/514, 2.4%) or clomiphene (19/397, 3.0%) found no difference.[171] Another comparing the incidence of birth defects in children born to mothers treated with letrozole or clomiphene to that in pregnancies conceived without treatment also observed no differences.[172]

In sum, the available data suggest that aromatase inhibitors may be as effective, but not more effective, than clomiphene as a first-line treatment for ovulation induction. Early trials all have used surrogate endpoints, which do not correlate consistently with live birth rates. Uniformly, they also have included adjuvant treatment with hCG, which likely is unnecessary (as in clomiphene-induced cycles) and certainly is not desirable because it requires serial ultrasonography, with all of the added costs and inconvenience. Aromatase inhibitors have great promise and appear to be associated with a lower risk for conceiving a multiple pregnancy. Their use seems certain to expand, but larger randomized trials are required to better define their efficacy and rightful place in the treatment of anovulatory infertility.

Laparoscopic Ovarian Drilling

Surgical treatments aimed at restoring ovulatory function in anovulatory infertile women date back to the classic bilateral ovarian wedge resection described originally by Stein and Leventhal in 1935.[198] The procedure understandably fell out of favor after the introduction

of clomiphene citrate and gonadotropins for ovulation induction. Advances in laparoscopic surgery sparked renewed interest in the procedure, with ovarian "drilling" now representing the modern equivalent of the classical wedge resection and another treatment option for clomiphene-resistant, hyperandrogenic, anovulatory women.

Several methods for ovarian drilling have been described, including electrocautery or laser vaporization (approximately four to six sites per ovary, avoiding surfaces near the tubo-ovarian interface to minimize the risk for adhesions that might adversely affect ovum capture) and multiple biopsy.[199–201] All are intended to cause focal destruction of the ovarian stroma in efforts to decrease both intraovarian and systemic androgen concentrations. There is no evidence for the superiority of any one method, but the most common technique involves electrocautery using a unipolar needle electrode insulated above the distal 1–2 cm.

Postoperative serum concentations of androstenedione and testosterone decrease, at least for a time,[202–204] and inhibin concentrations also decline.[205, 206] Both changes likely contribute to an associated increase in FSH levels. As with the classical surgical procedure, the principal risk associated with laparoscopic ovarian drilling is postoperative adnexal adhesion formation that may decrease overall fertility, although the risk and severity of adhesions are lower;[207–210] second-look laparoscopy and adhesiolysis do not appear necessary or useful.[211, 212] There is one report of unilateral ovarian atrophy after ovarian drilling with electrocautery.[213] Whether ovarian drilling might adversely affect ovarian reserve and predispose to early menopause has not been investigated specifically, but other data indicate that reductive ovarian surgery, including ovarian wedge resection, increases risk for early menopause.[214]

In numerous uncontrolled observational studies, 40–90% of women have ovulated after laparoscopic ovarian drilling and approximately half of those have conceived.[199–201, 215, 216] In truly clomiphene-resistant women, ovarian drilling can improve clomiphene sensitivity or response to exogenous gonadotropins when it does not restore spontaneous ovulatory cycles.[217] When considering laparoscopic ovarian drilling as a treatment option in clomiphene-resistant anovulatory infertile women, the most relevant data derive from randomized controlled trials comparing surgical treatment with ovulation induction using exogenous gonadotropins.[211, 218, 219] A 2007 systematic review and meta-analysis including nine trials found no evidence of a difference in live birth (OR=1.04, CI=0.59−1.85) or clinical pregnancy rate (OR=1.08, CI=0.69−1.71).[220] After 12 months, the ovulation rate achieved with drilling (52%) was similar to that with gonadotropin therapy (62%). However, as might be expected, laparoscopic ovarian drilling yields far fewer multiple pregnancies than gonadotropin treatment (1% vs. 16%; OR=0.13, CI=0.03−0.52); the miscarriage rates associated with surgical and medical treatment are comparable.[220]

The ideal candidate for ovarian drilling is not obese and has no other infertility factors. The procedure often is unsuccessful in obese women (BMI >30 kg/m^2),[221, 222] and whereas over 80% of women can be expected to conceive after surgery when clomiphene-resistant anovulation is the sole cause of infertility,[222, 223] only 15–30% achieve pregnancy when there is a coexisting tubal, male, or other infertility factor.[203, 223]

Laparoscopic ovarian drilling can be an effective therapeutic option for clomiphene-resistant anovulatory infertile women, but the temporary effects of treatment, the risk of postoperative adhesions, and the theoretical risk of adverse effects on ovarian reserve deserve careful consideration and discussion. The procedure is perhaps best reserved for women who are unable or unwilling to accept the costs and risks associated with gonadotropin therapy.

Exogenous Gonadotropins

Exogenous gonadotropins have been used to induce ovulation in gonadotropin-deficient women and those who fail to respond to other, less complicated forms of treatment for nearly 50 years. They are highly effective, but also very costly and associated with substantial risks including multiple pregnancy and ovarian hyperstimulation syndrome. ***Consequently, exogenous gonadotropins should be used only by clinicians having the training and experience necessary to provide safe and effective treatment.***

Gonadotropin Preparations

Gonadotropin preparations have evolved gradually over the years, from relatively crude urinary extracts, to more highly purified urinary extracts, to the recombinant preparations in common use today.[224, 225]

For almost 30 years, the only exogenous gonadotropins available were human menopausal gonadotropins (hMG, menotropins), an extract prepared from the urine of postmenopausal women containing equivalent amounts (75 IU) of FSH and LH per ampule or vial and requiring intramuscular injection. Originally, the urinary source was a single convent in Italy, but later collections were expanded to a number of centers in other countries.[226] Urinary menotropins also contain small but measurable and varying amounts of hCG, most of it added intentionally during the manufacturing process to provide the appropriate amount of LH activity and some derived from other sources.[227] Clinical use of hMG began in 1950 but the clinical trials did not begin until after 1960.[228, 229] Relatively crude gonadotropin extracts like traditional hMG also contained significant amounts of uncharacterized urinary protein that may be antigenic.[230] Contemporary hMG preparations are more highly purified than in the past and can be administered subcutaneously.[231]

Beginning about 25 years ago, more purified urinary FSH preparations (urofollitropin) were developed by removing LH from urinary extracts using immunoaffinity columns containing polyclonal anti-hCG antibodies.[232] Early preparations of purified urinary FSH (75 IU) contained less than 1 IU LH but a considerable amount of other urinary protein and still required intramuscular administration. Further purification using monoclonal antibodies specific for FSH yielded a preparation containing less than 0.1 IU LH and less than 5% unidentified protein. The even more highly purified products now in use contain less than 0.001 IU LH, very low levels of urinary protein, and can be administered subcutaneously.

Just over 15 years ago, the *in vitro* production of recombinant human FSH was achieved through genetic engineering. Briefly summarized, the process involves introduction of the genes encoding the α- and β-FSH subunits into the genome of a Chinese hamster ovary cell line, which then synthesizes and secretes a glycosylated bioactive dimeric FSH that is finally purified by immunochromatography using a specific anti-FSH monoclonal antibody. Recombinant FSH preparations contain less acidic FSH isoforms that have a shorter half-life than those derived from human urine but stimulate estrogen secretion as or even more efficiently.[233] The advantages of recombinant FSH preparations include the absence of urinary protein, more consistent supply, and less batch-to-batch variation in biologic activity. The two recombinant FSH preparations currently available are marketed

as follictropin alpha and follitropin beta. They are both structurally identical to native FSH and contain 1 alpha and 1 beta glycoprotein chain, but the post-translational glycosylation process and purification procedures for the two are different.[234] Despite the subtle differences in structure, they are functionally the same. The biological activity of all FSH preparations, including recombinant formulations, is ultimately confirmed using the classic Steelman-Pohley ovarian bioassay.[235]

Most recently, recombinant technology has been used to create a new chimeric gene containing the coding sequences of the FSH β-subunit and the C-terminal peptide of the hCG β-subunit (containing additional glycosylation sites). Co-expression of the α-subunit and the chimeric FSH β-subunit produces a new molecule called corifollitropin alpha, which has a prolonged half-life and enhanced *in vivo* bioactivity compared with wild-type FSH. Early studies in women suppressed by treatment with a long-acting GnRH agonist have confirmed the extended half-life of the compound and clinical trials have demonstrated that corifollitropin alpha can induce and sustain multifollicular growth for a week in women receiving ovarian stimulation for IVF. Corifollitropin alpha provides the means for simpler and more convenient treatment compared with conventional treatment regimens involving daily injections with FSH and GnRH agonist. Although the new recombinant gonadotropin is likely to find applications in assisted reproductive technologies where the goal is to induce multi-follicular development, it is not well suited for ovulation induction where unifollicular development is the objective.[236, 237]

A recombinant form of human LH having physicochemical, immunologic, and biologic acitivities comparable to those of human pituitary LH also is now available, supplied in vials with syringes designed to deliver 75 IU.[238–240] Combined use of recombinant LH and FSH (or hMG) helps to promote follicular development in women with hypogonadotropic hypogonadism who have a profound LH deficiency,[241] but otherwise does not appear necessary.[242, 243]

Traditionally, and still today, by virtue of its structural and biologic similarity to LH, hCG is used to simulate the LH surge and induce ovulation in gonadotropin-stimulated cycles once follicle development reaches maturity. LH/hCG promotes the final stages of follicular and oocyte maturation (from prophase I, the germinal vesicle stage, through meiotic maturation and metaphase II), which requires approximately 36 hours to complete, and ovulation generally occurs approximately 4 hours later. Although hCG extracted from human pregnancy urine and placental tissue still is in widespread use, a recombinant form of hCG also is available, produced using techniques similar to those described earlier for recombinant FSH.[244] The product became available in the U.S. in 2001 and has since grown rapidly in popularity. Although questions regarding the potency and dose equivalency of recombinant and urinary hCG remain, studies indicate that 250 μg of the recombinant product yields results comparable to those achieved with 5,000-10,000 IU of urinary hCG.[244–248]

The availability of recombinant FSH, LH, and hCG has done much to further our understanding of the specific actions of individual gonadotropins in follicular development and oocyte maturation.[249–251] Recombinant gonadotropins provide the capability to tailor ovarian stimulation regimens to the needs of the individual woman in an effort to optimize oocyte quality and cycle fecundity. Unfortunately, we still do not yet have the ability to accurately define what those specific needs are. It may someday be possible to design combinations of recombinant gonadotropins that will vary with the hormonal milieu of the individual, perhaps even within a cycle of stimulation, but for the present, our existing more generic treatment regimens must suffice.

Indications for Gonadotropin Treatment

Any discussion of ovulation induction with exogenous gonadotropins must first define the different clinical situations in which they may be used because the choice of gonadotropin preparation and treatment regimen vary with the type of ovulatory disturbance.

Hypogonadotropic Hypogonadism

Women with hypogonadotropic hypogonadism (hypothalamic amenorrhea, WHO Group I) are the most obvious candidates for ovulation induction with exogenous gonadotropins. Clomiphene and related medications typically are ineffective because their actions require an intact and functional hypothalamic-pituitary-ovarian axis. In a sense, gonadotropin therapy in women with hypogonadotropic hypogonadism may be viewed as hormone therapy intended to stimulate normal cyclic ovulation once fertility becomes a priority.

In women with hypogonadotropic hypogonadism, the drug of choice is menotropins because it contains both FSH and LH. Although follicular growth and oocyte maturation can be successfully stimulated with FSH alone,[252] LH also is required for normal steroidogenesis, luteinization, and ovulation[253–257]; endogenous LH levels typically are inadequate. ***Women with hypogonadotropic hypogonadism may respond to relatively low doses of gonadotropin stimulation, although treatment must nonetheless be carefully monitored and adjusted according to response. The objective, unifollicular ovulation, must be kept clearly in mind because hypogonadal women are otherwise normally fertile and at high risk for multiple pregnancy.***

The quality of luteal function after exogenous gonadotropin-induced ovulation in women with hypogonadotropic hypogonadism merits specific consideration. ***Although not always required,[258] luteal phase support with supplemental hCG (2,000–2,500 IU every 3–4 days)[259] or progesterone[260] generally is needed to compensate for low levels of endogenous LH secretion that can prove insufficient to support normal luteal function.*** Premenstrual spotting or a grossly short luteal phase suggests the possibility. Some have observed that supplemental hCG treatment can improve cycle fecundity,[259, 261] but its value has not been conclusively demonstrated, probably because endogenous LH levels vary in women with hypogonadotropic hypogonadism and only those with profoundly low LH concentrations (less than approximately 3 IU/L) may benefit from luteal phase support.[241, 262] Because supplemental hCG also increases the risk for ovarian hyperstimulation syndrome, treatment with hCG is best reserved for women who exhibit evidence of poor luteal function after ovulation induction; empiric treatment with progesterone is the obvious alternative.

Some women with secondary hypogonadotropic hypogonadism related to hyperprolactinemia become candidates for treatment with exogenous gonadotropins because they cannot tolerate dopamine agonist therapy. Consequently, it is important to know that hyperprolactinemia has no apparent adverse effect on the response to exogenous gonadotropins.[263]

Clomiphene-Resistant Anovulation

When clomiphene treatment fails to achieve ovulation, exogenous gonadotropins are an obvious option. Any of the alternative and adjuvant therapies discussed above also might be chosen in efforts to avoid the costs, logistical demands, and risks of

gonadotropin treatment, but failure with other such strategies is not a prerequisite for use of gonadotropins.

In women with hypogonadotropic hypogonadism, endogenous gonadotropin secretion is extremely low and menotropin (hMG) therapy provides the necessary gonadotropin stimulation. In contrast, serum gonadotropin concentrations in clomiphene-resistant anovulatory women with polycystic ovary syndrome (PCOS; WHO Group II) generally are normal and, in many, LH levels are relatively high. In this population of women, treatment with exogenous gonadotropins is superimposed on a background of erratic endogenous FSH and LH secretion. Purified FSH preparations offer a theoretical advantage over conventional menotropins because they avoid the risk of amplifying endogenous LH hypersecretion. However, in practice, there is no evidence that purified FSH has greater efficacy than hMG and either may be used. Numerous randomized controlled trials have compared purified urinary FSH with hMG therapy for ovulation induction in clomiphene-resistant anovulatory women with PCOS. A meta-analysis including 14 such trials found that purified urinary FSH was less likely than hMG to cause ovarian hyperstimulation (OR=0.20, CI=0.08−0.46), but has no other advantage.[264] Two other analyses of combined data from trials comparing recombinant FSH to purified urinary FSH or different recombinant FSH treatment regimens found no differences in the ovulation rate, pregnancy rate, miscarriage rate, multiple pregnancy rate, or incidence of ovarian hyperstimulation syndrome.[265, 266]

Like women with hypogonadotropic hypogonadism, clomiphene-resistant anovulatory women with PCOS generally respond to relatively low doses of gonadotropin stimulation. ***In many who are exquisitely sensitive, the therapeutic range is extremely narrow; doses only slightly higher than those proving ineffective can cause hyperstimulation.*** Treatment again must be carefully monitored and frequently requires small adjustments. Unifollicular ovulation remains the objective but often can be difficult to achieve. The risk for multiple pregnancy is high and risk of ovarian hyperstimulation is greater than in hypogonadal women.

Luteal phase support seldom is necessary after gonadotropin-induced ovulation in women with PCOS because endogenous LH levels typically are more than sufficient to support normal luteal function. However, in women also receiving treatment with a GnRH agonist to suppress endogenous gonadotropin secretion (discussed below)[267] and in others who may exhibit evidence of poor luteal function after otherwise successful ovulation induction, luteal phase support generally should be provided; considering the higher risk of ovarian hyperstimulation syndrome associated with hCG, progesterone therapy is preferable.[226, 268]

Unexplained Infertility

Exogenous gonadotropins can be used intentionally to stimulate the development and ovulation of more than one mature ovum in efforts to increase cycle fecundity in older subfertile women and those with otherwise unexplained infertility; superovulation is most effective when combined with timely IUI (Chapter 27). In this context, higher initial daily doses of exogenous gonadotropins are typically employed,[269] and because such women already ovulate normally and have no endocrinopathy, any of the available gonadotropin preparations can be used. Although superovulation is intended, careful monitoring is still required to avoid obviously excessive stimulation. The risk of multiple pregnancy is even greater than with ovulation induction in clomiphene-resistant anovulatory women, not surprising considering that superovulation is specifically intended. Luteal support is not required because the combined contributions of two or more corpora lutea may be reliably expected to yield supraphysiologic luteal phase serum progesterone concentrations.

Gonadotropin Treatment Regimens

Careful counseling and instruction are essential to the success of gonadotropin treatment. Couples must be thoroughly familiar with the medications prescribed, the methods for their preparation and injection, the need for frequent office visits to monitor response and reliable lines of communication, and the costs, prognosis, and risks associated with exogenous gonadotropin therapy.

Early retrospective studies established that daily treatment, frequently adjusted according to the clinical response, is the most effective treatment regimen.[270, 271] The dose and duration of gonadotropin treatment required to induce ovulation successfully varies among women, sometimes even among cycles within women, and must be determined empirically. Whereas many women are extremely sensitive to relatively low doses of gonadotropins (75–150 IU daily), others require substantially greater stimulation (300–450 IU daily). Although there is a direct relationship between body weight and dose requirement, the response threshold for a specific individual cannot be predicted reliably, even in the obese.[272] The treatment plan also must consider the intended goal, unifollicular ovulation or purposeful superovulation. ***Safe and effective ovulation induction with exogenous gonadotropins depends heavily on the experience and clinical judgment of the treating clinician.***

In both women with hypogonadotropic hypogonadism (WHO Group I) and those with clomiphene-resistant anovulation (WHO Group II), initial attempts to induce ovulation generally should begin with a low daily dose (75 IU daily) in a ***"step-up" treatment regimen*** designed to define the effective threshold of response. After 4 to 7 days of stimulation, a serum estradiol level, with or without transvaginal ultrasonography, provides the first measure of response. Thereafter, the dose of gonadotropins may be maintained or increased, as indicated. Once the serum estradiol level begins to rise, ovarian ultrasonography to determine the number and size of developing follicles becomes essential and the frequency of evaluation increases to every 1–2 days. When the mean diameter of the lead follicle reaches 16–18 mm, hCG is administered to trigger ovum release; ovulation generally may be expected to occur approximately 36–48 hours later. In subsequent stimulation cycles, the initial dose of gonadotropins should consider the response threshold and pattern of follicular development observed in previous cycles.

Because women with PCOS often are exquisitely sensitive to low doses of gonadotropin stimulation, early and frequent monitoring generally is wise. Such women typically have a larger number of small antral follicles poised to respond to FSH stimulation (recruitable follicles).[273] Ovarian hyperstimulation, higher risks of multiple pregnancy, and the expense and frustration associated with canceled cycles usually can be avoided by using a ***"low-slow" treatment regimen*** involving low doses (37.5–75 IU daily), small increments, and a longer duration of stimulation.[274–278] Although most gonadotropin stimulations span an interval of 7–12 days, low-dose stimulations in women with PCOS can take longer. Insulin-resistant women may be less sensitive to gonadotropin stimulation than those who are not.[279] In some such women, metformin treatment before and during gonadotropin stimulation can help to improve response, limit the number of smaller developing ovarian follicles,[280] and reduce the likelihood of cycle cancellation for excessive stimulation.[156]

The alternative ***"step-down" treatment regimen*** is designed to more closely approximate the pattern of serum FSH concentrations observed in spontaneous ovulatory cycles. Treatment begins with a higher dose (150–225 IU daily) and decreases gradually thereafter in an effort to promote continued development of only the more sensitive dominant follicle while withdrawing support from the less sensitive smaller follicles in the cohort. Considering that many anovulatory women are quite sensitive to low doses of exogenous gonadotropin stimulation, the step-down method generally is best applied only after the response threshold has been established in one or more previous stimulation cycles. However,

the two approaches can be effectively combined, first gradually increasing the dose of gonadotropins until a response is observed, and then decreasing the dose once a dominant follicle has emerged.

Recognition of the role that LH plays in the latter stages of follicular development, when FSH levels decline steadily, has suggested other approaches to ovulation induction with gonadotropins that may have particular value for women with PCOS, in whom standard treatment regimens too often result in multifollicular development and ovarian hyperstimulation. Although the selected dominant follicle is more sensitive to FSH than smaller follicles in the cohort, by virtue of it greater granulosa cell mass, FSH receptor content, and advanced microvascular development, the final stages of maturation are equally, if not more, dependent on low levels of LH.[250, 251, 281, 282] Whereas LH stimulates the theca (to produce androgens as substrate for estrogen synthesis) in all follicles, it also stimulates granulosa cells in larger follicles, via LH receptors induced by FSH and estrogen.[283–286] LH thus becomes the principal stimulus for the final stages of follicular maturation while declining concentrations of FSH starve the smaller, more FSH-dependent follicles into atresia.

Low doses of hCG[287] or recombinant LH[257] can selectively promote larger follicle growth while also hastening the regression of smaller follicles. To a limited extent, step-down gonadotropin treatment regimens, in which the amounts of FSH stimulation are gradually reduced, have exploited this phenomenon. The practice of "coasting," wherein FSH stimulation is withdrawn altogether during the latter stages of follicle development does so even more. In the latter instance, the largest follicles generally continue to function, most likely because their LH receptor expression renders them receptive to prevailing low concentrations of endogenous LH,[288] whereas estrogen levels plateau or decline and smaller follicles arrest or begin to regress.[289, 290] Continuing stimulation with low doses of hCG or recombinant LH after decreasing or discontinuing FSH treatment takes fullest advantage of the differential actions of LH in larger and smaller follicles by supporting continued development of the former[250, 281] and selectively excluding the latter,[282, 291] both by withdrawing FSH and by directly stimulating increased intrafollicular androgen concentrations.[292]

In women with hypogonadotropin hypogonadism or PCOS, recombinant LH treatment (225–450 IU daily) during the latter stages of follicular development can decrease the number of developing follicles.[291] In GnRH agonist-suppressed normal ovulatory women treated with 150 IU FSH daily for 7 days, a variety of treatment regimens involving combinations of decreasing FSH (50, 25, 0 IU) and increasing hCG (50, 100, 200 IU) have been observed to support the development of larger follicles and to speed the regression of smaller follicles.[293] Whereas either hCG or recombinant LH might be used, the longer half-life of hCG may help to provide a more stable level of LH activity between daily injections.[293] Interestingly, low-dose hCG treatment during the late stages of follicular development appears to have little effect on circulating progesterone or testosterone concentrations, at least in normal women, suggesting that the risk of causing premature luteinization or other adverse effects is low. By inducing the regression of smaller follicles, such treatment also may help to reduce the risk of ovarian hyperstimulation associated with exogenous gonadotropin therapy. The optimum sequence and relative amounts of FSH and LH/hCG to administer have not been defined and likely vary with the goals of treatment and the endocrinology of individual women.[291, 294, 295]

Some clomiphene-resistant anovulatory women can benefit from ***sequential treatment with clomiphene and gonadotropins***. The typical cycle involves a standard course of clomiphene treatment (50–100 mg daily), followed by low dose FSH or hMG (75 IU daily) beginning on the last day of clomiphene therapy or the next day; treatment is monitored and individualized thereafter as in standard gonadotropin-stimulated cycles. In most,[296–298] but not all studies,[299] cycle fecundity in sequential treatment cycles has approached or equaled that achieved with gonadotropins alone. In all, the dose and duration of gonadotropin therapy and the associated costs of monitoring were decreased significantly by 50%

or more. Logically, sequential therapy generally is useful only in women who respond to clomiphene, at least to some extent. Otherwise, treatment does not effectively begin until gonadotropin therapy starts.

The elevated endogenous LH levels in many clomiphene-resistant anovulatory women with PCOS predispose to premature follicular luteinization during exogenous gonadotropin stimulation[267, 300, 301] and have been implicated as a contributing factor in the higher incidence of spontaneous miscarriage observed in those who conceive.[302–305] ***Adjuvant treatment with a long-acting GnRH agonist*** before exogenous gonadotropin stimulation suppresses endogenous LH levels and continued GnRH agonist treatment during gonadotropin stimulation can prevent premature luteinization.[267, 301, 306] The risk that residual GnRH agonist-induced LH suppression might result in poor luteal function after ovulation induction appears more theoretical than real.[307]

Nonrandomized clinical trials have suggested that combined treatment with a GnRH agonist and exogenous gonadotropins can improve cycle fecundity in clomiphene-resistant anovulatory women.[301, 306, 308] However, randomized controlled trials comparing combined treatment with a GnRH agonist and exogenous gonadotropins to stimulation with gonadotropins alone have failed to demonstrate any differences in cycle fecundity or the incidence of ovarian hyperstimulation.[267, 307, 309, 310] Adjuvant GnRH agonist therapy also has no proven benefits for unselected subfertile women receiving gonadotropins to induce superovulation[311] and may even increase the amount and duration of gonadotropin stimulation required, at least in some.[267, 309] ***Although combined treatment with a GnRH agonist and exogenous gonadotropins is the established standard for controlled ovarian hyperstimulation in IVF cycles, it has no proven advantage over gonadotropin stimulation alone for ovulation induction.***

Monitoring Gonadotropin Therapy

To achieve ovulation but also avoid ovarian hyperstimulation and minimize the risk for multiple pregnancy, gonadotropin therapy must be carefully monitored with serial serum estradiol measurements and transvaginal ultrasonography. In effect, the clinician replaces the hypothalamus and pituitary in the feedback loop during treatment with exogenous gonadotropins. The chosen dose is administered, the ovarian response is measured and judged according to needs and expectations, and the gonadotropin dose is maintained or adjusted, re-evaluated, and readjusted as needed. Under normal circumstances, the hypothalamic-pituitary-ovarian axis performs the same task, constantly and repeatedly refining and coordinating the level of gonadotropin stimulation with the ovarian response. By contrast, the clinician can make no more than one such assessment daily, at most. Not surprisingly, the results achieved are relatively crude by comparison.

Serum Estradiol Levels

To best reflect the ovarian response to stimulation and provide for an efficient flow of information, gonadotropins generally are administered in the evening, typically between 5:00 and 8:00 p.m., and serum estradiol measurements are obtained early in the morning. Results usually are available for review by midday, and new instructions regarding the dose and duration of treatment and the next scheduled evaluation are communicated before the evening dose that day is due. In general, follicles less than approximately 10 mm in mean diameter produce relatively little measurable estrogen and larger follicles secrete progressively more as they grow and approach maturity. Usually, estradiol levels rise at a constant

exponential pace, doubling approximately every 2–3 days over the days before peak follicular development is achieved. A shallower or steeper slope of increase suggests the need to increase or decrease the level of stimulation.

In the natural ovulatory cycle, estradiol levels peak between 200 and 400 pg/mL just before the LH surge. Comparable levels of estradiol should be expected in gonadotropin-stimulated cycles, for each mature follicle observed. Clinical judgements also must consider the number and size of smaller follicles and their lesser but collective contributions to the serum estradiol concentration. ***Not surprisingly, cycle fecundability increases with serum estradiol levels; unfortunately, so do the risks of multiple pregnancy and ovarian hyperstimulation. With existing gonadotropin stimulation regimens, best results generally are obtained when estradiol concentrations peak between 500 and 1500 pg/mL; pregnancies are uncommon at levels below 200 pg/mL.***[312–316]

Ultrasonography

Ovarian ultrasonography defines the size and number of follicles contributing to the measured estradiol level. In the normal ovulatory cycle, the recruited cohort of antral follicles can be identified by cycle day 5–7, the dominant follicle emerges by day 8–12, grows approximately 1–3 mm per day thereafter (most rapidly over the 1–2 days immediately preceding ovulation), and measures approximately 20–24 mm in mean diameter when the LH surge occurs; lesser follicles rarely exceed approximately 14 mm in diameter.[92, 93] In 5–10% of spontaneous cycles, two preovulatory follicles may develop.

In exogenous gonadotropin-stimulated cycles, dominant follicles exhibit a similar linear growth pattern, but reach maturity at a smaller mean diameter and over a wider range of sizes. ***The likelihood of ovulation increases with follicular diameter.*** As judged by serial ultrasonography after hCG administration, follicles 14 mm and smaller occasionaly ovulate, but about 40% of those 15–16 mm, 70% measuring 17–18 mm, 80% measuring 19–20 mm in size, and virtually all larger follicles will ovulate.[317] The larger range of follicle size at maturity complicates clinical judgments. ***The risk of multiple gestation rises with the number of follicles likely to ovulate. Consequently, hCG generally should not be administered when the risk of multiple ovulation is high and the goal of treatment is unifollicular ovulation.*** A large number of intermediate and small follicles also increases risk for ovarian hyperstimulation syndrome.[318]

Baseline ovarian ultrasonography is prudent between consecutive cycles of stimulation with exogenous gonadotropins. In the absence of any significant residual ovarian cysts or gross enlargement, treatment can begin again immediately without need for an intervening rest cycle. Higher cycle fecundability and cumulative pregnancy rates have been observed in consecutive treatment cycles than with alternating cycles of stimulation and no treatment.[319, 320] When baseline ultrasonography reveals one or more residual ovarian cysts, it is usually best to briefly postpone further treatment. Stimulation cycles in the presence of ovarian cysts are less often successful,[321] possibly because newly emerging follicles can be difficult to distinguish from regressing cystic follicles, leading to errors in interpretation. Although many believe that suppressive therapy with a cycle of oral contraceptives helps to speed the regression of residual ovarian cysts, there is no evidence that such treatment is more successful than observation alone.

Studies of endometrial growth in exogenous gonadotropin-induced ovulatory cycles suggest that ultrasonographic measurements of endometrial thickness also have value. Cycle fecundity increases with endometrial thickness, which correlates with serum estradiol concentrations.[322] Few pregnancies result from cycles in which endometrial thickness is less than approximately 7 mm on the day of hCG-induced ovulation.[50, 313, 322, 323]

Results of Gonadotropin Treatment

Although exogenous gonadotropin therapy can successfully induce ovulation in over 90% of women with either hypogonadotropic hypogonadism (WHO Group I) or clomiphene-resistant anovulation (WHO Group II), the pregnancy rates achieved in the two populations differ significantly.[218, 324–327] ***In women with hypogonadotropic hypogonadism, cycle fecundity is approximately 25%, equal to or even greater than that observed in normal fertile women; cumulative pregnancy rates after up to six cycles of gonadotropin stimulation approach 90%. By comparison, cycle fecundity is significantly lower in clomiphene-resistant anovulatory women. Overall, cycle fecundity ranges between 5% and 15% and cumulative conception rates range between 30% and 60%; within the group, those with hyperandrogenic chronic anovulation have the poorest prognosis.***[218, 324–327] Although results generally do not vary with the duration of infertility or parity, pregnancy rates are significantly lower in women 35 years or older than in younger women.[326, 327]

The incidence of multifetal gestation is greatly increased in pregnancies resulting from exogenous gonadotropin-induced ovulation, even in anovulatory women where the goal of treatment is unifollicular ovulation. Whereas approximately 1 in 80 (1.25%) spontaneous pregnancies and 5–8% of those following clomiphene treatment are multiples,[102–104, 328] approximately 15% of all pregnancies following gonadotropin-induced ovulation in anovulatory women are multiples. [324, 326] Not surprisingly, the incidence of multiple gestation among subfertile women receiving gonadotropin stimulation for intentional superovulation is even higher and may approach 30% with nearly one-third being high-order multiple pregnancies (approximately 10% overall).[329] The higher frequency of multiple pregnancy after gonadotropin treatment obviously results from inadvertent or intentional multiple ovulation. Interestingly, however, there is some evidence to suggest that the normal frequency of monozygotic twinning (0.3–0.4%)[328] may be increased as much as 3-fold in pregnancy resulting from ovulation induction with exogenous gonadotropins.[330]

The overall incidence of spontaneous miscarriage in gonadotropin-induced conception cycles is approximately 20–25%,[218, 324–326] ***moderately higher than is generally observed (15%).*** A higher prevalence of advanced maternal age and obesity among women who receive gonadotropin therapy appear to contribute to the higher incidence,[331] but miscarriage rates also differ with the indication for treatment. In general, miscarriage rates are low in those with hypogonadotropic hypogonadism and significantly higher in clomiphene-resistant anovulatory women,[324–326] but not in all studies.[218] ***As with clomiphene, there is no evidence that gonadotropin therapy is associated with any increased prevalence of congenital anomalies.***[332]

Risks of Gonadotropin Treatment

In addition to the obviously greater costs and logistical demands involved, exogenous gonadotropin treatment also poses significant risks. Chief among these are the risks of multiple pregnancy and ovarian hyperstimulation syndrome. Neither can be avoided altogether, even by the most experienced clinician, but both risks can be reduced with careful management. As with any complicated form of treatment having significant intrinsic risks, thorough pretreatment counseling is essential.

Multiple Pregnancy

Twin births rose by 70% between 1980 and 2004, but have since plateaued (32.1 twins per 1,000 births in 2006); births of triplet and higher order multiple pregnancies more than

quadrupled between 1980 and 1998, but declined by 21% by 2006 (153.3 per 100,000 total births).[333] About 20% of the increase in multiple births, overwhelmingly twins, can be attributed to advanced maternal age and the societal trend towards older age at childbearing (older women are more likely to conceive a multiple pregnancy). The remainder, including almost all high-order multifetal gestations, results directly from the use of exogenous gonadotropins for ovulation induction, superovulation, and assisted reproductive technologies (ART).[334] The number of multiple pregnancies *conceived* is even greater because spontaneous and intentional multifetal pregnancy reductions do not appear in birth statistics.

Mutiple pregnancies are high risk pregnancies at any age because they are frequently complicated by preterm delivery, low birth weight, gestational diabetes, preeclampsia, and associated with high infant morbidity and mortality.[335, 336] Their clinical management often requires extended hospitalization, cesarean delivery, and neonatal intensive care; the associated health care costs are enormous, for both individual couples and society. ***In fact, evidence indicates that the combined costs associated with mutifetal pregnancies and their complications exceed those of all the treatments from which they derive.***[337] The less obvious social "costs" associated with multiple births also are high and include increased levels of parental stress, a higher incidence of maternal depression and child neglect or abuse, and a greater likelihood of behavioral problems among siblings.[338] Several factors contribute to the risks of multiple pregnancy associated with exogenous gonadotropin therapy. Although much of the attention in recent years has focused on embryo transfer practices in ART centers, less than half of all treatment-related multiple pregnancies results from IVF. The majority of multiple pregnancies, and the focus here, results from exogenous gonadotropin therapy for ovulation induction and superovulation.

Exogenous gonadotropins are an essential part of the therapeutic armamentarium with specific indications, and very real risks, including multiple pregnancy. ***Gonadotropins should be reserved for ovulation induction in infertile women with hypogonadotropic hypogonadism and clomiphene-resistant anovulation and for intentional superovulation in older subfertile women and those with otherwise unexplained infertility, including women who ovulate in response to treatment but ultimately fail to conceive.*** If unnecessary risk is to be avoided, the objective (unifollicular ovulation vs. intentional superovulation) must be kept clearly in mind; there is seldom an indication for superovulation in anovulatory but presumably otherwise fertile women.

Many infertile women seek the most aggressive forms of treatment simply because they offer the greatest chance for success, finding it hard to believe that any treatment could be *too* successful. Even those committed to avoiding excessive risks can find it very difficult to accept recommendations to cancel a treatment cycle, thereby forfeiting their investment of time and resources.[339] Financial pressures weigh heavily on the minds of even the most risk-averse patients and physicians. Cost considerations color perspectives and influence treatment decisions, tempting all involved to accept risks they otherwise would choose to avoid and later may regret. Some childless couples actually may hope for twins, but most are more circumspect when thoroughly counseled,[340] and none want triplets or more.

Multiple pregnancy is an intrinsic risk of intentional superovulation. In IVF cycles, the risk of multiple pregnancy relates to the number of embryos transferred, which the physician and patient control. However, the number of embryos that may implant is difficult to predict or control in superovulation cycles. Logically, risk might be reduced if ovulation simply was not triggered when the estradiol level or number of maturing follicles was excessive. Unfortunately, the response parameters that offer the best balance between increased cycle fecundity and the risks of multiple pregnancy and ovarian hyperstimulation have not been clearly defined and remain controversial.[341, 342] ***The risk of multiple pregnancy increases with serum estradiol concentrations, the total number of developing ovarian follicles, and with decreasing maternal age, but does not correlate well with the number of larger preovulatory follicles.***[49, 329, 343–346] Some have suggested varying cycle cancellation criteria

that might be used to guide treatment and limit risks of multiple pregnancies.[329, 343, 344] Whereas there is little doubt that withholding hCG when serum estradiol levels rise above approximately 900–1,400 pg/mL or ultrasonography reveals more than four to six follicles larger than 10–14 mm, application of such criteria also would dictate cancellation up to one-third of all exogenous gonadotropin-stimulated cycles.[346]

With relatively few exceptions among anovulatory women, exogenous gonadotropin therapy can be refined to achieve unifollicular ovulation with limited risk of multiple pregnancy and minimal risk of high-order multiple pregnancy. Conservative superovulation strategies likely can reduce the risk of multiple pregnancy associated with gonadotropin treatment, but any therapy wherein the specific objective is multifollicular ovulation logically cannot avoid the consequence. It is important to acknowledge that superovulation treatment exists primarily because practical considerations effectively prevent so many from pursuing the obvious alternative of IVF. In all likelihood, superovulation would fade into obsolescence if IVF was available to all those who need it, and few physicians or patients would lament its passing.

When ovarian stimulation exceeds its targeted goals, management options other than cycle cancellation include conversion to IVF and transvaginal aspiration of "excess" follicles. Unfortunately, the first of these is available only in centers where IVF also is offered. Most couples are unprepared for the change in plan and the substantial additional costs involved. The prognosis for success also may be poorer than in planned IVF stimulation cycles. Limited experience with the second option of aspirating excess follicles before hCG is administered to prevent ovulation of more than three ova suggests the strategy can effectively reduce the risk of multiple pregnancy and may be a legitimate alternative to cycle cancellation.[347, 348]

Multifetal Pregnancy Reduction

Women who conceive a high-order multiple pregnancy despite all efforts to avoid the complication must choose from among three difficult options. Termination of the entire pregnancy generally is unacceptable, particularly for those who have overcome infertility. Continuing the pregnancy carries the inherent risks of preterm birth and associated complications of increased neonatal morbidity and mortality and longer-term disability. Mutifetal pregnancy reduction sacrifices a portion of the pregnancy in efforts to save the whole, but for many it is no option at all, for a variety of personal, moral, ethical, or religious reasons. For most, a choice between the risks of carrying a high-order multiple pregnancy and those associated with pregnancy reduction presents a most difficult dilemma. Couples that ultimately choose pregnancy reduction experience a rapidly shifting tide of strong emotions. Anxiety falls with the diagnosis of pregnancy, rises to very high levels with recognition of a multiple pregnancy, decreases to some extent after consultation before reduction, increases sharply during the procedure, and falls to lower levels after its completion.[349] In retrospect, two-thirds of couples recall acute emotional pain, stress and fear, and almost 20% report feelings of guilt and anger.[350]

In most cases, multifetal pregnancy reduction is performed under transabdominal ultrasound guidance between 11 and 14 weeks gestation. By then, the possibility of spontaneous reduction has passed[351] and a limited screen for gross structural anomalies and features of aneuploidy can be performed to guide selection of fetuses for reduction.[352] There have been no randomized controlled trials comparing maternal and neonatal outcomes in high-order multiple pregnancies managed expectantly with those in which a multifetal reduction was performed, and it is unlikely there will be.[353] A 2001 report from an international registry including over 3,500 reductions performed at 11 centers indicated that multifetal pregnancy reduction had an overall pregnancy loss rate of nearly 10% with approximately 4% of subsequent deliveries occurring between 25 and 28 weeks gestation (severe prematurity).[354] Both results compare favorably with published outcomes for series of unreduced

high-order multiple pregnancies.[355–358] Outcomes correlated with the number of fetuses both before and after reduction and improve with the experience of the operator. Outcomes were better for triplet pregnancies (6% pregnancy loss rate, 3% severe prematurity) than for quadruplet and higher-order pregnancies (12–22% and 4–11%) and improved as the number of fetuses remaining after reduction decreased from three (20% and 6.5%), to two (9% and 4%), to one (9% and 1.6%).[354] A 2004 systematic review of prospective studies comparing outcomes from multifetal pregnancy reduction with those from twin pregnancies (conceived spontaneously or after ART) found no differences between women having a multifetal reduction and women with a twin pregnancy for pregnancy loss (RR=1.32, CI=0.42−4.16), preterm birth (before 34 weeks; RR=0.20, CI=0.01−3.18), stillbirth (RR=0.86, CI=0.05−13.48), or neonatal death (RR=0.86, CI=0.05−13.45).[359] A comparison of outcomes of reduction to twins with expectant management of triplets observed that women who had reductions were less likely to suffer a miscarriage (RR=0.44, CI=0.24−0.81), to deliver before 36 weeks (RR=0.35, CI=0.22−0.60), to have infants with very low birth weight (<1500 g; RR=0.26, CI=0.14−0.45), and had fewer neonatal deaths (RR=0.20, CI=0.06−0.64).[359] ***Multifetal pregnancy reduction is an effective management tool for the complication of high-order multiple pregnancy, but one which all would prefer to avoid.***

Ovarian Hyperstimulation Syndrome

Ovarian hyperstimulation syndrome (OHSS) is an iatrogenic complication of ovulation induction with exogenous gonadotropins. The disorder also can be observed occasionally in clomiphene-induced cycles. Rare cases of OHSS in spontaneous pregnancies generally have been associated with conditions characterized by supraphysiologic concentrations of hCG (multiple gestations, molar pregnancy). Cases of recurrent OHSS in spontaneous singleton pregnancies in individuals and families have been described and linked to germ line mutations in the FSH receptor resulting in the loss of ligand specificity that permits activation by hCG.[360–362]

OHSS has a broad pathophysiologic spectrum ranging from mild illness to severe disease. The syndrome normally is self-limited and resolves spontaneously within several days, but may persist for longer durations in conception cycles. The characteristic feature of OHSS is an increase in capillary permeability resulting in a fluid shift from the intravascular to extravascular spaces,[268, 363, 364] probably mediated by increased ovarian secretion of vasoactive substances including vascular endothelial growth factor (VEGF), elements of the rennin-angiotensin system, and other cytokines.[365–369]

Risk factors for OHSS include young age, low body weight, PCOS, higher doses of gonadotropins, and previous episodes of hyperstimulation.[318, 370, 371] Risk increases with serum estradiol levels and the number of developing ovarian follicles and when supplemental doses of hCG are administered after ovulation for luteal phase support.[372–374] Traditionally, OHSS has been classified as mild, moderate, or severe, but is perhaps best viewed as a continuum with a widely varying number and severity of symptoms.

Mild illness is characterized by ovarian enlargement, lower abdominal discomfort, and mild nausea and vomiting, diarrhea, and abdominal distention and occurs in up to one-third of superovulation cycles.[371] In general, only oral analgesics and counseling to alert affected women to the signs and symptoms of progressive illness are required; intercourse may be painful and is best avoided to limit the risk of ovarian rupture.

Persistent or worsening symptoms or ascites signal progressing illness and require treatment with anti-emetics and more potent oral analgesics. Outpatient management usually is still feasible but must include careful monitoring of daily weights and urinary frequency, serial clinical examinations to detect increasing ascites, and laboratory evaluation

of hematocrit, electrolytes, and serum creatinine.[375] Oral fluid intake should be maintained at no less than 1 L/day; electrolyte-supplemented commercial drinks generally are well-tolerated and can help to maintain electrolyte balance. Strenuous physical activity is best avoided to reduce the risk of ovarian torsion,[376] but light physical activity is preferable to bedrest, which can increase the risk for thromboembolism. Weight gain greater than approximately 2 pounds daily and decreasing urinary frequency are indications for prompt clinical and laboratory re-evaluation. ***Pregnant women with OHSS merit particularly close monitoring because rapidly rising hCG levels increase the risk for progression to severe illness***. The severity of symptoms, inadequate pain relief, or social considerations may require hospitalization.

Serious illness is uncommon but not rare, having an incidence of approximately 1%. Characteristic features include severe pain, rapid weight gain, tense ascites, hemodynamic instability, respiratory difficulty, progressive oliguria and laboratory abnormalities. Hypotension can result from vascular volume depletion, oliguria from reduced renal perfusion due to low vascular volume or tense ascites, and dyspnea from ascites or hydrothorax. Hemoconcentration, reduced peripheral perfusion, and inactivity increase the risk of thromboembolism. Renal failure, adult respiratory distress syndrome, hemorrhage from ovarian rupture, and thromboembolic phenomena are potential life-threatening complications of OHSS.[377–380]

Hospitalization for more careful monitoring and aggressive treatment warrants serious consideration in women with severe abdominal pain or peritoneal signs, intractable nausea and vomiting, severe oliguria, tense ascites, dyspnea or tachypnea, dizziness or syncope, severe hyponatremia (sodium less than 135 mEq/L) or hyperkalemia (potassium greater than 5 mEq/L), hemoconcentration (hematocrit greater than 45%), or abnormal renal functions (serum creatinine greater than 1.2 mg/dL; creatinine clearance less than 50 mL/min) or abnormal liver functions (elevated transaminases).[268, 375, 378, 379]

Recommended inpatient care for hospitalized women includes frequent evaluation of vital signs, daily weights, measurements of abdominal circumference and fluid intake and output, chest X-ray and echocardiogram when pleural or pericardial effusion is suspected, pulse oximetry for those with pulmonary symptoms, and serial hematocrits, electrolytes, renal and liver function studies.[375] Intravenous fluid management must restore an effective plasma volume but not contribute unnecessarily to the accumulation of extravascular fluid. After initial rehydration, fluids should be administered judiciously in the lowest volumes necessary to maintain adequate urine output and relieve hemoconcentration; because of the tendency to hyponatremia, saline is preferable to lactated Ringer's solution. When saline fails, slow infusions (over 4 hours) of albumin (25%; 50–100 g at 4–12 h intervals) can effectively expand plasma volume.[381] Premature or excessive use of diuretics is counterproductive. Intravenous fluid support can be reduced substantially after diuresis begins and oral intake is re-established. Hyperkalemia may require specific treatment to move potassium into the intracellular space (insulin/glucose, sodium bicarbonate) or to prevent cardiac dysrhythmias (calcium gluconate).

Ultrasound-guided paracentesis can be very helpful in women with painful ascites, pulmonary symptoms, or oliguria that does not respond to fluid management.[380, 382, 383] Fluid should be removed gradually to avoid consequences of sudden fluid shifts and repeated as necessary. In rare women with persistent bilateral or severe pleural effusions, thoracentesis also may be required to relieve pulmonary symptoms.[384] Full-length venous support stockings are recommended and prophylactic heparin therapy (5,000 units every 12 hours) merits consideration in severely hemoconcentrated patients. When symptoms prevent ambulation, the use of an intermittent pneumatic compression device can help to reduce the risk of thrombosis. Clinical signs and symptoms suggesting thromboembolism demand prompt additional diagnostic measures and therapeutic anticoagulation when the diagnosis is confirmed or strongly suspected.

In the severest cases of OHSS, intensive care may be required for management of thromboembolism, renal failure, or deteriorating pulmonary function. Women with severe hyperstimulation and ovarian torsion or a ruptured ovarian cyst with hemorrhage who require surgical management present a challenge to anesthesiologists who are understandably seldom familiar with the pathophysiology of OHSS.[385]

Knowledge and prompt recognition of the risk factors for ovarian hyperstimulation are essential for its prevention. Rapidly rising serum estradiol levels, concentrations over 2,500 pg/mL, and observations of a large number of small and intermediate sized ovarian follicles are high risk indicators and signals to proceed with great caution. Cycle cancellation and less aggressive stimulation in a subsequent cycle warrant consideration. Coasting without further gonadotropin stimulation and delaying administration of hCG for 1–3 days until estradiol levels plateau or decline can reduce the risk of hyperstimulation.[386–390] A lower dose of hCG (5,000 IU) also may help to reduce risk.[268] Alternatively, a GnRH agonist (leuprolide 0.5–1.0 mg) to trigger an endogenous LH surge[391] or recombinant LH can be administered to induce ovulation,[392] thereby avoiding the longer duration of action and further stimulation of hCG.[393] Evidence from a trial involving 69 oocyte donors at high risk for developing OHSS suggests that treatment with a dopamine agonist (cabergoline 0.5 mg) may decrease vascular permeability (mediated via VEGF) and decrease the risk for OHSS.[394] When luteal support is judged necessary, exogenous progesterone administered by injection (50 mg daily) or vaginally (suppositories (100 mg or 8% gel, daily) are preferable to supplemental doses of hCG.[372]

Breast and Ovarian Cancer

The evidence suggesting that ovulation-inducing drugs might be associated with an increased risk of breast or ovarian cancer was reviewed when discussing the potential risks associated with clomiphene treatment earlier in this chapter. In brief summary, a pooled analysis of results from eight case-control studies found that fertility drug use among nulliparous subfertile women was associated with an increased incidence of borderline serous ovarian tumors (OR=2.43, CI=1.01−5.88) but not with any invasive cancers (OR=1.60, CI=0.90−2.87).[395] Although most studies have found no evidence that fertility drug use increases overall breast cancer risk, the results of one case-control study suggested that prolonged or repeated use of exogenous gonadotropins (six cycles or more) may increase risk.[396] Overall, the available data are quite reassuring. ***No causal relationship between exogenous gonadotropin treatment and breast or ovarian cancer has been established, although longer-term studies are warranted and prolonged treatment is best avoided, especially when there is little hope for success.***

Pulsatile Gonadotropin-Releasing Hormone

Exogenous pulsatile GnRH therapy has been used successfully to induce ovulation since 1980.[397, 398] Compared to gonadotropin therapy, GnRH treatment has both advantages and disadvantages. Once established, the method is relatively simple to use, requires no extensive and costly monitoring, and is associated with low risks for both multiple pregnancy and ovarian hyperstimulation. However, because GnRH therapy requires maintenance of an in-dwelling intravenous catheter for an interval of 2–3 weeks or longer, many women fear needle displacement or other technical problems and are reluctant to use the method or reject the option outright. Pulsatile GnRH therapy is currently unavailable in the United States, but is used widely elsewhere in the world.

Where it is available, synthetic GnRH comes in a crystalline form that remains stable for at least 3 weeks at room temperature after reconstitution in aqueous diluent. GnRH is administered in a continuous pulsatile fashion using a portable, programmable mini-pump that must be worn constantly, around the clock, requiring some logistical ingenuity during bathing and sleep. Although the drug may be administered intravenously or subcutaneously, the intravenous route requires lower doses involving less cost (2.5–5.0 vs. 15–20 μg/pulse), is more physiologic, and more effective. The drug is metabolized rapidly and has a terminal half-life of 10–40 minutes after intravenous administration. Compared to the brief spikes in serum levels that result from intravenous administration and effectively mimic the normal pattern of pulsatile hypothalamic GnRH secretion, subcutaneous treatment creates a more continuous low level of GnRH stimulation without definite peaks.

In effect, pulsatile intravenous exogenous GnRH therapy represents an artificial hypothalamus. In women with hypogonadotropic hypogonadism who have absent or low levels of endogenous pulsatile GnRH secretion, treatment restores a normal pulsatile GnRH rhythm. In those with other forms of ovulatory dysfunction, treatment superimposes a normal rhythm on an existing but disorganized pattern of endogenous GnRH secretion. ***Importantly, exogenous pulsatile GnRH treatment generally stimulates only normal physiologic levels of pituitary gonadotropin secretion and allows normal feedback modulation of the pituitary response by ovarian steroids and peptides to operate. Consequently, follicular recruitment, selection, growth, and development in women using the GnRH pump progress as they do in the normal menstrual cycle.***[399–401]

Indications for Pulsatile GnRH Treatment

Anovulatory infertile women with hypogonadotropic hypogonadism are the best candidates for ovulation induction with exogenous GnRH because treatment is specific, physiologic, and highly effective; the GnRH pump provides the only instructional signals the pituitary gonadotropes are likely to receive. Although the drug also can be used in women with other ovulatory disorders, it is much less often effective, probably because the pituitary has more difficulty interpreting the mixed signals of endogenous and exogenous GnRH stimuli. As often may be observed in women with polycystic ovary syndrome (PCOS), an increased BMI (greater than 24), serum LH (greater than 15 IU/L), serum testosterone (greater than approximately 100 ng/dL), and fasting serum insulin (greater than approximately 15 U/mL) are associated with lower ovulation rates in response to exogenous GnRH and lower pregnancy rates per ovulatory cycle.[402, 403] The GnRH pump also can be effective in women with hyperprolactinemia and offers an alternative to exogenous gonadotropins when dopamine agonist treatment fails or cannot be tolerated.

Exogenous GnRH Treatment Regimens

Exogenous GnRH is most effective when administered intravenously in low doses (2.5–5.0 μg/pulse) at a constant interval (every 60–90 min).[402] Those who fail to ovulate may respond to a higher dose (10–20 μg).[404, 405] As with clomiphene and exogenous gonadotropins, treatment should begin with a low dose and gradually increase to meet the needs of the individual because the risk of multiple pregnancy increases with the pulse dose.[406] To a large extent, the dose and duration of exogenous GnRH treatment required to induce ovulation depend on the underlying endocrine milieu.[402, 407, 408]

In women with primary hypogonadotropic hypogonadism, a low dose (2.5 μg/pulse) can induce ovulation effectively, but follicular phase LH concentrations may remain lower than

normal and luteal phase progesterone concentrations often are reduced; both are typically normal when a higher dose of GnRH (5.0 μg/pulse) is used. Longer durations of treatment typically are required because available stores of pituitary gonadotropins are markedly reduced due to the historically low levels of endogenous GnRH secretion. In women with secondary idiopathic hypogonadotropic hypogonadism, treatment should begin with a low dose of GnRH (2.5 μg/pulse); the higher dose (5.0 μg/pulse) is associated with higher follicular and luteal phase LH and estradiol levels, a short follicular phase, multiple folliculogenesis, and a higher risk of multiple pregnancy, possibly because previous pituitary or ovarian priming confers a greater sensitivity to GnRH therapy.[402, 409]

The endocrine response of women with PCOS to pulsatile exogenous GnRH (5.0 μg/pulse) is markedly abnormal, but can be normalized by pre-treatment with a long acting GnRH agonist (daily subcutaneous administration) for 6–8 weeks immediately before starting pulsatile exogenous GnRH treatment.[401, 402, 410] Without GnRH agonist pretreatment, follicular phase FSH, LH, and estradiol levels and luteal phase estradiol concentrations all are abnormally elevated. After preliminary down-regulation with a GnRH agonist, the endocrine characteristics of induced cycles are much improved and closer to those observed in spontaneous ovulatory cycles in normal women. Unless similar down-regulation with a GnRH agonist also precedes subsequent cycles, the response to exogenous GnRH therapy again becomes abnormal. The benefits of GnRH agonist pretreatment probably result from suppression of intraovarian androgen levels[411] and improved (higher) FSH/LH ratios before GnRH therapy begins.[412]

After ovulation has been achieved, GnRH therapy can continue at the same or a slower pulse frequency (every 120–240 minutes).[402, 404] Whereas either can stimulate sufficient endogenous LH secretion to support normal corpus luteum function, a slower frequency more closely approximates the reduced endogenous pulse frequency observed in normal cycles during the luteal phase and may help to reduce the cost of treatment. However, it is simpler, much less costly, and just as effective to discontinue the pump after ovulation has occurred and to support the luteal phase with small doses of hCG (2000 IU every 3 days)[402] or exogenous progesterone.

One of the advantages that GnRH pump therapy has over exogenous gonadotropin treatment is that monitoring is not required once an effective treatment regimen has been established. Serial estradiol measurements and ovarian ultrasonography certainly can be used to monitor ongoing treatment, but are not necessary. Objective evidence of ovulation can be obtained by BBT recordings or periodic progesterone measurements. If needed, the time of ovulation can be estimated more accurately by monitoring urinary LH excretion as in spontaneous or clomiphene-induced ovulatory cycles.

Results of Exogenous GnRH Treatment

Overall, ovulation rates in response to pulsatile exogenous GnRH therapy vary between 50% and 80% and cycle fecundability ranges between 10% and 30% in ovulatory cycles. [402, 405, 413] ***Results are best in women with hypogonadotropic hypogonadism and worst in those with PCOS.***[402, 405, 413, 414] In the former, cycle fecundity equals that observed in normal fertile women and cumulative pregnancy rates can reach 80% or higher after 6–12 cycles of treatment.[402, 404, 413, 415] In the latter, cycle fecundability and cumulative pregnancy rates are moderately lower, when ovulation can be successfully induced, and pretreatment with a GnRH agonist improves ovulation rates.[402, 404, 413]

Overall, pulsatile GnRH therapy can achieve ovulation and pregnancy rates that compare with or even exceed those observed with exogenous gonadotropin treatment in women with PCOS,[414] but how the two treatments might compare in eugonadotropic clomiphene-resistant

anovulatory women, arguably the more clinically relevant question, is much less clear because no studies have examined their relative efficacy in such a selected population. One small randomized controlled trial in which pulsatile GnRH therapy (10–20 μg/90 min) after GnRH agonist suppression (nafarelin 400 μg daily for 3 weeks or longer) was compared directly with clomiphene citrate treatment (50–150 mg daily, cycle days 3–7) as first-line ovulation induction strategies over two to three cycles observed similar ovulation and pregnancy rates in the two groups.[405] These data serve to emphasize again that more complicated and costly ovulation induction regimens involving GnRH pump therapy or exogenous gonadotropin treatment are best reserved for those who fail to ovulate in response to clomiphene citrate.

In addition to requiring less or no monitoring after an effective treatment regimen has been established, another advantage that pulsatile GnRH therapy has over exogenous gonadotropin treatment is that it rarely results in multiple follicular development and ovulation; the risk for multiple gestation is therefore substantially lower and that for serious ovarian hyperstimulation is eliminated almost entirely. In the largest single collected series including over 100 pregnancies in 600 GnRH-stimulated cycles in nearly 300 women with a variety of ovulatory disorders, only four multiple pregnancies (4% incidence, one triplet and three twin pregnancies) and no cases of moderate or severe ovarian hyperstimulation were observed.[402] In most other similar but smaller series, the incidence of multiple pregnancy has ranged from 7% to 9%.[404, 413, 414] Overall, the risk of multiple pregnancy in GnRH-induced conception cycles is comparable to that associated with clomiphene treatment (5–8%) and 40–75% lower than that associated with exogenous gonadotropin therapy in anovulatory women (approximately 15%).

The overall incidence of spontaneous miscarriage in exogenous GnRH-induced conception cycles is approximately 30%.[402, 413] As has been observed in most studies of pregnancies resulting from exogenous gonadotropin treatment, [325, 326, 416] miscarriage rates are lowest in women with hypogonadotropic hypogonadism (less than 20%) and highest in those with PCOS (over 40%).[401, 402, 413]

Taken together, the results achieved with pulsatile exogenous GnRH therapy support its use as the drug of choice for treatment of anovulatory infertile women with hypogonadotropic hypogonadism.[402, 413] Unfortunately, few clinicians have experience with the method and relatively few women judge it an attractive choice after considering the available alternatives.

Dopamine Agonists

Hyperprolactinemia and its treatment with dopamine agonists are considered at length elsewhere in this text in the context of their association with amenorrhea (Chapter 11). Relevant details from that discussion are summarized briefly again here in an expanded discussion focused on the use of dopamine agonists for ovulation induction.

The two most common dopamine agonists in clinical use are bromocriptine and cabergoline. Both are ergot alkaloids that mimic the actions of dopamine via their binding to dopamine receptors. Serum concentrations peak 1–3 hours after an oral dose of bromocriptine and very little remains in the circulation 14 hours after administration; an oral dose of 2.5 mg generally lowers prolactin concentrations for up to 12 hours.[417] When administered vaginally, the same dose of bromocriptine has peak effects approximately 10–12 hours later that are sustained for up to an additional 12 hours.[418] Cabergoline is a longer-acting dopamine agonist with high affinity for the dopamine receptor. A single dose of cabergoline effectively inhibits prolactin secretion for 7 days or longer.[419]

Like endogenous hypothalamic dopamine, the agonists inhibit pituitary lactotrope prolactin secretion directly. By lowering serum prolactin levels into the normal range, dopamine agonist treatment allows the hypothalamic-pituitary-ovarian axis to escape from the suppressive influence that hyperprolactinemia has on pulsatile GnRH secretion and to resume normal operation, thereby restoring ovulatory function. Because even prolactin-secreting pituitary adenomas remain sensitive to the actions of dopamine, the agonists are effective in hyperprolactinemic women with and without a pituitary adenoma.[420]

Indications for Dopamine Agonist Treatment

Dopamine agonists are the treatment of choice for hyperprolactinemic infertile women with ovulatory dysfunction who wish to conceive. Although some hyperprolactinemic women will respond to clomiphene treatment, most do not, because the neuroendocrine consequences of hyperprolactinemia generally disrupt the mechanism by which clomiphene exerts its therapeutic action.

Dopamine agonist treatment can be highly effective in women who have galactorrhea but normal serum prolactin levels.[421] With few exceptions, the presence of galactorrhea can be regarded as a reliable indicator of excess prolactin secretion. Possible explanations for occult hyperprolactinemia include excess production of biologically active forms of prolactin not detected in all immunoassay systems and transient but exaggerated noctural prolactin secretion that goes unrecognized in randomly drawn blood samples.[421–425]

Up to 30% of women with polycystic ovary syndrome (PCOS) can exhibit mild hyperprolactinemia.[426, 427] Reduced levels of dopaminergic inhibition also have been implicated as a contributing cause of the elevated serum LH concentrations observed in women with the disorder.[426, 428] Consequently, dopamine agonists also have been advocated as adjuvant therapy for hyperprolactinemic anovulatory women with PCOS who require exogenous gonadotropin treatment. Limited evidence suggests that pretreatment with a dopamine agonist can temper the ovarian response to exogenous gonadotropins and may thereby help to decrease the risks of multiple pregnancy and ovarian hyperstimulation associated with such treatment.[427]

Dopamine Agonist Treatment Regimens

Because many hyperprolactinemic women are very sensitive to low doses of dopamine agonists, treatment generally should begin with a low dose and increase gradually until the dose required to restore and to maintain euprolactinemia has been established. Although the dose ultimately required roughly correlates with the degree of hyperprolactinemia, many women with very high prolactin levels respond to relatively low doses of dopamine agonists. The dose of dopamine agonist required to maintain euprolactinemia very often is lower than that needed to achieve it initially.[429]

With bromocriptine, treatment usually begins with a dose of 1.25–2.5 mg, administered at bedtime to more effectively suppress the normal nocturnal increase in prolactin secretion. A low initial dose also helps to minimize the frequency and severity of gastrointestinal and cardiovascular side effects related to dopamine receptor stimulation.[430] ***Prolactin levels decrease and stabilize shortly after treatment begins and a repeated serum prolactin measurement will demonstrate the effectiveness of any given dose in as little as a week.*** If needed, a second dose can be added, administered with breakfast or lunch. Although most women respond to 2.5–5.0 mg bromocriptine daily, some may require as much as 10 mg daily.[13]

Cabergoline treatment usually begins with a dose of 0.25 mg twice weekly, increasing gradually thereafter about every 4 weeks until the effective dose is established. Most women achieve normal prolactin levels with 0.5–1.0 mg weekly and doses greater than 2.0 mg weekly rarely are required.[13, 429] ***Cabergoline has proven effective in 70–85% of hyperprolactinemic women who are resistant to or cannot tolerate bromocriptine treatment.***[13, 429, 431]

Exogenous gonadotropins are an effective alternative for the few who do not respond to a dopamine agonist, alone or in combination with clomiphene.

Results of Dopamine Agonist Treatment

Overall, dopamine agonist treatment normalizes and maintains normal prolactin levels in approximately 60–85% of hyperprolactinemic women. Cyclic menses are restored in 70–90%, usually within 6–8 weeks after treatment begins, and ovulatory cycles return in 50–75% of treated women with or without tumors.[13, 14, 429] The probability of successful treatment is modestly lower in women with markedly elevated prolactin levels (greater than 100 ng/mL) than in those with lesser degrees of hyperprolactinemia. Breast secretions typically diminish noticeably within approximately 6 weeks and complete cessation of galactorrhea generally takes about twice as long to achieve. After discontinuation of dopamine agonist therapy, hyperprolactinemia and associated menstrual dysfunction return in 75–80% of women.

A randomized controlled trial involving over 450 hyperprolactinemic amenorrheic women found that cabergoline was more effective than bromocriptine in achieving and maintaining normal prolactin levels, restoring menses and ovulatory function, and was also better tolerated; compliance with a twice weekly treatment (cabergoline) is also better than with a twice daily regimen (bromocriptine).[13]

Side Effects of Dopamine Agonists

Overall, side effects of dopamine agonist treatment are common, are most severe during the first 2 weeks of therapy, but generally are well tolerated. Because bromocriptine stimulates both D1 and D2 dopamine receptors, most women will experience mild adrenergic side effects;[430] dizziness, nausea, vomiting, nasal stuffiness, and orthostatic hypotension are the most common. Although cabergoline has similar side effects, they generally are less frequent and severe because of the drug's higher affinity for D2 dopamine receptors. Side effects are severe enough to require discontinuation of treatment in approximately 12% of women treated with bromocriptine and in 3% of those treated with cabergoline.[13]

Side effects can be minimized by starting with a low dose, increasing gradually thereafter as needed and tolerated. Taking the medications with a snack or meal also improves tolerance. When necessary, vaginal administration of bromocriptine or cabergoline can help to reduce side effects and improve compliance.[432–434] Whereas much of an orally administered dose is not absorbed or rapidly metabolized in the first pass through the liver, vaginal absorption is more complete and avoids immediate hepatic metabolism. Consequently, therapeutic results often can be achieved with lower doses when the drugs are administered vaginally.

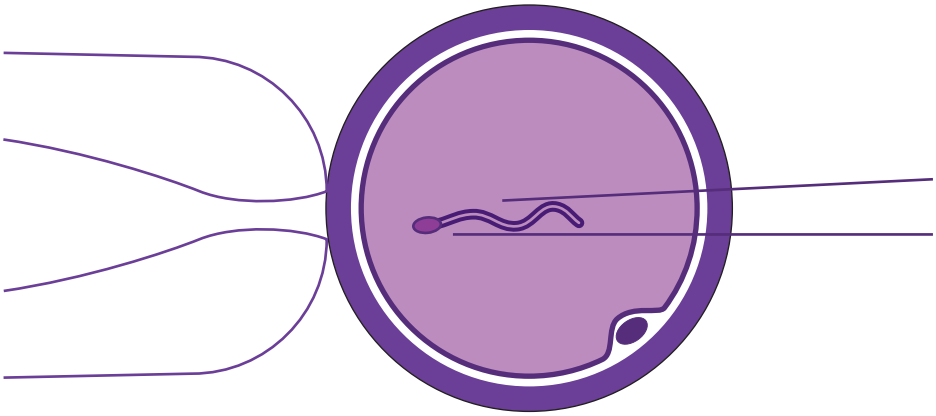
Risks of Dopamine Agonist Treatment

There is no evidence that dopamine agonists pose any increased risk for spontaneous miscarriage or birth defects. Numerous studies of women who have conceived during treatment have found no increase in the incidence of spontaneous miscarriage or congenital anomalies in pregnancies resulting from treatment with bromocriptine[13, 435–438] or cabergoline.[13, 418, 429, 431]

Treatment for Parkinson's disease with cabergline or another dopamine agonist, pergolide, has been associated with a 4- to 7-fold increased risk for valvular heart disease (mitral, aortic, or tricuspid regurgitation).[439, 440] The risk appears to result from mitogenic stimulation of normally quiescent valve cells via activation of serotonin (5-hydroxytryptamine, 5-HT) receptors (specifically, the 5-HT_{2B} receptor).[441] Bromocriptine, which has no 5-HT_{2B}-agonist activity, has not been associated with any increased risk for valvular heart disease. Although the doses of cabergoline used for treatment of hyperprolactinemia are less than 10% of those used for treatment of Parkinson's disease and have not been associated with the development of heart disease, it seems prudent to use the lowest effective dose of cabergoline for the shortest time required to achieve the goals of treatment. Current evidence does not support a recommendation for echocardiography before or during treatment with low doses of cabergoline in asymptomatic women.

All references are available online at: http://www.clinicalgynendoandinfertility.com

32 Assisted Reproductive Technologies



Assisted reproductive technologies (ART) encompass all techniques involving direct manipulation of oocytes outside of the body. The first and still most common form of ART is in vitro fertilization (IVF), but other related techniques also reside within the realm of ART. The success of modern ART has completely revolutionized both the evaluation and treatment of infertility. Some traditional diagnostic methods and treatments have been rendered obsolete and others have only limited applications because ART is simply more effective. The trend is clear and certain to continue.

IVF involves a sequence of highly coordinated steps beginning with controlled ovarian hyperstimulation with exogenous gonadotropins, followed by retrieval of oocytes from the ovaries under the guidance of transvaginal ultrasonography, fertilization in the laboratory, and transcervical transfer of embryos into the uterus. The first pregnancy resulting from IVF was reported in 1976, and was ectopic.[1] The first child resulting from IVF was born in 1978.[2] Over the more than 30 years since, ART has been greatly refined and expanded, resulted in millions of births worldwide, and now accounts for 1–3% of all births in the U.S. and Europe. ART includes methods for assisted fertilization by intracytoplasmic sperm injection (ICSI) using sperm isolated from the ejaculate or obtained by microsurgical epididymal sperm aspiration (MESA) or testicular sperm extraction (TESE), assisted embryo hatching, and preimplantation genetic diagnosis (PGD). In most cases, IVF is used to help an infertile couple conceive their own biological child, but donor sperm, donor oocytes, and gestational surrogates also play an important role in modern ART.

Other forms of ART include tubal transfer of oocytes and sperm (gamete intrafallopian transfer; GIFT), zygotes (zygote intrafallopian transfer; ZIFT), or embryos (tubal embryo transfer; TET) via laparoscopy. Whereas these more invasive techniques once had certain advantages over traditional IVF for some infertile couples, they now have only very limited indications.

A truly comprehensive discussion of ART is well beyond the scope of any single book chapter. The objective here is to provide an overview of the indications for ART, the most common methods for ovarian stimulation, oocyte retrieval, sperm recovery, fertilization, and gamete/embryo transfer, and the results and complications of ART, with emphasis on newly developing technologies and areas of controversy.

Indications for IVF

IVF was first developed as a method to overcome infertility resulting from irreparable tubal disease, but now is applied much more broadly for the treatment of almost all causes of infertility. IVF is most clearly indicated when infertility results from one or more causes having no other effective treatment; severe tubal disease relating to previous infection or advanced endometriosis and severe male factor infertility are the most obvious examples. IVF also is often the best treatment for couples with multifactor infertility because it can address or overcome all contributing causes at the same time. IVF is a legitimate treatment option for women with age-related or otherwise unexplained infertility and represents the treatment of last resort when other treatments fail.

In women with premature ovarian failure or reproductive aging and healthy women beyond normal reproductive age, IVF using oocytes from a young donor is highly successful. For women with normal ovaries but no functional uterus (müllerian agenesis, severe intrauterine adhesions, previous hysterectomy) and those with medical disorders that preclude pregnancy due to serious health risks, IVF with embryo transfer to a gestational surrogate still offers the possibility of genetic offspring. In couples who carry autosomal or sex-linked genetic disorders or balanced chromosomal translocations, IVF with preimplantation genetic diagnosis can avoid the risk of delivering an affected child.

Tubal Factor Infertility

Before the advent of IVF, women with irreparable bilateral tubal obstruction were essentially sterile, and the prognosis for those with less severe distal disease was only fair. In the modern era of ART, surgical treatments are declining in importance and the prognosis for women with tubal factor infertility has improved dramatically. Approximately 9% of patients using ART have a primary diagnosis of tubal factor infertility.[3] The relative advantages and disadvantages of surgery and IVF for the treatment of tubal factor infertility and the factors bearing on a choice between the two are discussed in depth in Chapter 27 and summarized here.

Reconstructive surgery remains a viable option for young women with mild distal tubal obstruction or peritubular adhesions (because postoperative live birth rates can exceed 50%),[4–6] but IVF is the treatment of choice for women with severe distal disease. Results achieved with surgery have varied, but success rates (10–35%) are generally lower than with IVF and the risk of ectopic pregnancy is higher (5–20%).[3, 7–10] In 2007, the overall IVF live birth rate (per cycle start) for U.S. women with tubal factor infertility (all ages) was 30.7%.[3] IVF is also the best treatment for women who remain infertile for more than a year after tubal surgery (the likelihood for success diminishes progressively with time after operation), for older women with significant distal tubal disease (cycle fecundity is low after distal tubal surgery and time is limited), and for women with recurrent distal tubal obstruction (repeated attempts to correct distal tubal occlusive disease are rarely successful).

Although not candidates for reconstructive surgery, women with severe distal tubal disease still can benefit from surgery before IVF. ***A substantial body of evidence indicates that communicating hydrosalpinges (proximal patency and distal occlusion) decrease the probability of both pregnancy and live birth after IVF by approximately one-half.*** The mechanism for the adverse effect of hydrosalpinges on IVF outcomes could involve mechanical interference with implantation or toxic effects on the embryo or endometrium.[11–15] A 2010 systematic review including five randomized trials involving 646 women observed that the odds of achieving an ongoing pregnancy were twice as great after laparoscopic salpingectomy for hydrosalpinges before IVF (OR=2.14, CI=1.23–3.73).[16] Laparoscopic proximal occlusion of the tubes also increased the odds of clinical pregnancy, compared to no intervention (OR=4.66, CI=2.47–10.01), and neither surgical procedure was superior.[16] Other treatments have been suggested, such as ultrasound-guided aspiration of hydrosalpinges after oocyte retrieval,[17] but are unproven, and evidence suggests the fluid re-accumulates rapidly.[18]

Proximal tubal occlusion observed during hysterosalpingography (HSG) often is not real and results from "cornual spasm" or other technical pitfalls of the procedure (Chapter 27). ***Efforts to confirm the diagnosis are justified; otherwise many women may needlessly undergo IVF.*** Common methods include repeated HSG[19] and laparoscopic "chromotubation."[20–22] Fluoroscopic or hysteroscopic selective tubal cannulation both establish the diagnosis and provide the means for successful treatment.[19, 20, 23–27] Microsurgical segmental resection and anastomosis is another proven treatment for true proximal tubal obstruction,[28–31] but requires uncommon technical expertise. IVF is the obvious alternative when cannulation is contraindicated (salpingitis isthmica nodosa) or technically unsuccessful, and when infertility persists for more than 6–12 months after the procedure.

Approximately 1 million U.S. women have an elective ***tubal sterilization*** procedure each year; up to 7% regret the decision and about 1% later request its reversal.[32, 33] The most commonly cited reasons for regret include new relationships, changes in family planning goals, and death of a child. Regrets are more common in younger women, those who were unaware of the spectrum of contraceptive options, women whose decision for sterilization was influenced by a third-party (partner, other family member, friend, or physician), and in those sterilized postpartum or after an abortion.[34, 35] Women 30 years old or younger are twice as likely as older women to express regret, 3.5 to 18 times more likely to request information about reversal of the procedure, and approximately eight times more likely to have a sterilization reversal or IVF.[36]

Young women sterilized using rings or clips and women having no other infertility factors have the best surgical prognosis; success rates are lower for older women, those sterilized by cautery (particularly multiple-burn techniques), and women with other infertility factors.[37–44] ***Although conception rates are quite good (45–82%) after microsurgical tubal anastomosis in properly selected candidates, IVF is a legitimate alternative to surgery, particularly for older women, those with a poor surgical prognosis or preferring to avoid surgery, and women who desire only one additional pregnancy.***

Endometriosis

The association between endometriosis and infertility and the pathogenic mechanisms involved are considered at length in Chapter 29. In brief summary, 20–40% of infertile women have endometriosis and accumulated evidence indicates that fertility decreases with the severity of the disease. Endometriosis may cause infertility by distorting adnexal anatomy and interfering with ovum capture,[45] or possibly by impairing oocyte development, early embryogenesis, or endometrial receptivity.[46–50] IVF should be expected to

overcome any anatomical obstacles, and although it would seem less likely to conquer functional disorders of oocyte, embryo, or endometrial development, outcomes in women with endometriosis suggest it can. Endometriosis is the primary diagnosis in approximately 5% of patients using ART.[3]

Treatment options for infertile women with advanced stages of endometriosis include conservative surgical treatment and IVF. For those with severe symptoms, surgery is the most logical treatment. Data from case series suggest that cumulative pregnancy rates 1–3 years after surgical treatment are approximately 50% for women with endometriomas,[51–54] and about 30% for women with complete cul-de-sac obliteration.[51, 55] Careful surgical technique is important because ovarian function can be compromised by excision of excessive tissue or damage to hilar vessels[56]; the risk of ovarian failure after excision of bilateral ovarian endometriomas is approximately 2.5%.[57] ***After surgical treatment, the choice between expectant management, empirical treatment, and IVF should be based on age, the surgical results, and the severity of any other coexisting infertility factors.***

Asymptomatic infertile women with advanced endometriosis, including those with ovarian endometriomas, can be treated surgically or proceed directly to IVF. There is no evidence to indicate that endometriomas have any important adverse effect on the response to ovarian stimulation or IVF outcomes.[58–65] Consequently, endometriomas can be left untreated before IVF. Aspiration of endometriomas before ovarian stimulation or at time of oocyte retrieval has been associated with an increased risk for developing an ovarian abscess,[66–68] although the risk appears quite low.[69]

Treatment options for asymptomatic women with known or suspected minimal or mild endometriosis and no other infertility factors include expectant management, surgical treatment, empiric treatment with clomiphene or exogenous gonadotropins and IUI, and IVF. In older women, those with other coexisting infertility factors, and women who have failed other forms of treatment, IVF is often the best overall choice.

Results of a 2006 systematic review including three randomized trials involving 165 infertile women with endometriosis of varying severity suggest that treatment with a GnRH agonist for 3–6 months before IVF can increase the odds of clinical pregnancy (OR=4.28, CI=2.0–9.15).[70] However, because prolonged treatment with a GnRH agonist also can decrease response to ovarian stimulation, most clinicians do not favor suppressive treatment before IVF.

Male Factor Infertility

Poor semen quality is the sole cause of infertility in approximately 20% of infertile couples and an important contributing factor in another 20–40% of couples with reproductive failure.[71, 72] Many infertile men have disorders than can be corrected medically or surgically if properly diagnosed and treated, allowing them to achieve natural conception with their partners.[72] In others, mild but important semen abnormalities can be overcome by IUI. ***When treatment is not possible or fails, and insemination with donor sperm is not an acceptable option, IVF and ICSI, using sperm isolated from the ejaculate or extracted from the epididymis or testis, offers realistic hope for success.*** The evaluation and treatment of male factor infertility are the focus of Chapter 30. Discussion here is limited to the indications for ART.[72]

The likelihood of male factor infertility is increased in men whose ejaculates consistently exhibit a sperm concentration under 15 million sperm/mL, less than 32% progressive motility, or fewer than 4% morphologically normal sperm (strict criteria, WHO III standard).[73]

The overall odds of male infertility increase with the number of abnormal parameters in the subfertile range; the probability is two to three times higher when one is abnormal, five to seven times higher when two are abnormal, and approximately 16 times greater when all three parameters are abnormal.[74] ***Additional genetic evaluation is indicated for men with severe oligospermia (sperm concentration <5 million/mL) whose sperm may be used for ICSI (Chapter 30).***

Medical or surgical treatment to improve or normalize poor semen quality is always the first and best option, when that is possible. When treatment is not feasible or proves unsuccessful, timely IUI can help to improve cycle fecundity in some couples with male factor infertility. ***Best results with IUI are achieved when the number of total motile sperm in the insemination specimen exceeds a threshold of approximately 10 million[75–77] and 14% or more of sperm have normal morphology (strict criteria; WHO III standard).[78]*** Higher counts do not further increase the likelihood for success[75, 79] and IUI is seldom successful when fewer than 1 million total motile sperm are inseminated.[80, 81] Success rates with IUI are best when 14% or more of sperm have normal morphology (strict criteria), intermediate with values between 4% and 14%, and generally quite poor when fewer than 4% of sperm are normal.[78] The likelihood of success with IUI also decreases with increasing female partner age and with coexisting infertility factors (ovulatory dysfunction, uterine and tubal factors).

When IUI is not possible, the prognosis for success with IUI is poor, or IUI proves unsuccessful and therapeutic donor insemination is rejected, IVF is the logical alternative. Approximately 18% of patients using ART have a primary diagnosis of male factor infertility.[3]

Conventional fertilization rates in IVF cycles are decreased when the total motile sperm count is less than 2–3 million (post-wash).[82] Numerous studies have observed that conventional fertilization rates also are decreased when less than 4% of sperm are morphologically normal.[83–87] Although severe teratospermia is widely accepted as an indication for assisted fertilization by ICSI, some observing no differences in fertilization, pregnancy, and live birth rates achieved with ICSI, compared with conventional fertilization, do not regard isolated teratospermia as an indication for ICSI.[88–91]

Ovulatory Dysfunction

For women with ovulatory disorders (hypogonadotropic hypogonadism, polycystic ovary syndrome, thyroid disorders, hyperprolactinemia), ovulation induction alone generally restores fertility (Chapter 31), but for some who require exogenous gonadotropins, ovulation induction proves difficult to achieve or consistently results in excessive ovarian stimulation and cycle cancellation for undue risk of ovarian hyperstimulation syndrome (OHSS) and high-order multiple gestation. For these difficult patients, IVF is an obvious treatment alternative, making their high sensitivity to gonadotropin stimulation an asset instead of a liability. Ovulatory dysfunction is the primary diagnosis in approximately 7% of patients using ART.[3]

Unexplained Infertility

The incidence of unexplained infertility ranges from 10% to as high as 30% among infertile populations, depending on diagnostic criteria.[92–94] For women with unexplained infertility, treatment options (cycle fecundability in parentheses) include expectant management (2–4%),[95] IUI (2–4%),[96, 97] empiric treatment with clomiphene (2–4%)[96, 98] or exogenous

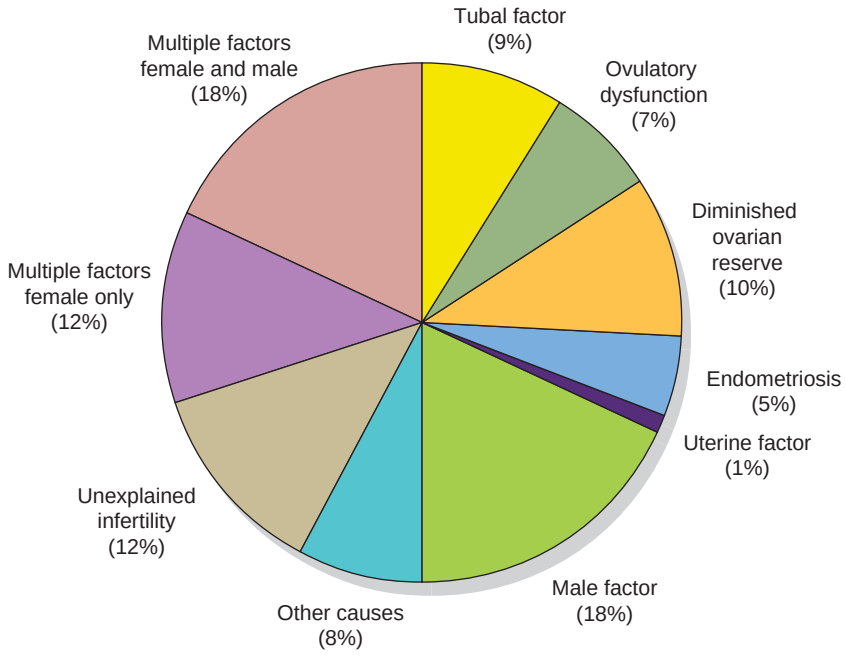
gonadotropins (5–7%),[99] combined treatment with IUI and clomiphene (5–10%)[100–102] or gonadotropins (7–10%),[99, 100, 103] and IVF (25–45%).[3, 104] As might be expected, success rates with all forms of treatment decline progressively with increasing age of the female partner.

Among couples with unexplained infertility, IVF is the preferred treatment for some and the treatment of last resort for others. ***In either case, there is no question that IVF is the most effective treatment for couples with unexplained infertility.*** A higher incidence of fertilization failure has been observed in several, but not all, studies of IVF outcomes in couples with unexplained infertility,[105–108] prompting many to recommend ICSI when IVF is planned. Approximately 12% of patients using ART have a diagnosis of unexplained infertility.[3]

Ovarian Failure and Diminished Ovarian Reserve

IVF using oocytes from a known or anonymous young donor was first developed for women with premature ovarian failure or menopause.[109] Now, oocyte donation is most commonly performed in women over age 42, those with grossly abnormal ovarian reserve test results, and women whose IVF cycles consistently yield poor quality embryos (Chapter 27). Approximately 10% of patients using ART have a primary diagnosis of diminished ovarian reserve.[3]

Diagnoses Among Couples Using ART[3]



Other Indications for IVF and Related Technologies

Although less commonly encountered, there are a number of other legitimate indications for IVF and related ART procedures.

Fertility preservation is fast becoming a more common indication for ART. Women with cancer or other illnesses requiring treatments (chemotherapy, radiation therapy) that pose a serious threat to future fertility may be candidates for urgent IVF and cryopreservation of

embryos before treatment begins, if time and health allow.[110] Oocyte cryopreservation is a viable option for women in similar circumstances having no male partner,[111] and is rapidly emerging as an option for young women at risk for premature ovarian failure,[112, 113] healthy aging women, and others who anticipate delayed childbearing.[111, 112, 114, 115]

For women with normal ovaries but no functional uterus, due to a developmental anomaly (müllerian agenesis), advanced disease (multiple myomas, severe intrauterine adhesions), or a previous hysterectomy, and for women with medical conditions that preclude pregnancy due to serious health risks, ***gestational surrogacy*** offers the opportunity to have their own genetic offspring.[116, 117]

For couples at risk for transmitting a specific genetic disease or abnormality to their offspring, IVF with ***preimplantation genetic diagnosis*** (PGD) provides the means to identify and exclude affected embryos and thereby avoid that risk. PGD is applied most commonly in couples who carry autosomal recessive and sex-linked disorders or harbor a balanced chromosomal translocation.[118] Women who carry a genetic disorder not amenable to diagnosis by PGD or who decline PGD may be candidates for oocyte donation. ***Preimplantation genetic screening*** (PGS) applies the same technology in couples having no known chromosomal or genetic abnormality in efforts to identify and exclude aneuploid embryos where the risk is increased, as in older women, those with a history of recurrent miscarriage, and in women with repeated unexplained IVF failure.[118] Although the technical limitations of current methods for PGS have so far prevented the technology from improving live birth rates in at-risk couples, more sophisticated and reliable methods now emerging hold promise.[119]

Prognostic Factors

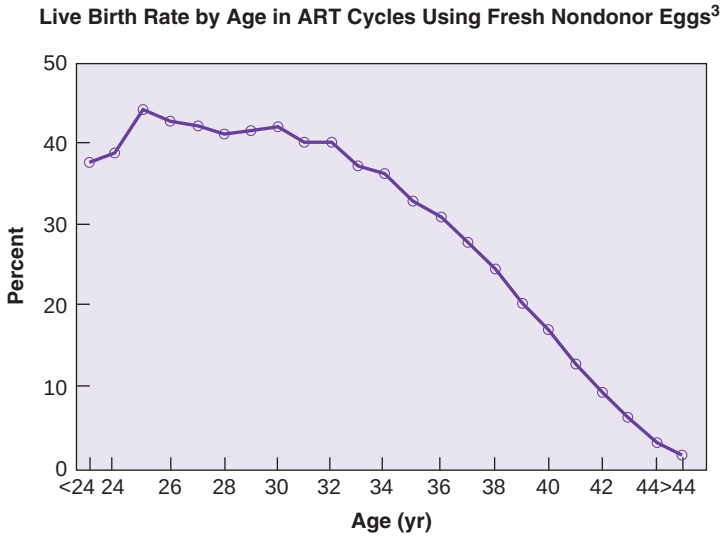
The probability for success with IVF relates to several factors, many of which are unfortunately not known until the treatment cycle is well underway (response to stimulation) or even nearing completion (number and quality of embryos). Before an IVF cycle begins, the primary prognostic indicators are maternal age, ovarian reserve, diagnosis, and past reproductive performance.

Maternal Age

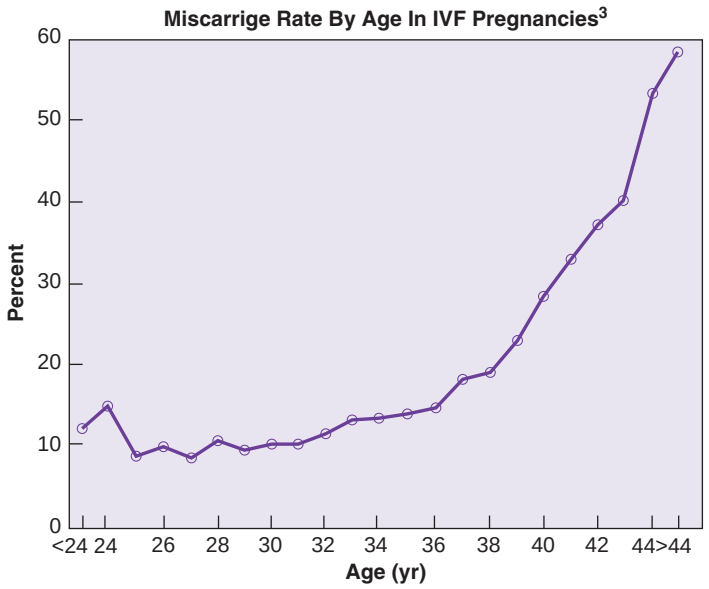
The relationship between maternal age and fertility and the physiology of reproductive aging are discussed in detail in Chapters 27 and 28 and only briefly summarized again here, where the focus is on the relationship between maternal age and IVF outcomes.

The average age of women using ART in the U.S. is 36 years.[3] ***Maternal age is the one most important factor in determining the likelihood for success with IVF.*** Although IVF can overcome most causes of infertility in younger women, it cannot negate or reverse the age-related decrease in biologic fertility in older women, particularly those over the age of 40.[120] The success rates achieved with IVF, like natural fertility rates, decline as maternal age increases. ***The pattern reflects a progressive decline in response to ovarian stimulation, resulting in fewer oocytes and embryos, and a decreased embryo implantation rate, due to declining oocyte quality.***[121–124] In 2007, the percentage of cycles that resulted in a

live birth from fresh nondonor oocytes, by maternal age, was 39.6% for women under age 35, 30.5% for ages 35–37, 20.9% for ages 38–40 yr, 11.5% for ages 41–42, and 5.4% for ages 43–44 years.[3] The pattern of decreasing success rates achieved with IVF parallels that associated with other, less complex, forms of treatment for infertility.[125]



Evidence from numerous lines of investigation indicates that the age-dependent decrease in success rates achieved with IVF relates primarily to an increasing prevalence of aneuploidy in aging oocytes,[126–128] which is reflected in the incidence of miscarriage in pregnancies achieved with ART: less than 14% for women under age 35, 19% at age 38, 28% at age 40, and nearly 60% for women over age 44.[3] In a case series of IVF cycles involving women ages 45–49, 70/231 cycles (30%) were cancelled before oocyte retrieval and 34/161 retrievals (21%) resulted in a pregnancy, but only 5/34 pregnancies (15%) and 5/231 cycles (2%) resulted in a live birth.[129]



Ovarian Reserve

The concept of ovarian reserve, generally defined as the size and quality of the remaining ovarian follicular pool, and the various methods for its measurement are discussed in detail in Chapter 27. In brief summary, the total number of oocytes in any given women is genetically determined and inexorably declines throughout life, from approximately 1–2 million at birth, to about 300,000 at puberty, 25,000 at age 40, and fewer than 1,000 at menopause.[128, 130–132] The rate of follicular depletion is not constant, but increases gradually as the number of follicles remaining decreases.[133–136] As the size of the remaining follicular pool decreases, circulating inhibin B levels (derived primarily from smaller antral follicles) decrease, resulting in lower levels of feedback inhibition and a progressive increase in serum follicle-stimulating hormone (FSH) levels, most noticeably during the early follicular phase.[137–145] Increasing inter-cycle FSH concentrations stimulate earlier follicular recruitment, resulting in advanced follicular development early in the cycle, an earlier rise in serum estradiol levels, a shorter follicular phase, and decreasing overall cycle length.[146–148]

The physiology of reproductive aging provides the foundation for all contemporary tests of ovarian reserve. In clinical practice, the basal early follicular phase (cycle day 2–4) FSH level is the most common test, but antimüllerian hormone (AMH) and antral follicle count are alternatives having significant potential advantages.

As basal FSH levels increase, peak estradiol levels during stimulation, the number of oocytes retrieved, and the probability for pregnancy or live birth decline steadily.[149–155] ***With current assays (using IRP 78/549), FSH levels greater than 10 IU/L (10–20 IU/L) have high specificity (80–100%) for predicting poor response to stimulation, but their sensitivity for identifying such women is generally low (10–30%) and decreases with the threshold value.***[156] Although most women who are tested have a normal result, including those with a diminished ovarian reserve (DOR), the test is still useful because those with abnormal results are very likely to have DOR. In a 2008 study, an FSH concentration above 18 IU/L had 100% specificity for failure to achieve a live birth.[157]

The basal serum estradiol concentration, by itself, has little value as an ovarian reserve test,[158–161] ***but can provide additional information that helps in the interpretation of the basal FSH level.*** An early elevation in serum estradiol reflects advanced follicular development and early selection of a dominant follicle (as classically observed in women with advanced reproductive aging), and will suppress FSH concentrations, thereby possibly masking an otherwise obviously high FSH level indicating DOR. When the basal FSH is normal and the estradiol concentration is elevated (>60–80 pg/mL), the likelihood of poor response to stimulation is increased and the chance for pregnancy is decreased.[162–165] When both FSH and estradiol are elevated, ovarian response to stimulation is likely to be very poor.

Antimüllerian hormone (AMH) derives from preantral and small antral follicles. Levels are gonadotropin-independent and vary little within and between cycles.[166–168] The number of small antral follicles correlates with the size of the residual follicular pool and AMH levels decline progressively with age, becoming undetectable near the menopause.[169–172]

Overall, lower AMH levels have been associated with poor response to ovarian stimulation and low oocyte yield, embryo quality, and pregnancy rates,[173–177] but studies correlating mean AMH levels with IVF outcomes have not yielded threshold values that can be applied confidently in clinical care.[158, 174, 175, 178] ***In the general IVF population, low AMH threshold values (0.2–0.7 ng/mL) have had 40–97% sensitivity, 78–92% specificity, 22–88% positive predictive value (PPV) and 97–100% negative predictive value (NPV) for predicting poor response to stimulation (<3 follicles, or <2–4 oocytes), but***

have proven neither sensitive nor specific for predicting pregnancy.[173, 179–181] In women at low risk for DOR, values of 2.5–2.7 ng/mL have had 83% sensitivity, 82% specificity, 67–77% PPV, and 61–87% NPV for clinical pregnancy.[159, 182] A study in women at high risk for DOR (involving older women, those with an elevated FSH, or history of poor response to stimulation) observed that an undetectable AMH had 76% sensitivity, 88% specificity, 68% PPV, and 92% NPV for three or fewer follicles.[174] A higher threshold value (1.25 ng/mL) had 85% sensitivity, 63% specificity, 41% PPV, and 57% NPV for cycle cancellation.[160]

The antral follicle count (AFC) is the total number of antral follicles measuring 2–10 mm in both ovaries during the early follicular phase and is a useful measure of ovarian reserve because it quantifies the number of follicles at the stage of development that responds to FSH stimulation.[183–187]

Several studies have observed a relationship between the AFC and response to ovarian stimulation in IVF cycles. In the general IVF population, including women at low and high risk for DOR, an AFC threshold value of three to four follicles has high specificity (73–100%) for predicting poor response to ovarian stimulation and failure to conceive (64–100%), but relatively low sensitivity for both endpoints (9–73% for poor response, 8–33% for failure to conceive).[160, 188–194] The PPV and NPV of AFC have varied widely in studies. ***A low AFC has high specificity for predicting poor response to ovarian stimulation and treatment failure, making it a useful test, but low sensitivity limits its overall clinical utility.***

In summary, none of the ovarian reserve tests currently in use is an accurate predictor of pregnancy in IVF cycles, unless extreme abnormal threshold values are applied, which results in very low sensitivity for identifying women having a poor prognosis.[157] The tests are adequate for predicting poor response, which does have prognostic value, although not as much in young women as in older women.[195–197] Although ovarian reserve tests have become a routine element of pre-treatment evaluation for couples planning IVF, it can be argued that routine testing has limited clinical utility in the large majority of patients and can be misleading, especially in women at low risk for having a diminished ovarian reserve.[198]

Ovarian reserve tests always should be interpreted with caution. Rigid application of test results risks inappropriate recommendations for treatment, or for no treatment, and both must be avoided. An abnormal test result does not preclude the possibility of pregnancy. ***Except perhaps when grossly abnormal, test results should not be used to deny treatment, but only to obtain prognostic information that may help to guide the choice of treatment and best use of available resources. Although the probability of pregnancy may be low, many with abnormal test results will achieve pregnancy if afforded the chance. Ultimately, regardless of the prognosis, the success rate for any individual woman will be zero or 100%.***

Diagnosis and Past Reproductive Performance

Although the average overall IVF live birth rate per cycle is approximately 29% for all women in the U.S., success rates vary, to some extent, with the cause of infertility. In 2007, the success rates for women with tubal factor infertility, ovulatory dysfunction, endometriosis, male factor, and unexplained infertility were above average, and those for women with multiple infertility factors, a uterine factor, and diminished ovarian reserve were below average. Whereas these data are useful, it is important to note that criteria for the different diagnoses are not standardized and likely vary among treatment centers.

Live Birth Rates in IVF Cycles, By Diagnosis, 2007[3]	
Infertility Diagnosis	*Live Births/Cycle*
Ovulatory dysfunction	37.3%
Male factor	35.8%
Endometriosis	34.3%
Unexplained infertility	31.8%
Tubal factor	30.7%
Multiple factors, female and male	27.5%
Uterine factor	26.9%
Multiple factors, female only	23.4%
Diminished ovarian reserve	15.3%

Women with a previous live birth are more likely to succeed with IVF than nulliparous women. In all age categories, success rates for women having one or more previous live births are modestly higher (2–4%) than for women with no previous live births.[3] A previous unsuccessful IVF cycle does not decrease the likelihood for success in subsequent cycles until approximately the fourth IVF cycle.[199] A history of an earlier unsuccessful pregnancy also has no effect on success rates.[3]

Other Prognostic Factors

As discussed earlier in the section focused on indications for ART, there is substantial evidence indicating that hydrosalpinges adversely affect IVF outcomes, and that salpingectomy or proximal tubal occlusion before IVF increases the likelihood for achieving a live birth by 2-fold.[16] A study evaluating the cost-effectiveness of preliminary salpingectomy concluded the procedure decreases the average cost per live birth, compared to no treatment.[200] ***Laparoscopic salpingectomy before IVF is generally recommended for women with hydrosalpinges.***

The effect of uterine myomas on IVF outcomes depends on their location. ***Submucosal myomas significantly decrease the likelihood for success, subserosal myomas have no significant impact, and the effect of intramural myomas is unclear.*** Overall, studies examining the effect of submucosal myomas on IVF outcomes indicate they decrease clinical pregnancy rates and delivery rates by approximately 70%,[201–207] and increase risk for miscarriage by more than 3-fold.[206, 207] A 2009 systematic review of studies examining outcomes after submucosal myomectomy concluded that clinical pregnancy rates achieved with IVF were 2-fold higher after surgery than in women with submucous myomas *in situ*, and comparable to those observed in women without myomas.[207] Results of individual studies examining the effects of intramural myomas on IVF outcome are inconsistent, with some observing an adverse effect,[208–212] and others not.[206, 213–218] ***Although systematic reviews have concluded that intramural myomas have significant negative impact on implantation rates and live birth rates,[204, 205, 219] there is no compelling evidence that their removal improves outcomes.[220]***

All smoking women should be strongly encouraged to stop smoking before IVF because smoking decreases the likelihood for success by up to one-half.[221–223]

Evaluation Before IVF

Individuals and couples planning IVF require additional specific evaluation before a treatment cycle begins. At a minimum, evaluation generally includes a test of ovarian reserve, a current assessment of semen quality, infectious disease screening, a trial transfer, and imaging of the uterine cavity.

Ovarian reserve tests (basal FSH, AMH, antral follicle count) have value for predicting response to gonadotropin stimulation and therefore can be helpful in planning treatment. If a threshold value with high specificity for detecting diminished ovarian reserve is applied, the test can accurately identify women at high risk for poor response and treatment failure.

Semen quality should be assessed not long before the treatment cycle is scheduled to start, even when earlier diagnostic evaluation revealed normal semen parameters, to ensure there has been no appreciable change that might affect the choice between conventional fertilization and intracytoplasmic sperm injection (ICSI). Evaluation of sperm morphology, as judged by "strict" criteria (WHO III standard), also may help to determine whether ICSI should be planned (Chapter 30).[83–87] Sperm cryopreservation is prudent when semen quality is severely abnormal or there is reason to anticipate difficulty with obtaining a fresh specimen on the day of oocyte retrieval. Although fertilization rates achieved with frozen thawed sperm may be somewhat lower than when fresh sperm are used, pregnancy rates are comparable.[224, 225]

Infectious disease screening is recommended for both partners for human immunodeficiency virus (HIV), hepatitis B (hepatitis B surface antigen), hepatitis C (hepatitis C antibody), and syphilis (rapid plasma reagin), for the protection of medical and laboratory staff, the protection of any fetus that may result from IVF, and protection against the risk for cross-contamination of cryopreserved embryos in storage. Whereas some advocate routine testing for chlamydia and gonorrhea in the female partner, others choose to limit evaluation to women with tubal factor infertility or other risk factors.

A ***trial transfer*** helps to determine the technique required to achieve an atraumatic embryo transfer and to identify women whose transfer may be difficult to accomplish, although the orientation of the uterus can change when the ovaries are enlarged after stimulation.[226, 227]

Imaging of the uterine cavity a short time before a cycle of treatment identifies submucosal myomas or endometrial polyps that may interfere with implantation or have an adverse effect on pregnancy outcome. An HSG performed earlier during the diagnostic evaluation may suffice if entirely normal and relatively recent (within approximately 6 months), but sonohysterography and hysteroscopy are the more sensitive and preferred methods. Routine office hysteroscopy before IVF can be expected to identify potentially significant abnormalities such as polyps, myomas, adhesions, or septa in 10–20% of patients without symptoms.[228–230] Many prefer sonohysterography to hysteroscopy because it is easier to perform, highly sensitive, and also can detect hydrosalpinges and unsuspected ovarian pathology.[231–233]

Ovarian Stimulation Regimens

The ideal ovarian stimulation regimen for IVF should have a low cancellation rate, minimize drug costs, risks and side effects, require limited monitoring for practical convenience, and maximize singleton pregnancy rates. Numerous regimens have been described, ranging

from no stimulation (natural cycles), to minimal stimulation (clomiphene citrate) or mild stimulation (sequential treatment with clomiphene citrate and low dose exogenous gonadotropins), to aggressive stimulation (high dose exogenous gonadotropins, alone or in combination with a gonadotropin-releasing hormone agonist or antagonist). Ovarian stimulation has been a basic element of IVF for more than 25 years, but concerns about multiple pregnancies and the costs of IVF have sparked renewed interest in natural cycle IVF and mild stimulation regimens.

Natural Cycle

The first birth resulting from IVF derived from a single oocyte collected in a natural ovulatory cycle.[2] Compared to stimulated IVF cycles, natural cycle IVF offers a number of attractive advantages. Natural cycle IVF involves only monitoring the spontaneous cycle and retrieving a single oocyte before the midcycle LH surge occurs. It is physically less demanding, requires little or no medication, decreases costs by 75–80%,[234, 235] and all but eliminates risks for multiple pregnancy and ovarian hyperstimulation syndrome (OHSS). The chief disadvantages of natural cycle IVF are high cancellation rates due to premature LH surges and ovulation, and the comparatively low success rate, which is approximately 7%.[236]

When oocyte retrieval is based on detection of the midcycle rise in LH, careful and frequent monitoring is required and procedures are difficult to schedule efficiently. Alternatively, exogenous human chorionic gonadotropin (hCG) can be administered when the lead follicle reaches a size consistent with maturity, thereby better defining the optimum time for oocyte retrieval.[235] Adjuvant treatment with a GnRH antagonist also can be used to prevent a premature LH surge, but requires "add-back" treatment with exogenous FSH, and success rates are still quite low, ranging up to 14% per cycle in non-randomized trials.[237–240] In one large cohort study involving 844 treatment cycles in 350 good prognosis patients, the cancellation rate was 13%, the pregnancy rate was 8% per cycle and the cumulative pregnancy rate after three "modified natural IVF cycles" was 21%.[241] In a cohort of infertile couples with male factor infertility, success rates in modified natural cycles have reached as high as 13% per cycle, with a cumulative pregnancy rate of 44% after six treatment cycles.[242]

Clomiphene Citrate

Clomiphene citrate was the first method of ovarian stimulation used in IVF,[243, 244] but now has been almost entirely replaced by more effective stimulation regimens using human menopausal gonadotropins (hMG) or FSH, in combination with a GnRH agonist or antagonist.[245]

Clomiphene (100 mg daily) usually is administered for 5–8 days, beginning on cycle day 3, and induces development of two or more follicles in most normally ovulating women,[246–248] although egg yields (1–3) are only slightly greater than in unstimulated cycles and substantially lower than in cycles stimulated with exogenous gonadotropins.[248–250] Cycle cancellation rates are somewhat lower than in natural cycles and the numbers of oocytes retrieved, embryos transferred, and pregnancy rates are greater. As in natural cycles, exogenous hCG is administered when the lead follicle reaches mature size and a GnRH antagonist can be used to prevent a premature endogenous LH surge.

Sequential treatment with clomiphene (100 mg daily for 5 days) and modest doses of exogenous gonadotropins (150–225 IU daily beginning on the last day of clomiphene treatment or the day after) stimulates multifollicular development more effectively than treatment with

clomiphene alone.[251–253] Drug costs and monitoring requirements are moderately higher, but still substantially less than in standard stimulation regimens involving higher dose gonadotropin treatment after down-regulation with a long-acting GnRH agonist (described below).[254, 255] In one comparative trial, higher cancellation rates and lower pregnancy rates were observed in sequential clomiphene/gonadotropin cycles.[255] In another, the sequential stimulation regimen yielded fewer oocytes and embryos, but pregnancy rates were similar and the risks of OHSS were lower.[254] Adding a GnRH antagonist to the treatment regimen can prevent premature LH surges and improve outcomes, but also increases costs. In a randomized trial, sequential clomiphene/gonadotropin stimulation and GnRH antagonist treatment yielded a pregnancy rate comparable to that achieved with a more aggressive standard treatment protocol,[256] confirming the results of two earlier retrospective studies,[257, 258] but contrasting with those of another observing lower pregnancy rates.[259]

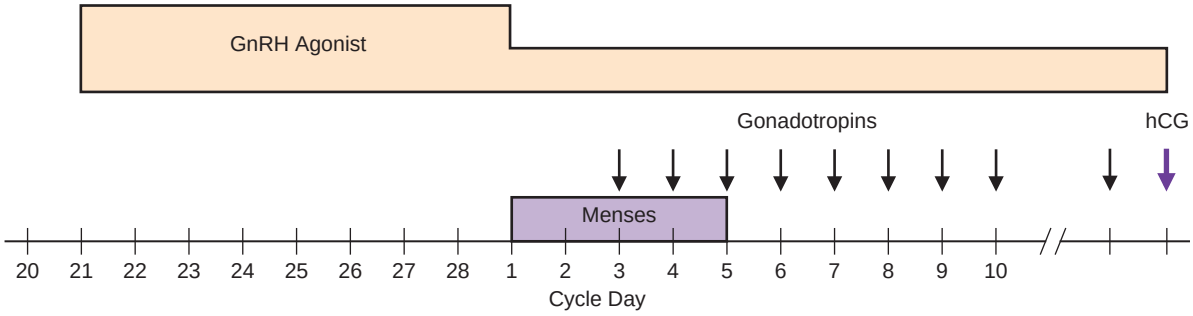
GnRH Agonist Down-Regulation Gonadotropin Stimulation—The "Long" Protocol

The introduction of long-acting GnRH agonists in the late 1980s revolutionized the approach to ovarian stimulation in ART by providing the means to suppress endogenous pituitary gonadotropin secretion and thereby prevent a premature LH surge during exogenous gonadotropin stimulation. Adjuvant treatment with a GnRH agonist eliminated the need for frequent serum LH measurements and assuaged fears of premature luteinization which previously had required cancellation of approximately 20% of all IVF cycles before oocyte retrieval.[260–262] Because fewer than 2% of cycles are complicated by a premature LH surge after down-regulation with a GnRH agonist,[263] stimulation could continue until follicles were larger and more mature. Numerous clinical trials subsequently demonstrated that egg yields and pregnancy rates were significantly higher than in cycles stimulated with exogenous gonadotropins alone.[264, 265] Moreover, GnRH agonist treatment offered the welcome additional advantage of scheduling flexibility, allowing programs to coordinate cycle starts for groups of women simply by varying the duration of GnRH agonist suppression. Not surprisingly, the "long protocol" quickly became the preferred ovarian stimulation regimen for all forms of ART. Its only disadvantages are that GnRH agonist treatment sometimes blunts the response to gonadotropin stimulation and increases the dose and duration of gondotropin therapy required to stimulate follicular development. The combined costs of the additional gonadotropins and the agonist itself increase the total cost of treatment substantially. ***Nevertheless, because GnRH agonists have more advantages than disadvantages, the long protocol became and has remained the standard ovarian stimulation regimen in IVF cycles.***

In the typical cycle, GnRH agonist treatment begins during the midluteal phase, approximately 1 week after ovulation, at a time when endogenous gonadotropin levels are at or near their nadir and the acute release of stored pituitary gonadotropins in response to the agonist, known as the "flare" effect, is least likely to stimulate a new wave of follicular development.[266, 267] GnRH agonist treatment can be scheduled to begin on cycle day 21 (assuming a normal cycle of approximately 28 days duration), but most prefer to first confirm that ovulation has occurred by measuring the serum progesterone concentration. In women who do not cycle predictably, oral contraceptives (OC) can be used to control the onset of menses, starting GnRH agonist treatment 1 week before their discontinuation. In the U.S., leuprolide acetate (administered by s.c. injection) is the most commonly used GnRH agonist. In Europe and elsewhere, buserelin acetate (administered by s.c. injection or intranasal spray) and triptorelin (administered subcutaneously) are more common[268]; all work equally well. For leuprolide, the usual treatment regimen begins with 1.0 mg daily for approximately 10 days or until onset of menses or gonadotropin stimulation, decreasing to

0.5 mg daily thereafter until hCG is administered. A single dose of a longer-acting depot form of GnRH agonist (leuprolide, goserelin) offers greater convenience, but evidence indicates the total dose and duration of gonadotropin stimulation required are increased significantly when depot forms of the agonists are used.[269]

Gonadotropin stimulation begins after confirming that effective pituitary down-regulation has been achieved (serum estradiol level <30–40 pg/mL, no follicles >10 mm in diameter). Some women require longer durations of treatment to achieve suppression or may develop an ovarian cyst.[260] The significance of an ovarian cyst has been controversial. Whereas some investigators have observed that baseline cysts are associated with a poorer response to gonadotropin stimulation, decreased numbers of oocytes and embryos, and lower overall IVF success rates,[270–272] others have not.[273–277] ***Overall, the weight of available evidence suggests that women who develop cysts or require longer durations of GnRH agonist treatment to achieve suppression are more likely to respond poorly to gonadotropin stimulation and less likely to achieve pregnancy.*** Cyst aspiration immediately before stimulation does not appear to adversely affect response[278] and may even improve response in the aspirated ovary,[279] but probably is not warranted in women with a normal contralateral ovary.



The initial dose of exogenous gonadotropins must be tailored to the needs of the individual woman. Typical starting doses range between 150 and 300 IU of urinary FSH (uFSH), recombinant FSH (rFSH), or urinary menotropins (hMG) daily, depending on age, the results of ovarian reserve testing, and the response observed in any previous stimulation cycles. Either a "step-up" (beginning with a low dose, increasing as necessary based on response) or a "step-down" (beginning with a higher dose, decreasing as necessary based on response) can be used, but the latter approach is generally preferred. All contemporary gonadotropin preparations, including hCG, can be administered subcutaneously.

Numerous clinical trials and meta-analyses have compared outcomes in cycles stimulated with uFSH, rFSH, or hMG, with or without GnRH agonist pretreatment, concluding that there is no compelling evidence to indicate the superiority of one gonadotropin preparation over others.[280–283] However, a 2008 systematic review including seven trials comparing outcomes in cycles stimulated with rFSH or hMG, involving 2,159 patients, observed a significant increase in live birth rate with hMG (RR=1.18, CI=1.02–1.38); the pooled risk difference for live birth was 4%.[284]

Recombinant DNA technology has been used to develop a new longer-acting form of rFSH. Corifollitropin alpha is the product of a chimeric gene containing the sequences of the FSH-β subunit and the C-terminal peptide of the hCG-β subunit, which bears four O-linked glycosylation sites, and has a half-life three times longer than standard rFSH (95 vs. 32 hours).[285, 286] A single dose (100 µg for women ≤60 kg, 150 µg for those >60 kg) can induce and sustain multi-follicular growth for a week in women receiving ovarian stimulation for IVF. Corifollitropin has shown considerable promise in phase II trials, currently is being evaluated for safety and efficacy in large phase III trials, and is the first step

towards a new generation of recombinant gonadotropins.[287] The first live birth resulting from treatment with corifollitropin was reported in 2003.[288]

The low levels of LH secretion remaining after down-regulation with a GnRH agonist are sufficient to support normal follicular development in most women stimulated with uFSH or rFSH alone,[289] because only about 1% of LH receptors must be occupied to sustain normal levels of steroidogenesis.[290] However, in some women treated only with FSH, LH levels are markedly suppressed (<1 IU/L) and may be inadequate.[291, 292] In such cycles, the dose and duration of gonadotropin stimulation required is higher, peak estradiol levels are lower, and the numbers of oocytes and embryos may be reduced.[293, 294] Extremely low LH levels also may adversely affect fertilization, implantation, and pregnancy rates,[295–299] and have been associated with a higher incidence of biochemical pregnancy and early pregnancy loss.[300, 301] A 2007 systematic review and meta-analysis including 11 trials comparing stimulation with rFSH alone or in combination with recombinant LH (rLH) after GnRH agonist-induced down-regulation in IVF and ICSI cycles observed no significant differences in clinical or ongoing pregnancy rates.[302] However, in three trials including only poor responders, the pregnancy rate was higher in those receiving combined stimulation with rFSH and rLH.[302] ***In sum, the evidence indicates there may be a subgroup of women who could benefit from supplemental rLH or hMG during ovarian stimulation.*** In the absence of any reliable method for identifying such women, and in light of recent evidence suggesting that use of hMG may increase live birth rates,[284] many clinicians favor combined stimulation with FSH and hMG over stimulation with FSH alone.

The response to stimulation is monitored with serial measurements of serum estradiol and transvaginal ultrasonography. The first estradiol level usually is obtained after 3–5 days of stimulation to determine whether the chosen dose of gonadotropins requires adjustment. Thereafter, serum estradiol concentrations and sonography are obtained every 1–3 days, based on the quality of the response and the need to evaluate the impact of any further adjustments in the dose of gonadotropin treatment. In general, stimulation continues until at least two follicles measure 17–18 mm in mean diameter, when others typically measure 14–16 mm and the serum estradiol concentration reflects the overall size and maturity of the cohort. Most women require a total of 7–12 days of stimulation. It is important to emphasize that these parameters only approximate the goals of stimulation. In clinical practice, follicle measurements vary among observers and estradiol assays vary in their performance characteristics. Ultimately, each program must empirically establish its own thresholds, based on its own experience.

The endometrium is monitored during stimulation by measuring the endometrial thickness or "stripe" (the sum thickness of the two layers, measured in the mid-sagittal plane). Numerous studies have examined the prognostic value of endometrial thickness and pattern in ART cycles, but the issue remains unsettled. Many have suggested that results are best when endometrial thickness measures 8–9 mm or greater or appears "trilaminar," and poor when the endometrium is less than 6–7 mm in thickness or appears homogeneous on the day of hCG administration.[303–308] However, numerous others have failed to observe any clear correlation between endometrial thickness or appearance and outcomes.[309–314] Some have suggested that excessive endometrial growth (>14 mm) also is a poor prognostic indicator,[305, 315] but that too has been refuted.[316, 317] Overall, although measurements of endometrial growth are routine, their utility remains unclear. Consequently, changes in stimulation regimens and cycle cancellations based on endometrial thickness or appearance alone are difficult to justify.[318]

When the cohort of ovarian follicles reaches maturity, hCG (5,000–10,000 IU) is administered to stimulate the final stages of follicular development. The equivalent dose of the recombinant form of hCG now available is 250 µg.[319, 320] A 2005 systematic review including seven trials comparing recombinant and urinary hCG observed no differences in clinical outcomes.[321] The predictive value of the serum progesterone concentration

on the day of hCG administration has been debated vigorously, with some arguing that pregnancy rates were substantially lower when levels exceeded 0.9–1.0 ng/mL,[322–327] and others refuting the contention.[328–332] It is now clear that mildly increased progesterone levels are relatively common in women who respond well to gonadotropin stimulation and are a poor prognostic indicator only in poor responders.[333] ***Whereas it may be tempting to delay hCG administration in poor responders to afford smaller follicles the opportunity to further mature, the strategy is not likely to succeed and may be detrimental.***

Approximately 7–18% of stimulation cycles are cancelled before oocyte retrieval, most for lack of adequate response, and some for excessive response.[3] When the ovaries become grossly enlarged, containing large numbers of follicles of all sizes, and serum estradiol concentrations are markedly elevated (>5,000 pg/mL), the risk for OHSS increases substantially.[334–336] Management options in "high responders" include all of the following:

- Cycle cancellation.
- "Coasting," in which GnRH agonist treatment continues but without further gonadotropin stimulation for 1–3 days, administering hCG after estradiol levels moderate.
- Proceeding with oocyte retrieval and fertilization but freezing all embryos in lieu of transfer.
- Delaying transfer until 5 days after retrieval, while observing for clinical signs and symptoms of developing OHSS.

Canceling the cycle and starting anew using a more conservative stimulation regimen may ultimately decrease overall costs and maximize the chances for success.[337] The prognosis for high responders in subsequent cycles is generally very good. Dual suppression with both an OC (one pill daily for 21 days or more) and a GnRH agonist (leuprolide 1.0 mg s.c. daily, beginning 1 week before discontinuation of contraceptive treatment) can attenuate the response to subsequent lower dose gonadotropin stimulation.[338] Coasting allows larger follicles to continue growing but withdraws support from small and intermediate-sized follicles.[339, 340] Although approximately 20–30% of coasted cycles are ultimately cancelled, the strategy can help to reduce the risks for developing severe OHSS and avoid cancellation.[339, 341] Proceeding to oocyte retrieval and fertilization and freezing all embryos can salvage the cycle but avoid the greater risks of serious or prolonged OHSS observed in conception cycles.[342, 343] Delaying transfer until after symptoms abate and freezing all embryos when they persist is another option.[344]

The challenges presented by "poor responders" are far greater. Poor responders include women who develop few follicles (<3–5) despite high doses of gonadotropin stimulation or have relatively low peak estradiol levels (<500–1,000 pg.mL); there are no consensus criteria defining a poor responder. The prognosis is relatively poor for such women[345–347] and the important decision centers on whether to attempt stimulation again using a different or more aggressive treatment regimen.[348] Some of the more commonly employed options include the following:

- The long protocol, beginning with higher doses of gonadotropin stimulation.
- Decreasing the doses of GnRH agonist or discontinuing agonist treatment immediately before or soon after gonadotropin stimulation begins.
- A short follicular phase GnRH agonist treatment regimen using a standard or microdose "flare" protocol (described below).
- Using a GnRH antagonist (described below) instead of a long-acting agonist.

Higher doses of gonadotropin stimulation may generate a somewhat more vigorous follicular response, but doses greater than 450 IU daily generally have little or no additional benefit.[349–352] Decreasing the dose or discontinuing GnRH agonist treatment early

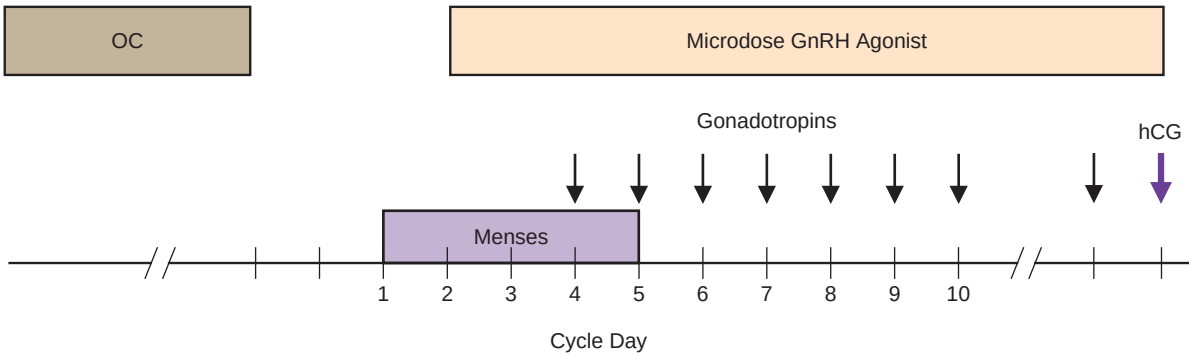
or completely may help to improve the quality of response.[353–357] A standard or microdose GnRH agonist "flare protocol" (described below) may stimulate an improved response in some poor responders.[358–360] Stimulation regimens employing a GnRH antagonist instead of a long-acting agonist eliminate any suppressive effects of the agonist altogether.[361] Other strategies have included efforts to increase androgen concentrations by treatment with dehydroepiandrosterone (DHEA)[362] or an aromatase inhibitor,[363] and the addition of growth hormone to the stimulation regimen.[364] A 2009 systematic review and meta-analysis of randomized trials comparing different stimulation regimens in poor responders found insufficient evidence to support the routine use of any particular intervention.[365] A 2010 systematic review including 10 trials involving eight different comparison groups reached the same conclusion.[345]

GnRH Agonist "Flare" Gonadotropin Stimulation Protocol

The "short" or "flare" protocol is an alternative stimulation regimen designed to exploit both the brief initial agonistic phase of response to a GnRH agonist and the suppression that results from longer-term treatment.[359, 366] In a typical standard short protocol, leuprolide acetate (1.0 mg daily) is administered on cycle days 2–4, continuing thereafter at a reduced dose (0.5 mg daily), and gonadotropin stimulation (225–450 IU daily) begins on cycle day 3. Later adjustments in the dose of gonadotropin stimulation, if needed, are based on response and indications for hCG administration are the same as in the long protocol (described above).

An early meta-analysis including seven clinical trials comparing the short and long GnRH agonist treatment regimens determined that the two protocols yielded similar cancellation and pregnancy rates.[264] A 2000 systematic review including 22 trials concluded that pregnancy rates achieved with the long protocol were superior to those using the flare regimen (OR=1.27, CI=1.04–1.56) overall,[265] but the analysis did not control for diagnosis and other prognostic factors and results may not apply to all women, or to poor responders in particular. Whereas some have observed improved follicular response and lower cycle cancellation rates in poor responders treated with a flare protocol, pregnancy and live birth rates remained low.[367, 368] Decreased scheduling flexibility is a distinct disadvantage of the flare protocol, unless the onset of menses is controlled by preliminary treatment with an OC. The regimen also can result in a significant increase in serum progesterone and androgen levels, presumably resulting from late corpus luteum rescue,[369, 370] which may adversely affect oocyte quality and fertilization and pregnancy rates.[371]

The "OC microdose GnRH agonist flare" stimulation regimen is a variation of the standard short protocol involving 14–21 days of preliminary ovarian suppression with an OC (one pill daily), followed by microdose leuprolide treatment (40 µg twice daily) beginning 3 days after discontinuation of OC treatment, and high-dose gonadotropin stimulation (300–450 IU daily) starting on day 3 of leuprolide therapy. Indications for later gonadotropin dose adjustments and hCG administration are the same as in other stimulation regimens. Its primary advantage over the standard short protocol is that it does not induce any increases in serum progesterone or androgen concentrations,[360] possibly because the doses of GnRH agonist administered are much lower, but likely also because preliminary OC treatment all but eliminates the possibility there may be a corpus luteum left to respond.[372, 373] ***The OC-microdose GnRH agonist flare protocol may be useful in previous poor responders, in whom it can stimulate increased endogenous FSH release and may yield lower cancellation rates and higher peak serum estradiol levels, transfer rates, and pregnancy rates.***[360, 374, 375]



GnRH Antagonist Gonadotropin Stimulation Protocol

The introduction of GnRH antagonists into clinical practice provided another option for ovarian stimulation in ART. In contrast to the long-acting agonists, which first stimulate and later inhibit pituitary gonadotropin secretion by desensitizing gonadotropes to GnRH via receptor down-regulation, the antagonists block the GnRH receptor in a dose-dependent competitive fashion and have no similar flare effect[376, 377]; gonadotropin suppression is almost immediate.

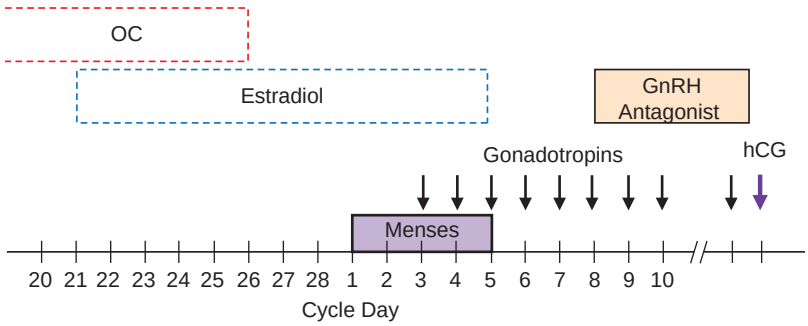
GnRH antagonists offer several potential advantages over agonists. First, the duration of treatment for an antagonist is substantially shorter than for an agonist. Since its only purpose is to prevent a premature endogenous LH surge and its effects are immediate, antagonist treatment can be postponed until later in follicular development (after 5–6 days of gonadotropin stimulation), after estradiol levels are already elevated, thereby eliminating the estrogen deficiency symptoms that may emerge in women treated with an agonist.[378] Second, because any suppressive effects that agonists may exert on the ovarian response to gonadotropin stimulation also are eliminated, the total dose and duration of gonadotropin stimulation required is decreased.[378, 379] For the same reason, GnRH antagonist stimulation protocols may benefit women who are poor responders when treated with a standard long protocol.[378, 380] Third, by eliminating the flare effect of agonists, GnRH antagonists avoid the risk of stimulating development of a follicular cyst. Finally, the risk of severe OHSS associated with use of antagonists also appears lower than with agonists.[381–383]

GnRH antagonists have some potential disadvantages. When administered in small daily doses, strict compliance with the prescribed treatment regimen is essential.[378] Antagonists suppress endogenous gonadotropin secretion more completely than agonists. Whereas the low levels of LH observed during agonist treatment are usually sufficient to support normal follicular steroidogenesis during stimulation with uFSH or rFSH, the even lower concentrations in women treated with an antagonist may not be. Indeed, serum estradiol levels may plateau or fall when antagonist treatment begins.[299, 378, 384] Although follicular growth appears unaffected, most prefer to add or substitute a low dose of hMG (75 IU) at the same time if it was not already part of the stimulation regimen. Evidence also suggests that pregnancy rates in antagonist treatment cycles may be modestly lower than in cycles using agonists in the long protocol.[385]

The two GnRH antagonists available for clinical use, ganirelix and cetrorelix, are equally potent and effective. For both, the minimum effective dose to prevent a premature LH surge is 0.25 mg daily, administered subcutaneously.[299, 386] Either can be administered in a series of small daily doses (0.25 mg). The treatment protocol may be fixed and begin after 5–6 days of gonadotropin stimulation,[299, 386, 387] or tailored to the response of the individual, starting

treatment when the lead follicle reaches approximately 13–14 mm in diameter.[388, 389] The individualized treatment regimen generally requires fewer total doses and may yield better overall results.[388] Alternatively, a single larger dose of cetrorelix (3.0 mg) will effectively prevent an LH surge for 96 hours. If given on day 6–7 of stimulation, the interval of effective suppression will encompass the day of hCG administration in most women (75–90%); the remainder may receive additional daily doses (0.25 mg) as needed, ending on the day of hCG treatment.[390–392] The single dose antagonist treatment regimen also can be withheld until the lead follicle reaches 13–14 mm in diameter.[393]

A common variation of the antagonist stimulation regimen uses preliminary treatment with an OC to control the onset of menses, typically ending approximately 5 days before the scheduled start, which also may help to synchronize the follicular cohort before stimulation begins. Another variation advocated for poor responders uses micronized estradiol (2 mg twice daily, administered orally, beginning on day 21 of the preceding cycle) to suppress FSH during the late luteal phase for the same purpose, ending on the day before gonadotroins stimulation begins,[393, 394] or continuing through the first 3 days of gonadotropin stimulation.[347] The improved follicular dynamics observed are similar to those achieved by down-regulation with a GnRH agonist in the long protocol. The rebound increase in endogenous FSH levels that follows the discontinuation of estradiol treatment also may synergize with exogenous gonadotropins to promote multifollicular development.[395, 396]



Results of a number of early trials comparing a fixed antagonist treatment protocol to the standard long protocol suggested that the two stimulation regimens yielded similar pregnancy rates.[379, 391, 397, 398] However, a 2006 systematic review and meta-analysis including 27 trials comparing different antagonist stimulation protocols with the long GnRH agonist protocol observed a significantly lower clinical pregnancy rate (OR=0.84, CI=0.72–0.97) and ongoing pregnancy/live birth rate (OR=0.82, CI=0.69–0.98). Overall, the total dose and duration of gonadotropin stimulation required, peak serum estradiol levels, and the number of follicles and oocytes were lower in antagonist cycles.

The explanation for the modestly lower pregnancy rates observed in antagonist treatment cycles is not clear. It is possible, but unlikely, that GnRH antagonists may have adverse effects on oocytes, embryos, or the endometrium.[399, 400] It is far more likely that early results reflected inexperience and improved with time and further refinements in the treatment regimen like those described above. Many of the advantages originally envisioned for GnRH antagonists already have been realized. Whether antagonists ultimately will replace agonists and become the standard ovarian stimulation regimen in ART cycles remains to be seen, but their place in the therapeutic arsenal already is firmly established.

Women with polycystic ovary syndrome (PCOS) characteristically exhibit high tonic LH secretion and are predisposed to premature LH surges when treated with standard ovulation induction regimens. Women with PCOS also are at increased risk for developing OHSS when aggressively stimulated with exogenous gonadotropins. Whereas both GnRH agonists and antagonists can suppress elevated circulating LH concentrations, the smaller

follicular cohorts observed in antagonist cycles may help to reduce the risk of OHSS in women with PCOS who tend to be high responders. ***The use of antagonists, rather than agonists, provides the opportunity to use an agonist instead of hCG to induce final oocyte maturation, thereby possibly further decreasing the risk of OHSS.***[401] Whereas a single bolus injection of an agonist (leuprolide 0.5 mg, triptorelin 0.2 mg) triggers a physiologic LH surge that lasts less than 24 hours, hCG levels remain elevated for several days and stimulate markedly higher estradiol and progesterone concentrations.[402]

The antagonist treatment regimens currently in use have potential disadvantages for women with PCOS. Their tonically elevated LH levels will remain high until antagonist treatment begins. Consequently, LH levels may rise prematurely, particularly if antagonist treatment is withheld until the lead follicle reaches 14 mm or more. Moreover, evidence indicates that increased LH exposure during early follicular development may be detrimental and predispose to lower pregnancy rates.[403–406] In theory, pretreatment with an OC might prove quite useful by suppressing LH and androgen levels before stimulation begins, decreasing exposure during early follicular development and the risk of rising LH levels before antagonist treatment starts. Preliminary OC suppression and later antagonist treatment may help to limit the follicular response to gonadotropin stimulation while preserving the option to use an agonist to trigger final oocyte maturation. These considerations simply serve to illustrate that GnRH antagonists are not a panacea and are not necessarily the best choice even for women with PCOS.

Antagonist stimulation protocols are advocated for poor responders, primarily because they avoid the suppressive effects that agonists can have on follicular response and can prevent the premature LH surges observed commonly in women stimulated with gonadotropins alone.[407] However, evidence is insufficient to indicate they yield results consistently better than other stimulation regimens.[345, 365]

Oocyte Retrieval

Oocyte retrieval is generally performed approximately 34–36 hours after hCG administration. Modestly longer intervals do not substantially increase the risk of ovulation or adversely affect oocyte quality, fertilization rates, or overall results in GnRH agonist downregulated stimulation cycles,[408–411] but earlier retrieval may yield fewer mature oocytes.[412]

Whereas oocyte retrieval once was performed via laparoscopy, transvaginal aspiration guided by ultrasonography under intravenous sedation is now the standard technique. Deep sedation (propofol) is most common, but most women tolerate the procedure very well with "conscious sedation" using short-acting narcotics (fentanyl) and benzodiazepines (midazolam), administered in small doses, as needed. There is no compelling evidence to indicate any difference in patient satisfaction or outcomes.[413] Constant monitoring by automated blood pressure recordings and pulse oximetry is essential to ensure that the proper plane of sedation is maintained and not exceeded. Specific reversal agents for narcotics (naloxone) and benzodiazepines (flumazenil) should be readily available.

Prophylactic antibiotic treatment (doxycycline 100 mg or cefoxitin 2 g), administered intravenously 30–60 minutes before retrieval is common but controversial because of the low incidence of infectious complications following retrieval (0.3–0.6%).[414, 415] Alternatively, oral antibiotics may be started immediately following the procedure (tetracycline, doxycycline), reserving prophylactic intravenous antibiotics for women at increased risk for infection (history of pelvic inflammatory disease, endometrioma).

Antiseptics (povidine iodine) are toxic to oocytes and limited evidence suggests their use may be associated with lower pregnancy rates.[416] When used to prepare the vagina before retrieval, thorough irrigation with sterile saline should follow, but repeated irrigation with saline alone is generally sufficient to cleanse the vagina. The bladder can become distended as a result of intravenous fluid administration, but can be drained immediately before retrieval; an indwelling catheter is unnecessary.

A vaginal probe (5–7 MHz) in a sterile plastic sheath with an attached needle guide is used to image the ovaries and to align the guide with the follicles in their largest diameter. A specially designed disposable 16–17 gauge needle is used to enter each follicle, in turn, and to aspirate the follicular fluid and oocytes. At the proper vacuum pressure (approximately 100 mm Hg), the follicle walls collapse but do not obstruct the needle lumen. Whereas some have observed that flushing and re-aspiration of follicles using a double-lumen needle can increase oocyte yield,[417] it is generally unnecessary and increases both operating time and analgesic requirements.[418] Efforts to minimize the arc swept by the needle within the ovary help to limit discomfort and ovarian trauma. In general, all follicles within the ovaries greater than 10 mm in diameter can be aspirated with no more than one to three separate entries on each side. Flushing the needle and attached tubing with media after each withdrawal helps to maximize oocyte yield. Abdominal pressure can sometimes stabilize a mobile ovary or move an ovary into a more convenient location for aspiration. Ovaries adherent to the posterior uterus often are more easily approached from the contralateral side but may be difficult to enter without traversing a portion of the uterus.[419] It may be more prudent to simply abandon some follicles, particularly when the number of oocytes already retrieved is sufficient.

The "empty follicle syndrome," characterized by a failure to retrieve oocytes despite apparently normal multi-follicular development, occurs in up to 0.5–1% of cycles.[420–422] The phenomenon can be observed when hCG is administered later than scheduled[423] or forgotten altogether,[421] and might rarely result from reduced biological activity in some lots of commercially prepared hCG.[424–426] The serum hCG concentration 36 hours after injection generally ranges between 100 and 300 IU/L.[425]

Serious complications of oocyte retrieval are uncommon. Limited vaginal hemorrhage from a puncture site is relatively common (8%) and usually can be controlled with a brief interval of direct pressure, but sometimes may require a suture.[415] Acute hemorrhage from the ovary and hemorrhage or hematomas resulting from injury to the uterine, ovarian, or iliac vessels are rare (0.04–0.07%).[415] The incidence of postoperative pelvic infections is quite low even without prophylactic antibiotic treatment (0.3–0.6%) and almost half present as tubo-ovarian abscesses, 1–6 weeks after retrieval.[414, 415] Women with ovarian endometriomas and those with past history of salpingitis are at highest risk.[66, 67, 427, 428] Other rare reported complications include the rupture of a dermoid cyst,[429] laceration of a sacral vein,[430] and lumbo-sacral osteomyelitis.[431] Potential complications include pelvic infection, adnexal torsion, and even vertebral osteomyelitis.[432]

Oocyte Maturation

Up to 20–30% of retrieved oocytes may be immature at the time of retrieval, reflecting the varying size and maturity of follicles in the cohort at the time hCG is administered. An accurate assessment of oocyte maturity is important to the timing of fertilization, even more so when ICSI is to be performed.

Like the LH surge in natural cycles, hCG triggers the resumption of meiosis in primary ooyctes previously arrested at prophase I of the first meiotic division. Oocyte maturity

generally can be judged by the expansion of the cumulus mass, radiance of the corona cells, the size and cohesiveness of granulosa cells, and the shape and color of the oocyte. When the cumulus mass is removed, as it is in preparation for ICSI, the oocyte can be further evaluated according to the presence or absence of the first polar body and germinal vesicle (nuclear membrane).

A mature (metaphase II) oocyte has extruded the first polar body and is in the resting phase of meiosis II. The cumulus cells are typically expanded and luteinized and the corona radiata exhibits a sunburst pattern. A metaphase I oocyte of intermediate maturity has no polar body and denser cumulus cells, but the germinal vesicle and nucleolus have faded. Metaphase I oocytes require additional time in culture before fertilization and must be examined periodically to document extrusion of the first polar body. A prophase I oocyte is grossly immature and exhibits a compact corona containing relatively few cumulus cells and a prominent germinal vesicle and nucleolus; dissolution of the germinal vesicle signals the resumption of meiosis I.

In Vitro Maturation

Human oocytes reach full size (100–200 μm) during the early antral stage of follicular development. The ability of an oocyte to resume and complete meiosis relates to follicular diameter.[433] Although immature oocytes collected from small antral follicles can mature with time in culture (the majority within 46–48 hours), even those that reach meiosis II do not necessarily acquire developmental competence, which requires synchronous maturation of both the nucleus and cytoplasm. Consequently, although they frequently fertilize, immature oocytes yield embryos that often develop poorly and exhibit low implantation potential.[434, 435] Nuclear maturation involves germinal vesicle breakdown, normally induced by the LH surge, followed by resumption of meiosis and, finally, extrusion of the first polar body. Cytoplasmic maturation is more difficult to define, but involves a number of factors that prepare the cytoplasm for fertilization and subsequent embryonic development.[436] Epigenetic processes are involved in both nuclear and cytoplasmic maturation and influence development after fertilization.[437, 438]

Technically, the term in vitro maturation (IVM) describes the maturation of immature oocytes in culture after their retrieval from follicles ***not*** exposed to exogenous LH or hCG *in vivo*. In efforts to improve the relatively low efficiency of classical IVM, new methods involving preliminary "follicular priming" have been developed.[439–441] One method involves FSH treatment for 3–6 days, followed by retrieval on cycle day 9–10. Another involves a single injection of hCG (10,000 IU), administered when the largest follicle reaches 10–12 mm in size and 36 hours before retrieval. A third method combines the two techniques, involving sequential treatment with FSH and hCG before oocyte retrieval.

Numerous studies have explored methods for IVM using oocytes obtained from normal women[442–447] and from women with PCOS.[442, 443, 448–454] Studies examining the effects of follicular priming *in vivo* have yielded inconsistent results, but embryos derived from primed oocytes generally have yielded higher implantation and pregnancy rates than those derived from oocytes collected from unstimulated antral follicles.[450, 452, 454] In one large trial examining the efficiency of IVM in women with normal ovaries, 400 women were randomly allocated to receive no priming or priming with hCG, FSH, or FSH and hCG.[455] The overall maturation rate and total number of available metaphase II oocytes were significantly higher in the groups receiving hCG than in those not receiving hCG. The overall clinical pregnancy rate per transfer was 18.3% and the implantation rate was 10.6%. Among the groups, the clinical pregnancy rate was higher in the group receiving both FSH and hCG priming (30%) than in all others.[455]

The best timing and method for efficient retrieval of immature oocytes from small follicles (<10 mm in diameter) have not been established.[447, 452, 456] Aspiration of follicles greater than 13 mm in size generally has yielded fewer oocytes, possibly because such follicles are already atretic.[457] A variety of aspiration vacuum pressures (80–300 mm Hg) and needles (16–20 guage) has been described.[447, 458] Whereas evidence is insufficient to warrant recommendation of any one method, extremely high vacuum pressures appear to adversely affect oocyte development *in vitro*.[459] The culture media composition that best supports IVM and the best method for fertilization of oocytes subjected to IVM also remain to be established; ICSI has achieved higher fertilization rates, but embryos derived from oocytes fertilized by conventional methods have exhibited higher implantation rates and yielded higher clinical pregnancy rates,[443] suggesting that ICSI is not required.

Although the clinical pregnancy rates achieved in IVM trials have been reasonably good, they do not approach those of standard IVF and have been achieved by transfer of a larger number of embryos. Implantation rates for embryos derived from IVM oocytes (5–22%) also are lower than those expected in similar women (age <35 years) receiving treatment with conventional IVF (34%).[3] Although the number of children derived from oocytes subjected to IVM is still quite small, preventing confident conclusions, the incidence of malformations and developmental abnormalities has thus far not differed from those in children resulting from traditional IVF or ICSI.

In summary, the results achieved thus far with IVM after follicular priming in vivo suggest the methods have real clinical promise. However, numerous questions must be answered before IVM can be recommended for wider clinical application.[460] Women with polycystic ovary syndrome (PCOS) having large numbers of antral follicles and greatest risk for developing ovarian hyperstimulation syndrome (OHSS) represent one population that could benefit from IVM, because purposeful retrieval of immature oocytes would require fewer days of gonadotropin stimulation. Women with cancer represent another, because many require immediate treatment that does afford them the time to pursue established methods of fertility preservation, which require ovarian stimulation, oocyte retrieval and oocyte or embryo cryopreservation.

Fertilization

Fertilization can be achieved by conventional microinsemination or by ICSI when there is a known or suspected male factor and poor or failed fertilization is a concern. In fact, male factor infertility is the one most common diagnosis among couples who undergo IVF. In the U.S. national ART summary for 2007, 18% of all cycles were performed for male factor indications and a male factor was one of multiple infertility factors in another 18% of cycles.[3]

A semen sample should be obtained by masturbation immediately before or after retrieval. The two methods most commonly used for sperm preparation before fertilization, the "swim-up" procedure and density gradient centrifugation, are described in detail in Chapter 30. Whereas both methods can successfully isolate a population of highly motile sperm for insemination, density gradient centrifugation also appears to select sperm with normal morphology and is widely regarded as the better choice when semen parameters are abnormal.[461–464] The isolated sperm are then incubated in media supplemented with a high concentration of protein for 0.5–4.0 hours to achieve capacitation.

In general, each oocyte is incubated with 50–100 thousand motile sperm for an interval of 12–18 hours at 37°C in 5% carbon dioxide in air at 98% relative humidity. The acrosome

reaction, which enables sperm to penetrate the zona pellucida, is initiated by contact between the sperm and the zona. In turn, sperm penetration triggers the cortical reaction which involves exocytosis of cortical granules from the ooplasm and renders the zona pellucida relatively refractory to penetration by more than a single sperm (polyspermy). Conventional IVF typically achieves fertilization rates ranging between 50% and 70%.

Sperm penetration also activates the oocyte and stimulates the second meiotic division, resulting in segregation of chromatids between the oocyte and the second polar body. Oocytes are evaluated for evidence of fertilization at approximately 18 hours after insemination. A normally fertilized oocyte exhibits two distinct pronuclei, one derived from the oocyte and the other from the sperm, and two polar bodies in the peri-vitelline space. The zygotes must be carefully inspected for the presence of extra pronuclei because polyploid embryos may cleave normally and go unrecognized at later stages of development. Polyploidy can be observed in up to 5–10% of embryos overall, but is far more prevalent in immature oocytes (up to 30%) than in mature oocytes (1–2%).[465, 466] Besides polyspermy, polyploidy may result from digyny (fertilization of a diploid oocyte), due to meiotic spindle errors or failure to extrude a polar body, which are more commonly associated with immature, aging, or postmature oocytes.[467, 468] The fertilization process requires approximately 24 hours and ends with the first mitotic division (cleavage).

Past failure of fertilization or severe male factor infertility requires ICSI, which yields pregnancy rates in couples with male factor infertility that compare favorably with those in couples without a male factor.[469] In the absence of a male factor, ICSI offers no clinical advantage over conventional IVF[470, 471]; in fact, evidence suggests that standard IVF yields higher implantation and clinical pregnancy rates.[3, 470]

When there is no ejaculate (aspermia) or only rare or no sperm (azoospermia) in the ejaculate, a variety of methods can be used to retrieve sperm for fertilization. Donor sperm also can be used, by design or as a contingency should efforts to retrieve sperm on the day of oocyte retrieval fail. Men with ejaculatory failure have no ejaculate or retrograde ejaculation. Ejaculatory failure may result from neurologic dysfunction or injury to the sympathetic outflow tracts that control emission and ejaculation (spinal cord injury, diabetes mellitus, multiple sclerosis, retroperitoneal surgery) or can be psychogenic in origin. Azoospermia may relate to ductal obstruction (obstructive azoospermia) or result from Sertoli cell-only syndrome, maturation arrest, or hypospermatogenesis (non-obstructive azoospermia). The diagnostic evaluation for aspermic and azoospermic men is described in detail in Chapter 30.

Sperm Retrieval Techniques

In the past, men with non-obstructive azoospermia were considered sterile and untreatable by any means other than the use of donor sperm. However, testis biopsy specimens in such men often demonstrate sperm,[472] suggesting low level production of sperm that do not survive epididymal transit to reach the ejaculate.[473] ***Whereas conventional wisdom was that sperm must traverse the male reproductive tract to acquire the ability to fertilize an oocyte, success with ICSI using epididymal or testicular sperm has demonstrated otherwise.*** Even grossly immature sperm (round spermatid nuclear injection; ROSNI) have been used to achieve fertilization, albeit with limited success.[474]

It is important to emphasize again that genetic evaluation and counseling are indicated for men with severe seminal abnormalities before their sperm are used for ICSI. Men with congenital bilateral absence of the vas deferens (CBAVD) or less severe forms of vasal aplasia, and their female partners, should be screened for cystic fibrosis gene mutations

before any attempts at pregnancy via ART to determine the risk for transmitting cystic fibrosis or CBAVD to offspring.[475–477] Men with non-obstructive azoospermia or severe oligospermia (less than 5 million/mL) should be offered karyotyping and screening for Y chromosome microdeletions.[477]

Sperm Recovery in Men with Retrograde Ejaculation

Men with documented retrograde ejaculation may be treated with sympathomimetics directed at control of the internal sphincter (imipramine 25 mg twice daily or 50 mg at bedtime, pseudoephedrine 60 mg, ephedrine 25–50 mg four times daily, phenylpropanolamine 50–75 mg twice daily). When medical treatment proves unsuccessful, sperm can be recovered directly from the bladder after masturbation; best results are achieved when urine pH and osmolality (300–380 mOsm/L) are carefully controlled by alkalinizing the urine (sodium bicarbonate 650 mg four times daily, beginning 1–2 days before collection) and controlling fluid intake.[478, 479] Alternatively, the bladder can be filled with buffered medium immediately before ejaculation.

Vibratory Stimulation and Electroejaculation

In men with psychogenic ejaculatory failure or spinal cord injuries below the T6 level, vibratory stimulation often can succeed in producing an ejaculate. Rectal probe electrical stimulation (electroejaculation) is recommended for men who fail vibratory stimulation and those with previous retroperitoneal surgery.[480, 481] Induced ejaculations may be retrograde and further require the procedures described above. Because electroejaculates frequently exhibit asthenospermia and teratospermia, ICSI is often necessary.

Epididymal Sperm Aspiration

Sperm can be obtained by microsurgical epididymal sperm aspiration (MESA) at the time of vasoepididymostomy or as an isolated procedure in men with CBAVD or uncorrectable obstructions. The technique involves incision of an isolated dilated tubule, gradually moving more proximally, if necessary, until sperm are obtained.[482, 483] Sperm are collected into a micropipette by capillary action with gentle compression of the testis and epididymis and flushed into a container with a small volume of IVF culture medium. Recovered sperm are cryopreserved in multiple aliquots for use in IVF cycles, if required.[484]

Percutaneous epididymal sperm aspiration using a fine needle also has been used successfully to obtain sperm and achieve pregnancy,[485, 486] but the technique is less reliable, the small quantities of sperm obtained are sometimes inadequate to allow cryopreservation, and the pregnancy rates achieved generally have been lower than with the open technique.

Testicular Sperm Extraction and Aspiration

In men with non-obstructive azoospermia and those in whom epididymal sperm aspiration techniques fail or do not apply, sperm can be retrieved using any of three other techniques. Open microsurgical testicular sperm extraction (TESE) yields the greatest number of sperm with potential for cryopreservation. Percutaneous core biopsy or aspiration of the

testis has also been described but is most applicable in men with normal spermatogenesis and obstructive azoospermia.[487, 488]

Using the preferred open microsurgical technique, sperm can be retrieved from the majority of men, even those with non-obstructive azoospermia. Magnification minimizes the risk of injury to the testicular blood supply, increases the probability of retrieving a blood-free biopsy specimen, and allows identification of larger caliber tubules that are more likely to yield sperm.[489, 490] Normal pregnancies have been achieved even in those with congenital or acquired testicular failure,[491] post-chemotherapy azoospermia,[492] and Klinefelter syndrome.[493]

In men with non-obstructive azoospermia, TESE is best performed on the day of or day before oocyte retrieval, when possible, and no earlier than approximately 6 months after any previous biopsy or TESE procedure, for several reasons. First, up to one-third of men with apparent non-obstructive azoospermia may exhibit sperm in their ejaculate on the day of planned retrieval and will not require TESE.[494] Second, sperm retrieved from men with non-obstructive azoospermia may not be motile or even viable after cryopreservation and thawing and ICSI using immotile sperm may yield poorer results than when performed with motile sperm.[495] Finally, the likelihood of successful retrieval of viable sperm for ICSI is significantly reduced when TESE is performed soon after a testis biopsy or previous TESE.[495] Matched donor sperm should be available in case they are needed, because TESE yields viable sperm in only about half of men with non-obstructive azoospermia.[489, 496, 497] When TESE cannot be performed near the time of oocyte retrieval, elective TESE can be performed and the recovered sperm cryopreserved; the risk of having no viable sperm after thawing is real but relatively small, and donor sperm can be used if needed.[498–500]

Intracytoplasmic Sperm Injection (ICSI)

Assisted fertilization techniques were developed to circumvent the need for sperm to penetrate the zone pellucida. A variety of methods have been described, but the success of ICSI has rendered all others obsolete.[501, 502] Earlier methods, including zona "drilling" (using a micropipette and acidified Tyrodes's solution or laser),[503, 504] partial zona dissection (opening the zona with a microneedle),[505] and subzonal insertion or insemination (injection of sperm beneath the zona in the perivitalline space),[506] still required sperm to interact with the oolemma and did not prevent polyspermic fertilization, but ICSI solved those problems.[507]

In the ICSI procedure, a single selected sperm is first immobilized by compressing the sperm tail with an injection pipette (inner diameter 5–7 μm), then drawn into the pipette. The oocyte is stabilized, usually with the polar body at the 6 or 12 o'clock position, and entered at the 3 o'clock position. The pipette pierces the zona and oolemma and the sperm is injected directly into the ooplasm. ICSI does not require sperm to undergo the acrosome reaction or to fuse with the oocyte membrane as occurs with natural fertilization. Instead, the mechanical disruption of the ooplasm and sperm membranes, facilitated by the sperm immobilization procedure and the gentle aspiration and reinjection of oocyte cytoplasm, triggers oocyte activation.[508–512] ***In most cases, ICSI achieves fertilization rates comparable to those observed with conventional IVF in the absence of male factors (50–70%).***

ICSI can damage the meiotic spindle even if the area adjacent to the first polar body is avoided, because the second meiotic spindle varies in position and is not always located immediately beneath the first polar body.[513, 514] A polarizing optical system that images the meiotic spindle can help to reduce the risk of spindle damage.[515]

The principal indication for ICSI is male factor infertility. Threshold semen parameters vary among centers but typically include severe oligospermia (<5 million sperm/mL),

asthenospermia (<5% progressive motility), or teratospermia (<4% normal forms by strict criteria). ICSI also is indicated when using surgically retrieved sperm (because the number of mature sperm is relatively limited) or treatment includes preimplantation genetic diagnosis (because conventional insemination may result in extra sperm attached to the zona, which may contaminate the sample for diagnosis by polymerase chain reaction), and for couples with previous failed or poor fertilization with conventional IVF. Other circumstances where low fertilization efficiency or fertilization failure is anticipated may be viewed as an indication for ICSI. To guard against the potential consequences of an undiagnosed sperm function abnormality, some centers perform ICSI on at least a portion of the oocytes retrieved from women with unexplained infertility.[107, 516, 517] ICSI also may yield higher fertilization rates for oocytes matured *in vitro*[518–520] and cryopreserved oocytes[521, 522] which often exhibit a hardened zona (resistance to digestion by proteases).[523–526]

Embryo Culture

Although much attention has focused on culture media formulations, other components of the culture system are equally important, including the carbon dioxide concentration (4–7%), incubation volume (10–50 μL), embryo group size (1–4) and the type of protein supplement (human serum albumin, recombinant albumin, synthetic serum substitute).[527–529]

Although the first human birth after IVF resulted from transfer of a blastocyst,[2] most transfers since then have involved earlier cleavage-stage embryos (day 2 or 3 after fertilization), primarily for the lack of culture media that could reliably sustain embryos during the compaction (morula) and blastocyst stages of development. However, the identification of key regulators and a greater understanding of the changing physiologic requirements of growing embryos have fostered the development of "sequential" media which vary in composition with the stage of embryo development.[530] ***Whereas pre-compaction embryos prefer pyruvate as a nutrient and non-essential amino acids (found in higher concentrations in the oviduct), post-compaction embryos favor glucose and essential amino acids (found in higher concentrations in the uterus).***[531, 532] Commercially available media provide the opportunity for any program to incorporate extended culture into its practice.

Extended culture and blastocyst transfer offer several potential advantages over the transfer of cleavage-stage embryos:

- Better assessment of true viability, after activation of the embryonic genome.
- Better synchronization between the stage of embryonic development and the endometrial environment.
- The opportunity to perform preimplantation genetic diagnosis (PGD), when it is indicated.
- Higher implantation rates, allowing transfer of fewer embryos, decreasing the risk for multiple pregnancy.

Extended culture is a more reliable test of viability and developmental potential because few embryonic genes are transcribed before the 8-cell stage and early measures of quality relate almost exclusively to the quality of the oocyte.[533–536] Post-compaction embryos also possess a transporting epithelium and can therefore better regulate their intracellular physiology and adapt to their environment.[537–539] Although pronuclear and cleavage-stage embryos *can* adapt to relatively hostile environs, survive, and successfully implant, those demands generate stresses that may compromise viability.[532, 540] Extended culture also

may also help to minimize any adverse effects of an abnormal hormonal milieu on uterine receptivity and contractility, in the aftermath of ovarian stimulation.[541–544]

The strongest argument in favor of extended culture is that the implantation rate for blastocysts (30–60%) is significantly higher than for cleavage-stage embryos (12–20%).[545–549] Almost certainly, the higher implantation rate of blastocysts merely reflects better selection of the most viable embryos, as there is no evidence that extended culture improves the intrinsic quality of embryos. Consequently, there is little point to extended culture in cycles yielding few or only poor quality embryos.[550, 551] In fact, already low implantation rates for lesser quality embryos exhibiting slow development or significant fragmentation may be further reduced if time in culture is extended, and many may not survive the challenge.[552]

Not all have embraced the trend to wider use of extended culture, which also has at least two potential disadvantages:

- Embryos of lesser quality that may implant if transferred on day 3 may fail to reach the blastocyst stage *in vitro*, increasing the risk there may be no embryos for transfer.
- Whereas higher blastocyst implantation rates permit transfer of fewer embryos, multiple pregnancy rates after transfer of two blastocysts are the same or higher than those observed after transfer of larger numbers of cleavage-stage embryos.

The results of recent systematic reviews illustrate both the principal advantage and disadvantage of extended culture. A 2007 systematic review of randomized trials comparing blastocyst and cleavage-stage embryo transfers observed a higher live birth rate per couple after blastocyst transfer (36% vs. 29.4%; OR=1.35, CI=1.05–1.74) that was most evident for "good prognosis patients" randomized on day 3 of culture and receiving an equal number of embryos.[553] The embryo cryopreservation rate was lower for blastocysts (OR=0.45, CI=0.36–0.56), and the overall risk of having no embryos for transfer was higher for blastocysts (OR=2.85, CI=1.97–4.11), but not in patients having a good prognosis.[553] A 2008 meta-analysis observed that the live birth rate for blastocyst transfer was higher than for cleavage-stage transfer only when patients were randomized on day 2 or 3 of culture (as opposed to earlier) or when an equal number of embryos were transferred (as opposed to transferring a greater number of cleavage-stage embryos than blastocysts).[554] Nine of the 18 trials included in the analysis compared outcomes in a good prognosis population (as defined by age, number of previous failed cycles, response to ovarian stimulation, and quality of embryos). Among these, clinical pregnancy rates achieved with cleavage-stage embryo and blastocyst transfer were not different (1,315 patients, OR=1.21, CI=0.96–1.51), but live birth rates were significantly higher with blastocyst transfer.[554]

In unselected populations,[555–564] and among couples having one or more previous failed cycles,[565] pregnancy rates and live birth rates after blastocyst and cleavage-stage embryo transfer are similar. In a trial involving 54 patients with three or more previous failed cycles (after transfer of cleavage-stage embryos) who were randomized to receive cleavage-stage or blastocyst transfer, the implantation rate (21% vs. 6%) and clinical pregnancy rate (22% vs. 13%) were higher with blastocyst transfer, but the live birth rates were not different (10% cleavage-stage vs. 13% blastocyst) because some randomized to blastocyst transfer had no embryo to transfer.[565] ***Altogether, these observations indicate that blastocyst transfer yields a higher live birth rate in good prognosis patients, particularly when the decision to extend culture occurs on day 3, but does not improve live birth rates for poor prognosis patients.***

The principal disadvantage of extended culture and blastocyst transfer is the higher risk for cancelled transfer. Although the emerging new "omic" technologies (genomic, transcriptomic, proteomic, or metabolomic profiling) hold promise for helping to identify developmentally competent embryos, none has proven ready for application in clinical practice.[566, 567]

Evidence suggests that clinical measures (age, parity, antral follicle count)[568, 569] and laboratory parameters (fertilization method, number of blastomeres, and the degree of fragmentation observed on day 3)[549, 551, 570, 571] can predict potential for blast formation, but the ability to generate blastocysts *in vitro* varies widely among individuals[572] and prediction models have not yet been tested in multi-center trials. ***The risk of cancelled transfer associated with extended culture is quite real in unselected populations (OR=2.85, CI=0.79–2.84), but no different from that of cleavage-stage embryo transfer in good prognosis patients (OR=1.50, CI=0.79–2.84).***[554]

In some studies examining the outcomes of extended culture and blastocyst transfer, dizygotic twin rates up to 50% have been observed after transfer of two blastocysts. ***In good prognosis patients, elective single blastocyst transfer significantly reduces the incidence of twins without reducing the overall pregnancy rate.***[573, 574] A study in donor oocyte recipients found that single blastocyst transfer yields a somewhat lower overall pregnancy rate, compared to transfer of two blastocysts, but reduces the twin rate dramatically.[574] It is disappointing that blastocyst transfer has not yet delivered on its promise to reduce the incidence of multiple pregnancy, primarily because few have been willing to transfer only a single blastocyst.[575–577]

Most,[578, 579] but not all studies,[580] have observed a 2- to 5-fold increase in the incidence of monozygotic twinning after blastocyst transfer. The cause is unknown, but culture-induced changes in the zona pellucida or embryo hatching have been implicated.[581–583] Most, but not all,[584] also have observed that blastocyst transfer shifts the sex-ratio, favoring males, compared to that observed in children conceived naturally,[585] or resulting from cleavage-stage embryo transfer.[547, 579, 586–588] The phenomenon may reflect the more rapid development of male embryos (at least in animals),[589] and the tendency to select the most advanced embryos for transfer.

Predictably, patients receiving blastocyst transfer have fewer spare embryos available for cryopreservation than those receiving a cleavage-stage embryo transfer (OR=0.28, CI=0.14–0.55).[590] Since the cumulative live birth rate (including births resulting from both fresh and frozen embryo transfers) is the most appropriate measure for comparison between blastocyst and cleavage-stage embryo transfer, having fewer cryopreserved embryos could negate some of the benefits of extended culture.

A number of reports have raised concern that longer duration of embryo culture may predispose to a higher risk of epigenetic (imprinting) mutations,[591–595] although subsequent studies examining the question have been reassuring.[596, 597] The mechanism is unknown, but the methionine component of culture medium has been implicated.[598] Evidence from animal studies that developmental programming during the preimplantation interval can be influenced by manipulations *in vitro*[599, 600] suggest that efforts to define and standardize culture conditions are justified and that careful long-term studies of children resulting from blastocyst transfer are warranted.

Preimplantation Genetic Testing

Preimplantation genetic testing broadly describes procedures involving the removal of one or more nuclei from polar bodies (oocytes) or cells (blastomeres, trophoectoderm) from embryos to test for mutations or evaluate their chromosomal complement.[118] Preimplantation genetic *diagnosis* (PGD) describes testing for a known genetic abnormality carried by one or both parents to determine whether it has been transmitted to the oocyte or embryo. Preimplantation genetic *screening* describes testing for oocyte or embryo aneuploidy when the parents are known or presumed to be normal.[118]

Preimplantation Genetic Diagnosis (PGD)

PGD is indicated for couples at risk for transmitting a specific genetic abnormality to their offspring. The risk of transmission is 50% for carriers of autosomal dominant disorders (e.g., Marfan syndrome), 25% for carriers of autosomal recessive disorders (e.g., cystic fibrosis), and 25% (half of male embryos) for female carriers of X-linked disorders (e.g., hemophilia A). PGD also can be used to detect genetic mutations that predispose to a disease (early onset Alzheimer's disease,[601] familial adenomatous polyposis coli,[602] p53 tumor suppressor gene mutations[603]), to detect an unbalanced chromosomal translocation in the embryos of a couple harboring a balanced translocation, and for human leukocyte antigen (HLA) matching of embryos to an existing child of the same parents (bone marrow transplantation).[604]

PGD can be performed on polar bodies removed from oocytes before fertilization (preconception diagnosis)[605] or on blastomeres or trophoectoderm removed from embryos before transfer. The equipment and techniques required for PGD are the same as for ICSI and related procedures (assisted hatching). After creating an opening in the zona pellucida using a laser or acid Tyrode's solution, the polar body or blastomere(s) is extracted for genetic analysis.

Preconception diagnosis is cumbersome and often requires sequential removal of both the first and second polar bodies to avoid misdiagnoses.[606] Although most aneuploidies result from errors in meiosis I, the composition of the oocyte cannot be confidently inferred from that of the first polar body, because of recombination events. Abnormalities can also arise in meiosis II, requiring examination of the second polar body. Even then, subsequent mitotic errors and any resulting from paternal inheritance cannot be detected. When polar body analysis is inconclusive, embryo biopsy must be performed. The cumulative trauma likely exceeds that of a single cleavage-stage embryo biopsy.

To detect abnormalities in embryos, one or two nucleated cells are removed, typically on the third day after fertilization (the 6–8 cell stage), before compaction when the blastomeres become more tightly adherent.[607, 608] After biopsy, the embryo can be placed in extended culture to develop to the blastocyst stage, or cryopreserved until the results of genetic analysis can be completed. Embryo biopsy also can be performed later, at the blastocyst stage. Blastocysts have more cells for genetic analysis and are less likely to be injured by biopsy,[609, 610] but later sampling leaves little time for analysis before the embryo must be transferred or frozen. Although biopsied embryos are more sensitive to the rigors of freezing and thawing, technical modifications in cryopreservation techniques have largely overcome the limitation.[611, 612]

For detection of specific gene mutations, individual cells are placed in small test tubes for DNA analysis using mutation-specific primers and the polymerase chain reaction (PCR) to amplify the segment of DNA containing the gene of interest. For disorders involving multiple mutation sites, multiplex PCR or whole-genome amplification can be used to allow simultaneous analysis of the different loci.[613–615] After DNA amplification with PCR, a variety of techniques can be used to detect the targeted mutation. Most involve separation of the amplification products by electrophoresis for direct comparison to normal reference DNA. Restriction enzyme analysis can be applied when the altered sequence results in a loss or gain of a restriction site (sickle cell disease). For detection of mutations that do not affect a restriction site, specially-designed primers can be used to selectively amplify the abnormal or the normal sequence to determine the presence or absence of the mutation. Fluorescent PCR is a modification useful for detecting deletions (the common ΔF508 cystic fibrosis mutation), insertions (the Tay Sachs Disease 1278ins4 mutation), and mutations involving multiple loci. Real-time PCR is a method that allows continuous

measurement of the accumulation of a specific amplification product and eliminates the need for electrophoresis.[616] Liquid chromatography can be used to sequence PCR product directly.

For detection of numerical and structural chromosomal abnormalities, PGD is usually performed using fluorescence in situ hybridization (FISH), which uses probes labeled with colored fluorochromes that bind to specific DNA sequences unique to each chromosome. FISH can detect an excess or missing piece of chromosomal material in oocytes (when the female is the carrier) or embryos (when either parent is a carrier).[617] After removal, the cells are fixed on glass, the cytoplasm is dispersed, and the fluorescent probes are applied and allowed to hybridize with complimentary DNA sequences on targeted chromosomes. The different colored fluorescent signals can be observed with microscopy using filters of the appropriate wavelength. The copy number of each chromosomal segment of interest is defined by the number of fluorescent signals detected. For couples harboring a balanced chromosomal translocation, PGD decreases the risk of miscarriage (if pregnancy is achieved).[618, 619] Although most couples with balanced translocations ultimately will achieve a successful pregnancy without IVF and PGD, the time to delivery can be prolonged and most will suffer additional miscarriages in the interim.[620–622]

The PGD procedure presents a number of technical challenges, primarily relating to the short time and limited amount of genetic material available for analysis. Polar body or embryo biopsy is a delicate procedure requiring extraordinary skill to minimize trauma to the oocyte or embryo. When PCR-based methods are used, diagnostic errors can result from anucleate cells, failed or partial amplification,[623–625] or external contamination. The estimated risk of misdiagnosis is approximately 2% for recessive disorders and 11% for dominant disorders,[626] but the true risk is unknown. When FISH is used, approximately 10% of cells yield no results, or results that are not confirmed when the remaining cells in the embryo are analyzed. The risk for inconclusive or inaccurate results relates to the number of cells and chromosomes included in the analysis. If a probe fails to hybridize, no result is available for the corresponding chromosome; commercially available probes are more than 95% efficient, but that still leaves room for error. Overlapping chromosomes can yield fused or split signals that can be misinterpreted.

PGD offers couples who carry serious genetic disorders the opportunity to have a healthy child without the practical and ethical problems associated with terminating an affected pregnancy after traditional prenatal diagnosis (chorionic villus sampling, amniocentesis). However, careful counseling is required and must include the following:

- The possibility of diagnostic error or inconclusive results
- The possibility of abnormalities arising from the biopsy procedure
- The possibility that the chance for success may be reduced, compared to that expected when PGD is not performed, due to embryo trauma and the smaller number of embryos available after abnormal embryos are excluded.
- The need for conventional prenatal diagnosis to confirm the accuracy of PGD

Preimplantation Genetic Screening (PGS)

Aneuploidy is common in human embryos, most resulting from meiotic errors in the oocyte, which increase in prevalence with advancing age. Although aneuploidy is more common in morphologically abnormal embryos, even embryos with normal morphology and developmental progress may be aneuploid.[627, 628] Logically, embryo biopsy, aneuploidy screening, and transfer of proven euploid embryos would be expected to improve implantation efficiency and to reduce the incidence of miscarriage in pregnancies resulting from

IVF. Since most aneuploidy derives from the oocyte, even polar body screening should be informative.[629, 630] Older women are the most obvious potential candidates for PGS. Others include women with a history of recurrent pregnancy loss, those with repeated IVF failure despite transfer of morphologically normal embryos, and couples with severe male factor infertility.

Aneuploidy screening is usually performed using FISH to identify the copy number of selected chromosomes, but the number of chromosome pairs from each nucleus that can be evaluated is limited. As many as nine chromosomes, including those involved in most aneuploidies (X, Y, 13–16, 18, 21, 22), can be examined with two sequential FISH analyses in a single cell.[631–634] Alternatively, all 23 chromosome pairs can be amplified using random primers for analysis by comparative genomic hybridization (CGH)[635] The technique involves simultaneous amplification of test and reference samples using red (test sample) and green (reference sample) fluorochromes, which then are allowed to hybridize with a normal male metaphase chromosome spread. Image-processing software is used to analyze the relative amounts of red and green signal to determine the chromosome numbers.

PGS has several inherent limitations. First, fewer than half of the chromosome pairs can be evaluated with FISH, and studies comparing results obtained with FISH and CGH have demonstrated that up to 25% of aneuploid embryos escape detection by FISH because the abnormal pair was not among those analyzed.[611] Unfortunately, conventional CGH cannot be completed in the short time available between biopsy on day 3 and transfer on day 5 or 6 after fertilization, requiring that biopsied embryos be cryopreserved for transfer in a subsequent cycle after the results are known. More rapid methods of CGH are under development, but may not prevent diagnostic errors relating to early embryonic mosaicism, which is common and increases in prevalence with maternal age.[636–640] In fact, analysis of all 24 chromosomes (22 autosomes, X, and Y) might increase the number of false-positive diagnoses and the number of potentially normal embryos discarded. A mosaic embryo can be identified only if two or more cells are removed and analyzed and cannot be excluded unless all of the cells are analyzed, which destroys the embryo. One study of embryos yielding discordant results after two cells were analyzed found that half were euploid when all of the cells were analyzed.[641] The observation suggests that, in some cases, biopsy may remove an abnormal cell and "correct" an abnormality. However, mosaic embryos must be considered abnormal because the proportion of euploid cells required for normal development is unknown. Results from other studies suggest that up to half of all abnormal cleavage-stage embryos that survive to become blastocysts "self-correct."[639, 642, 643] Alternatively, the abnormal cell line might fail to propogate, or the original diagnosis may have been incorrect.

Unfortunately, PGS has not yet delivered on its promise. Four randomized trials have examined the impact of PGS with FISH on outcomes in women of advanced maternal age; all failed to demonstrate any benefit and two provided evidence that PGS decreased live birth rates.[644 645–647] No randomized trials have evaluated the clinical utility of PGS in women with recurrent miscarriage, repeated IVF failure, or with severe male factor infertility. A study comparing outcomes in a group of women with recurrent pregnancy loss to those in a group having PGD performed for X-linked disease (controls) observed no difference in the ongoing/delivered pregnancy rates between the groups.[648] In another comparing outcomes after PGS in younger (age <37 years) and older women (age ≥37 years) with a history of recurrent miscarriage, pregnancy rates were not different from those achieved in the general IVF population.[649] A PGS study comparing outcomes in a group of patients with an average of 4.2 previous failed cycles with those in a group having PGD for X-linked disease (controls) observed no difference in pregnancy rates between groups.[650] Another observed no differences in implantation or pregnancy rates in women with three or more previous failed IVF cycles who chose either PGS or assisted hatching on day 3 before day 5 transfer.[632] Finally, two randomized trials evaluating PGS for selection

of embryos in good prognosis patients have failed to demonstrate benefit.[651, 652] No studies have evaluated the clinical utility of PGS for couples with male factor infertility.

Given the disappointing results of PGS using FISH, those of a small observational study using CGH for PGS are intriguing. The study recruited 45 infertile volunteers whose embryos were cultured to the blastocyst stage before biopsy of trophectoderm for PGS using CGH, with all blastocysts cryopreserved for later transfer if euploid.[567] Overall, 51% of blastocysts screened abnormal, 100% of euploid blastocysts later thawed for transfer survived, and 69% of frozen-thawed euploid blastocysts yielded an ongoing pregnancy or live birth. These promising observations await confirmation in larger randomized trials, but suggest PGS may yet have the potential to improve outcomes in IVF.

Assisted Hatching

"Hatching" of the blastocyst from the zona pellucida is a natural process in which the embryo expands and emerges before implantation. Under culture conditions, the embryo erupts, leaving behind an empty zona, but *in vivo*, the mammalian zona normally dissolves. Evidence suggests that hatching *in vivo* results from embryo-uterine interactions with the embryo secreting an activator of zona lysins in the uterine fluid.[653] Zona thickness and relative resistance to enzyme digestion correlate with embryo quality and implantation potential.[654–657]

"Assisted hatching" describes a variety of techniques for artificially thinning or opening the zona. The procedure is intended primarily to improve implantation potential. Assisted hatching also offers the opportunity to remove cytoplasmic fragments from the perivitelline space,[658] but evidence indicates that removing fragments has no impact on implantation, clinical pregnancy rates, or live birth rates.[659] A wide assortment of methods has been used for assisted hatching, including zona drilling with acidified Tyrode's solution,[660–663] partial zona dissection with a glass microneedle,[664, 665] laser photoablation,[666–668] enzymatic thinning,[669, 670] and the use of a piezo-micomanipulator.[671]

The idea that assisted hatching might improve implantation and pregnancy rates arose from observations that embryos subjected to zona drilling during early experience with assisted fertilization exhibited increased implantation efficiency.[672] Results of subsequent clinical trials varied widely with some suggesting that assisted hatching improved results in selected individuals having a relatively poor prognosis (advanced maternal age, previous failed IVF cycle, poor embryo morphology, thickened zona),[660, 663, 665–668, 671] and others observing no demonstrable benefits, particularly when hatching was more broadly applied.[661, 662, 664, 673] A 2009 systematic review and meta-analyses of combined data from 28 randomized trials involving 3,646 women found that assisted hatching increased clinical pregnancy rates (OR=1.29, 95% CI=1.12–1.49) and multiple pregnancy rates (12 trials, OR=1.67, CI=1.24–2.263), but had no impact on miscarriage rates (14 trials, OR=1.13, CI=0.74–1.73) or live birth rates (seven trials, OR=1.13, CI=0.83–1.55), and concluded that data are insufficient to determine the impact of assisted hatching on outcomes.[674]

The varying results of clinical trials employing different techniques do not allow confident conclusions regarding the value of assisted hatching. ***On balance, the weight of available evidence suggests that assisted hatching may have benefit for selected individuals. However, routine or universal hatching is not warranted, particularly because the procedure also has potential risks. Hatching may cause damage to embryos and may increase the risk of multiple pregnancy and monozygotic twinning.***[675–678]

Embryo Transfer

Although embryos have been transferred successfully at any stage of early development, from zygote to blastocyst, transfer is most commonly performed 3 days after oocyte retrieval and fertilization. The relative advantages and disadvantages of extended culture to the blastocyst stage are discussed above.

Systems for grading the quality of embryos vary among programs, but the morphological features on which grading is based are similar and include cell number, symmetry and shape of the blastomeres, the extent of cytoplasmic fragmentation in the perivitelline space, and the rate of cleavage. The ideal day 3 cleavage-stage embryo has six to eight blastomeres of equal size and no cytoplasmic fragmentation. Embryos of lesser quality may exhibit fewer cells, blastomeres of unequal size, or varying degrees of fragmentation.

The essential features of embryo transfer have not changed significantly since the procedure was first described in 1984.[679] Although the impact of transfer technique on results is difficult to study, most clinicians believe it is as important as embryo quality.[680] Most studies,[681–687] but not all,[688, 689] have observed higher pregnancy rates after "easy" transfers than after "difficult" transfers. The measure is subjective and difficult to quantify, but not necessarily invalid.

Embryo transfer has a number of potential pitfalls. Mucus within the cervical canal may plug the catheter tip, resulting in retained embryos or improper placement.[683, 690, 691] Cervical mucus also can be a source of bacterial contamination of the endometrial cavity, adversely affecting results.[692, 693] Although any obvious or excess cervical mucus is best removed before transfer, there is no evidence that vigorous cervical lavage is helpful.[686] Blood on the catheter tip after transfer suggesting trauma to the endocervical mucosa or endometrium has been associated with reduced pregnancy rates.[690] Embryos adhering to the outside of the catheter after transfer can be relocated or removed inadvertently when the catheter is withdrawn.[686, 694] Microscopic examination of the catheter immediately after transfer identifies retained embryos requiring a second transfer procedure; whether the second transfer procedure decreases success rates is unclear.[683, 689, 690]

Transfer catheters vary widely in design. They may be relatively stiff or quite "soft" and open on the end or on the side; many also have a malleable outer sheath. Stiff catheters and those with a rigid outer sheath are easier to insert but more traumatic than soft catheters, which can better follow the contours of the endocervix and endometrium.[695] Although soft catheters yield better results than stiff catheters,[696–698] no one catheter has proven superior.[682, 699–701] Syringes that may be used with a transfer catheter also vary in design and performance characteristics. Some require careful controlled injection to avoid a sudden expulsion and others having a plunger with a compressible end may cause an inadvertent re-aspiration of embryos after release of pressure. Re-injection after withdrawing the catheter approximately 1 cm may help to prevent retrograde flow of transfer medium that may result from capillary action.[702]

Uterine contractions at the time of embryo transfer are best avoided, to the extent possible. Careful analysis of real-time ultrasound scans has revealed that implantation and pregnancy rates decrease as the frequency of myometrial contractions increases.[543] The manipulations associated with technically difficult transfers or the use of a cervical tenaculum stimulates uterine contractions that may propel embryos up into the fallopian tubes or down into the cervix.[703, 704]

Larger volumes of transfer media (greater than 20–50 μL) or air above the column of media may increase the risk that embryos may be expelled from the uterus or propelled into the fallopian tube.[694, 705] The concentration of protein and the viscosity of transfer medium do not appear to affect results.[706, 707] Limited evidence suggests that best results are obtained

when the catheter tip does not touch the fundus and transfer occurs at a level approximately 0.5–1.0 cm lower.[708, 709] Transfers higher in the fundus may increase the risk of ectopic pregnancy,[709, 710] and low transfers may result in cervical implantations.[711]

A trial transfer before the cycle begins can identify women with cervical stenosis or an acutely angled cervico-uterine junction that can make transfer technically difficult to perform.[681] Several studies have suggested that, when required, cervical dilation is best performed before the cycle begins[683, 712–714]; shorter intervals of time between dilation and transfer may be insufficient to allow the endometrium to recover from the trauma or bacterial contamination and are associated with significantly lower pregnancy rates. Dilation with laminaria in advance of the treatment cycle also can be effective.[715] When transfer proves difficult despite these preparations, a malleable stylet can be used to introduce the outer sheath of a soft catheter beyond the internal cervical os, then replaced by the soft inner catheter containing the embryos ("after-loading"). Transmyometrial embryo transfer under ultrasound guidance has been described,[716, 717] but yields lower pregnancy rates,[718] probably relating to uterine contractions,[719] and rarely should be necessary.

Embryo transfer under trans-abdominal ultrasonography offers a number of potential advantages over a blind technique. ***Ultrasonography facilitates the insertion of soft catheters, confirms correct positioning, and avoids inadvertent trauma to the fundal endometrium.***[686, 720] Urine in the bladder also may help to straighten the plane of the cervico-uterine junction.[721, 722] Uterine position and orientation often change in the interval between the mock transfer and the actual transfer, primarily due to the ovarian enlargement.[723] Combined, these factors probably explain why numerous comparative clinical trials have found that pregnancy rates are higher after ultrasound-guided transfers than after blind transfers.[685, 724–728] A 2007 systematic review including 20 trials comparing outcomes of transfers guided by ultrasonography or performed using the "clinical touch" technique observed a significantly higher live birth rate after guided transfers (OR=1.78, CI=1.19–2.67].[729] A 2008 single-operator randomized trial reached the same conclusion (41% vs. 28%, OR=1.75, CI=1.14–2.69).[730]

A few studies have compared outcomes of transfers performed by different clinicians, after controlling for other obvious and important variables. Whereas some have found that "physician factors" influence results,[731, 732] others have found no correlation when results are compared in an "ideal" patient population.[733]

Although bedrest for 30 minutes or more often is recommended after transfer, there is no evidence that it improves outcomes.[734–737] Subsequently, patients can resume normal daily activities; physical activity and diet have no demonstrable effect on outcomes. Mild intermittent cramping and bloating are normal symptoms, but moderate or severe pain require evaluation to exclude infection, ovarian torsion, ovarian hyperstimulation syndrome, and other causes of abdominal pain.

In summary, the goal of embryo transfer is to deliver embryos to the uterus in an accurate and atraumatic fashion. ***Whenever possible, mucus, blood, and uterine contractions should be avoided. A preliminary trial transfer can identify women who may benefit from cervical dilation before treatment begins, and transfers in small volumes using soft catheters guided by ultrasonography produce the best results.***[226, 686]

Embryo Transfer Guidelines

The goal of IVF is to maximize pregnancy rates while, at the same time, minimizing multiple gestations, high-order multiple gestations in particular. The likelihood of success increases with the number of embryos transferred, to a point, beyond which only the risk of *multiple* pregnancy increases.[738–740] Strict regulations on the number of embryos

transferred, as defined by law in some countries, reduce the number of multiple pregnancies and all but eliminate high-order multiple gestations,[738] but do not allow treatment to be individualized, considering unique patient characteristics and circumstances (age, the number and quality of embryos, the opportunity for cryopreservation, and the outcome of any previous cycles), or to be adjusted according to new clinical data. Registry data from the U.S. strongly suggest that regulations ignoring the unique circumstances of individual women inevitably reduce the chance of pregnancy for many.[739]

The data generated by individual programs are the best guide for determining the optimal number of embryos to transfer in women of varying age and clinical characteristics. In their absence, the Society for Assisted Reproductive Techology (SART) and the American Society for Reproductive Medicine (ASRM) have offered guidelines. First published in 1998,[741] the guidelines have been revised several times, based on new clinical data reflecting steady advances in ART indicating that fewer embryos can be transferred without adversely affecting the likelihood for success.[742, 743] The guidelines issued in 2009 are summarized below:

2009 SART/ASRM Guidelines for the Number of Embryos Transferred[744]

	Age			
Prognosis	*<35 yrs*	*35–37 yrs*	*38–40 yrs*	*41–42 yrs*
Cleavage-stage embryos				
Favorable*	1–2	2	3	5
All others	2	3	4	5
Blastocysts				
Favorable*	1	2	2	3
All others	2	2	3	3

***Favorable prognostic characteristics:**

- First cycle of IVF
- Good embryo quality
- Excess embryos available for cryopreservation
- Previous successful IVF cycle

Maternal age and embryo quality are the most important factors influencing the implantation potential of each embryo.[745] When surplus high quality embryos are available for cryopreservation, permitting a more discriminating selection, fewer embryos can be transferred because higher implantation efficiency can be anticipated.[739, 746] For the same reason, fewer blastocysts than cleavage-stage embryos can be transferred.[545–549] Women with the best prognostic features (age under 35, first or previously successful IVF cycle, surplus good quality embryos) have an increased risk for multiple pregnancy and are candidates for single embryo transfers.[747–749] ***Overall, the weight of available evidence indicates that an optimal balance between pregnancy rates and the risk of multiple pregnancy can be achieved with a flexible embryo transfer policy based on maternal age, embryo quality, and the availability of surplus high quality embryos.***

Luteal Phase Support

Controlled ovarian hyperstimulation with exogenous gonadotropins generally yields multiple corpora lutea that might well be expected to sustain supraphysiologic serum concentrations of estradiol and progesterone during the luteal phase of IVF cycles. Co-treatment

with GnRH analogs for prevention of premature LH surges and luteinization effectively suppresses endogenous LH secretion, as intended. Unfortunately, even though agonist and antagonist treatment ends abruptly on the day of hCG administration, residual suppression of endogenous LH does not. ***Abnormally low levels of LH during the luteal phase may be insufficient to stimulate and maintain the level of luteal function required to promote timely endometrial maturation in preparation for implantation or to support an early pregnancy once established.*** Endogenous LH secretion can remain suppressed for as long as 10 days after treatment with a GnRH agonist ends and luteal function is frequently inadequate in amount or duration.[750] Although antagonists have a much shorter duration of action, they often have the same consequence. Integrated estradiol and progesterone levels are abnormally low and luteal phase duration is grossly short in GnRH antagonist treatment cycles, particularly when a GnRH agonist rather than hCG is used to stimulate the final stages of ooycyte maturation.[751] Because there is no way to predict who may or may not require luteal support in any given cycle, some form of treatment must be provided for all.

Progesterone supplementation generally begins on the day of oocyte retrieval, or at the time of embryo transfer.[752–754] Numerous clinical trials have compared clinical, ongoing, or delivered pregnancy rates or spontaneous abortion rates between groups receiving treatment with different luteal phase support regimens, with varying results. Progesterone has been administered orally (300–800 mg daily), vaginally as a bioadhesive 8% gel (90 mg daily), cream or tablet (100–600 mg daily), and by intramuscular injection (25–50 mg daily; 17α-hydroxyprogesterone 341 mg every 3 days); supplemental doses of hCG generally have been administered every 3 days (1500–2500 IU). There is no evidence that any one treatment regimen is superior,[753–755] although results achieved with oral progesterone have been inconsistent. The optimal duration of supplementation also has not been established. Treatment regimens vary widely, from discontinuation of supplementation at the time of the pregnancy test (positive or negative), to continuation throughout the first trimester.[756] Supplemental natural progesterone is not associated with any increased risks of birth defects.[752] Although supplemental estradiol also is commonly administered, there is no evidence that it improves outcomes, compared to those achieved with progesterone supplementation alone.[757, 758]

Embryo Cryopreservation

The first pregnancy resulting from transfer of a cryopreserved human embryo was reported in 1983.[759] In the years since, advances in cryobiology have made embryo cryopreservation an integral part of modern ART. Success with frozen embryo transfer cycles significantly increases the overall cumulative pregnancy rate per retrieval. Cryopreservation of all embryos also may be an effective management strategy for women at high risk for OHSS.[760, 761]

The cryopreservation process has two distinct stages, freezing and thawing. The objective of freezing is to avoid ice crystallization of intracellular water, which can result in cellular damage. Freezing protocols vary with the stage of embryo development, which affects cellular permeability. There are two basic methods for embryo cryopreservation, the "slow-freeze" technique and "vitrification." In both, cellular water is gradually replaced by cryoprotectants (dimethyl sulfoxide, propanediol glycerol) via osmosis by passage through increasing concentrations of the cryopreservative. In the slow-freeze method, embryos are sealed in ampules or vials, cooled to temperatures between –30°C and –110°C in a programmed 2-step process, and then stored in liquid nitrogen. The first phase of the freezing process is rapid to prevent ice crystal formation (more likely to occur with gradual cooling), and the second phase more gradual. In the vitrification method, embryos are flash frozen

by immersion into liquid nitrogen, creating a solid glass-like state.[762, 763] After thawing, the process is reversed, gradually passing the embryo through decreasing concentrations of the cryoprotectant, followed by an interval of culture before transfer.

Embryos can be frozen at any stage, from zygote to blastocyst, and remain viable for at least several years, perhaps indefinitely.[764] Numerous studies have compared embryo thaw survival, implantation, and pregnancy rates among embryos frozen at different stages of development. In general, post-thaw survival rates after slow-freezing range between 50% and 90% and are higher for zygotes than for cleavage-stage embryos and blastocysts.[765–770] Implantation rates (5–15%) and pregnancy rates (10–30%) after transfer of slow-frozen thawed zygotes, cleavage-stage embryos, and blastocysts have varied among studies, but not dramatically. Early experience with vitrification suggests the method is associated with consistently high survival rates (90–100%), and may yield higher implantation and pregnancy rates.[762, 763]

Overall, success rates for frozen embryo transfer cycles are approximately one-half to two-thirds those observed in fresh transfer cycles in most centers, at least in part because the highest quality embryos are generally selected for fresh transfer. Results achieved with embryos derived from conventional IVF and ICSI are comparable. Embryos that resume cleavage and survive longer in culture are more likely to result in pregnancy, and cryopreserved sibling embryos derived from successful cycles yield higher success rates than those derived from unsuccessful cycles, probably reflecting overall better embryo quality.

Frozen-thawed embryos can be transferred in a monitored natural cycle in women with normal ovulatory function.[771–773] Alternatively, transfer can occur in an artificial cycle in which endometrial development is carefully controlled by programmed sequential treatment with exogenous estrogen (oral micronized estradiol 4–6 mg daily or transdermal estradiol 0.2–0.4 mg) and progesterone (50–100 mg daily intramuscularly, 8% vaginal gel, administered twice daily), beginning with or shortly before the onset of menses.[774–776] Preliminary down-regulation with a GnRH agonist also can be used, as is typical in donor oocyte recipients.[260, 773, 774, 777] A 2010 systematic review including 22 randomized trials comparing different endometrial preparation regimens found no evidence that one method of endometrial preparation was superior to others.[754] In both natural (days after ovulation) and artificial cycles (days of progesterone treatment), the time of transfer is synchronized with the stage of embryo development, as in donor oocyte recipients (discussed below).

Outcomes of IVF

IVF outcomes have improved steady throughout the years since its introduction into clinical practice. Early on, IVF offered only a modest chance for success and was reserved appropriately for couples who had no option or had failed all other available forms of treatment. As technology and outcomes improved, IVF became a realistic and attractive option for an increasing number of couples. The advent of ICSI revolutionized the treatment of severe male factor infertility and greatly contributed to the growth of ART. Now, ART often is the first and best option for a large proportion of infertile couples.

IVF success rates may be expressed in several ways, using different numerators and denominators. The two most common numerators are pregnancy and live birth, the latter being the most relevant measure. ***Approximately 18% of pregnancies result in miscarriage (15.8%), induced abortion (1.0%) stillbirth (0.6%) or an ectopic pregnancy (0.7%).*** Pregnancy or live birth rates may be calculated as a percentage of cycle starts, retrievals, or transfers. Overall, approximately 11% of cycles are discontinued before retrieval, due to

an inadequate (80.6%) or excessive response (5.4%) to stimulation, a concurrent medical illness (1.0%), or a patient's own personal reasons (13.0%).[3]

For the year 2007, the U.S. registry recorded a total of 142,530 cycles of ART nationwide. Among these, 101,897 (71.5%) were IVF cycles involving fresh, nondonor oocytes (64.3% with ICSI), with 35.4% resulting in a pregnancy. The live birth rates were 29.0% per cycle, 32.7% per retrieval, and 35.9% per transfer. The live birth per retrieval rates for IVF cycles (33.3%) and for ICSI cycles (32.5%) were similar. Another 23,133 (16.2%) were frozen embryo transfer cycles, yielding a live birth rate of 29.9% per transfer. In all age groups, success rates were higher for women with a previous live birth and those in their first ART cycle than for nulliparous women and those with a previous failed ART cycle.[3]

Multiple Gestation

The risk of multiple gestation is increased substantially in ART cycles. In 2007, 31.2% of all births in the U.S. resulting from ART were multiples, a rate 10 times higher than the 3% multiple-infant birth rate in the general population; 29.4% of live births were twins and 1.8% were triplets or more.[3] The higher maternal and neonatal risks associated with multiple pregnancies, their greater financial and social costs, and the many different factors that contribute to the high incidence of multiple births are reviewed in detail in Chapter 31. Consequently, discussion here is limited to issues specifically relating to multiple gestations that result from ART.

Success rates increase with the number of embryos transferred, to a point, beyond which only the multiple pregnancy rate further increases.[738, 739] The number of embryos corresponding to that threshold generally defines the maximum number of embryos that should be transferred. As reflected in the 2009 revised SART/ASRM embryo transfer guidelines outlined above, that number increases as age increases. Based on data from the 2007 national summary of ART results,[3] when one embryo is transferred, 97.5% of resulting births are singletons. When two embryos are transferred, 65.9% are singletons, 33.3% are twins, and 0.7% are high-order multiple births (triplets or more). When three embryos are transferred, 66.7%% of births are singletons, 29.4% are twins, and 4% are triplets or more; the distribution of singleton, twin, and high-order multiple births does not change further with transfer of four or more embryos.

Age and the number of embryos generated and available for transfer are almost if not as important for predicting success as the number of embryos transferred.[738, 739, 778] Younger women tend to have both higher success rates and higher multiple-infant birth rates. Data from the 2007 U.S. national ART summary illustrate the point.

Live Births per Transfer and Percentages of Multiple-Infant Births for Women Under Age 35 with Surplus Embryos Suitable for Cryopreservation, by Number of Embryos Transferred[3]

Embryos Transferred	Live Births per Transfer	Singletons	Twins	Triplets +
1	50.2%	97.6%	2.3%	0.1%
2	55.9%	59.2%	40.0%	0.8%
3	50.1%	55.5%	37.2%	7.3%
4	41.6%	51.8%	42.0%	6.3%
5+	35.2%	47.4%	47.4%	5.3%

Some have argued that extended culture to the blastocyst stage facilitates selection of the highest quality embryos having the greatest implantation and developmental potential and thereby reduces both the number of embryos needed to maximize success rates and the risk for multiple birth. In 2007, 33.1% of all transfers occurred on day 5 or 6 after fertilization (blastocysts). For all age groups, the live birth per transfer rate for blastocysts was higher than for cleavage-stage embryos; the differences ranged from 4.0% to 10.8% and were greatest for women under age 35 (53.0% vs. 42.2%) and least for those over age 44 (6.8% vs. 2.8%).[3] Overall, 35.7% of live births resulting from day 5 transfers were multiples (33.9% twins, 1.8% triplets and more), compared to 28.5% of day 3 transfers (26.6% twins, 1.9% triplets and more). When two blastocysts are transferred, the incidence of high-order multiple gestation is markedly reduced but not altogether eliminated, because the incidence of monozygotic twinning may be increased after blastocyst transfer,[779] and the incidence of twins is no lower than that associated with transfer of greater numbers of cleavage-stage embryos.[575, 577]

As previously discussed, the SART/ASRM guidelines on the number of embryos transferred have been revised several times since first issued in 1998, in efforts to reduce the incidence of multiple pregnancy, high-order multiple pregnancy in particular. The average numbers of embryos transferred in the U.S. began decreasing in 1997, with the steepest decline observed between 1998 and 1999 after the first guidelines were issued. Whereas the number of pregnancies and live births per cycle have increased steadily, the percentage of high-order multiple pregnancies has decreased, with the steepest decline (20.8%) again occurring between 1998 and 1999.[780] Unfortunately, the percentage of twin pregnancies has remained at the same relatively high level. The 2009 guidelines, recommending that no more than two cleavage-stage embryos be transferred in women under age 35, and only one (cleavage-stage or blastocyst) in those with the best prognostic features, should help to decrease the percentage of twin gestations.

Offspring of IVF

Studies of the offspring resulting from IVF have raised concerns that the children may be at increased risk for birth defects, prematurity, low birth weight, delayed neurological development, genetic and epigenetic abnormalities, and cancer.

Preterm Birth and Low Birth Weight

Singleton pregnancies resulting from IVF, with or without ICSI, are at increased risk for preterm birth and low birth weight (LBW, ≤2,500 g), compared with naturally conceived pregnancies.[781–790] In 2007, 12.5% of all singleton births from ART cycles were preterm births, as were 19.9% of singletons from multi-fetal pregnancies, 61.7% of twins, and 96.1% of triplets and or more.[3] The corresponding percentages of LBW infants from ART cycles were 8.4%, 16.5%, 56.2%, and 91.5%, respectively. A population-based study comparing over 42,000 infants conceived with ART from 1996–97 with more than 3 million births in the general population observed a higher prevalence of LBW (6.5% vs. 2.5%, RR=2.6, CI=2.4–2.7) and preterm LBW (6.6% vs. 4.7%, RR=1.4, CI=1.3–1.5) among singleton, but not twin, infants conceived with ART that persisted after adjustment for maternal age, and parity, gestational age at delivery, multifetal reduction procedures, and cause of infertility.[781] The absence of an association between IVF and preterm birth in twins probably reflects primarily the overall profound effect of twinning on pregnancy outcome, obscuring any effect relating to IVF.

A 2004 meta-analysis of data from 15 studies involving 12,283 IVF and 1.9 million naturally conceived singletons found that IVF pregnancies were associated with significantly higher odds of perinatal mortality (OR=2.2, CI=1.6–3.0), preterm delivery (OR=2.0, CI=1.7–2.2), LBW (OR=1.8, CI=1.4–2.2), very LBW (OR=2.7, CI=2.3–3.1), and small for gestational age (OR=1.6, CI=1.3–2.0).[784]

These data suggest that ART is associated with an approximate 2-fold increased risk of preterm birth and LBW in singleton pregnancies, which could be related to IVF (medications, manipulation of gametes, or culture) or to infertility, independent of treatment. Overall, the incidence of adverse pregnancy outcomes is higher among infertile women who conceive than in the general population.[791–794] The results of a population-based cohort study comparing the outcome of a pregnancy conceived with ART to the outcome of an earlier or subsequent pregnancy in the same women, and to outcomes in the general population, were revealing; their infants conceived naturally and with ART were of similar gestation age and birth weight, but were delivered earlier and had lower birth weight than infants in the general population.[795]

Congenital Anomalies

Whereas numerous studies have observed no increase in the incidence of congenital anomalies among children conceived with ART (with or without ICSI), others have suggested an increased risk of specific abnormalities, including hypospadias (and other genitourinary defects), neural tube defects, cleft lip and cleft palate, gastrointestinal malformations, musculoskeletal and chromosomal defects. A 2005 meta-analysis of data from four prospective cohort studies including 5,395 children conceived with ICSI found no overall increase in the risk for major birth defects, or for any specific defect, compared to that for children conceived with standard IVF, but no comparison was made with children conceived naturally.[796] A study comparing the prevalence of birth defects in 301 infants conceived with ICSI and 837 conceived with IVF to the prevalence of defects in 4,000 infants conceived naturally, based on an assessment at 12 months of age using a standardized classification system, observed an approximate 2-fold increase in the prevalence of single defects among children conceived with ICSI (9.0%) and IVF (8.6%), compared to that in children conceived naturally (4.2%); the prevalence of multiple anomalies in ICSI children (2.0%) and IVF children (1.6%) was also higher than in naturally-conceived children 0.5%).[797] A 2005 meta-analysis of data from 25 studies found a smaller, but still significant, increase in the prevalence of birth defects among children conceived with ART (OR=1.40, CI=1.28–1.53).[798]

As with the higher observed incidence of preterm birth and LBW among children conceived with ART, it is uncertain whether the apparent excess risk of congenital anomalies observed in ART children relates to treatment, to the infertility requiring such treatment, or to other factors. A 2006 cohort study comparing the prevalence of birth defects in children of infertile couples who conceived, naturally or with treatment, to the prevalence among children of fertile couples found that singleton children of infertile couples had a higher prevalence of congenital malformations, regardless whether they were conceived naturally (HR=1.20, CI 1.07–1.35) or after treatment (HR=1.39, CI 1.23–1.57), and that the overall prevalence of birth defects increased with increasing time to pregnancy.[799] An analysis restricted to singleton children conceived by infertile couples observed that children conceived after treatment had an increased prevalence of genital organ malformations (HR=2.32, CI=1.24–4.35), compared to children conceived naturally. Whereas these data suggest that hormonal treatment for infertility may increase the risk for genital anomalies, they also suggest that some of the excess risk

for birth defects observed among children conceived with ART relates to underlying infertility or its causes.

Chromosomal, Genetic, and Epigenetic Abnormalities

Limited evidence suggests that the prevalence of chromosomal abnormalities in children conceived with ART is not different from that in children conceived naturally. Nonetheless, concerns persist that the use of sperm from infertile men, and ICSI itself, might increase the risk for conceiving a child with a chromosomal or genetic defect, because infertile men (and women) are more likely than fertile men (and women) to have a chromosomal abnormality that may contribute to their infertility and that their children may inherit.

Genomic imprinting describes the process that allows only one parental allele (maternal or paternal) to be expressed. For a substantial number of genes involved in early embryonic growth and placental and neurologic development, transcription is normally limited to one allele. Usually, the maternal allele is active in imprinted genes involved in fetal development, and the paternal allele is active in genes involved in placental growth. Imprinting disorders can arise via several mechanisms, including mutations in an imprinted gene, uniparental disomy (both copies of a gene coming from one parent), and changes in DNA methylation. There is reason for concern that elements of ART might predispose to imprinting disorders, because imprints are established during meiosis and both meiotic divisions in the oocyte, the first occurring at ovulation and the second at fertilization, are exposed to treatments and manipulations during ART. Any excess risk for imprinting disorders that might relate to ART is difficult to detect because the disorders are quite rare (1 in 12,000 births). Nonetheless, three of the nine known disorders have been linked to ART, including Beckwith-Wiedemann syndrome,[592–594, 800–802] Angelman syndrome,[591, 803, 804] and maternal hypomethylation syndrome.[599] In each, the epigenetic defect involves hypomethylation of the maternal allele. Once again, whether the excess risk for these imprinting disorders observed among children conceived with ART relates to treatment, to underlying infertility, or to factors predisposing to infertility is unknown.

Development

Neurodevelopmental outcomes in children conceived with ART appear normal,[805] but studies suffer from a number of limitations relating to size, duration of follow-up, possible selection bias, and choice of controls. Moreover, ART is associated with an increased risk of preterm birth and LBW, which are important risk factors for neurodevelopmental problems.

Although a meta-analysis of data from three studies including more than 19,000 children conceived with IVF and from 430,000 children conceived naturally observed an association between IVF and cerebral palsy (OR=2.18, CI=1.71–2.77), most of the excess risk appeared related to the relatively high prevalence of multiple gestation, preterm birth, and LBW in the IVF group.[806] One case-control study comparing infants with intraventricular hemorrhage (IVH) to controls matched for gestational age, birth weight, and multiple gestation observed a significant increase in the risk of grade III/IV IVH in infants conceived with IVF.[807]

In studies of children conceived with ART up to 18 years of age, their emotional development, behavior, self-esteem, family relationships, and cognitive development appear similar to those in children conceived naturally.[808–814] Likewise, studies using measures of cognitive and motor development have observed no significant differences between children conceived with ART and those conceived naturally.[815–818]

Cancer

There is concern that ART might increase the risk for some childhood cancers, because some have a suspected link to defective genetic imprinting, which, in turn, has been linked with ART. However, several studies have failed to find any substantive evidence for an increase in cancer risk among children conceived with ART.[819–822] Although one observed an increased risk of retinoblastoma among IVF children (RR=2.54, CI=1.02–5.23), the absolute risk was extremely low (seven cases among all IVF births in the Netherlands from 1995 to 2007).[595, 823] Results of an analysis of pooled results from four cohort studies examining cancer incidence in ART children found no evidence for an increased risk; the standardized incidence ratio (SIR, events observed/events expected) was 1.03 (CI=0.61–1.63).[824] A 2005 meta-analysis of data from 11 cohort studies of childhood cancer risk among children conceived with ART reached the same conclusion (SIR=1.33, CI=0.62–2.85).[825] Nonetheless, the question remains and awaits the results of larger studies with longer-term follow-up.

SUMMARY

Concerns about the health and welfare of children born after ART are reasonable and understandable. The available data indicate that ART is associated with an increased risk of multiple gestation, congenital anomalies, preterm delivery, low birth weight, and the complications associated with these outcomes. The concerns are justified, but are not cause for undue alarm.

Oocyte Donation

Until approximately 25 years ago, women with ovarian failure were understandably considered irreversibly sterile, but advances in ART have changed that view forever. Oocyte donation now offers women with premature ovarian failure, those with premature reproductive aging, and even women beyond their normal reproductive years a very realistic chance for pregnancy

A successful pregnancy established in one women (the recipient) using oocytes from another (the donor) was first reported in 1983. The original technique involved intracervical artificial insemination of a normal volunteer with sperm from the male partner of an infertile woman, uterine lavage during the preimplantation interval, and transfer of the recovered embryo to the uterus of the infertile female partner who received a programmed regimen of hormone replacement designed to synchronize endometrial and embryo development.[826] Numerous ethical and technical problems prevented wide

application. That same year saw the first report of a pregnancy established by ovum donation, IVF, and transfer to a cycling recipient.[827] Within another year, the first successful pregnancy resulting from oocyte donation and IVF in a woman with ovarian failure was reported.[828] The U.S. national ART registry recorded 17,405 oocyte donation cycles in 2007.[3]

Oocyte donation is now commonly achieved by IVF using oocytes retrieved from healthy young donors after controlled ovarian hyperstimulation and the sperm of the recipient's partner, with the resulting embryos then transferred to the uterus of the recipient.[829] Success with donor oocytes also has been achieved using tubal transfer techniques. Although straightforward in concept, the requirements for successful ovum donation IVF are many and complicated. The unique and key features of a donor oocyte IVF cycle relate to the need for embryo/endometrial synchronization and exogenous hormonal support of early pregnancy until the luteal-placental shift. Other important issues relate to donor recruitment, selection, and screening.

Indications

There are five accepted indications for ovum donation IVF—ovarian failure, genetically-transmitted disease, declining or absent ovarian function, advanced reproductive age, and persistent poor oocyte quality in IVF cycles. Women with ovarian failure from any cause (X chromosome abnormalities; idiopathic gonadal dysgenesis or premature oocyte depletion; previous surgery, irradiation or chemotherapy; autoimmune disease) are candidates. So are women who carry specific heritable disorders not amenable to PGD or who reject PGD, and women with a diminished ovarian reserve due to age or other causes who have a poor prognosis for successful IVF using their own oocytes. Rare women with severe pelvic adhesive disease and inaccessible ovaries also are occasionally encountered.

Donor Oocyte Recipients

With a few exceptions, the pretreatment evaluation and screening of couples seeking oocyte donation is virtually identical to that recommended before conventional IVF. Psychological counseling is an important element of the evaluation and helps to identify couples with unresolved concerns or fears and to ensure that both partners are fully committed to the effort.

Women with Turner syndrome may be considered candidates for ovum donation and deserve specific mention. ***Evidence indicates that pregnancy may pose unique and serious risks for women with Turner syndrome, who often have cardiovascular malformations involving the aortic root.*** Like women with Marfan syndrome, women with Turner syndrome are at increased risk for aortic dissection during pregnancy, presumably relating to the increased cardiovascular demands. ***The maternal risk of death from rupture or dissection of the aorta in pregnancy may be 2% or higher.***[830] Women with Turner syndrome expressing interest in oocyte donation should be carefully evaluated, to include echocardiography or magnetic resonance imaging, with any significant abnormality best regarded as a contraindication to oocyte donation. In general, even women with normal evaluations should be discouraged, because aortic dissection still may occur. Those who choose to proceed require careful observation and frequent re-evaluation during pregnancy.[831]

Controlled Endometrial Development

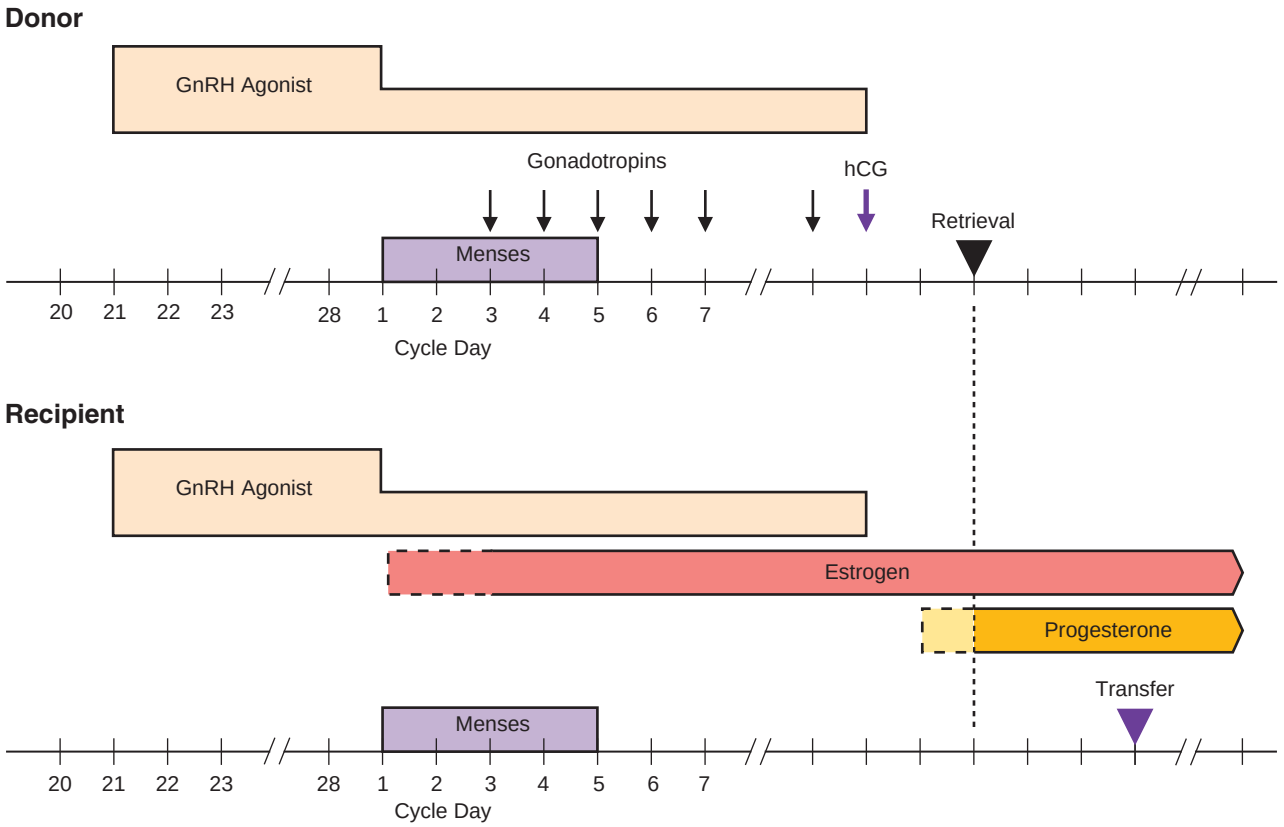
In spontaneous cycles, endometrial proliferation and secretory maturation are closely coordinated with follicular growth, ovulation, and luteal function; development of the endometrium and the embryo are naturally synchronized. In ovum donation cycles, that same careful synchronization must be orchestrated. The "window of endometrial receptivity," the interval during which implantation normally occurs, is relatively narrow and spans approximately 3 days, perhaps as much as 5 days.[832, 833] The opening and duration of the implantation window is controlled primarily by the duration of progesterone exposure. The length of the preceding proliferative phase is extremely flexible and can vary widely,[832] as occurs naturally in oligo-ovulatory women.

To synchronize endometrial development with the embryos to be transferred, recipients with functioning ovaries are first down-regulated with a long-acting GnRH agonist, treatment that women with ovarian failure obviously do not require. In either case, a programmed regimen of sequential estrogen and progesterone replacement is used to simulate the natural cycle and to promote normal endometrial growth and maturation, the same way as in frozen embryo transfer cycles (discussed above). A wide variety of treatment regimens have been used successfully to achieve controlled endometrial development and maturation.

Estrogen can be administered orally (micronized estradiol 4–6 mg daily) or transdermally (estradiol 0.2–0.4 mg daily). Both routes of administration are effective and neither has proven superior, despite the widely varying serum estradiol levels that can result.[109, 834–837] Oral and transdermal estrogen treatment regimens are designed to achieve serum levels that approximate those observed in the late follicular phase in natural cycles (200–400 pg/mL); equivalent doses of vaginal estrogen achieve markedly higher serum and tissue concentrations.[838] The duration of estrogen therapy is quite flexible and can range from as little as 7 days to as long as 3 weeks or more.[835] Progesterone can be administered intramuscularly, in doses designed to achieve serum concentrations approximating 20 ng/mL (50–100 mg daily),[839, 840] or vaginally, in the form of suppositories, tablets, or gel (180–600 mg daily). Intramuscular administration yields substantially higher serum concentrations, but endometrial tissue levels are highest after vaginal treatment.[834, 836]

The methods and extent of monitoring vary widely among programs. Many use transvaginal ultrasonographic measurements of endometrial thickness, aiming to achieve a thickness greater than 6–7 mm, at a minimum.[303, 841, 842] In the occasional woman who does not achieve the desired endometrial thickness in response to the standard replacement regimen, vaginal estrogen administration can help to promote additional proliferation.[843]

To maximize the probability for successful implantation, embryo transfer must be carefully timed. To achieve the same coordination of embryo and endometrial development that occurs in natural conception cycles, progesterone treatment in the recipient should begin on the day the donor undergoes retrieval.[754] Day 2 embryos (2 days after retrieval and fertilization) are transferred on the third day of progesterone therapy, day 3 embryos on the fourth day, and day 5 embryos on the sixth day.[109] Although the effective "transfer window" is wider than a single day, synchronous transfer provides a margin of safety and compensates for any minor variations in the speed of endometrial maturation. Flexibility in the duration of preliminary estrogen treatment in the recipient facilitates convenient scheduling. In general, estrogen therapy begins at or near the time that stimulation begins in the donor, allowing ample time to achieve the desired degree of endometrial proliferation before the donor's retrieval.



Luteal Phase and Early Pregnancy Support

In naturally conceived early pregnancies, the rapidly rising levels of hCG first "rescue" and then stimulate the corpus luteum to maintain the high levels of estrogen and progesterone secretion necessary to ensure endometrial stability in support of early embryonic growth and development until the emerging placenta achieves the capacity to assume that responsibility. The donor oocyte recipient has no corpus luteum. Consequently, exogenous luteal support must be provided for the requisite interval. Normally, the luteal-placental transition is completed between 7 and 9 weeks of gestation (menstrual dates), between 5 and 7 weeks after embryo transfer.[844] ***Exogenous estrogen and progesterone treatment therefore must continue until at least 7 weeks, and many recommend treatment until approximately 10 weeks gestation, for added safety.*** Some prefer to monitor serum estradiol and progesterone concentrations during the early weeks of pregnancy, decreasing the dose of exogenous hormone treatment by half after observing a sharp rise in levels, and discontinuing treatment after an additional week if concentrations continue to rise normally.

Oocyte Donors

The limited availability of suitable oocyte donors is the greatest obstacle in maintaining an active donor oocyte program. Donors may be a known relative or acquaintance of the recipient,[845] but most are anonymous, young healthy volunteers recruited from the local

population. In most metropolitan areas in the U.S., oocyte donors are compensated for their time, inconvenience, and assumption of risk, with payments generally ranging from $2,500 to $8,000. Outside of the U.S., such compensation is discouraged and, in some countries, is illegal.

The American Society for Reproductive Medicine has provided detailed guidelines for the appropriate screening of candidate oocyte donors.[846] In brief summary, all donors should be between 21 and 34 years of age, have a thorough medical history and physical examination to exclude those at high risk for sexually-transmitted infections or genetically-transmissible disease, and undergo standard preconception testing. In accordance with U.S. federal law, candidate donors also must be thoroughly screened for sexually-transmitted infections, including syphilis, hepatitis B (surface antigen and core antibody), hepatitis C (antibody), and HIV-1/HIV-2, using tests performed in a laboratory approved by the U.S. Food and Drug Administration (FDA) for donor screening. Screening also includes tests for gonorrhea and chlamydia, and all tests must be performed within the 30 days immediately preceding oocyte retrieval. False-positive results exclude donors and repeated testing is not permitted. Written documentation of donor eligibility is also required. Specific genetic screening is performed for donors at risk for genetic illnesses, according to race and ethnicity. Psychological evaluation by a qualified mental health professional is recommended.

Recent advances in cryobiology have greatly improved the efficiency of oocyte cryopreservation, making the possibility of egg banking a likely near-term reality. Successful egg banking would simplify oocyte donation dramatically, by eliminating the need to synchronize donors and recipients and would have the added benefit of serving to decrease the number of unused frozen embryos resulting from conventional oocyte donation cycles. The feasibility of egg banking already has been demonstrated in a report describing a series of 10 donors, 20 recipients, and 15 pregnancies.[847]

Outcomes of Oocyte Donation

Experience with ovum donation has provided important insights into the mechanisms involved in the age-related decline in female fertility. The oocyte donation model effectively dissociates oocyte and uterine age. Success rates with conventional IVF decline steadily as age increases, most noticeably after age 35, and viable pregnancies are infrequent beyond age 42. ***In contrast, the live birth rate in oocyte donation cycles varies little across all age groups.*** These data demonstrate that the declining developmental potential of aging oocytes is the limiting factor.

Data from the 2007 U.S. national ART summary indicate that among 10,321 fresh donor oocyte cycles ending in embryo transfer across all age groups, 55.1% of transfers resulted in a live birth, with an average of 2.2 embryos transferred. Among 5,633 transfers of frozen embryos derived from donor oocytes, 31.9% resulted in a live birth, with an average of 2.3 embryos transferred.[3]

There are no unique problems associated with pregnancy after oocyte donation. However, because most recipients are over age 35, their pregnancies may be considered high-risk pregnancies. Multiple pregnancies are common, and are associated with well-known and specific risks. In 2007, 42.6% of pregnancies and 40.3% of live births resulting from oocyte donation were multiples.[3] Gestational hypertension is relatively common, especially in those over the age of 40 years,[848–853] and is associated with intrauterine fetal growth restriction. In one study of outcomes in 74 women ages 45 to 56 years, the incidence of antenatal complications was 38%, including preterm labor, gestational hypertension, gestational

diabetes, preterm rupture of membranes, placenta previa, placenta accreta, preeclampsia, HELLP syndrome, and carpal tunnel syndrome.[854]

Gestational Surrogacy

Gestational surrogacy offers women without a functional uterus the opportunity to have genetic offspring. The techniques involved are no different that those applied in other forms of ART, but the ethical, legal, and psychosocial issues involved are complex.

Gestational surrogacy involves transfer of embryos to the uterus of a woman willing to carry a pregnancy on behalf of an infertile couple. ***Surrogacy is an option for couples wherein the female partner has no uterus (congenital, hysterectomy), an irreparably damaged uterus (congenital malformation, severe intrauterine adhesions), or a medical condition for which pregnancy may pose a life-threatening risk.*** The host carrier may be a relative, a friend, or someone with no attachment to the couple who may or may not be compensated for her service. Regardless what the circumstances are, candidates for surrogacy should have previously given birth and undergo thorough psychological evaluation. The legal status of gestational surrogacy varies widely among different states, and even where it has recognition, a formal legal contract is required to formalize agreements between the infertile couple and the surrogate.

Gamete and Zygote Intrafallopian Transfer

Gamete intrafallopian transfer (GIFT) and zygote intrafallopian transfer (ZIFT) are alternatives to IVF in which oocytes and sperm (GIFT) or zygotes (ZIFT) are transferred to the fallopian tubes via laparoscopy instead of into the uterus. Once commonly performed because they offered significantly higher success rates than IVF to women with normal tubal anatomy, both procedures are now relatively rare.

Combined data from the U.S. and Canada indicate that as recently as 1995, the delivery rates per retrieval for GIFT (27.0%) and ZIFT cycles (27.9%) were still higher than for IVF (22.5%).[855] Over subsequent years, IVF success rates steadily improved and surpassed those achieved with GIFT and ZIFT. Accordingly, the procedures now have few indications that can justify the greater costs and risks associated with surgical transfers. In 2007, the U.S. national ART registry reported that ZIFT accounted for 0.1%, and GIFT for less than 0.1%, of all ART procedures.[3]

In the GIFT procedure, stimulation and oocyte retrieval proceed as in a conventional IVF cycle. Very soon thereafter, laparoscopy is performed and oocytes and sperm are drawn into a specially designed catheter (separated by air) and deposited into the fallopian tubes at a level approximately 4 cm proximal to the fimbria. The techniques involved in ZIFT are the same as for GIFT, except that conventional IVF is first performed in the laboratory, and zygotes are transferred on the following day. Today, GIFT and ZIFT are reserved entirely for women in whom a uterine transfer cannot be accomplished for technical reasons. Whereas fertilization is documented in ZIFT, in GIFT it is not. With GIFT, fertilization occurs *in vivo* rather than in vitro. For that reason, GIFT may be the only option for couples with personal, moral, ethical, or religious objections to conventional IVF. Predictably, the risk of ectopic pregnancy is higher for GIFT and ZIFT than for conventional IVF; the risk of multiple pregnancy is comparable.

Oocyte and Ovarian Tissue Cryopreservation

Each year, cancer occurs in approximately 100 per 100,000 women under age 50 in the United States. Chemotherapy and radiation therapy for malignant and non-malignant systemic disease very often results in ovarian failure. Women with cancer and other serious illnesses requiring treatments that pose a serious threat to their future fertility have relatively few options. In some cases, the ovaries may be moved out of the radiation field. Treatment with GnRH agonists has been suggested as a way to protect the gonads from the insult of chemotherapy, but there is no convincing evidence for its efficacy. Although embryo banking is effective, the time required for stimulation and retrieval are often prohibitive. With recent advances in cryobiology, oocyte and ovarian tissue cryopreservation hold promise as methods to preserve reproductive potential.

Oocyte Cryopreservation

Although the first pregnancy resulting from oocyte cryopreservation was reported in 1986,[856] success rates achieved with the technology were historically very low, and only recently improving. The primary obstacle was the poor survival of oocytes, which are fragile due to their size, high water content, and chromosomal arrangement; the meiotic spindle is easily damaged by intracellular ice formation during freezing or thawing.[857, 858] Germinal vesicle stage oocytes are hardier,[859] but progress with *in vitro* maturation of immature oocytes has been slow. Another obstacle was hardening of the zona pellucida, which interfered with normal fertilization.

The improved survival of cryopreserved oocytes today relates primarily to modifications in the sucrose and sodium concentrations in traditional "slow-freeze" protocols,[860–865] changes in the initial temperature of the cryoprotectant,[866] and seeding temperature.[867] Survival rates have been further improved with vitrification, a technique that uses high concentrations of cryoprotectant and rapid freezing by immersion in liquid nitrogen, preserving oocytes in a solid glass-like state without ice formation.[868, 869] With the use of intracytoplasmic sperm injection (ICSI), the hardened zona is not a barrier to fertilization.[870, 871]

Survival, fertilization, and pregnancy rates achieved with cryopreserved oocytes are rapidly improving and approaching those achieved with fresh oocytes.[111, 115, 847, 872] A randomized comparison of results achieved with slow-freeze and vitrification observed that vitrification resulted in better oocyte survival (81% vs. 67%), fertilization (77% vs. 67%), and clinical pregnancy rates per thawed oocyte (5.2% vs. 1.7%).[873] A study examining outcomes achieved with vitrified donor oocytes observed 87% thaw survival, 87% fertilization, and 68% blastocyst formation, with 15/20 recipients (75%) achieving pregnancy after embryo transfer.[115] Another using both slow-frozen and vitrified oocytes observed 92% survival, 79% fertilization, 42% implantation, and a 57% ongoing pregnancy rate.[111]

Although the number of pregnancies and deliveries resulting from oocyte cryopreservation is still somewhat small, the number is rapidly increasing, and early perinatal outcomes data are reassuring. The incidence of chromosomal abnormalities in human embryos derived from cryopreserved oocytes is no different from that observed in control embryos derived from fresh oocytes.[874, 875] A study comparing outcomes in 200 infants derived from vitrified oocytes and in infants resulting from conventional fresh IVF found no differences in birth weight or in the incidence of birth defects.[876] A review of over 900 live births resulting from

IVF of cryopreserved oocytes also observed no increase in the prevalence of congenital anomalies, compared to that in the general population.[877]

Oocyte cryopreservation is a viable fertility preservation strategy for women without partners seeking to preserve their fertility. Unfortunately few cancer patients have sufficient time to undergo ovarian stimulation before their treatment begins. The technology also holds enormous promise as a means to simplify oocyte donation, via egg banking, and is rapidly emerging as an elective fertility preservation strategy for women anticipating delayed childbearing and concerned about their future fertility. Currently, elective oocyte cryopreservation to defer reproductive aging is controversial, primarily because the great majority of outcomes data have come from experience with cryopreserved oocytes obtained from healthy young oocyte donors and cannot be extrapolated to older women who represent the majority of those expressing interest in elective oocyte cryopreservation.[878, 879] However, when age-stratified outcomes data become available, allowing women to be accurately informed about their prognosis for success, elective oocyte cryopreservation may realistically offer women the means to set their "biological clock."

Ovarian Tissue Cryopreservation

At least in theory, ovarian tissue cryopreservation offers the means to freeze thousands of primordial follicles for later *in vitro* maturation or to store tissue for xenografting into an animal host or later autotransplantation.[880] Currently, autologous transplantation of ovarian tissue seems the most practical and effective approach because the technique has successfully restored fertility to women with ovarian failure resulting from cancer chemotherapy.[881–885]

Ovarian tissue is removed surgically via laparoscopy or laparotomy and frozen using either a slow-cool or vitrification technique, before the insult expected to result in ovarian failure. Later, it can be thawed and transplanted back into the patient in or near its original location (orthotopic transplantation) or to another site, such as the forearm or abdominal wall (heterotopic transplantation). The advantage of orthotopic transplantation is that pregnancy might be achieved without assistance, whereas heterotopic transplantation requires IVF.[880]

Live births have been achieved after transplantation of frozen-thawed ovarian tissue in sheep,[886–888] and the first live birth in a primate after a fresh heterotopic ovarian transplantation has been reported.[889] Human oocytes have been obtained from heterotopic transplants and fertilized *in vitro* to yield embryos for transfer, resulting in a biochemical pregnancy.[890] The only human pregnancy achieved after heterotopic transplantation was achieved without assistance, indicating that the oocyte from which it arose came from the patient's existing ovary rather than from the transplant. [891]

Orthotopic transplantation has been successfully achieved in humans. A number of live births have been reported after autologous orthotopic transplantation of cryopreserved ovarian tissue. Frozen ovarian tissue also has been transplanted successfully between monozygotic twin sisters after the receiving twin developed premature ovarian failure.[892] A 2008 systematic review identified 25 reports describing a total of 46 cases of ovarian tissue transplantation for treatment of premature ovarian failure or infertility, although most involved transplantation of fresh rather than frozen ovarian tissue.[893] The mean time to return of ovarian function was 120 days (range 60–244 days) and data were insufficient to evaluate function beyond 1 year. Fresh grafts were more likely to succeed, and in 25 women who sought pregnancy, eight conceived nine pregnancies.

At least one potential risk of ovarian tissue cryopreservation and auto-transplantation is re-seeding of tumor cells in women with malignancies. Future research focusing on defining patient suitability, tissue collection methods, and cryopreservation protocols is certainly warranted, but until effective techniques and the possibility for success can be defined, ovarian tissue cryopreservation will remain investigational and cannot be justified solely for the purpose of future use in healthy women.

All references are available online at: http://www.clinicalgynendoandinfertility.com

33 Ectopic Pregnancy

Ectopic pregnancy remains an important cause of maternal morbidity and mortality. However, because modern diagnostic methods now permit early recognition of most ectopic pregnancies, contemporary treatments are more conservative than in the past. The focus of attention has shifted from emergency surgery for the control of life-threatening hemorrhage to medical treatments aimed at avoiding surgery and preserving reproductive anatomy and fertility. This chapter will review the history, epidemiology, and pathogenesis of ectopic pregnancy and discuss current methods for diagnosis and treatment.

History of Ectopic Pregnancy

The modern management of ectopic pregnancy is arguably one of medicine's greatest success stories. Ectopic pregnancy was first described in the 11th century and, for centuries thereafter, often was a fatal complication of pregnancy. In medieval times, ectopic implantation was viewed as the consequence of violent emotion, usually fright or surprise, during coitus in the cycle of conception.[1] The first documented unruptured ectopic pregnancy was described in the results of an autopsy performed on a female prisoner condemned to death and executed in 1693. Ectopic pregnancy and infertility were first linked in 1752, in the report of an extrauterine pregnancy in an infertile prostitute. By the mid 19th century, autopsy observations had raised suspicion that ectopic pregnancy might be related to pelvic infection, but treatment was not yet available for either.

Early treatments were designed to kill the ectopic conceptus and included starvation, purging, bleeding, and even treatment with strychnine. Attempts to surgically disrupt or to pass electrical current into an ectopic gestational sac frequently resulted in sepsis and death. Isolated reports of abdominal surgical procedures in women with ectopic pregnancies first appeared in the early 17th century, but not again until more than 100 years later. The first known surgical procedure for ectopic pregnancy in the 18th century was performed in France, in 1714. In the United States, John Bard of New York (1759) and William Baynham of Virginia (1791) were the first to perform abdominal surgery for ectopic pregnancy.[1] However, during the first 80 years of the 18th century, only 5 of 30 women who underwent abdominal surgery for ectopic pregnancy survived; those not treated had a greater chance for survival (1 in 3)!

In 1849, W.W. Harbert of Kentucky was the first to suggest early surgical intervention to stop fatal hemorrhage.[2] Unfortunately, the diagnosis of ruptured ectopic pregnancy came too late for most. In 1876, John Parry of Philadelphia aptly described the prognosis for women with ectopic pregnancy in his era.[3]

> …when one is called to a case of this kind, it is his duty to look upon his unhappy patient as inevitably doomed to die, unless he can by some active measure wrest her from the grave already yawning before her.

After witnessing the death and autopsy of several women with ectopic pregnancies, Robert Lawson Tait of London discovered the source and means to control hemorrhage in women with ruptured ectopic pregnancies and performed the first deliberate laparotomy to ligate bleeding vessels in 1884.[4] Within little more than a year, Tait accumulated a relatively large and successful experience with the procedure.

Over subsequent years, the advent of aseptic techniques, anesthesia, antibiotics, and blood transfusions combined to save the lives of many women, but late diagnosis and intervention were still common. Even during the first half of the 20th century, the maternal mortality from ectopic pregnancy in the United States was between 2% and 4%. Although immediate salpingectomy and blood transfusion dramatically improved outcomes in women with ectopic pregnancy, the impact of modern methods for diagnosis and treatment developed over the last 25 years has been far greater. As recently as the 1970s, approximately 15% of women with ectopic pregnancies presented in hypovolemic shock, but by the early 1980s, fewer than 5%. Accordingly, attention shifted from saving lives to preserving fertility.

Epidemiology of Ectopic Pregnancy

The incidence of ectopic pregnancy is approximately 1.5–2% of all pregnancies. The incidence increased dramatically, by about 6-fold, between 1970 and 1992. At the same time, the risk of death related to ectopic pregnancy decreased by almost 90% (from 35.5 to 3.8 deaths/10,000 ectopic pregnancies).[5] In 1989, less than 2% of all pregnancies were ectopic, but related complications accounted for 4–10% of all pregnancy-related deaths, and were *the* leading cause of maternal death during the first trimester.[5, 6] Recognizing the increasing trend towards outpatient surgical and medical management of ectopic pregnancy,[7-9] the Centers for Disease Control and Prevention combined data from the National Hospital Discharge Survey and the National Hospital Ambulatory Medical Care Survey and estimated the incidence of ectopic pregnancy at 19.7/1,000 reported pregnancies in 1992,[6] the last time that U.S. national data were reported.

Efforts to determine accurately more recent trends in the incidence of ectopic pregnancy have proven very difficult, because the numbers of ectopic pregnancies managed in the

outpatient setting (not captured in hospital discharge surveys) and involving multiple health care visits (confounding data from ambulatory care surveys) have increased dramatically.[10] Moreover, the incidence of ectopic pregnancy is expressed as the number of ectopic pregnancies per 1,000 pregnancies, but pregnancies not resulting in delivery or hospitalization are not counted and the large majority of failed pregnancies now are managed entirely in the outpatient setting. Nonetheless, available evidence suggests strongly that the incidence of ectopic pregnancy is now relatively stable and no longer increasing.[11-13] Whereas the use of ovulation inducing drugs and assisted reproductive technologies (ART) has increased significantly in recent years (both known risk factors for ectopic pregnancy) and contemporary methods permit more accurate and earlier diagnosis than in the past (resulting in fewer ectopic pregnancies escaping detection), advances in detection and treatment of sexually-transmitted infections have, at the same time, helped to prevent or limit related damage to the fallopian tubes (a major risk factor for ectopic pregnancy).[12, 14, 15]

Ectopic pregnancy rates are higher for blacks and other minorities than for whites in all age groups. For all races, the ectopic pregnancy rate increases progressively with age and is three to four times higher for women ages 35–44 than for those ages 15–24.[5, 16, 17]

Risk Factors

Many women with ectopic pregnancies have one or more recognized risk factors, but half of all women with ectopic pregnancy have none.[18, 19] A comprehensive analysis of case-control and cohort studies, using women with intrauterine pregnancies or nonpregnant women as controls, has helped to define their relative importance.[20, 21 22]

Risk for ectopic pregnancy is increased as much as 10-fold for women with a previous ectopic pregnancy, compared to the general population. The overall risk for recurrence is approximately 15%, reflecting both the underlying tubal pathology that led to the first ectopic and the damage or trauma resulting from its treatment. Data from several studies indicate that the overall risk of recurrence is approximately 10% for women with one previous ectopic pregnancy and at least 25% for women having two or more.[12, 23-27] In a study comparing the recurrence risks associated with medical and surgical treatment for previous ectopic pregnancy, the risk of recurrence was approximately 8% after medical treatment with methotrexate (single-dose regimen), 10% after a salpingectomy, and 15% after a linear salpingostomy.[28] Approximately 60% of women who have an ectopic pregnancy will subsequently achieve a successful intrauterine pregnancy.[27, 29, 30]

Risk for ectopic pregnancy is increased at least 3-fold for women with documented tubal pathology.[20, 21, 36] In most cases, the damage results from sexually-transmitted infections, gonorrhea and chlamydia being the most common. Salpingitis damages the endosalpingeal mucosa, causing agglutination of mucosal folds and intraluminal adhesions that may entrap a migrating embryo, leading to ectopic implantation. Risk for ectopic pregnancy is increased 2-fold for women with circulating chlamydia antibodies and the majority of women with ectopic pregnancies have high levels.[37-40] In a retrospective cohort study of women with previous documented chlamydia infection, the risk of hospitalization for ectopic pregnancy was increased more than 2-fold for women with two previous infections and more than 4-fold for those having three or more.[41] Overall, women with surgically documented salpingitis have a 4-fold increased risk of ectopic pregnancy; the risk is approximately 10% after one episode of pelvic infection and increases progressively with each subsequent infection.[42]

Risk for ectopic pregnancy is increased at least 2-fold for women exposed to diethylstilbestrol (DES) in utero.[20, 43, 44] Numerous abnormalities of tubal anatomy have been observed in DES-exposed women, including shortened and convoluted tubes, constricted

fimbria, and paratubal cysts,[45, 46] but whether such abnormalities relate directly to the increased risk for ectopic pregnancy is unknown. DES was banned from future use in 1971 after its association with vaginal clear cell adenocarcinoma was recognized,[47] but the youngest DES-exposed women are still in their late reproductive years and occasionally may be encountered.

The incidence and absolute risk of ectopic pregnancy is reduced with all methods of contraception.[48-50] For any method, the ectopic pregnancy rate (pregnancy rate multiplied by the proportion of pregnancies with ectopic implantation) is lower than for women using no contraception (2.6 ectopic pregnancies/1,000 women-years). Among the most common methods of contraception, estrogen-progestin contraceptives and vasectomy are associated with the lowest absolute incidence of ectopic pregnancy (0.005 ectopic pregnancies/1,000 women-years). Rates are still very low, but about 60 times higher for tubal sterilization (0.32/1,000 women-years) and 200 times higher for the intrauterine device (IUD; 1.02/1,000 women-years).[49-51] The risk for ectopic pregnancy in women who conceive while using barrier methods or oral contraceptives is no different than in pregnant controls.[50]

Most of the available data relating to the risk of ectopic pregnancy associated with the IUD derives from older studies involving IUDs no longer in use.[21, 52] Only two IUDs currently are marketed in the U.S., a copper-bearing device and the levonorgestrel intrauterine system (LNG-IUS). Both are highly effective in preventing intrauterine and ectopic pregnancy with cumulative 5-year pregnancy rates that compare favorably with those observed in women after tubal sterilization (0.5–1%).[53-58] ***However, if pregnancy does occur with an IUD in situ, the risk for ectopic pregnancy is high.[59, 60]*** Logically, the IUD should be expected to protect better against intrauterine than extrauterine implantation. Consequently, a greater proportion of pregnancies that occur will be ectopic. In one study of outcomes in 64 documented pregnancies in women with a LNG-IUS, one-half were ectopic.[59]

The U.S. Collaborative Review of Sterilization, involving more than 10,000 women who had a tubal sterilization, found that the 10-year cumulative probability of pregnancy after sterilization was 18.5/1,000 procedures.[61] ***Data from the same cohort and others indicate that approximately one-third of all pregnancies resulting from sterilization failure are ectopic.[62-64]*** The overall 10-year cumulative risk of ectopic pregnancy after tubal sterilization is approximately 7.3/1,000 procedures, but risk varies with the surgical method. Bipolar coagulation is associated with the highest risk (17.1/1,000 procedures) and postpartum partial salpingectomy with the lowest (1.5/1,000 procedures). For all methods other than postpartum partial salpingectomy, the probability of ectopic pregnancy is greater for women sterilized under age 30 than for older women. The 10-year cumulative probability of ectopic pregnancy for women sterilized by bipolar coagulation before age 30 (31.9/1,000 procedures) is more than 25 times the rate for postpartum partial salpingectomy at any age, which might be attributed to an increased incidence of tubo-peritoneal fistula at the distal end of the proximal tubal segment. For all methods combined, only about 20% of pregnancies occurring within 3 years after tubal sterilization are ectopic, but more than 60% of those occurring after 4 or more years are ectopic pregnancies.[62-65]

Ectopic pregnancies have been reported following emergency oral contraception. The best available evidence indicates that emergency contraception acts primarily by preventing or delaying ovulation or by preventing fertilization, rather than by inhibiting implantation. In theory, progestational agents may inhibit tubal motility and predispose to ectopic implantation, but none of the emergency oral contraceptive regimens in use appears to increase the risk.[66-69]

Risk for ectopic pregnancy is increased approximately 2-fold for infertile women.[20, 70-74] The association between infertility and previous pelvic infection and tubal pathology offers one obvious explanation. Ovulation-inducing drugs also are associated with increased risk, but whether unrecognized co-existing tubal factors or altered tubal function in stimulated cycles may be responsible is unclear.[36, 71, 75]

The risk of ectopic pregnancy may be increased as much as 2-fold in women who conceive via ART.[76, 77] Indeed, it is interesting to remember that the first pregnancy achieved with in vitro fertilization (IVF) and embryo transfer was ectopic.[78] Although the mechanisms responsible have not been defined, natural migration into the tube and inadvertent direct tubal embryo transfer are the logical explanations. Women with tubal factor infertility or history of a previous ectopic pregnancy are at highest risk, presumably because embryos that migrate or are transferred into the fallopian tube inadvertently are less likely to return to the uterus before implantation. However, risk is increased even among women without tubal damage.[12] It is possible that elevated hormone levels in IVF cycles may adversely affect tubal transport function.[75, 79] Higher volumes of transfer media or deep catheter insertion may predispose to accidental tubal transfer.[76, 80, 81] Technically difficult transfers have been identified as another independent risk factor.[82] Heterotopic pregnancies, in which one or more embryos implant both in the uterus and elsewhere, are rare in naturally conceived pregnancies (approximately 1 in 4,000–10,000 pregnancies), but far more common in infertile women who conceive after ovulation induction or IVF.[77, 83-85] The most recent data from the U.S. ART registry suggest that the risk of ectopic pregnancy associated with ART has decreased in recent years and now is no greater than in naturally conceived pregnancies; in 2007, only 0.7% of pregnancies resulting from ART using fresh nondonor oocytes or embryos were ectopic pregnancies.[86]

Overall, the risk for ectopic pregnancy is increased approximately 2-fold among women who smoke. Compared to never smokers, risk is increased by approximately 50% for past and light smokers (one to nine cigarettes daily) and rises progressively with heavier daily consumption.[21, 87] Studies in animals suggest that the mechanism responsible may involve a lower efficiency of oocyte-cumulus complex capture or a decreased tubal ciliary beat frequency induced by chemical components of cigarette smoke.[88, 89] There is no evidence to indicate any relationship between ectopic pregnancy and exposure to other chemical or physical agents.[90]

Early age at first intercourse and the number of lifelong sexual partners are associated with a mildly increased risk of ectopic pregnancy, presumably because of the higher probability of exposure to sexually-transmitted infections.[20, 21, 91] Numerous studies have found an association between vaginal douching and ectopic pregnancy.[92-97] Understandably, most have attributed the observation to an associated increased risk of ascending infection, but others have suggested that women with symptoms of genital infection are simply more likely to douche;[98] a causal relationship between douching and ectopic pregnancy has not been established.

Pathogenesis of Ectopic Implantation

The fallopian tube is by far the most common site of ectopic implantation, accounting for more than 98% of all ectopic pregnancies. ***Overall, 70% of ectopic pregnancies are located in the tubal ampulla, 12% in the isthmus, 11% in the fimbria, and 2% in the interstitial (cornual) segment.***[99, 100] ***Ectopic pregnancies in other sites are relatively rare and divided between ovarian, cervical, and abdominal sites.***[99, 100]

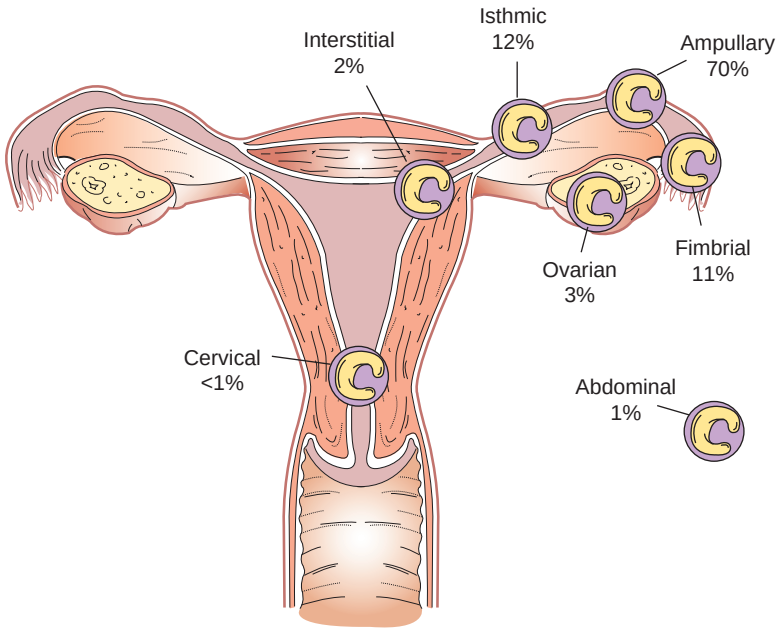
Anything that interferes with normal tubal transport mechanisms may predispose to ectopic pregnancy. It is possible, but unproven, that endocrine factors predisposing to premature implantation also may contribute to the pathogenesis of ectopic pregnancy.[101] In histopathologic studies, post-inflammatory changes (chronic salpingitis, salpingitis isthmica nodosa) have been observed in up to 90% of fallopian tubes excised for ectopic pregnancy.[102-104] Other abnormalities have included diverticula and foci of persistent decidual transformation. Any such underlying tubal pathology remains after conservative treatments, medical or surgical,

and may predispose to recurrence. The ectopic implantation itself may further damage the tube, depending on the extent of trophoblastic invasion. The endometrium in women with ectopic pregnancies usually exhibits decidual change, but also can have secretory or even proliferative histologic characteristics.[105]

The histopathology of ectopic pregnancies varies with the site of implantation. In approximately half of ampullary ectopic pregnancies, trophoblastic proliferation occurs entirely within the tubal lumen and the muscularis remains intact.[106] In the remainder, the trophoblast penetrates the tubal wall and proliferates in the loose connective tissue between the muscularis and the serosa.[106-108] In most cases, the characteristic segmental dilation of the tubal ampulla is comprised mostly of coagulated blood rather than trophoblastic tissue. In contrast, ectopic implantations in the tubal isthmus typically penetrate the tubal wall relatively early,[106] probably because the more muscular segment is less distensible. All ectopic pregnancies are not destined to rupture. In fact, many will resolve without intervention, presumably by spontaneous regression *in situ* or tubal abortion (expulsion via the fimbria).[109]

A number of studies have suggested that the prevalence of chromosomal abnormalities is increased in ectopic pregnancies and that intrinsic genetic abnormalities might in some way predispose to extrauterine implantation. However, more careful studies have failed to corroborate the finding.[110] In fact, the prevalence of chromosomal aberrations among ectopic pregnancies appears almost identical to that expected (approximately 5%) when maternal and gestational age are considered.[110-112]

Sites of Ectopic Pregnancy[100]



Diagnosis of Ectopic Pregnancy

Ectopic pregnancy is associated with a classic triad of symptoms—delayed menses, vaginal bleeding, and lower abdominal pain—observed in women with both ruptured and unruptured ectopic pregnancies. In a series of 147 patients with ectopic pregnancy, the clinical presentation included abdominal pain in 99%, amenorrhea in 74%, and vaginal bleeding in

56%.[113] Other symptoms associated with ectopic pregnancies include shoulder pain (from irritation of the diaphragm by blood in the peritoneal cavity), lightheadedness, and shock (from severe intra-abdominal hemorrhage). Unfortunately, there are no physical findings unique to ectopic pregnancy; similar symptoms are observed commonly in women with failing intrauterine pregnancies.[114, 115] Thankfully, the symptoms associated with advanced or ruptured ectopic pregnancy now are seldom seen, because most women present with mild pain or vaginal spotting well before tubal rupture and are identified promptly using methods that are more sensitive and specific than in the past.

Clinical suspicion, based on an awareness of risk factors and the early symptoms of ectopic pregnancy, is the key to identifying women who merit prompt and careful evaluation. ***The ready availability of highly sensitive and specific assays for the β-subunit of human chorionic gonadotropin (hCG) has narrowed the differential diagnosis of ectopic pregnancy to include only pregnancy-related problems, including threatened, missed, complete, and incomplete abortions.*** In almost all women, one or more serum β-hCG determinations and transvaginal ultrasonography can establish or exclude the diagnosis of ectopic pregnancy within a short time, if not immediately. Serum progesterone measurements and uterine curettage (when an early viable intrauterine pregnancy can be confidently excluded) also can be useful. Laparoscopy remains an important treatment option but rarely is any longer necessary for diagnosis alone.

Numerous diagnostic algorithms have been proposed for women suspected of having an ectopic pregnancy. All are based on the same basic concepts. Outpatient evaluation has been shown to be safe and effective for establishing a diagnosis of viable or nonviable intrauterine pregnancy or ectopic pregnancy. Accurate diagnosis is important because management of the three conditions is distinctly different.

The Serum β-hCG Concentration

HCG is secreted by the syncytiotrophoblast and becomes detectable in maternal serum as early as 8–10 days after ovulation in normal conception cycles. At or near the time of the first missed menses, serum levels between 50 and 100 IU/L are typical. Modern assays for the β-subunit of hCG are highly specific and sensitive, with detection limits below 5 IU/L. Consequently, virtually all women with suspected ectopic pregnancy who are not truly pregnant will have a negative test (no detectable hormone). False-negative tests are quite rare but, in the past, have been described in women with documented ectopic pregnancies.[116-118] False-positive tests are equally rare and most often result from endogenous heterophilic antibodies, which bind to the animal antibodies (mouse, rabbit, goat) used in commercial immunometric assay systems and thereby mimick hCG immunoreactivity.[119-122]

Although rare, heterophilic antibodies are important to understand and to recognize because persistent false positive tests may be misinterpreted as evidence of ectopic pregnancy or gestational trophoblastic disease and lead to inappropriate evaluation and treatments having serious potential consequences.[119] A false-positive hCG usually remains at the same level over time, neither increasing or decreasing. When the clinical presentation is uncertain or inconsistent with the test result, a true-positive hCG can be confirmed by 1) obtaining a similar result with a different assay method; 2) demonstrating hCG in the urine; and 3) obtaining parallel results with serial dilutions of the hCG standard and the patient's serum.

Serum β-hCG concentrations rise predictably, at an exponential pace, during the early weeks of normal intrauterine pregnancy. In general, levels double every 1.4–2.1 days in early pregnancy and peak between 50,000 and 100,000 IU/L at 8–10 weeks of gestation.[123-125] The rate of rise slows gradually as gestational age and β-hCG concentrations

increase,[125] but during the brief interval when diagnosis of ectopic pregnancy is most important (from 2 to 5 weeks after ovulation), the pattern is essentially linear. Evidence from studies performed before 1990 indicated that β-hCG levels should increase at least 66% every 2 days at concentrations below 10,000 IU/L in viable early intrauterine pregnancies, that few normal pregnancies (3–10%) ever exhibit an abnormal pattern, and most of those only transiently.[123, 124, 126, 127] However, in a more recent and careful analysis of data derived from evaluation of 287 women presenting with pain or bleeding and non-diagnostic ultrasonography who ultimately proved to have viable intrauterine pregnancies, the slowest or minimal rise observed was 24% after 1 day and 53% after 2 days.[128] The median rise in β-hCG levels was 50% after 1 day, 124% after 2 days, and 400% after 4 days. ***These data indicate that the minimal normal increase in β-hCG concentrations for women with a viable intrauterine pregnancy (50% over 2 days) is "slower" than previously reported and that the criteria used to diagnose and treat abnormal pregnancies must therefore be more conservative than has been recommended in the past.***[128]

Compared to the pattern observed in viable intrauterine pregnancies, β-hCG levels increase at a slower rate in most, but not all, ectopic and nonviable intratuterine pregnancies.[123, 129] In an analysis of data derived from evaluation of 200 women presenting to an emergency department who ultimately proved to have ectopic pregnancies, the median rise in serum β-hCG levels was 25% over 2 days, with 60% of patients having an increase and 40% exhibiting a decrease in β-hCG concentrations over that interval.[129] Among those with rising levels, the mean increase (75% over 2 days) was smaller than the average for women with viable intrauterine pregnancies, and among those with declining concentrations, the decrease (27% over 2 days) was smaller than the average for women with completed spontaneous abortions. However, in 21% of women with ectopic pregnancies, the rise in β-hCG levels was greater than or equal to the minimal rise defined for women with viable intrauterine pregnancies, and in 8% of women with declining levels, the decrease was similar to that in women with completed spontaneous abortions.[129] When levels do not rise normally, or fall, the pregnancy is almost certainly not viable, but may be ectopic or intrauterine. Serum hCG concentrations generally fall more rapidly in spontaneous abortions than in ectopic pregnancies, but the rate of decrease varies with the initial β-hCG concentration and is slower when levels are lower;[130] overall, a decrease of less than 21% after 2 days or 60% after 7 days suggests retained products of conception or an ectopic pregnancy. ***Normally rising (>50% over 2 days) or rapidly falling (>20% over 2 days) β-hCG concentrations generally are reassuring, but do not exclude the possibility of ectopic pregnancy.***[129, 131, 132]

When the risk of multiple gestation is relatively high, as in pregnancies resulting from ovarian stimulation or IVF, serial serum β-hCG determinations are more difficult to interpret confidently because the usual standards established for naturally conceived singleton pregnancies may not apply.[133] In most multiple pregnancies, β-hCG levels are higher than in singleton gestations of the same age, reflecting the combined contributions of all gestations, but still rise at a normal rate.[134] However, spontaneous pregnancy reductions are common in multiple gestations[135] and heterotopic pregnancies are not altogether rare in stimulated women.[77, 84] One or more normally progressing intrauterine gestations may produce normal or increased levels of hCG, but a coexisting failing intrauterine or ectopic gestation likely will not. The β-hCG level at any one point in time will reflect the sum contributions from all gestations, normal and abnormal, intrauterine and ectopic. The normal hCG production from a viable intrauterine pregnancy easily can obscure the abnormally small contribution from another abnormal gestation. Alternatively, falling levels of hCG production from an ill-fated intrauterine or ectopic gestation may yield a slower than expected increase in serum concentrations even though a coexisting viable intrauterine gestation is progressing normally.

It also is important to note that inter-assay variation in β-hCG measurements ranges between 10% and 15% in most laboratories. Consequently, for most confident interpretation, serial concentrations should be performed in the same laboratory whenever possible. Alert to all possible scenarios, serum β-hCG concentrations must be interpreted very cautiously.

In sum, paired serum hCG concentrations alone cannot reliably distinguish ectopic pregnancies from abnormal or even normal intrauterine pregnancies. Consequently, the diagnostic evaluation of women with suspected ectopic pregnancy also must include transvaginal ultrasonography.

Transvaginal Ultrasonography

Numerous studies have helped to define the ultrasonographic characteristics of normal and abnormal early pregnancies.[136-145] A gestational sac is the first ultrasonographic landmark in early intrauterine pregnancy. The sac consists of a sonolucent center with a thick echogenic ring, formed by the surrounding decidual reaction. Modern high frequency transducers (greater than 5 MHz) can detect a gestational sac earlier than older, lower frequency probes.[146, 147] ***Today, in pregnancies 5.5 weeks of gestation or greater, transvaginal ultrasonography should identify a viable intrauterine pregnancy with almost 100% accuracy.***[148-150] The absence of an intrauterine gestational sac 38 days or more after onset of menses or 24 days after conception is strong presumptive evidence for a nonviable pregnancy (ectopic or intrauterine).[151] The criterion is useful when the menstrual history is well documented or conception occurs under close observation, but has little practical value when broadly applied, because irregular bleeding so often confounds attempts to define gestational age.[152]

When ultrasonography reveals no obvious intrauterine pregnancy, careful examination of the adnexal regions and cul-de-sac can provide additional useful information. ***Observation of a gestational sac with a yolk sac, embryo, or cardiac activity outside of the uterus establishes the diagnosis of ectopic pregnancy and justifies immediate treatment.*** Evidence of an extrauterine gestation can be identified in up to 80–90% of ectopic pregnancies.[145, 149, 153-155] A complex adnexal mass (not a simple cyst) or fluid in the cul-de-sac increases the probability of ectopic pregnancy but does not, by itself, make the diagnosis or justify immediate treatment.[156, 157] Any other result is simply inconclusive. Some have suggested that measurements of endometrial thickness have predictive value because the endometrium is thinner in women with ectopic pregnancy than in those with viable or nonviable intrauterine pregnancies.[158] However, others have observed wide variations in endometrial thickness among women with suspected ectopic pregnancy or differences too small to have clinical utility.[152, 159] In a study involving 576 women presenting to an emergency department with complaints of pain and/or bleeding, the mean endometrial thickness was 9.56 ± 4.87 mm for women with ectopic pregnancies, 12.12 ± 6.0 mm for those with intrauterine pregnancies, and 10.19 ± 6.10 mm for women with spontaneous abortions.[160] Although the extent of overlap among groups makes endometrial thickness a poor diagnostic test, a thickness greater than 21 mm with no evidence of an intrauterine gestational sac excludes ectopic pregnancy with 96% specificity.[160]

In many cases, transvaginal ultrasonography alone can establish the diagnosis in women with suspected ectopic pregnancies by revealing either an intrauterine or an extrauterine gestational sac. In the emergency room or other acute setting, ultrasonography is diagnostic in 70–90% of women with suspected ectopic pregnancy.[15, 143, 144, 148, 150, 161, 162] ***When neither an intrauterine nor an extrauterine gestational sac is observed, defining a "pregnancy of unknown location," the possibilities include an intrauterine pregnancy in which the gestational sac has not yet developed, collapsed, or completely aborted, and an ectopic pregnancy that is too small to be detected or has aborted.*** In some cases, uterine anomalies, fibroids, or a hydrosalpinx can obscure an intrauterine or extrauterine pregnancy; obesity also can prevent confident interpretation. ***Overall, at least 25% of women with an ectopic pregnancy present first with a pregnancy of unknown location,***[15, 144, 162] ***and 7–20% of women with that initial diagnosis prove ultimately to have an ectopic pregnancy.***[144]

When available, color and pulsed Doppler ultrasonography can improve diagnostic accuracy. A small intrauterine gestational sac sometimes can be difficult to distinguish from the "pseudosac" (blood in the uterine cavity) observed in approximately 10% of women with ectopic pregnancy.[163] The local vascular changes associated with a true gestational sac can help to differentiate the two.[164, 165] Vascular pulses and arterial flow velocity increase in early intrauterine pregnancy. The extent of peri-trophoblasic arterial flow correlates with gestational sac size and serum β-hCG concentrations. Blood flow in the arteries of the fallopian tube containing an ectopic pregnancy is 20–40% greater than in the contralateral tube.[165, 166] Similarly, adnexal masses may be distinguished by the characteristics of blood flow surrounding them. For example, the resistive index of ectopic pregancies also is higher than for corpus luteum cysts.[167] However, the method has numerous diagnostic pitfalls, requires substantial technical expertise, and use of standard transvaginal ultrasonography and serum β-hCG concentrations usually is sufficient.[168]

When ultrasonography is inconclusive, serum β-hCG concentrations can serve as a surrogate marker for gestation age and help determine whether an intrauterine gestational sac should or should not be present. The concept of a "discriminatory zone," the minimum serum β-hCG concentration above which a gestational sac always should be detected in a viable intrauterine pregnancy, revolutionized the diagnostic approach to women with suspected ectopic pregnancy. When the concept was first introduced in 1981, transabdominal ultrasonography was the standard and the discriminatory zone was 6,000–6,500 IU/L.[169] ***With the development of endovaginal transducers of higher frequency, the discriminatory zone decreased progressively and now generally ranges between 1,500 and 3,000 IU/L.***[27, 157, 170-172] In one study, 185 of 188 intrauterine pregnancies (98%) among women with a β-hCG concentration greater than 1,500 IU/L were imaged.[15] ***In any given center, the discriminatory zone or value will depend on the experience of the examiner and on the type of equipment in use.***[157, 170, 171] ***Using a higher, more conservative threshold value (e.g., 2,000 or 2,500 IU/L) helps to minimize the risk of diagnostic error, but also may delay diagnosis of an ectopic pregnancy.*** The threshold value of 2,000 IU/L is suggested in the algorithm appearing in this chapter. Identification of an intrauterine gestational sac excludes the diagnosis of ectopic pregnancy, except in circumstances wherein a heterotopic pregnancy or a pregnancy in a rudimentary uterine horn must be considered.

When the β-hCG concentration is clearly above the discriminatory value, attention should focus on establishing the location of the pregnancy, now deemed nonviable by virtue of having failed to observe an intrauterine gestational sac.[27, 143, 173] ***The absence of an intrauterine gestational sac is strong, but not conclusive, evidence for an ectopic pregnancy.***[15] Other possibilities must be considered before treatment begins. In incomplete abortions, an intrauterine gestational sac may be absent or difficult to recognize. In very recent complete abortions, serum β-hCG levels may be declining rapidly, but still elevated. Even a viable intrauterine pregnancy cannot be excluded entirely when there is good reason to suspect a multiple gestation. ***Consequently, even when the β-hCG concentration is above the discriminatory value, a repeated β-hCG measurement in 1–2 days merits consideration in women at low risk for ectopic pregnancy with few or no symptoms, to identify those who might otherwise receive unnecessary treatment, or worse, harmful treatment.*** A rapidly falling β-hCG level indicates a resolving nonviable pregnancy and can be followed until undetectable. A normally rising β-hCG concentration indicates the need for repeated ultrasonography to exclude the possibility of a viable intrauterine pregnancy not detected previously and that otherwise might be exposed to methotrexate, resulting in inadvertent termination or in severe embryopathy (intrauterine growth restriction, microcephaly, and facial, cranial, and skeletal abnormalities).[174-179]

In patients with initial β-hCG values below the discriminatory zone, the absence of an intrauterine gestational sac is inconclusive; clinical signs (hemodynamic instability), symptoms (pain), and other sonographic findings (extrauterine gesational sac, complex adnexal mass, cul-de-sac fluid) must guide clinical management.[180] In some women, clinical

circumstances will demand an immediate and definitive surgical diagnosis. Women with no or few symptoms require close follow-up and serial evaluations until the possibility of ectopic pregnancy can be excluded.[148, 157, 181, 182] Repeated measurements of serum β-hCG at 2-day intervals help to distinguish nonviable pregnancies from early viable intrauterine gestations not yet large enough to detect.

In women with normally increasing β-hCG levels below the discriminatory value, ultrasonography should be performed or repeated when levels have risen above the discriminatory value. After 2–7 days, ultrasonography can be expected to demonstrate the location of the pregnancy in the large majority of cases.[144] Occasionally, a new adnexal abnormality (complex mass) or increasing clinical symptoms may demand a definitive surgical diagnosis when a desired viable intrauterine pregnancy cannot be confidently excluded.[157, 181, 182] ***Women with rapidly decreasing β-hCG concentrations warrant only continued observation because the likelihood of ectopic pregnancy is low.***[183] Nonetheless, serial measurements should be obtained until levels are no longer detectable, which can take up to as much as 6 weeks.[12] Those in whom an intrauterine pregnancy has not been documented remain at risk for rupture of an ectopic pregnancy until β-hCG is no longer detectable.[148] ***Slowly declining or abnormally rising β-hCG levels indicate a nonviable pregnancy that still may be ectopic or intrauterine, but virtually exclude the possibility of a viable intrauterine pregnancy. The same is true when β-hCG levels rise to a concentration clearly above the discriminatory value and ultrasonosgraphy is again inconclusive (no intrauterine or extrauterine gestational sac). In either case, medical treatment could be offered safely, but a presumed diagnosis of ectopic pregnancy will be inaccurate, and treatment unnecessary, in up to 40% of women.***[184] Therefore, many prefer to perform curettage to distinguish the remaining 2 possibilities (discussed below).

A conservative approach to pregnancies of unknown location prevents inappropriate intervention in a viable intrauterine pregnancy. Although it risks a modest delay in diagnosis of ectopic pregnancy and a small possibility of rupture, evidence from several studies indicates that a conservative diagnostic approach rarely compromises the care of women with pregnancies of unknown location.[150, 162, 181, 185-189] Consequently, every reasonable effort should be made to establish a definite diagnosis.

The Serum Progesterone Concentration

Serum progesterone concentrations generally are lower in ectopic pregnancies than in viable intrauterine pregnancies.[190-193] The most logical explanation is that ectopic pregnancies are almost always accompanied by abnormally low levels of hCG production. Whereas the hCG secreted by ectopic gestations is chemically and biologically indistinct from that in intrauterine pregnancies,[194, 195] production rates are lower, primarily because ectopic trophoblast proliferates more slowly and is less biologically active.[196, 197] Corpus luteum progesterone production in early pregnancy is regulated primarily by the rate of change in serum hCG concentrations.[195] Under normal circumstances, the exponential increase in hCG levels ensures that LH/hCG receptors are occupied to the extent corresponding with maximal stimulation as the corpus luteum matures and the number of available receptors increases. In contrast, few ectopic pregnancies exhibit normal hCG production rates for very long. Consequently, progesterone secretion may increase normally at first but inevitably slows, resulting in lower serum concentrations.[198, 199] There is no evidence to support the alternative hypothesis that poor luteal function in ectopic pregnancy results from reduced or absent production of other trophic feto-placental factors distinct from hCG.[195]

Serum progesterone levels associated with early normal and abnormal intrauterine pregnancies and ectopic pregnancies vary widely and overlap to a large extent. Consequently,

whereas a grossly low serum progesterone concentration is unlikely to be associated with a viable intrauterine pregnancy, it cannot distinguish an ectopic pregnancy from a failed intrauterine gestation.[9, 14, 200-204] The probability of a viable intrauterine pregnancy increases with the serum progesterone concentration. Levels greater than 20 ng/mL almost always indicate a normal intrauterine pregnancy. Conversely, concentrations less than 5 ng/mL almost always indicate a nonviable pregnancy, which may be either ectopic or intrauterine.[9, 14] Unfortunately, 50% of ectopic pregnancies, nearly 20% of spontaneous abortions, and almost 70% of viable intrauterine pregnancies are associated with serum progesterone levels *between* 5 and 20 ng/mL.[205, 206] Moreover, because exceptions to the norm can and do occur, neither threshold value is entirely reliable in an individual woman. Only about 0.3% of women with viable intrauterine pregnancies have a serum progesterone level under 5 ng/mL, but approximately 3% of those with ectopic pregnancies have progesterone concentrations above 20 ng/mL.[201, 206] The utility of serum progesterone measurements in the evaluation of women with suspected ectopic pregnancy is further limited when conception results from treatments involving ovarian stimulation. Higher than usual progesterone levels logically may be expected because treatment often yields more than a single corpus luteum.[207]

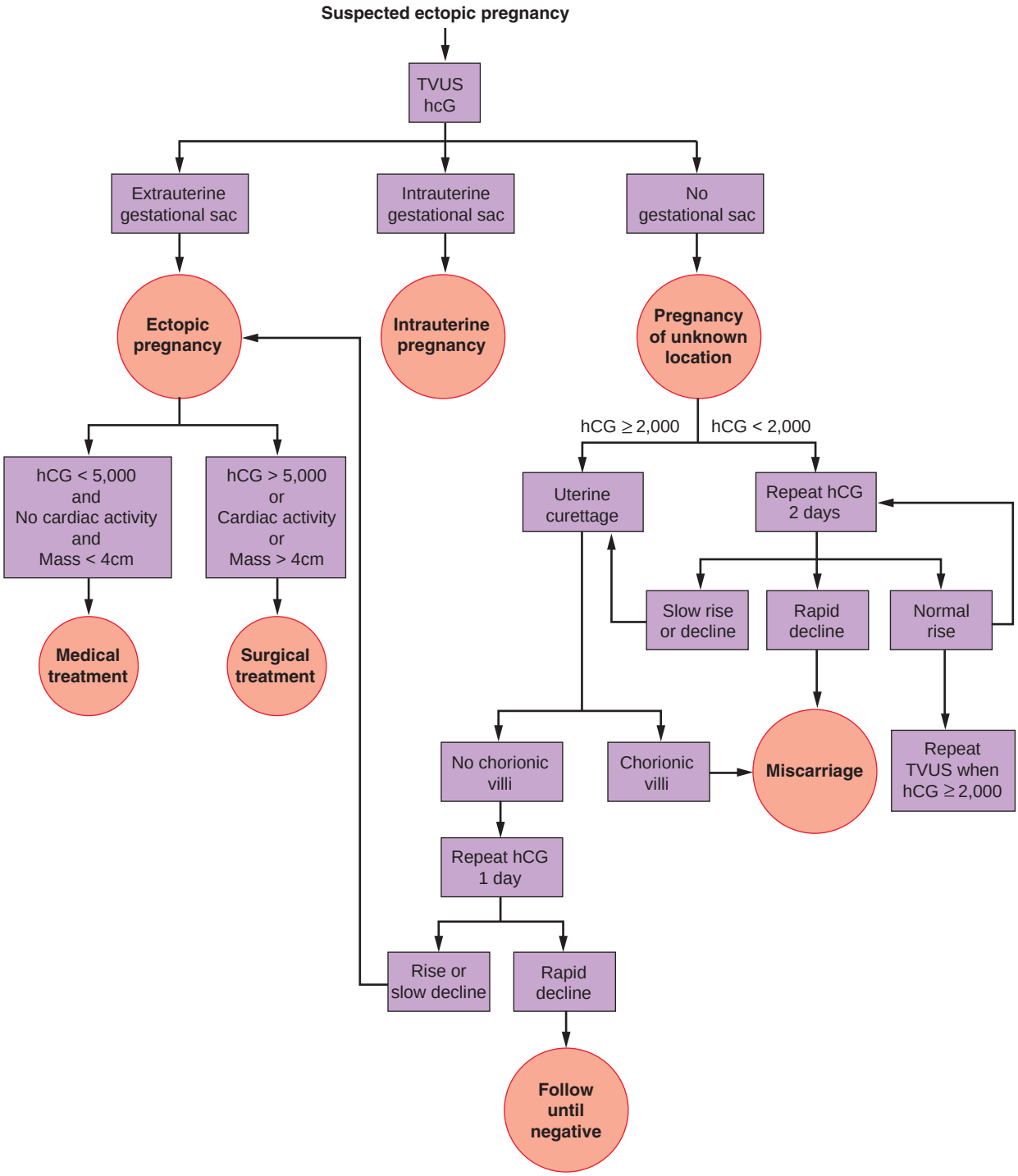
Some have suggested that progesterone levels less than 5 ng/mL identify women in whom uterine curettage can be performed safely in search of chorionic villi to distinguish ectopic pregnancies from spontaneous abortions.[9, 208] Using that approach, the probability of inappropriate intervention in a viable intrauterine pregnancy is indeed very low, but even that small risk may be unacceptably high. ***In sum, serum progesterone measurements generally add very little to the diagnostic evaluation of women suspected of having an ectopic pregnancy.***

Uterine Curettage

When ultrasonography is inconclusive and β-hCG concentrations are above the discriminatory zone, or below the threshold value and rise abnormally, plateau, or fall, the possibility of a viable intrauterine pregnancy is all but excluded; an early multiple pregnancy or an error in performing or interpreting ultrasonography are the only exceptions. ***Uterine curettage can help to distinguish ectopic from nonviable intrauterine pregnancies, but still should be applied selectively.*** Curettage clearly is inappropriate when there is any possibility of interrupting a viable intrauterine pregnancy and is unnecessary in women with rapidly falling β-hCG levels. ***Recovery of chorionic villi excludes ectopic, but not heterotopic pregnancy.*** The absence of villi makes the diagnosis of ectopic pregnancy likely, although a very recent complete abortion or a technical failure to obtain or to recognize villi also is possible; chorionic villi are not detected by histopathology in 20% of elective terminations of pregnancy.[209] ***If curettage is not performed and ectopic pregnancy is presumed, misdiagnosis leading to unnecessary treatment will occur in approximately 40% of women.***[184] ***Therefore, uterine curettage is recommended for women with nonviable pregnancies of unknown location, to distinguish an ectopic pregnancy that requires treatment from a nonviable intrauterine pregnancy that does not.***

If they are present, gross inspection of the curettings in saline will reveal obvious chorionic villi about half of the time. Frozen section and histologic examination will demonstrate villi in 80–90% of specimens obtained from women with spontaneous abortion, and is recommended when available, to help avoid delays in establishing a definite diagnosis. Otherwise, a postoperative serum β-hCG may be obtained. A 20% or greater decrease in the β-hCG level within 12–24 hours suggests strongly that the patient had a nonviable intrauterine gestation that was removed.[209, 210] Conversely, a slower rate of decrease, or an increase, strongly suggests an ectopic pregnancy.[12] Women with declining levels can be monitored with serial β-hCG concentrations until no longer detectable or until the

pathology report confirms the presence of chorionic villi. In a series of 111 women with nonviable pregnancies of unknown location that had curettage, villi were detected in 37% overall and in 51% of those whose initial β-hCG level was greater than 1,500 IU/L.[184]



Uterine curettage is costly, requires an operating room and anesthesia in some institutions, and is associated with a small risk of a complication. By comparison, uterine aspiration using a pipelle is simple to perform in the outpatient setting and minimally invasive. The results obtained with pipelle aspiration and curettage generally correlate extremely well when they are performed for suspected endometrial pathology (hyperplasia or carcinoma).

Unfortunately, pipelle biopsy is not an effective substitute for curettage in the evaluation of women with suspected ectopic pregnancy. The sensitivity of pipelle biopsy for detecting chorionic villi is unacceptably poor, ranging between 30% and 60%.[211, 212] If treatment is based on results obtained with a pipelle biopsy, up to one in three women with a miscarriage may receive unnecessary and inappropriate medical or surgical treatment. The higher sensitivity of curettage (80–90%) yields a misdiagnosis of ectopic pregnancy in no more than approximately 2 in 10 women with a spontaneous abortion.[212-214] Although rare, it is useful to remember that chorionic villi sometimes may be found in curettings from women with an ectopic pregnancy.[215]

Some advocate empiric medical treatment for all women with nonviable pregnancies of unknown location, viewing treatment as more practical and less invasive than curettage.[187, 216] ***A decision analysis found neither approach superior, but also observed that empiric treatment yields little savings, does not reduce complications, and clouds the prognosis for future fertility and the risk of repeat ectopic pregnancy.***[217] We favor curettage over empiric medical treatment, preferring to avoid unnecessary treatment and the uncertainties that result from a presumptive diagnosis of ectopic pregnancy.

Screening for Ectopic Pregnancy

Transvaginal ultrasonography and serum β-hCG determinations have proven diagnostic value in the evaluation of symptomatic women with suspected ectopic pregnancy. Not surprisingly, some have advocated applying the same diagnostic tools to screen asymptomatic women at increased risk for ectopic pregnancy. In practice, women at risk might be instructed to contact their clinician as soon as pregnancy is suspected and, if confirmed, receive careful monitoring with serial β-hCG determinations and timely ultrasonography. The alternative is to evaluate only those in whom clinical symptoms of pain or vaginal bleeding emerge. The rationale for screening at-risk women is that early diagnosis of ectopic pregnancy allows early intervention and non-invasive treatment that may help to minimize tubal damage and to reduce costs.[218, 219] However, widespread screening of symptom-free women is costly and increases the likelihood of false-positive diagnoses of ectopic pregnancy that may result in unnecessary or inappropriate medical or surgical treatment.[219, 220]

From both a clinical and economic perspective, the cost-effectiveness of screening depends on the prevalence or risk of ectopic pregnancy in the population chosen for screening. If the risk is low, very few ruptured ectopic pregnancies will be prevented and the costs of screening far exceed the benefits and savings resulting from early diagnosis and medical instead of surgical treatment. If uterine curettage is performed for all nonviable pregnancies identified through screening, the costs are even greater because the large majority will be spontaneous abortions that otherwise might be managed expectantly.[220-223] If the risk is high, the benefits and savings realized through prevention of ruptured ectopic pregnancies are proportionately greater and better justify the costs associated with screening.

Results of a decision analysis suggest that screening probably is justified when the risk of ectopic pregnancy is approximately 8% or higher. At that risk level, screening may be expected to prevent one to two ruptured ectopic pregnancies and to yield less than one false positive diagnosis for every 100 women screened.[220] ***Accepting a 2% background rate of ectopic pregnancy and considering the increased incidence associated with certain risk factors, screening seems justified for women with previous tubal surgery or ectopic pregnancy and those with known tubal pathology or who conceive with an IUD in situ or after a***

sterilization procedure. Screening is more difficult to justify for women in whom a history of infertility or pelvic infection is the only risk factor.

Expectant Management of Ectopic Pregnancy

If untreated, an ectopic pregnancy may end in tubal abortion, spontaneous regression, or tubal rupture. Unfortunately, the outcome cannot be predicted reliably. Because the risk of tubal rupture and potential associated morbidity and mortality are significant, almost all women with a diagnosis of ectopic pregnancy receive medical or surgical treatment. However, in selected women, expectant management of a presumed ectopic pregnancy is an option with a reasonably high probability for success.[224, 225]

Expectant management does not equate with simple observation alone. Rather, expectant management includes careful monitoring of clinical symptoms, serum β-hCG concentrations, and transvaginal ultrasonography. In essence, expectant management is identical to the diagnostic approach recommended for women with suspected ectopic pregnancy in whom ultrasonography is inconclusive and β-hCG levels are below the discriminatory zone.

In the absence of any significant change in clinical status, women with known or presumed ectopic pregnancies can be observed as long as β-hCG levels are steadily decreasing. Overall, approximately 25% of ectopic pregnancies are associated with declining β-hCG concentrations and almost 70% of these (about 18% of all ectopic pregnancies) will resolve spontaneously without medical or surgical treatment.[189, 225, 226] The likelihood of success with expectant management is high when there is no demonstrable extrauterine gestational sac and β-hCG levels are relatively low.[225, 226] When the baseline β-hCG concentration is less than 1,000 IU/L and falling, almost 90% of ectopic pregnancies regress without treatment.[227] Expectant management succeeds in 60% of women with decreasing β-hCG concentrations under 2,000 IU/L, but fails in over 90% with higher baseline levels.[226]

The only randomized trial involving expectant management of women with known or suspected ectopic pregnancy involved 60 women who received either oral methotrexate (2.5 mg daily for 5 days) or placebo.[228] The mean β-hCG concentrations in the 2 groups were not different; in the placebo group, the mean β-hCG was 211 IU/L (range 30–1343 IU/L). In both groups, 77% of patients were managed successfully without surgery, over an interval of less than 4 weeks. Given the absence of evidence that oral methotrexate treatment is effective, it is perhaps not surprising that outcomes in the 2 groups were similar. Whereas these data demonstrate that expectant management often can succeed, the success rates achieved with both medical and surgical treatment are higher.

Considering the potentially serious risks of tubal rupture and hemorrhage and the established safety and effectiveness of both medical and surgical treatment of ectopic pregnancy, it seems prudent that expectant management should be reserved only for asymptomatic patients with very low and falling β-hCG levels. The acceptable upper limit of β-hCG has not been established, but a threshold value of 200 IU/L has been suggested.[19] It is important to remember that tubal rupture has been observed even in women with low and declining β-hCG concentrations.[229] If expectant management is elected, close follow-up is essential and β-hCG levels should be followed until undetectable. Any plateau or rise in the β-hCG concentration should prompt medical or surgical treatment.

Limited evidence indicates that long term outcomes (subsequent intrauterine and ectopic pregnancies) after successful expectant management are comparable to those achieved with medical and surgical treatment.[230]

Medical Treatment of Ectopic Pregnancy

Methotrexate, potassium chloride, hyperosmolar glucose, actinomycin-D, and prostaglandins all have been used successfully as medical treatments for ectopic pregnancy.[231-236] Direct local injections into the ectopic gestational sac under ultrasonographic or laparoscopic guidance have been described but generally are reserved for the treatment of ectopic pregnancies in unusual sites (abdominal, cervical, interstitial) or conditions (heterotopic pregnancy), as discussed below. Systemic methotrexate therapy is simpler and less invasive. ***Methotrexate has been studied extensively, and now is established as a safe and effective alternative to surgical treatment for ectopic pregnancy.***[7, 12, 237] Medical management avoids the inherent morbidity of anesthesia and surgery and reduces costs.[238] Success rates and future reproductive performance also are comparable to those observed with surgical management.[9, 238, 239]

Methotrexate is a folic acid antagonist that inactivates the enzyme dihydrofolate reductase, thereby depleting available stores of tetrahydrofolate, an essential cofactor in DNA and RNA synthesis during cell multiplication. Rapidly proliferating tissues like trophoblasts are particularly vulnerable to its actions. Considering that methotrexate has long been used successfully for the treatment of gestational trophoblastic disease, the drug was a logical choice for the treatment of ectopic pregnancy. Even in women with undesired viable early intrauterine pregnancies, methotrexate treatment decreases the rate of increase in serum β-hCG levels; as a direct result, progesterone and 17-hydroxyprogesterone concentrations also fall.[240] Methotrexate is cleared rapidly, via the kidneys; 90% of an intravenous bolus dose is excreted unchanged within 24 hours.[241]

Prognostic Indicators

Numerous studies have sought to define ultrasonographic and biochemical characteristics that might reliably predict success or failure with medical treatment for ectopic pregnancy.[242-247] The size of any demonstrable extrauterine gestational mass, the presence or absence of embryonic heart activity or cul-de-sac fluid, and baseline serum β-hCG and progesterone concentrations all have been examined as potential prognostic indicators.

Ultrasonographic Characteristics

An ectopic gestational mass greater than 3–4 cm is widely considered as a relative contraindication to medical treatment with methotrexate.[19, 248-251] There are few data to justify the recommendation, primarily because virtually all studies have limited treatment to women with ectopic pregnancies measuring less than 3.5–4 cm. In one study of 44 women treated with intratubal injection of methotrexate, ectopic pregnancies measuring 2 cm or less regressed more often (76%) than larger masses (52%).[252] However, larger studies involving systemic methotrexate therapy have observed no correlation between size and treatment success,[155, 247, 253] possibly because the ectopic pregnancy cannot always be distinguished from surrounding blood clot, or because size does not accurately predict viability. ***In reality, the distinction has limited importance because relatively few ectopic pregnancies imaged with ultrasonography exceed 4 cm in size.***[247, 252-254]

Embryonic heart activity and free peritoneal fluid also are considered relative contraindications to medical treatment.[19, 243, 249-252, 254] Treatment can succeed, but fails significantly

more often when heart activity is present (OR=9.09; CI=3.8−22).[253, 255] Free peritoneal fluid, presumably blood, has been viewed as evidence suggesting tubal rupture and intra-abdominal hemorrhage. In older studies, 70–80% of ectopic pregnancies were associated with a positive culdocentesis (recovery of nonclotting blood), even though only 40–50% were ruptured.[256, 257] Peritoneal blood also can result from tubal abortion. ***More contemporary studies involving ultrasonography indicate that free peritoneal fluid may be observed in almost 40% of women with early unruptured ectopic pregnancies and that the presence or absence of cul-de-sac fluid does not accurately predict the success or failure of medical treatment.***[247] Thus, cul-de-sac fluid, by itself, has relatively little prognostic value.

Serum β-hCG Concentrations

Whereas ultrasonographic observations have limited prognostic value, serum β-hCG concentrations are quite useful. ***The likelihood of failed medical treatment correlates directly with the initial serum β-hCG concentration; as the level rises, the probability of success decreases.***[247, 258] A systematic review and analysis of data derived from five observational studies involving 503 women found a substantial and significant increase in failure rate for patients with initial β-hCG concentrations greater than 5,000 IU/L, compared with that for women with levels less than 5,000 IU/L (OR=5.45; CI=3.04–9.78).[258]

Serum β-hCG (IU/L)	*Women Treated Successfully (%)*[258]
< 1,000	98.5
1,000–1,999	94.4
2,000–4,999	96.2
5,000–9,999	85.7
10,000–150,000	81.2

The correlation between serum β-hCG concentrations and treatment success is understandable. High or normally increasing serum β-hCG levels suggest an advanced ectopic pregnancy that is still viable and growing.[12] Conversely, low serum β-hCG concentrations and absent embryonic heart activity are more likely to be associated with a very early or failing ectopic pregnancy having greater sensitivity to methotrexate therapy. Not surprisingly, the prevalence of embryonic heart activity increases with the β-hCG concentration. Heart activity is present in only 5% of ectopic pregnancies associated with a β-hCG level less than 5,000 IU/L, but observed in 27% with concentrations between 5,000 and 10,000 IU/L, in 41% with levels between 10,000 and 15,000 IU/L, and in 50% of ectopic pregnancies associated with a serum β-hCG level over 15,000 IU/L.[247] The correlation between β-hCG concentrations and embryonic heart activity explains why medical treatment fails more often when heart activity is observed.

In general, as one might logically expect, serum progesterone concentrations correlate closely with the β-hCG level. The serum progesterone level therefore has no significant added predictive value.

Indications and Contraindications

Logically, the best candidates for medical treatment of ectopic pregnancy are those in whom treatment is most likely to succeed, although there are other important practical considerations. The ideal candidate has the following characteristics.

Absolute Requirements

- Hemodynamic stability
- No evidence of acute intra-abdominal bleeding
- Reliable commitment to comply with required follow-up care
- No contraindications to methotrexate treatment (see below)

Preferred Characteristics

- Absent or mild symptoms (pain)
- Serum β-hCG concentration less than 5,000 IU/L
- Absent embryonic heart activity
- Ectopic mass measuring less than 4 cm in diameter

The first four criteria are the most important, for obvious reasons. Among the second four criteria, the first is a practical consideration; it is difficult to justify medical treatment in women with severe or unrelenting pain, which can be signs of impending or ongoing rupture. The next two predict a high probability for success. ***Medical treatment is not contraindicated for ectopic pregnancies associated with serum β-hCG concentrations greater than 5,000 IU/L or embryonic heart activity, but the likelihood of failure and the risk of tubal rupture are increased substantially.***[237, 242, 247] The last criterion conforms to convention in the absence of any data to indicate the safety of medical treatment in women with larger ectopic pregnancies.

Women who do not meet the absolute requirements listed above and those having a specific contraindication to methotrexate are not candidates for medical treatment.[19, 259]

Contraindications to Methotrexate Treatment

- Breastfeeding
- Immunodeficiency states
- Hematologic abnormalities (severe anemia, leukopenia, thrombocytopenia)
- Known sensitivity to methotrexate
- Active pulmonary disease
- Active peptic ulcer disease
- Alcoholism
- Clinically important hepatic or renal dysfunction

Systemic Methotrexate Treatment

Early trials of methotrexate treatment for ectopic pregnancy were modeled after regimens already in wide use for the treatment of gestational trophoblastic disease. Consequently, the first suggested treatment regimen involved multiple alternating daily doses of methotrexate and folinic acid (also known as leukovorin or citrovorum).[248, 260] As experience with medical treatment expanded, a single-dose regimen was introduced in efforts to simplify treatment, improve compliance, and reduce side effects and costs.[237, 243, 250] A 2-dose treatment regimen also has been described, aimed at maximizing success rates while minimizing the number of injections and visits required.[261]

Regardless which treatment regimen is selected, pre-treatment evaluation should include the following:

- Complete blood count
- Blood type and Rh(D)

- Serum creatinine
- Liver function tests
- Transvaginal ultrasonography

In addition, patients should be advised to discontinue any folic acid supplements they may be taking and to avoid sun exposure (to decrease risk of methotrexate dermatitis), use of non-steroidal anti-inflammatory drugs (interaction with methotrexate may cause bone marrow suppression or gastrointestinal toxicity), and intercourse or strenuous physical activity (to decrease the risk of tubal rupture).

Although Rh(D) immunoglobulin treatment is recommended commonly for all Rh-negative women who have an ectopic pregnancy or early spontaneous abortion,[262, 263] evidence to support the recommendation is weak. Overall, the likelihood of Rh sensitization after ectopic pregnancy is extremely small because few ectopic pregnancies are sufficiently advanced to have a blood volume large enough to pose a significant risk. It is entirely possible that treatment may be necessary only for ectopic pregnancies that reach at least 8 weeks gestation.[264] Nevertheless, current recommendations are to administer at least 50 μg of Rh immune globulin to all non-sensitized Rh-negative women with an ectopic pregnancy or early spontaneous abortion (protection against a feto-maternal hemorrhage of up to 2.5 mL).[262] Because side effects and complications of treatment are rare, the overall balance of risks and benefits favors treatment.

It is important to note that the single-dose, 2-dose, and multi-dose treatment regimens are named for the number of doses *intended*, rather than the number actually administered. Moreover, the first day of treatment has not been consistently defined, designated in some studies as day 0, and in others as day 1. In the absence of evidence that it matters, and for clarity, the different treatment regimens are described here as beginning on day 1.

In the "single-dose" treatment regimen, methotrexate is administered in a single dose (50 mg/m^2, day 1) only if the serum β-hCG declines 15% or more between days 4 and 7. When that occurs, treatment is deemed successful and β-hCG concentrations are monitored weekly thereafter until undetectable. In the majority of women (85%), serum β-hCG concentrations rise somewhat between days 1 and 4.[237, 265] The observation is normal and does not necessarily indicate failed treatment. However, any subsequent further increase in β-hCG levels, or a decrease of less than 15% between days 4 and 7, is indication for a second dose (on day 7), using the same criteria for judging response (on day 11). If required, a third dose can be administered on day 11 and the response evaluated by measuring β-hCG again on day 14. Although a fourth dose can be administered (on day 14), surgical treatment generally is recommended after 2 weeks of failed medical treatment.

In the 2-dose treatment regimen, methotrexate is administered on days 1 and 4 (50 mg/m^2). If the serum β-hCG declines 15% or more between days 4 and 7, levels are monitored weekly until levels become undetectable. If the β-hCG concentrations decreases by less than 15% between days 4 and 7, a third dose of methotrexate is administered (on day 7) and the same criterion is applied in judging the response to treatment on day 11.[261] If necessary, a fourth dose is administered (on day 11) and the serum β-hCG concentration is assessed again on day 14.

In the multi-dose treatment regimen, methotrexate (1 mg/kg i.m.) and leukovorin (0.1 mg/kg i.m.) are administered on alternate days, up to a maximum of four doses, until the serum β-hCG level declines by 15% from the previous value. Thereafter, β-hCG concentrations are monitored on a weekly basis until levels become undetectable.

Treatment Day	Single-Dose Regimen	Two-Dose Regimen	Multi-Dose Regimen
1	β-hCG	β-hCG	β-hCG
	MTX 50 mg/m^2 i.m.	MTX 50 mg/m^2 i.m.	MTX 1 mg/kg i.m.
2			LEU 0.1 mg/kg i.m.
3			β-hCG
			If ≥ 15% decrease (day 1 to 3), repeat β-hCG weekly until undetectable
			If < 15% decrease, MTX 1 mg/kg i.m.
4	β-hCG	β-hCG	LEU 0.1 mg/kg i.m. (if MTX on day 3)
		MTX 50 mg/m^2 i.m.	
5			β-hCG
			If ≥ 15% decrease, (day 3 to 5), repeat β-hCG weekly until undetectable
			If < 15% decrease, MTX 1 mg/kg i.m.
6			LEU 0.1 mg/kg i.m. (if MTX on day 5)
7	β-hCG	β-hCG	β-hCG
	If ≥ 15% decrease (day 4 to 7), repeat β-hCG weekly until undetectable	If ≥ 15% decrease (day 4 to 7), repeat β-hCG weekly until undetectable	If ≥ 15% decrease, (day 5 to 7), repeat β-hCG weekly until undetectable
	If < 15% decrease, MTX 50 mg/m^2 i.m.	If < 15% decrease, MTX 50 mg/m^2 i.m.	If < 15% decrease, MTX 1 mg/kg i.m.
8			LEU 0.1 mg/kg i.m. (if MTX on day 7)
11	β-hCG	β-hCG	β-hCG
	If ≥ 15% decrease, repeat β-hCG weekly until undetectable	If ≥ 15% decrease, repeat β-hCG weekly until undetectable	If ≥ 15% decrease, repeat β-hCG weekly until undetectable
	If < 15% decrease, MTX 50 mg/m^2 i.m.	If < 15% decrease, MTX 50 mg/m^2 i.m.	If < 15% decrease, surgical treatment
14	β-hCG	β-hCG	
	If ≥ 15% decrease, repeat β-hCG weekly until undetectable	If ≥ 15% decrease, repeat β-hCG weekly until undetectable	
	If < 15% decrease, surgical treatment	If < 15% decrease, surgical treatment	

Routine serial sonography during medical treatment is not useful. After methotrexate treatment, half or more of ectopic masses enlarge, probably due to formation of hematomas, but such observations do not predict treatment failure and most women remain asymptomatic.[155, 266] ***In women with complaints of increasingly severe abdominal pain, repeat ultrasonography should be performed to detect an obvious increase in peritoneal fluid suggesting rupture of a tubal pregnancy.*** Moderate to severe pain, fluid above the uterine fundus or surrounding the ovary, and a hemoglobin concentration less than 10 g/dL suggest a significant hemoperitoneum; one study found that the probability of a hemoperitoneum amounting to 300 mL or more was 93% when two or more of the criteria were met.[267]

It is important to note that symptoms of pain commonly emerge or increase over the days following methotrexate treatment.[237] The cause is uncertain but most likely reflects tubal

abortion (expulsion via the fimbria) or peritoneal tension resulting from a hematoma.[210, 268] ***A complaint of increasingly severe pain should prompt thorough re-evaluation to determine whether observation and medical treatment can be continued safely, but is not, by itself, an indication for immediate surgery.*** Although "separation pain" may be the most likely cause, tubal rupture also is a possibility. Most women can be reassured and continue outpatient management. Some may require hospitalization for more careful observation. Even those with severe or unrelenting pain after methotrexate treatment often can be managed with analgesics and serial hematocrits as long as they remain hemodynamically stable.[268] In a review of 56 women with abdominal pain severe enough to require re-evaluation or hospitalization, only eight ultimately required surgical treatment.[268] However, stubborn adherence to medical management can be difficult to justify in women with severe pain. ***Surgical treatment is indicated when a ruptured ectopic pregnancy is suspected or diagnosed, or when the patient elects not to continue medical treatment.***

Outcomes of Medical Treatment

In numerous independent case series, both the single-dose and multi-dose methotrexate treatment regimens have achieved excellent success rates (75–95%).[210, 255] A meta-analysis including 26 nonrandomized studies involving 1,327 cases of ectopic pregnancy treated with systemic methotrexate found that both single-dose treatment (1,067 women; 88% successful) and multi-dose therapy (260 women; 93% successful) were highly effective.[255] ***However, when viewed from the opposite perspective, the failure rate of single-dose treatment (127/1,067; 12%) was significantly higher (OR=1.7; CI=1.04–2.82) than for multi-dose therapy (19/260; 7%). After adjustment for other known prognostic indicators (β-hCG concentration, embryonic heart activity), the higher risk of failure with single-dose treatment was even more striking (OR=4.8; CI=1.8–12.6).[255]*** The gross difference in success rates between the two treatment regimens (approximately 5%) is comparable to that observed by other investigators[253, 269, 270] and suggests that for every 20 women treated with multi-dose rather than single-dose therapy, at least one surgical procedure can be avoided. The results of a small randomized trial are consistent with those of the meta-analysis; treatment was successful in 48/54 women (89%) receiving single-dose therapy, and in 50/54 who received multi-dose therapy (93%).[271]

The meta-analysis observed that single-dose treatment often involves more than one dose of methotrexate (15%) and multi-dose therapy frequently requires less than four doses (50%).[255] Success rates also were higher in women who received a second "single-dose" (50 mg/m^2) treatment, and lower in those who required more than four doses during multi-dose (1 mg/kg) therapy. These observations suggested the optimal treatment regimen likely would involve at least two doses of methotrexate.

The efficacy of a planned 2-dose regimen was evaluated in a study involving 104 women with a diagnosis of ectopic pregnancy (ultrasonographic imaging of an extrauterine gestational sac containing a yolk sac and/or fetal pole, no products of conception observed in frozen curettings, or an increase in β-hCG 12–24 hours after curettage).[261] Ultimately, three patients were diagnosed with a failed intrauterine pregnancy after products of conception were identified on final pathologic examination of uterine curettings. Among the remaining 101 women, 88 were treated successfully (87%); three required only a single dose of methotrexate, 73 received the planned two doses of treatment, seven required a third dose, and five required four doses. Medical treatment failed in 13/101 patients (13%) and all received surgical treatment; three were diagnosed with a ruptured ectopic pregnancy, four elected surgical treatment after medical therapy began, four had a plateau in β-hCG levels after an initial decline and chose surgery over additional medical treatment, and two received surgical treatment because of increasing β-hCG or liver transaminase levels.[261] Whereas the results achieved with the

2-dose treatment regimen are comparable to those reported for single-dose and multi-dose treatment regimens, the success rate may be a conservative estimate, primarily because all patients met fairly strict criteria for diagnosis of ectopic pregnancy. ***Patients included in previous series and trials often did not receive curettage, making it likely that many did not, in fact, have an ectopic pregnancy, resulting in falsely inflated estimates of the success of medical treatment.***

Ipsilateral tubal patency rates after successful medical treatment of ectopic pregnancy are generally comparable to those observed after conservative laparoscopic surgical treatment (linear salpingostomy) and range between 60% and 85%.[29, 237, 269, 272-274] Among those who seek another pregnancy, subsequent reproductive performance and pregnancy outcomes also are similar. ***In general, 50–80% of women treated with methotrexate later achieve an intrauterine pregnancy and 10–20% experience a recurrent ectopic pregnancy.***[28, 29, 275] Women with ectopic pregnancies associated with an IUD typically have a better prognosis, probably because their ectopic pregnancy was less likely related to tubal pathology.[275] As might be expected, older women,[21, 276-278] previously infertile women,[279-281] and those with past history of ectopic pregnancy or tubal damage[21, 276, 279] have a poorer prognosis. ***In essence, fertility after ectopic pregnancy depends more on already established patient characteristics than on the method of treatment.*** [282, 283]

Occasionally, the gestational mass may persist for weeks after successful medical treatment. Whereas it seems prudent to postpone new attempts to conceive again until the mass resolves completely, whether the delay is necessary or may reduce risk for another ectopic pregnancy is unknown.

Evidence from studies of pregnancy outcomes in women treated with methotrexate for ovarian germ cell tumors,[284] gestational trophoblastic disease,[285] or ectopic pregnancy[286, 287] do not suggest that treatment has any adverse effect on the outcomes of future pregnancies, or on ovarian reserve.[288]

Side Effects and Complications

Although medical treatment is a viable option for many women with unruptured ectopic pregnancies and, on average, costs significantly less than surgical treatment,[289] it is not necessarily the best option for all women. Medical treatment avoids anesthesia and invasive surgery, but also has its drawbacks. If thoroughly counseled and given the choice, some women prefer surgical to medical treatment.[290]

Whereas surgery usually is definitive and followed by a prompt return to normal function, medical treatment can become tedious and inconvenient. The results of both medical and conservative surgical treatment must be monitored with serial serum β-hCG determinations to ensure that treatment has been successful, but levels typically fall about twice as fast after surgery.[291] ***When medical treatment is successful, the time to resolution (undetectable serum β-hCG) generally correlates with the initial serum β-hCG concentration; the average time to resolution is approximately 5 weeks.***[237, 291] Unfortunately, some women may require weekly monitoring for up to 3 months or more before β-hCG is no longer detectable.[237, 253]

Side effects of methotrexate are relatively common, but also usually minor and transient; their prevalence is somewhat higher with multi-dose therapy than with single-dose treatment.[255] Elevated hepatic transaminases are the most common. Nausea, vomiting, and diarrhea may result from a drug-induced gastritis or enteritis. Stomatitis, reversible alopecia, and pneumonitis are uncommon but do occur. Serious side effects are rare and

include severe bone marrow suppression and hepatotoxicity. When necessary, treatment with leukovorin can help to reduce their severity and speed resolution.

Compared to women who receive surgical treatment, medically treated women more often experience prolonged vaginal bleeding and perceive more depression and limitations in physical and social function.[273, 292] ***Tubal rupture requiring emergent surgery during medical treatment can and does occur, even when serum β-hCG levels are falling; tubal rupture has been observed as long as 6 weeks after initiation of medical treatment.***[210] Evidence from randomized trials comparing systemic methotrexate and conservative surgical treatment indicates that approximately 15% of women first treated medically ultimately require surgery, about half because of tubal rupture.[272, 273] Isthmic tubal ectopic pregnancies are at higher risk for rupture, but cannot be accurately differentiated from more common ampullary implantations without laparoscopy. ***Fortunately, tubal rupture does not appear to have any independent adverse effect on subsequent fertility or pregnancy outcomes.***[36]

Local Medical Treatment by Direct Injection

Methotrexate also can be administered by direct local injection (1 mg/kg) into an ectopic gestational sac under laparoscopic or ultrasonographic guidance.[254, 291, 293] The method delivers a high concentration of the drug to the site of implantation and achieves circulating concentrations comparable to those seen with systemic therapy.[294] A large experience with direct local injection of methotrexate has accumulated, mostly in Europe. Overall, results have been somewhat inconsistent but generally comparable to those achieved with systemic therapy.[29] Direct local injection also is more invasive, more costly, and requires greater technical skill. With those disadvantages, and no clear advantages, systemic methotrexate treatment is the more logical choice.

The efficacy and safety of direct local injection of other medications (potassium chloride, hyperosmolar glucose) and their long-term impact on fertility are not well established because experience is limited,[29] However, intratubal injection of potassium chloride or hyperosmolar glucose does have one important and specific niche application. In heterotopic pregnancies, the method can ablate the ectopic implantation without compromising a coexisting viable intrauterine pregnancy while also avoiding surgery and its attendant potential complications.[295]

Surgical Treatment for Ectopic Pregnancy

The contemporary management of ectopic pregnancy has moved away from surgical treatment. Nonetheless, many women still receive surgical treatment, by choice, or because they are poor candidates for medical treatment. Traditionally, ectopic pregnancies were treated surgically and salpingectomy was the most common operation performed. As modern methods for early diagnosis of unruptured ectopic pregnancies emerged, surgical treatment shifted gradually to more conservative procedures like linear salpingostomy and segmental resection. At first, most operations still were performed via laparotomy, but soon thereafter, laparoscopic surgery became standard treatment for unruptured ectopic pregnancies. With the instrumentation available today, even most ruptured ectopic pregnancies can be successfully managed laparoscopically.

Indications

The following generally are accepted as indications for choosing surgical treatment over medical therapy:

- Hemodynamic instability
- Rupture of an ectopic mass
- Coexisting viable intrauterine pregnancy
- Unwillingness or inability to comply with required follow-up after medical treatment
- Lack of ready access to a hospital for management of tubal rupture
- Desire for permanent sterilization
- Contraindications to medical treatment
- Failed medical treatment

Surgical treatment also merits serious consideration for women having clinical characteristics known to be associated with an increased risk for failed medical therapy, such as a serum β-hCG concentration over 5,000 IU/L or demonstrable embryonic heart activity. In women who are stable hemodynamically, surgical treatment should be reserved for those having a demonstrable extrauterine gestational sac or an adnexal mass consistent with the diagnosis of ectopic pregnancy. Otherwise, there is a high probability that no ectopic pregnancy will be identified at surgery. Such women are best managed expectantly until the diagnosis is more firmly established, via repeated β-hCG measurements and timely ultrasonography, or by performing curettage (when a viable intrauterine pregnancy can be confidently excluded). Alternatively, they can be treated empirically with medical therapy, realizing that many will receive unnecessary methotrexate treatment and thereafter carry a presumptive diagnosis that clouds their prognosis and decisions regarding future treatment.

Surgery does have some distinct advantages over medical therapy, primarily including a shorter time to resolution of the ectopic pregnancy, with no need for prolonged monitoring. Surgery also allows for an accurate assessment of the pelvic anatomy, which helps in subsequent counseling and treatment planning.

Surgical Techniques

For the treatment of ectopic pregnancy, laparoscopy has several advantages over laparotomy—less blood loss, fewer adhesions, less operating time, a shorter hospital stay (usually less than 24 hours), reduced postoperative analgesic requirements, and a more rapid convalescence—all of which also help to reduce costs.[29, 296-303]

The recommended surgical procedure for an unruptured ampullary ectopic pregnancy is linear salpingostomy. The technique involved is straightforward. A longitudinal incision is made on the antimesenteric surface of the fallopian tube directly over the bulging mass using an electrosurgical needle or scissors, or a laser, and the products of conception are gently removed with forceps or irrigation. Preliminary injection of a dilute solution of aqueous vasopressin (1 unit/mL)[304] or oxytocin (20 units)[305] into the subjacent mesosalpinx can help to minimize the amount of electrocautery needed to achieve hemostasis The incision may be closed, but usually is left open to heal by itself; long-term reproductive outcomes are the same with either technique.[306] Because the bulk of the ectopic mass usually resides just beneath the serosa, the endosalpingeal mucosa most often can be left relatively undisturbed. Attempts at linear salpingostomy are successful in approximately 80% of women; in the remainder, persistent bleeding may require salpingectomy. Unsuccessful salpingostomy procedures are associated with higher serum β-hCG concentrations,[307] possibly because they are more advanced and possess a more highly developed neovasculature.

Fimbrial expression of an ectopic pregnancy may risk causing greater damage to the tube and is best reserved for those already protruding through the fimbria.[308] The typical subserosal propagation of ampullary ectopic pregnancies explains why "milking" a more proximal ectopic is not recommended. ***Isthmic ectopic pregnancies probably are best managed by segmental excision, with the option of later microsurgical tubo-tubal anastomosis.*** The lumen of the isthmic segment of the fallopian tube is much narrower than in the ampulla, more likely to be damaged by salpingostomy, and more prone to postoperative obstruction. The involved segment is simply excised, taking care to ensure that hemostasis is achieved.

Conservative surgical treatment is not always the best or most appropriate option. In certain circumstances, salpingectomy is the more appropriate choice. Salpingectomy is most clearly indicated in the following circumstances:

- Completed childbearing
- Recurrent ectopic pregnancy in the same fallopian tube
- Uncontrolled bleeding
- Extensive damage to the involved tube, with a normal contralateral tube

In all cases, every effort should be made to preserve the adjacent ovary. The only indication for salpingo-oophorectomy (which rarely arises) is bleeding that cannot be controlled by more conservative measures.

Outcomes of Surgical Treatment

Observational studies indicate that cumulative intrauterine pregnancy rates are significantly higher after salpingostomy than after salpingectomy (73% vs 57%), but the incidence of recurrent ectopic pregnancy also is higher (15% vs. 10%).[27, 29, 30] Parous women are more likely to conceive again than nulliparous women[281, 309] Women with a normal contralateral tube have a higher probability of achieving an intrauterine pregnancy and a lower risk for recurrent ectopic pregnancy than those with contralateral tubal pathology.[303, 309, 310]

Conservative surgical treatment is successful (requires no additional treatment) in approximately 90% of women with unruptured ectopic pregnancies. The combined results of two randomized trials [298, 299, 301]indicate that salpingostomy is less often successful when performed laparoscopically than via the open surgical approach (OR=0.28; CI=0.09–0.86), because the incidence of persistent trophoblast is higher (OR=3.5; CI=1.1–11).[29] ***Among women who desire future fertility, tubal patency rates (80–90%) and intrauterine (55–75%) and recurrent ectopic pregnancy rates (10–15%) after laparoscopy or laparotomy are similar.***[29]

Persistent Ectopic Pregnancy

Persistent ectopic pregnancy is the most common complication of conservative surgical treatment for ectopic pregnancy.[311] The phenomenon was first described in 1979.[312] The reported incidence of persistent ectopic pregnancy has ranged widely from 3% to almost 30%,[313-316] in part because of differences in definition (postoperative increase in serum β-hCG vs. continued growth requiring additional treatment).[315, 317] The risk for persistent ectopic pregnancy may be increased when surgery is performed early (before 6 weeks gestation) and for small ectopic pregnancies (less than 2 cm in diameter) which may be more difficult to identify and excise completely.[313, 314, 318] ***Clinicians must remain alert to the possibility because unrecognized persistent ectopic pregnancies frequently rupture during the postoperative period.***[319]

The most effective way to identify persistent ectopic pregnancies is to monitor serum β-hCG levels during the postoperative period. Recommendations have varied from every 3 days to 2 weeks. The postoperative day 1 serum β-hCG concentration has significant predictive value; the greater the fall from preoperative levels, the lower the incidence of persistent ectopic pregnancy.[320] A 50% decline is the threshold value offering the best blend of sensitivity (42%) and specificity (88%). Women in whom the serum β-hCG concentration falls less than 50% are more than 3 times as likely to have a persistent ectopic pregnancy (RR=3.51; CI=1.25–6.68), but risk is quite low when levels fall by 80% or more.[320] ***In light of these observations, it seems logical to recommend measuring serum β-hCG on postoperative day 1 and every 3–7 days thereafter, depending on the rate of decline, until levels are undetectable.*** For those who exhibit evidence of persistent ectopic pregnancy (rising or slowly falling postoperative β-hCG concentrations), the single-dose methotrexate treatment regimen (50 mg/m^2 i.m.) is highly effective.[321]

A decision analysis comparing observation and prophylactic postoperative methotrexate, involving 1,000 hypothetical women treated with linear salpingostomy, concluded that prophylactic treatment results in fewer cases of tubal rupture, fewer procedures, and lower costs, but also invites complications relating to medical treatment.[322] Treatment can be offered routinely or reserved for those in whom the postoperative day 1 serum β-hCG does not fall by more than 50%. Regardless whether prophylactic treatment is or is not used, serum β-hCG levels should be monitored until no longer detectable. Intrauterine pregnancy rates after treatment of persistent ectopic pregnancy are similar to those after primary treatment.[323]

Unusual Types of Ectopic Pregnancy

Heterotopic pregnancies and abdominal, ovarian, interstitial, cervical, and cesarean scar ectopic pregnancies present unique challenges and often require individualized treatment.

Heterotopic Pregnancy

A heterotopic pregnancy involves coexisting pregnancies at 2 different implantation sites. The most common combination is an intrauterine and an extrauterine gestation, most of which are in the tube (90%), but implantations in the cervix, ovary, interstitial segment, abdomen, and previous cesarean scar have been reported.[324-330] The often quoted approximate incidence of 1/30,000 pregnancies was derived from calculations over 50 years ago, based on the observed incidence of ectopic pregnancy (0.37%) and dizygotic twinning (0.8%) at the time.[83] The incidence of heterotopic pregnancy has risen substantially due to increasing use of exogenous gonadotropins and assisted reproductive technologies (ART). Today, the best overall current estimate is approximately 1/3,900 pregnancies,[27, 331, 332] but the incidence is about 1.5/1,000 pregnancies resulting from ART.[333]

Heterotopic pregnancies often escape early recognition because both serum β-hCG concentrations and ultrasonography can be misleading. Normally rising levels of β-hCG derived from a normally developing intrauterine pregnancy usually obscure the abnormal pattern typically observed in ectopic pregnancies, and when ultrasonography reveals an intrauterine pregnancy, the possibility of an ectopic pregnancy generally is excluded. Consequently, diagnosis is delayed and over half of all heterotopic pregnancies are recognized only after tubal rupture occurs.[326] The signs and symptoms of heterotopic pregnancy are similar to those of ectopic pregnancy but, unfortunately, are often dismissed.

Treatment of heterotopic pregnancies is complicated by the coexisting intrauterine pregnancy. Expectant management is inappropriate because neither serum β-hCG levels nor ultrasonography can accurately determine the fate of the ectopic pregnancy and the risk for rupture. Systemic methotrexate treatment is contraindicated when the intrauterine pregnancy is viable and desired. Even direct local injection of methotrexate[334] may be unwise because, ultimately, it too is systemic.[294] Surgical treatment by salpingostomy or salpingectomy is effective and generally is considered the treatment of choice.[325, 326] Selective embryo reduction by direct local injection of potassium chloride[295, 335] or hyperosmolar glucose[336, 337] into the ectopic gestational sac is another viable treatment option. However, in a review of 11 cases of heterotopic pregnancy treated with potassium chloride injection, 6 patients failed medical therapy and required surgical treatment.[325]

Abdominal Pregnancy

Abdominal ectopic pregnancies, involving implantation in the peritoneal cavity, are rare, with an estimated incidence of approximately 1/10,000 pregnancies and 1/100 ectopic pregnancies.[338, 339] Implantation sites include the omentum, pelvic sidewall, broad ligament, cul-de-sac, the spleen, bowel, liver, diaphragm, and the serosa of the uterus. Whether abdominal pregnancies result from primary peritoneal implantation or secondary implantation of a tubal abortion is unknown. One reported case after in vitro fertilization in a patient without tubes may have resulted from uterine perforation at the time of embyo transfer.[339]

The symptoms most frequently encountered are abdominal pain, nausea and vomiting, general malaise, and painful fetal movements, and the most frequent physical findings are abdominal tenderness, abnormal fetal lie, and a displaced uterine cervix.[340] In rare cases, diagnosis may follow a failed induction of labor.[341] Ultrasonography is the most accurate method for diagnosis, but fewer than half of all abdominal pregnancies are recognized before intra-abdominal hemorrhage occurs. The classic ultrasonographic sign is the absence of any myometrium between the maternal bladder and the pregnancy.[342]

Immediate surgery is the treatment of choice, except perhaps in rare cases of advanced abdominal pregnancy involving a fetus nearing viability. If recognized early, an abdominal pregnancy may be amenable to laparoscopic surgery;[324, 343, 344] in one reported attempt, medical treatment with methotrexate was unsuccessful.[345] Although delivery of viable infants has been reported,[342, 346] the likelihood is very low and the risk of maternal complications is very high. The maternal mortality associated with abdominal pregnancy (at least 5/1,000 cases) is the highest of all types of ectopic pregnancy.

Whereas delivery of the fetus is accomplished easily, management of the placenta is more complicated. When technically feasible, the placenta should be removed because complications (hemorrhage, abscess, sepsis, intestinal obstruction, amniotic fluid cysts, hypofibrinogenemia) are otherwise common.[347] However, removal of the placenta also can result in hemorrhage that can be difficult to control.[348] When the placenta is left *in situ*, arterial embolization and systemic methotrexate treatment may help to speed its involution.[342, 349, 350]

Ovarian Pregnancy

Ovarian pregnancy accounts for less than 3% of all ectopic pregnancies. The clinical signs and symptoms are the same as for more common tubal ectopic pregnancies.[351, 352] Although ultrasonography may suggest the possibility, diagnosis usually is made only at surgery, or by the pathologist, because an ovarian ectopic pregnancy frequently is confused with a hemorrhagic corpus lutuem at the time of surgery,[351, 353] Diagnosis of ovarian pregnancy has specific

criteria, but they are largely academic: 1) an intact ipsilateral tube, separate from the ovary; (2) a gestational sac occupying the position of the ovary; (3) a gestational sac connected to the uterus by the ovarian ligament; and (4) ovarian tissue in the wall of the gestational sac.[354] Treatment of almost all known ovarian ectopic pregnancies has been surgical. Case reports have described successful methotrexate therapy,[355-357] and it seems likely that many more of the innumerable ectopic pregnancies treated medically have been ovarian implantations.

Interstitial Pregnancy

No more than approximately 2% of tubal ectopic pregnancies implant in the interstitial tubal segment, which lies within the uterine wall and measures 1–2 cm in length. In the largest single series of 32 interstitial pregnancies collected by the Society for Reproductive Surgeons, 13 were recurrent ectopic pregnancies, 12 of those in women with a previous ipsilateral salpingectomy, and 11 were in women who conceived after IVF.[358]

Conventional gynecologic wisdom has held that interstitial pregnancies rarely rupture before approximately 12 weeks of gestation because the myometrium surrounding the interstitial tubal segment is more distensible than the muscularis of more distal portions of the fallopian tube. However, in the case series alluded to above, 14/32 (44%) interstitial pregnancies ruptured before diagnosis and all before 12 weeks (mean 6.9 weeks; range 5–12 weeks).[358]

Most cases of interstitial pregnancy are diagnosed after the typical symptoms of ectopic pregnancies appear. The unusual location of interstitial pregnancies makes diagnosis difficult, but their ultrasonographic characteristics also are somewhat unique. A careful and experienced examiner may observe an eccentric gestational sac or heterogeneous mass, abnormal thinning of the myometrial mantle, or an abnormally prominent interstitial tubal segment ("the interstitial line"), the last criteria having the greatest diagnostic sensitivity (80%) and specificity (98%).[359] Other proposed criteria include the combination of an empty uterine cavity, a gestational sac separate and at least 1 cm from the lateral edge of the cavity, and a thin (<5 mm) myometrial layer surrounding the sac.[360] When doubt persists, laparoscopy can help to accurately differentiate true interstitial pregnancies from normal but eccentric or "angular" intrauterine pregnancies.[361]

Historically, the traditional treatment for interstitial pregnancy has been hysterectomy or cornual resection by laparotomy, primarily because most interstitial pregnancies were not recognized before rupture and often were associated with massive hemorrhage.[327, 362-364] Now, earlier diagnosis of interstitial pregnancy offers the opportunity for more conservative surgical or medical treatment. Some advocate laparoscopic "cornuostomy" with excision of the interstitial portion of the tube, if necessary.[358, 362] Others prefer primary medical therapy, typically using a multi-dose methotrexate treatment regimen, reserving surgery for those who fail medical management; in one series, 16/17 (94%) were treated successfully, including four cases in which embryonic heart activity was present.[365] Successful hysteroscopic resection of interstitial pregnancies and selective arterial embolization (alone or in combination with medical treatment) have been described.[366-369] Direct local injection of potassium chloride is another option that may have particular value for the treatment of heterotopic interstitial pregnancies,[362, 370] A recently proposed "best practice" recommends systemic methotrexate therapy for hemodynamically stable women with interstitial ectopic pregnancies, and laparoscopic surgical resection for those who are hemodynamically unstable or prefer surgery over medical treatment.[364]

The risk of uterine rupture in a subsequent pregnancy after successful treatment for an interstitial pregnancy is uncertain, but cases of uterine rupture have been reported.[371, 372] Consequently careful monitoring of any later pregnancy is essential and cesarean delivery is recommended.

Cervical Pregnancy

Cervical pregnancy is a rare type of ectopic pregnancy in which implantation occurs in the endocervical canal. The reported incidence ranges between 1/2,500 and 1/10,000 pregnancies.[373, 374] Cervical pregnancy is somewhat more common among pregnancies conceived via ART, with an incidence of approximately 1/1,000 pregnancies resulting from IVF.[375] The cause is unknown, but over two-thirds of women with cervical pregnancies have had a previous uterine curettage or cesarean delivery, suggesting some association with surgical trauma to the uterus or cervix.[373, 374]

Painless vaginal bleeding is the classical symptom of cervical pregnancy. The cervix is usually enlarged or distended, appearing hyperemic or cyanotic, soft, and enlarged out of proportion to the uterus ("hour-glass" cervix).[373] Most often, the diagnosis has been incidental to routine ultrasonography or curettage for an assumed incomplete abortion. Early diagnosis is based on ultrasonographic observations of an endocervical gestational sac with associated trophoblastic invasion below a closed internal cervical os, with a normal endometrial thickness;[373, 376, 377] focally increased blood blow detected by color and pulsed Doppler ultrasonography or cardiac activity confirms the diagnosis.[376, 378, 379]

In hemodynamically stable women with cervical pregnancies, conservative treatment is aimed at preserving the uterus and reducing the substantial risks of catastrophic hemorrhage. A wide variety of management strategies have been employed successfully. Local and/or systemic treatment with methotrexate has been successful in over 80% of treated women with cervical pregnancies,[373, 380] but experience is limited to case reports and small series.[381-383] Other medical treatment strategies include direct local injection of potassium chloride when embryonic heart activity is present.[384-386]

Traditional treatment for cervical pregnancy was curettage and hysterectomy, when necessary for control of hemorrhage. Methods aimed at minimizing that risk have included cervical cerclage, intracervical injection of vasopressin, and transvaginal ligation of the cervical branches of the uterine arteries. Similar measures, as well as intracervical balloon tamponade and bilateral uterine or internal iliac artery ligation have been used to control postoperative bleeding. More recently, uterine artery embolization (UAE) has been used both preoperatively and postoperatively to prevent or control hemorrhage,[373, 384, 385] but experience to date is insufficient to warrant its recommendation.

Cesarean Scar Pregnancy

Ectopic pregnancies in a cesarean scar account for approximately 6% of all ectopic pregnancies among women with a previous cesarean delivery.[387-389] They are presumed to result from migration of the embryo through a defect in the scar.[390]

The clinical presentation of women with cesarean scar ectopic pregnancies varies widely, from vaginal bleeding, with or without pain, to uterine rupture and hemorrhagic shock.[391] Diagnosis usually is made by ultrasonography in the first trimester, revealing an enlarged hysterotomy scar with an associated mass that extends beyond the external contour of the uterus.[392, 393] Associated findings include the absence of fetal parts inside of the uterus or myometrium between the gestational sac and the bladder. When identified, efforts should be directed to defining the extent to which the pregnancy involves surrounding structures.[388, 390]

The best management for cesarean scar ectopic pregnancies has not yet been defined. Consequently, treatment must be individualized, according to the patient' desire for future fertility and the size and gestational age of the pregnancy. Management options include

resection via the vagina, laparotomy or laparoscopy, local potassium chloride injection, and either local or systemic methotrexate treatment.[387-391, 394] Uterine artery embolization (UAE) also has been used to decrease the risk of hemorrhage.[394-397] In a randomized trial involving 72 women with cesarean scar pregnancy treated with UAE or methotrexate before dilation and curettage, UAE was associated with significantly less blood loss and a shorter hospital stay.[398] Although some cesarean scar pregnancies may reside partially in the uterus and progress normally, most do not. Expectant management generally is not recommended due to the substantial risk of rupture and massive hemorrhage.[399] After successful treatment, recurrent cesarean scar implantation, successful intrauterine pregnancies, placenta accreta, and uterine rupture (resulting in maternal death) all have been reported.[400-402] Consequently, early ultrasonography to establish the site of implantation is indicated in subsequent pregnancies.[403]

All references are available online at: http://www.clinicalgynendoandinfertility.com

Appendix I: Interpreting Epidemiologic Reports

Clinical practice is the ultimate distillate of evidence, judgment, and experience. The safety, side effects, and benefits of treatments are established by epidemiologic studies. The clinician must determine whether the data derived from epidemiologic studies are clinically relevant and useful. The incorporation of the data into clinical practice depends upon that determination. In this appendix, we provide a guide for interpreting epidemiologic reports, a guide intended to help clinicians make appropriate determinations regarding epidemiologic data, and ultimately to apply this information properly in clinical practice.

The Hierarchy (in descending order) of Epidemiologic Studies

I. Randomized Controlled Trials

A randomized trial is a true clinical experiment in which an intervention is compared with a standard treatment, no treatment, or a placebo, with allocation to treatment by chance. More than one comparison can be made within a study. Participants theoretically have a random (an equal and unbiased) chance of being assigned to each group in the study, and the participant characteristics should be nearly if not totally the same in each group. In *crossover trials*, participants are randomly assigned to one treatment group and later switched to the other group, and thus the participants serve as their own controls.

Advantages:	Provides scientific, epidemiologic proof.
Disadvantages:	Very expensive and time-consuming. Only a limited number of hypotheses can be evaluated in any one study.
Example:	The Women's Health Initiative

II. Observational Studies (Non-experimental Studies: Observation Without Intervention)

Cohort studies: A prospective follow-up over a long period of time of a large group of individuals, also referred to as longitudinal or follow-up studies. Exposure information is collected from all subjects who are disease-free, and subjects are followed over time to determine who develops disease. An historical cohort study is retrospective, following a cohort in past time, not from current time onwards.

Advantages:	A relatively accurate estimation because of large numbers, can evaluate changes over time, avoids recall bias.
Disadvantages:	Expensive, lengthy in time, and subject to biases (particularly selection bias and surveillance bias) making the two groups being compared unequal.
Example:	The Nurses' Health Study

Case-control studies: A retrospective comparison of a group of individuals with a condition or problem compared with a carefully selected control group. Subjects are selected according to specific inclusion and exclusion criteria. The exposure history of those with disease and those with no disease is collected and compared.

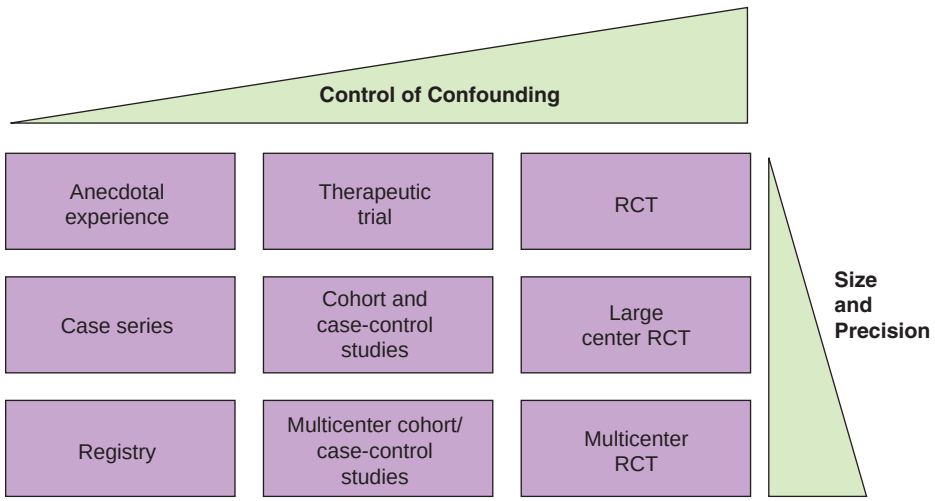
Advantages:	Relatively quick and inexpensive because of small sample sizes.
Disadvantages:	Subject to biases and errors.
Example:	WHO Collaborative Study of Cardiovascular Disease and Steroid Hormone Contraception

Cross-sectional studies: A description of a group of individuals at one point in time.

Advantages:	A reliable method to estimate prevalence, quick and inexpensive.
Disadvantages:	Cannot assess changes over time and very susceptible to sampling error (the group is not representative of the actual population of interest).
Example:	The Health and Nutritional Examination Survey

III. Clinical Reports

A case series:	A collection of similar cases that suggests more than a chance or coincidental occurrence.
A case report:	An anecdotal report that serves to bring attention to a possible problem or condition.



Possible Confounders and Biases of Importance

Confounders: Factors associated with the disease and the exposure, such as age, body weight, smoking, family history, duration of contraceptive use, preferential prescribing, healthy user effect.

Biases: Errors due to study design.

Detection or Surveillance Bias: Systematic errors in methods of ascertainment, diagnosis, or verification of cases. Not everyone in the study population has equal access to or utilization of medical interventions and diagnostic tests.

Publication Bias: Negative (null) studies and studies that confirm old results tend not to be published. An important source of bias in meta-analysis.

Reporting or Recall Bias: Inaccurate memory and dishonesty introduce errors.

Selection Bias: Differences in characteristics between those selected for study (cases) and those in the control group, such as preferential prescribing, family history, preferential referral of patients, healthy user effect. For case-control studies, the source of the controls is important. Hospital-based controls are less likely to be representative of the general population than population-based controls. It is best to choose controls by random selection, but this is not always possible. Selection bias in a cohort study can result in differences between exposed and unexposed groups.

Information or Observer Bias: A flaw in measuring exposure or outcome that produces different results between comparison groups. Nonresponse by subjects or patients lost to follow-up can produce differences in cohort studies.

A Guide To Epidemiologic Terms Commonly Used

Relative Risk

The ratio of the risk among those exposed to the risk among the unexposed or the ratio of the cumulative incidence rate in the exposed and the unexposed. Also called risk ratio. In its simplest definition, relative risk compares the rate of disease in two groups, one of which has been exposed to something that is believed to either increase or decrease the risk of that disease, usually in a prospective study.

Odds Ratio

The odds ratio is the measure of association calculated in case-control studies when the prevalence of disease events is low; the estimate and interpretation are similar to relative risk.

Confidence Interval (CI)

The range of relative risk that would include 95% of the subjects being studied; the range of relative risk within which the true magnitude of effect lies, given the study data, with a certain degree of assurance. To be statistically significant, a reduced relative risk (a beneficial effect) requires the larger number (the right hand number) to be less than 1.0 (thus, both numbers are less than 1.0). An increased relative risk (an adverse effect), to be statistically significant, requires the smaller number (the left hand number) to be greater than 1.0 (thus, both numbers are greater than 1.0).

The tighter (more narrow) the range, the more precise the conclusion. The wider the CI, the more imprecise the conclusion, usually because of small numbers of study subjects.

P Value

By convention, the P value is significant if below 0.05. This is the probability of obtaining the relative risk or odds ratio by chance. The lower the P value the more likely a result is real. A P value of 0.05 means that there is a 5% probability that the result occurred by chance.

Attributable Risk

The difference in actual incidence between exposed and unexposed groups, providing a realistic estimate of the change in incidence in a given population. A modest increase in relative risk will produce only a small number of cases when clinical events are rare, such as venous thromboembolism and arterial thrombosis in young women. If the absolute risk is very low, a statistically significant increase in relative risk may mean little or nothing in practical, real numbers.

Number Needed To Treat

The number of individuals that must be treated, usually over a 1-year time period, to produce one instance of either a positive or a negative effect.

Important Points

Epidemiology is a tool to detect disease patterns in large populations. Epidemiologic studies do not prove causation; they identify associations between diseases and certain factors.

A relative risk in the range of 1.0–2.0 represents an increased risk, but a weak association.

The clinical significance of an increase in risk is influenced by the rate of the disease in the general (unexposed) population (attributable risk). If the rate of the disease in the unexposed population is 10% and the relative risk is 1.4, an exposed person has a disease risk of 14%. If the rate of disease in the unexposed population is only 1%, then the same relative risk of 1.4 increases the actual disease risk by only 1.4%.

Criteria that strengthen the conclusion that an epidemiologic finding is clinically real include the following:

1. **The strength of the association** (the larger the relative risk, the more likely it is real).
2. **Consistency, uniformity, and agreement** among many studies.
3. **A dose-response relationship** (either with dose of a drug or an increasing effect with increasing time of exposure).
4. **Biological plausibility** of the finding (known mechanisms by which exposure could cause or influence disease).
5. **An appropriate temporal relationship** (the amount of time between exposure and development of disease is appropriate according to the pathogenesis of the disease).

U.S. Preventive Task Force Evidence Grading Scheme

Quality of Evidence

Level I Evidence from at least 1 properly designed randomized controlled trial.
Level II-1 Evidence from well-designed, non-randomized, controlled trial.
Level II-2 Evidence from well-designed cohort or case-control studies.
Level II-3 Evidence from cross-sectional studies, or uncontrolled studies.
Level III Evidence from descriptive case reports, case series, or expert/committee opinions.

Strength of Recommendation

A Good and consistent scientific evidence to support recommendation, a substantial net benefit.
B Limited evidence to support a recommendation, a moderate net benefit.
C Not enough evidence to make a recommendation, possibly a small net benefit.
D Good evidence against a recommendation, no net benefit, possibly greater harm.
I Insufficient evidence, no conclusion.

Appendix II

Laboratory Values for Selected Measurements in Urine

Substance	*Conventional Units*	*Conversion Factor*	*SI Units*
Cortisol, free	10–90 µg/24 hr	2.759	28–250 nmol/24 hr
Estrogens, total	5–25 µg/24 hr	3.67	18–92 nmol/24 hr
17-Hydroxycorticosteroids	2–6 mg/24 hr	2.759	5.5–15.5 µmol/24 hr
17-Ketosteroids	6.0–15 mEq/24 hr	3.467	21–52.5 µmol/24 hr

SI Prefixes and Their Symbols

10^{9}	giga	G
10^{6}	mega	M
10^{3}	kilo	k
10^{2}	hecto	h
10^{1}	deka	da
10^{-1}	deci	d
10^{-2}	centi	c
10^{-3}	milli	m
10^{-6}	micro	µ
10^{-9}	nano	n
10^{-12}	pico	p
10^{-15}	femto	f
10^{-18}	alto	a

Laboratory Values for Selected Measurements in Blood, Plasma, and Serum

Substance	*Conventional Units*	*Conversion Factor*	*SI Units*
ACTH, adrenocorticotropin hormone			
6:00 AM	10–80 pg/mL	0.2202	2.2–17.6 pmol/L
6:00 PM	<50 pg/mL	0.2202	<11 pmol/L
Androstenedione	60–300 ng/dL	0.0349	2.1–10.5 nmol/L
Calcium, total	8.5–10.5 mg/dL	0.25	2.1–2.6 mmol/L
Cholesterol	<200 mg/dL	0.0259	<5.2 mmol/L
LDL-cholesterol	60–130 mg/dL	0.0259	1.6–3.4 mmol/L
HDL-cholesterol	30–70 mg/dL	0.0259	0.8–1.8 mmol/L
Cortisol			
8:00 AM	5–25 µg/dL	27.6	140–690 nmol/L
4:00 PM	3–12 µg/dL	27.6	80–330 nmol/L
10:00 PM	<50% of AM value	27.6	<50% of AM value
DHAS, Dehydroepiandrosterone sulfate	80–350 µg/dL	0.0027	2.2–9.5 µmol/L
11-Deoxycortisol	0.05–0.25 µg/dL	28.86	1.5–7.3 nmol/L
11-Deoxycorticosterone	2–10 ng/dL	30.3	60–300 pmol/L
Estradiol	20–400 pg/mL	3.67	70–1500 pmol/L
Estrone	30–200 pg/mL	3.7	110–740 pmol/L
FSH, reproductive years	5–20 mIU/mL	1.0	5–20 IU/L
Glucose, fasting	70–110 mg/dL	0.0556	4.0–6.0 mmol/L
Growth hormone	<10 ng/mL	1.0	<10 µg/L
17-Hydroxyprogesterone	100–300 ng/dL	0.03	3–9 nmol/L
Insulin, fasting	5–20 µU/mL	7.175	35–145 pmol/L
Insulin-like growth factor-I	0.3–2.2 U/mL	1000	300–2200 U/L
LH, reproductive years	5–20 mIU/mL	1.0	5–20 IU/L
Progesterone			
Follicular phase	<3 ng/mL	3.18	<9.5 nmol/L
Secretory phase	5–30 ng/mL	3.18	16–95 nmol/L
Prolactin	1–20 ng/mL	44.4	44.4–888 pmol/L
Testosterone, total	20–80 ng/dL	0.0347	0.7–2.8 nmol/L
Testosterone, free	100–200 pg/dL	0.0347	35–700 pmol/L
TSH, thyroid stimulating hormone	0.4–4.5 µU/mL	1.0	0.4–4.5 mU/L
Thyroxine, free T_4	0.8–2.3 ng/dL	1.29	10–30 nmol/L
Triglycerides	40–250 mg/dL	0.0113	0.5–2.8 mmol/L
Triiodothyronine, T_3, total	80–220 ng/dL	0.0154	1.2–3.4 nmol/L
Triiodothyronine, T_3, free	0.13–0.55 ng/dL	15.4	2.0–8.5 pmol/L
Triiodothyronine, reverse	8–35 ng/dL	15.4	120–540 pmol/L

Index